

이질적 전자기록물로 구성된 복합전자기록물 장기보존에 관한 연구 조사 및 분석

A Survey on Long-Term Preservation of Composite Electronic Records with Heterogeneous Electronic Records

황윤영(Yun-Young Hwang)*, 이규철(Kyu-Chul Lee)**

초 록

우리나라 정부부처 및 공공기관에서는 매우 많은 양의 전자기록물을 생산하고 있는데, 그 종류도 다양하다. 또한, 이질적인 전자기록물이 결합하여 하나의 전자기록물을 이루기도 하는데, 이 논문에서는 이를 복합전자기록물이라고 부른다. 복합전자기록물의 대표적인 예로는 소셜 네트워크 서비스(Social Network Service) 기록물이 있다. 현재, 우리나라 대부분의 정부부처에서는 페이스북을 이용하여 각 정부부처의 정책 공지 및 국민의견을 수렴하는 등 그 활용도와 가치가 높다. 이 외에도 많은 복합전자기록물이 우리나라 정부부처로부터 생산 및 활용되고 있는 실정이나, 이들에 대한 보존 방안이 제대로 마련되어 있지 않아 많은 정보가 유실되거나 방치되고 있다. 이 논문에서는 현재 우리나라 정부부처에서 생산 및 유통되는 복합전자기록물의 종류 및 특성을 살펴보고, 복합전자기록물 보존 관련연구를 조사·분석한 뒤 장기보존 전략을 제안한다.

ABSTRACT

Since the introduction of e-Government in Korea, government departments and public sectors produced a diversity of electronic records. In addition, they generated new electronic records (we are called composite electronic records(CERs)) which are composed of disparate electronic records. Typical examples of CERs are the records of Social Network Services (SNS). Currently, most of the government departments in Korea uses the Facebook Page to announce their policy and to collect feedback from the people. In addition, many CERs are created and produced. However, the government of Korea has no preservation policies about CERs. It causes problems with losing or neglecting a lot of information. Therefore, we presented the type and the characteristics of CERs which are produced by government departments of Korea. Moreover, we analyze and represent the researches of CERs preservation. Finally, we proposed the mechanism to preserve of CERs in Korea.

키워드 : 전자기록물, 복합전자기록물, 장기보존

Electronic Records, Composed Electronic Records, Digital Archiving

본 연구는 국가기록원의 2011년 기록관리분야 연구 개발 사업의 “복합 전자기록물 장기보존 아키텍처 연구”에 의해 수행되었음.

* Korea Institute of Science and Technology Information(yyhwang@kisti.re.kr)

** Corresponding Author, Dept. Computer Engineering, Chungnam National University(kclee@cnu.ac.kr)

2013년 07월 02일 접수, 2013년 08월 22일 심사완료 후 2013년 09월 4일 게재확정.

1. 서 론

전자정부사업백서[21]에 따르면, 우리나라는 1987년에 전자정부(e-Government)를 도입하여 국가 주요 행정정보의 데이터베이스 구축, 부처·기능별 업무 시스템 전산화, 범정부 전자정부 핵심 기반과 서비스 구축을 위한 전자정부 11대 과제 추진 등 지속적으로 정보화를 추진하였다. 이 결과, 우리나라의 기능별·부처별 단위업무 정보화는 고도화 단계에 진입하였으며, 민원, 세금, 조달, 통관 등 여러 영역이 전자적으로 처리되고, 국민을 위한 온라인 통합 서비스도 제공되기 시작하였다.

전자정부의 도입 이후 각 정부부처 및 공공기관에서는 종이기록물 대신 전자기록물을 생산·유통하기 시작하였다. 이로 인해 전자기록물 장기보존에 대한 필요성이 대두되어, 우리나라 국가기록원은 전자기록물 장기보존을 위한 우리나라 보존 포맷 및 보존 메타데이터 표준[24]을 제정하였다. [25]에서는 종이기록물의 구조인 칠/건 구조를 그대로 가지며, 문서 보존 포맷으로는 PDF/A-1(ISO 19005-1 : 2005 Part 1 : Use of PDF 1.4)[24]을 규정하고 있다. [24, 25]의 전자기록물은 우리나라 「공공기록물 관리에 관한 법률 시행령 제1장 제2조 2항」 [22]의 정의를 따르고 있는데, 전자기록물이란 “정보처리 능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신수신 또는 저장되는 전자문서, 웹 기록물 및 행정정보 데이터 세트 등의 기록정보자료”이다.

우리나라 전자정부 도입의 성공과 전자기록물 장기보존에 대한 많은 노력은 우리나라 정부기관 및 공공기관의 전자기록물이 체계

적으로 보존·관리될 수 있도록 하였다. 그러나 이러한 노력에도 불구하고 우리나라 정부부처 및 공공기관의 몇몇 전자기록물들은 아직까지 장기보존되고 있지 못하는 실정이다. 대표적인 예로 우리나라 정부부처 및 공공기관에서 오랫동안 생산되고 CAD 도면류는 장기보존 대상이지만, 아직까지 대부분의 CAD 도면류가 프린트되어 보존되고 있다. 또한 웹 2.0시대가 도래하면서 소셜 네트워크 서비스와 같은 새로운 형태의 서비스가 등장하였고, 우리나라 정부부처 및 공공기관에서 소셜 네트워크서 비스를 이용한 대국민 서비스를 하고 있다. 따라서 소셜 네트워크 서비스 기록물도 장기보존되어야 하지만 아직까지 이를 위한 장기보존 대책 및 표준이 국내·외 모두 미비한 실정이다. 이처럼 우리나라 정부부처 및 공공기관의 전자기록물 중 아직까지 장기보존되지 않는 전자기록물에는 CAD 도면류, 전자도서(e-Book), 전자문서 등이 있으며, 최근 신기술의 등장으로 인해 새롭게 생성되기 시작한 전자기록물에는 소셜 네트워크 서비스 기록물, HTML5 문서, 이러닝 및 스마트TV 콘텐츠 등이 있다. 이 논문에서는 이들을 모두 복합전자기록물(composite electronic records)이라 부른다.

복합전자기록물 장기보존과 관련하여 국외에서는 몇몇 연구가 진행되고 있으나 우리나라 정부에서 사용하고 있는 복합기록물의 장기보존에 바로 적용하기에는 한계점이 있다. 호주[30], 미국[16, 34], 캐나다[33], 영국[35] 등은 복합전자기록물의 한 종류인 CAD/CAM 도면류, 전자도서(전자저널 포함), 웹 기록물에 대한 보존의 필요성을 언급하고 있으나, 아직까지 구체적인 보존 방안은 제시하고 있지

못한다. LOTAR(Long Term Archiving and Retrieval) 프로젝트[20]는 CAD 도면류 장기 보존 프로젝트로, 현재 2D CAD 도면류에 대한 보존방안을 제시하고 있으며, 3D 도면류에 대한 장기보존 연구를 계속 진행하고 있다. 이 외에도 많은 복합전자기록물 장기보존 연구가 진행되고 있는데, 이들 대부분이 구체적인 보존 방안을 제시하고 있지 못하다. 이들에 대한 상세한 내용은 제 3장에서 자세히 설명한다.

이처럼 복합전자기록물에 대한 장기보존 방안은 국내·외 모두 미비한 상황에서 우리나라 정부부처 및 공공기관에서 생산·유통되는 복합전자기록물에 대한 장기보존 방안이 요구되고 있다. 따라서 이 논문에서는 우리나라 정부부처 및 공공기관에서 생산·유통하는 복합전자기록물의 종류 및 특성을 파악하고, 각 복합전자기록물 장기보존 관련 연구를 살펴본 뒤, 우리나라 복합전자기록물 장기보존 전략을 도출한다.

이 논문은 다음과 같은 구성을 갖는다. 제 2장은 복합전자기록물을 정의하고, 우리나라 정부부처 및 공공기관에서 생산·유통·관리되는 복합전자기록물의 종류 및 특성을 살펴본다. 제 3장에서는 복합전자기록물 장기보존 관련연구를 조사·분석하며, 제 4장에서는 우리나라 복합 전자기록물 장기보존 전략을 제시하고 제 5장에서는 결론 및 향후연구를 기술한다.

2. 복합전자기록물

이 장에서는 국제표준 및 우리나라 표준의

전자기록물 정의에 따라 복합전자기록물을 정의하고, 우리나라 정부부처 및 공공기관에서 생산·유통·보존되는 복합전자기록물 종류 및 특성을 살펴본다.

2.1 복합전자기록물의 정의

국제 표준인 ISO 15489에서는 기록물(Records)을 “내용(Content) + 메타데이터”로 정의하고 있다[13]. 여기서 내용이란 기록물 자체를 의미하며, 메타데이터란 기록물의 구조와 기록을 구성하는 요소간의 관계를 유지하면서 기록물의 맥락 기록을 유지할 수 있도록 하는 구성요소이다. 미국 NARA에서는 ISO 15489의 기록물 정의에 기반을 두어 전자기록물을 정의하고 있는데, “전자기록물은 컴퓨팅 플랫폼에서 생성되는 모든 종류의 전자기록물”로 정의하고 있다. 우리나라의 전자기록물 정의는 다음과 같다. 「공공기록물 관리에 관한 법률 시행령 제1장 제2조 2항」 [22]에 따르면 “정보처리 능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신·수신 또는 저장되는 전자문서, 웹기록물 및 행정정보 데이터 세트 등의 기록정보자료”로 정의하고 있다. 복합전자기록물과 관련한 정의로는 PANIC [14]에서 복합 디지털 오브젝트(Composite Digital Objects)를 “여러 이질적인 디지털 리소스로 구성된 객체”로 정의하고 있었으며, 이질적인 디지털 리소스에는 이미지, 비디오, 오디오, 텍스트, 하이퍼링크, 소프트웨어 프로그램 등이 해당한다.

이를 기반으로 이 논문에서는 [22]의 전자기록물을 재정의하고, 재정의된 전자기록물을 기반으로 <Figure 1>과 같이 기존의 전자기

록물 정의에 기반 한 복합전자기록물을 정의하였다. 이 논문에서 정의하고 있는 복합 전자기록물은 ISO 15489의 기록물 정의에 기반 하여 우리나라 「공공기록물 관리에 관한 법률 시행령 제1장 제2조 2항」에서 정의한 전자기록물을 포함하였다. 현재 우리나라 「공공기록물 관리에 관한 법률 시행령 제1장 제2조 2항」에서 정의된 전자기록물을 협의의 전자기록물이라 재정의하고, PANIC에서 정의된 복합 디지털 오브젝트의 개념을 포함하는 광의의 복합전자기록물을 정의하였다.

<i>The broad definition of a composite electronic records</i>	
The narrow definition of a composite electronic records	PANIC
Act on the Management of Public Archives	
ISO 15489	

〈Figure 1〉 The Relationship Between the Broad Definition of Composite Electronic Records and the Definition of Other Electronic Records

이 논문에서 정의한 전자기록물, 복합전자기록물, 광의 복합전자기록물의 정의는 다음과 같다.

정의 1 : 전자기록물의 정의

정보처리 능력을 가진 장치에 의하여 전자적인 형태로 생성되어 송신·수신 또는 저장되는 기록정보자료

정의 2 : 복합전자기록물의 정의

여러 상이한 형태의 전자기록물이 복합되어 구성되는 전자기록물

정의 3 : 광의 복합전자기록물의 정의

복합 전자기록물과 메타데이터로 구성되며, 메타데이터는 아래와 같은 조건을 만족해야 함

- 전자문서 형태의 기록에 초점을 둔 현재 메타데이터를 다양한 전자기록물의 특성을 기술하도록 확장해야 함
- 전자환경에서 기록의 연속적 관리(seamless flow)를 위해 생애주기별 메타데이터 연계 강화 필요

2.2 복합전자기록물 생산 및 유통 현황

복합전자기록물을 장기보존을 위한 요구사항 추출을 위해서는 우리나라 정부부처에서 생산·유통·보존되는 복합전자기록물의 종류 및 유통 현황을 파악할 필요가 있다.

현재 우리나라 정부부처 및 공공기관에서 생산·유통되는 복합전자기록물에는 CAD 도면류, 전자도서, 소셜 네트워크 서비스 기록물, 전자문서 등이 있다.

2.2.1 CAD 도면류

CAD(Computer Aided Design)는 설계에서 제품제조에 이르는 모든 공정을 컴퓨터로 관리하는 기술이다. CAD에 의하여 설계된 내용은 바로 CAM으로 연결되고, CAM을 통해 NC(수치제어) 공작기계에 정확한 작업동작지시를 하게되며 생산, 가공, 조립, 검사 등의 제조과정을 컴퓨터로 관리하여 작업의 신

속성과 제품의 정밀성을 기하게 된다. 근래에는 설계도면을 입체적으로 나타낼 수 있는 3차원 CAD · CAM 시스템이 개발되어, 입체형상을 컴퓨터 화면에 완벽하게 재현시켜줄 뿐만 아니라, 그 대상 입체의 겉넓이 · 부피 · 무게 · 강도 등 물리적 성질까지 자동으로 계산해서 적합한 형태로 설계하여 준다.

일반적으로 CAD는 사용 목적에 따라 다음과 같은 종류로 나뉜다.

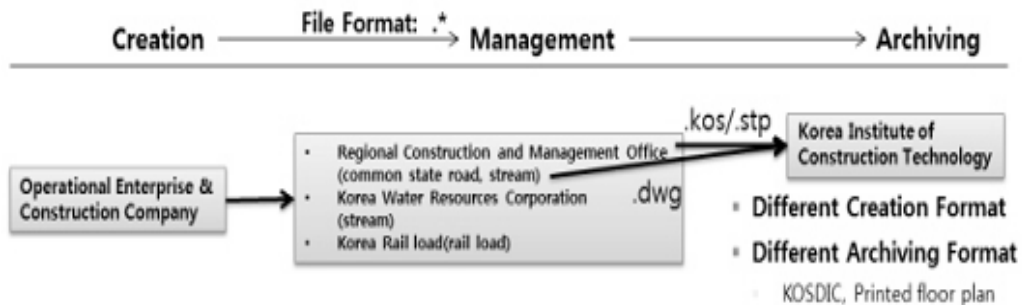
- 기계설계(Mechanical) CAD : MCAD
- 전기전자(Electric/Electronic) CAD : ECAD
혹은 EDA(Electronic Design Automation)
- AE&C(Architecture, Engineering, and Construction) CAD : AE&C CAD
- 매핑(Mapping) CAD : 지도 작성

이중 가장 많이 사용하는 CAD는 기계설계 CAD, 즉 MCAD로, 대부분의 회사에서 사용한다. AE&C CAD는 엔지니어링 회사나 중공업 회사에서 주로 사용하며, ECAD는 전자제품이나 반도체를 만드는 회사에서 사용한다.

우리나라 공공기관에서 생산되는 CAD 도면류 생애주기는 <Figure 2>와 같다. 각 시

공업체나 건설업체는 설계시점 또는 준공시점마다 자신들이 설계한 CAD 도면류를 발주처인 각 기관에 보내게 된다. 그 사업 내용에 따라 일반 국도나 지방하천의 경우 지방국토관리청이 CAD 도면을 관리하게 되며, 항만에 관한 CAD 도면은 한국 항만공사가 관리하고, 한국수자원공사는 하천을, 한국철도공사, 지하철공사, 한국공항공사, 한국도로공사는 각각 철도, 지하철, 공항, 고속국도와 관련된 CAD 도면을 관리한다. 각 관리기관으로 이관된 CAD 도면 중 지방국토관리청에서 관리하는 CAD 도면만이 한국건설기술연구원으로 이관되고 있다. 지방국토관리청은 우리나라 건설분야 도면정보 교환표준인 KOSDIC(Korean Standard of Drawing Information in Construction)[17]을 따라 CAD 도면을 관리하고 이관한다. 항만 CAD 도면은 AutoCAD 파일 포맷(.dwg)으로 한국건설기술연구원으로 이관되고 있으며, 다른 기관이 관리하는 CAD 도면들은 보관이 되고 있지 않은 실정이다. 단지 필요기관에서 요청시에만 도면을 인쇄하여 전달하고 있었다.

한국건설기술연구원으로 이관되고 있는 CAD 도면류는 한해 30~40건으로 건당 수백 개의



<Figure 2> The Lifecycle of CAD Records

도면이 포함되어 있다.

현재 우리나라 정부기관에서 발주된 사업에서 생산된 CAD 도면의 유통 및 관리 현황은 장기보존의 측면에서 다음과 같은 문제점을 가지고 있다.

- 각 시공사 및 건설업체 마다 생산되는 CAD 파일 포맷이 다름

각 시공사 및 건설업체 마다 보유하고 있는 CAD 소프트웨어가 상이함에 따라 그들이 생산하는 CAD 전자 파일 포맷이 다르다. 이렇게 다양한 포맷으로 생산된 CAD 전자파일들은 발주처로 이관이 되고 다시 장기보존기관으로 이관되면서 특정 포맷으로 변환되는 데, 이때 많은 정보들이 유실될 수 있다. 또한, 전자 파일 형태로 이관되지 않고 인쇄물로 이관될 경우, 3D CAD 전자파일들의 데이터들이 손실되는 것은 당연한 결과일 것이다.

- 각기 다른 기관에서 다른 형태로 보관
<Figure 2>에서 보여지는 것과 같이 사업에 따라 발주기관이 다르고, 그들이 시공사나 건설업체에게 이관시 요구하는 전자 파일 포맷이 상이하다. 또한, 어떤 기관은 장기보존기관으로 이관조차 하지 않고 있다.

2.2.2 소셜 네트워크 서비스

소셜 네트워킹 웹 사이트란 온라인상에서 공통의 관심사를 가진 사용자 간의 관계맺기를 지원하고, 축적된 지인 관계를 통해 인맥 관리, 정보 공유 등 다양한 커뮤니티 활동을 할 수 있도록 하는 서비스를 의미한다. 온라인 커뮤니티와 혼용되어 사용되기도 하나, 커뮤니티가 주로 비슷한 관심사의 사람들이 한

장소에 모여 활동하는 그룹 중심의 커뮤니티 서비스를 지칭한다면, 소셜 네트워킹 웹 사이트에는 개인이 중심이 되어 다른 개인과 관계를 맺고, 이러한 개개인의 관계가 모여 축적됨으로서 더 큰 네트워크를 형성하는 구조를 띄고 있다. 가장 대표적인 소셜 네트워크 서비스로는 페이스북, 트위터, 미투데이 등이 있다.

소셜 네트워크 서비스는 기본적으로 관계 맺기 및 프로필 제공 기능을 중심으로 다음과 같은 기능들을 제공하고 있다.

- 관계 맺기

친구/지인 맺기, 팬 되기 등 개인들이 관계를 맺고 시스템에 축적할 수 있는 기능

- 프로필

사진과 신상정보, 취미, 관심사 등 개인 자신의 아이덴티티를 밝힐 수 있는 기능

- 커뮤니케이션

이메일, 쪽지, 채팅, 메신저 등 사용자 간의 커뮤니케이션을 지원하는 기능

- 콘텐츠 생산

프로필 이외에 블로그, 포토, 동영상 등 사용자가 시스템 안에서 콘텐츠를 생산할 수 있는 기능

- 네트워크의 활용

사용자 생산 콘텐츠나 외부 콘텐츠를 공유하거나 추천, 배포하고 함께 협업등 축적된 관계 네트워크를 활용하는 기능

우리나라 정부부처 및 공공기관 대부분에서는 페이스북(Facebook), 트위터(Twitter), 미투데이(M2Day)를 활용해서 국민과의 대화를 시도하고 있으며, 국민들의 의견에 빠르게 대응하고 있다. 2012년 우리나라 15부 2처

폴리씨 · 정부부처 SNS 지도
정부 단위

〈Figure 3〉 The Map of Korea Government Ministry

18정 모든 정부부처에서 운영하고 있는 페이스북 제공 현황은 <Figure 3>과 같다.

<Figure 3>은 폴리씨[27] 페이스북 페이지에서 제공하는 정부부처 소셜 네트워크 서비스 지도다. 정부부처의 소셜 네트워크 서비스 사용 목적은 각 정부부처의 정책을 알리고 해당 정책에 대한 국민들의 의견을 수렴하는 등 국민과의 대화를 목적으로 적극 활용하고 있다.

이처럼 중요한 가치를 가진 소셜 네트워크 서비스 기록물에 대한 장기보존 방안이 국내·외 모두에 없으며, 소셜 네트워크 서비스 제공업체들이 모두 개인 기업체여서 영구적으로 서비스된다는 보장이 없다. 따라서 이에 대한 장기보존 방안이 시급한 실정이다.

2.2.3 전자도서(e-Book)

한국전자출판협회에 따르면 전자도서는 “도서로 간행되거나 또는 도서로 간행될 수 있는 저작물의 내용을 디지털 데이터로 CD-

ROM, DVD 등의 전자도서 기록매체, 또는 저장장치에 수록하고, 유무선 정보통신망을 경유하여 컴퓨터 또는 휴대용 단말기 등을 이용해 그 내용을 읽고, 보고, 들을 수 있는 것을 말한다”라고 정의하고 있다. 따라서 현재는 PC를 통해 보는 CD롬 타이틀도 전자도서로 볼 수 있고, 개인휴대 단말기(PDA)도 전자도서의 한 종류로 볼 수 있다. 그러나 협의로 보았을 때 전자도서는 인터넷 언어인 HTML과 XML을 응용하여 만든 책과 같은 고안물로서 일반적으로 인터넷을 통해 다운로드 받는 것은 물론 전용 뷰어를 통해 PC나 단말기로 볼 수 있는 디지털 영역을 통칭한다.

전자도서 산업은 세 가지 영역으로 구분된다. 전자도서의 콘텐츠를 전자도서로 부르고, 전자도서 콘텐츠를 보는 소프트웨어를 전용 뷰어(Viewer), 하드웨어는 전자도서 단말기(e-book Device)로 불리어지고 있다. 그러나 콘텐츠와 소프트웨어, 하드웨어를 합쳐 전자도서라 하기도 한다. 즉 전자도서는 인터넷을 통해 다운로드를 받은 것은 물론 전용 뷰어

를 통해 PC나 전용 단말기로 볼 수 있는 디지털 출판영역이라 할 수 있다.

위키피디아[39]에 따르면 <Table 1>과 같이 전자도서 포맷은 18종으로 소개되고 있다. 이들 간에도 서로 다른 전자포맷 형태로 생산되고 있으며, 이들을 아우르는 공개 표준도 없는 것으로 파악되었다.

전자도서 역시 우리나라 정부부처 및 공공기관에서 각 기관의 정책을 홍보하고 국민들에게 관련지식을 전달하기 위해 사용된다. 전자도서를 지원하는 정부부처는 고용노동부,

교육과학기술부, 국토해양부, 문화체육관광부, 법무부, 외교통산부 등이 있다. 이들은 각기 상이한 전자도서 소프트웨어를 사용하고 있었으며, 그로 인해 상이한 포맷으로 전자도서를 생성하고 보급하고 있었다. <Table 1>에서 볼 수 있듯이 전자도서를 생산하는 소프트웨어에 따른 파일 포맷이 생성되고 유통되고 있다. 또한 공식적인 납본 절차가 없으며, 이들을 영구보존하기 위한 보존 포맷도 없는 상황이어서 많은 중요한 기록들이 보존되지 못하고 있다.

<Table 1> The Comparison Between e-Book Formats

Format	Extension	DRM	Image	Table	Sound	Interactivity	Word wrap	Open Standard	Embedded Anotation	Book Marking
ArgghosReader	.aeh	Yes	Yes	?	No	No	Yes	No	?	Yes
DjVu	.djvu	?	Yes	Yes	No	No	No	Yes	Yes	Yes
EPUB(IDFF)	.epub	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FictionBook	.fb2	No	Yes	?	No	No	Yes	Yes	Yes	?
HTML	.html	No	Yes	Yes	No	No	Yes	Yes	No	No
Kindle	.azw	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
Microsoft Reader	.lit	Yes	Yes	?	No	No	Yes	No	?	Yes
Mobipocket	.prc, .mobi	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes
Multimedia Ebook	.exe	Yes	Yes	?	Yes	Yes	No	Yes	Yes	Yes
eReader	.pdb	Yes	Yes	?	No	No	Yes	No	Yes	Yes
Plain text	.txt	No	No	No	No	No	Yes	Yes	No	No
Plucker	.pdb	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes
Portable Document Format	.pdf	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
PostScript	.ps	No	Yes	?	No	No	No	Yes	?	?
Repligo	.rgo	?	Yes	Yes	No	No	Yes	No	No	No
TealDoc	.pdb	Yes	Yes	?	No	No	Yes	Yes	?	Yes
Tome Raider	.tr2, .tr3	Yes	Yes	?	No	No	Yes	No	?	?
WOLF	.wol	Yes	Yes	?	No	No	No	No	?	?

〈Table 2〉 The Purpose of Use E-Book and The Present Condition of Supported Format and Software

	Dept. 1	Dept. 2	Dept. 3	Dept. 4	Dept. 5	Dept. 6
Format	HTML	PDF	Fbook	XDF	XDF	PDF
Software	FlexPaper	PDF Reader	Fbook	Espresso Reader™ EPYRUS	Espresso Reader™ EPYRUS	WebBook Maker 3.0 - YESeBOOK
OS	Window	Window	Window	Window	Window	Window
Purpose	Webzine	Webzine	Guideline	Promotion	Guideline	Promotion

또한, <Table 1>에 우리나라 정부부처에서 사용하는 XDF 포맷, FBook 자체 포맷 등은 포함되어 있지 않다. 이를 미루어보아 이것 외에도 많은 전자도서 포맷이 존재할 것으로 추정된다.

3. 복합전자기록물 장기보존 관련 연구

제 2장에서 설명된 우리나라에서 생산·유통되는 복합기록물을 장기보존하기 위한 요구사항을 파악하기 위해 그와 관련된 장기보존 관련 연구를 살펴본다.

3.1 복합전자기록물의 장기보존 관련연구

3.1.1 CAD 도면류 장기보존

CAD 영구보존과 관련한 연구로는 LOTAR (Long Term Archiving and Retrieval) 프로젝트[20]가 있다. LOTAR 국제 프로젝트의 목적은 3D CAD나 PDM 같은 디지털 데이터를 영구보존하기 위한 표준 개발에 있다. 이 프로젝트는 OEM이 주도하고, ASD-STAN,

AIA, PDES Inc. ProSTEP iViP Association 같은 항공 우주방제 산업체의 지원을 받아 진행된다.

<Figure 4>에 나타난 것처럼 LOTAR 프로젝트는 3D CAD 영구보존을 목적으로 하였으나, 처음에는 수치 데이터를 제외한 CAD 데이터만 영구보존 하였다. 그러나, 최종 목표는 수치 데이터(PMI)까지 모두 포함한 CAD 데이터를 영구보존 하는데 있다.

LOTAR 프로젝트는 다음과 같은 3단계 프로세스를 갖는다.

- Phase 1 : EN/NAS 표준 문서 및 실제 상황에서 3D CAD 데이터의 개념, 프로세스 고려사항을 생성
- Phase 2 : EN/NAS 표준 문서 및 3D CAD, PMI, PDM, 3D CAD 조합 등 장기보존을 위한 고려사항을 생성
- Phase 3 : EN/NAS 표준 문서 및 시스템 설계 시뮬레이션 등을 위한 고려사항 생성

LOTAR 프로젝트는 아직 연구 중반 단계로 2D 형태의 데이터까지 영구보존을 마무리한 상태이다. 그러나, 이 역시 특정 포맷에만

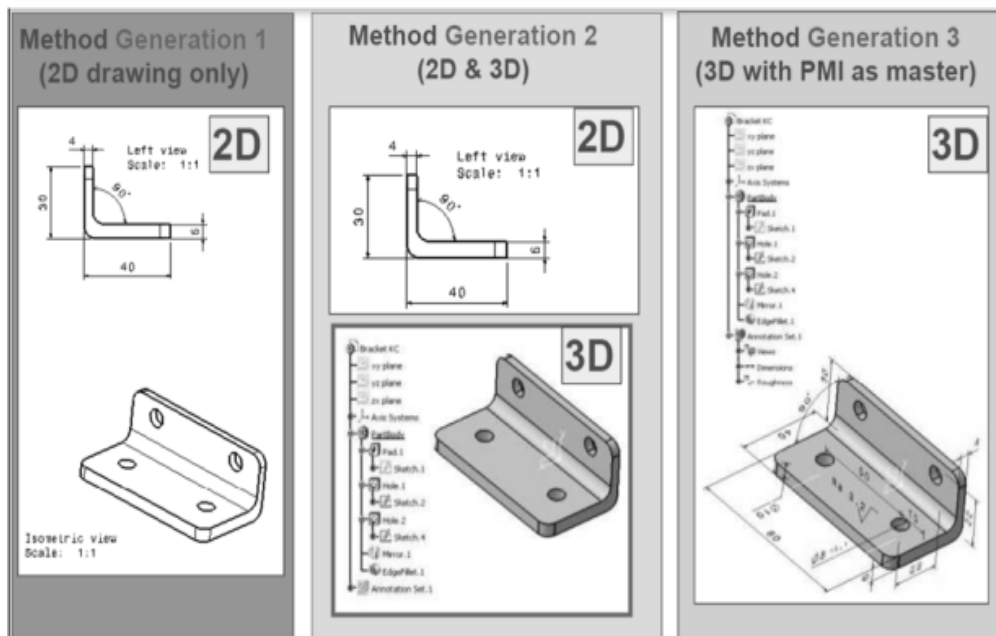
국한된 형태로 영구보존이 마련되어 있어, 우리나라에서 생산되는 CAD 전 기록물에 적용할 수 없다. 따라서 우리나라 교환 표준인 KOSDIC에 기반하여 3D CAD 기록물 영구보존 할 수 있는 방안이 마련되어야 할 것으로 보인다.

3.1.2 소셜 네트워크 서비스 기록물 장기보존

소셜 네트워크 서비스 중 페이스북/트위터의 기록물 장기보존을 위한 연구로는 다음과 같은 것들이 있다

- 페이스북 기록물 장기보존 관련 연구
- 페이스북 기록물 장기보존 관련 연구로는 NASA의 Archive-It[4], FacebookArchive[11], ArchiveFacebook[3], SocialSafe[32] 등이 있다. NASA Archive-It은 모든 우주·항공 기

판의 소셜 미디어 행위를 저장하기 위한 목적을 가지는 프로젝트로, 웹 아카이빙 방식으로 소셜 네트워크 서비스 기록물을 저장한다. Archive-It은 페이스북 외에도 유튜브(YouTube), Flickr, Ustream, MySpace 등을 저장하며, 향후 타 기관의 디지털 콘텐츠를 수집하고 보존하는 기능을 포함할 계획을 가진다. Facebook-Archive는 페이스북에서 공식적으로 제공하는 아카이브 도구다. Facebook Archive는 페이스북 웹 페이지 스타일로 페이스북 기록물을 로컬 디스크에 저장하는데, 이 파일은 압축된 파일 형태로 저장된다. 파이어폭스(FireFox)에서 제공하는 Archive Facebook 1.1은 개인의 하드디스크에 페이스북 기록물을 저장하기 위한 방법을 제공한다. SocialSafe는 API 기반의 기록물 백업 방법을 이용하여 페이스북 기록물을 저장·관리하고 있으며, 기록물 검색 기



〈Figure 4〉 The Domai n of LOTAR Project

능까지 제공한다.

이들은 아직까지 아카이빙 방법이라기 보다는 페이스북 데이터를 저장하기 위한 도구라 할 수 있다. 기존 연구들은 단순 데이터만을 저장하고, 그들 간의 시간적 관계성, 기록간의 선후 관계, 연관관계 등을 표현할 수 없어, 향후 복원 시 그 한계를 드러내고 있다.

• 트위터 장기보존 관련 연구

트위터 장기보존 관련 연구로는 미국 NDIIPP에서 진행하는 트위터 아카이빙 프로젝트[19], SAA 프로젝트[31], TwitSCAN[38], Twapper Keeper[36], TweetDoc[37], Archivist Desktop[5] 등이 있다.

NDIIPP(National Digital Information Infrastructure and Preservation Program)는 트위터에 대한 아카이빙 프로젝트를 진행한다. 이 프로젝트는 2010년 4월 14일에 시작된 프로젝트로 케이스 스터디, 개인정보 이슈, 트위터 기록물에 대한 접근, 사용, 관리 방법에 대한 연구를 진행한다. SAA(Society of American Archivists)에서는 트위터를 이용하여 트위터 아카이빙에 대한 설문조사를 실시하였다. 트위터 기록물 아카이빙에 대한 방법론을 제공하지만, 구체적인 방안은 아직까지 내놓지 못하고 있다. TwitSCAN은 트위터에서 제공하는 기록물 백업 도구로 트위터의 모든 종류의 기록물을 저장할 수 있으나, 최대 최근 1,000개까지의 메시지만 저장가능하다는 단점을 가진다. TwapperKeeper는 온라인/오프라인 형태의 아카이빙 서비스를 제공한다. 온라인 아카이빙 서비스는 웹 기반의 트윗 아카이빙 서비스를 제공하는데, 모든 트윗을 저장할 수 없다는 제약점을 가지며, 오프라인

아카이빙 서비스는 모든 트윗을 저장할 수 있지만 프로그램을 개인 컴퓨터에 설치해야 한다. TweetDoc은 사용자가 설정한 특정 이벤트나 용어에 대한 모든 트윗을 문서 형태로 생성하는 서비스로 트위터 기록물에 대한 재현은 고려하지 않았다. 마지막으로 Archivist Desktop은 데이터마이닝 및 분석을 위해 트윗을 아카이빙 하는 윈도우 애플리케이션으로 최근 3, 4주의 내용 중 최대 1,500개의 개수만 저장 가능하다는 제약점을 가진다.

3.1.3 전자도서 장기보존

복합전자기록물 중 전자도서는 가장 활발히 장기보존이 진행되고 있는 분야이다.

오래전부터 전자기록물과 관련하여 연구를 진행해온 호주[30], 미국[17, 34], 캐나다[33], 영국[35]은 복합전자기록물인 CAD 도면류, 전자도서(전자저널), 웹 기록물에 대한 보존의 필요성을 언급하고 있다. 그러나 아직까지는 구체적인 보존 방안은 제시하고 있지 않고 있다.

네덜란드[15, 16]는 전자도서(전자저널) 보존 방안을 제시하고 있다. 이들이 제시하는 보존 방안은 마이그레이션 방법으로써 이미 생성된 전자도서와 웹 기록물에서 메타데이터를 수집한 후 원본과 함께 보존하는 방안을 취한다. 그러나 네덜란드에서도 보존 포맷을 제안하고 있지는 않다.

복합전자기록물 영구보존에 대한 연구를 진행하고 있는 공동 연구로는 Portico[28, 29]가 있었다. Portico는 2002년 설립된 디지털 보존 커뮤니티로, 네덜란드 미국 캘리포니아 전자도서관, 스탠포드 대학, 영국 국립 도서관,

런던 대학 도서관 등이 참여하고 있다. Portico는 도서관, 출판사 전문가 그룹의 반복적인 토론에 바탕해서 e-콘텐츠의 보존 및 접근 방법에 대한 지침을 개발한다. e-콘텐츠란 텍스트, SGML, XML, 그래픽, 수식, 이미지 등이 복합적으로 융합된 것을 의미한다. 현재, Portico는 전자저널과 전자도서를 대상으로 보존 방안에 관한 연구를 진행하고 있다. Portico는 보존 메타데이터로 METS(Meta-data Encoding and Transmission Standard)[10] 기반의 메타데이터를 설계하였으며, 이와 관련하여 XML 스키마 및 DTD를 제공한다.

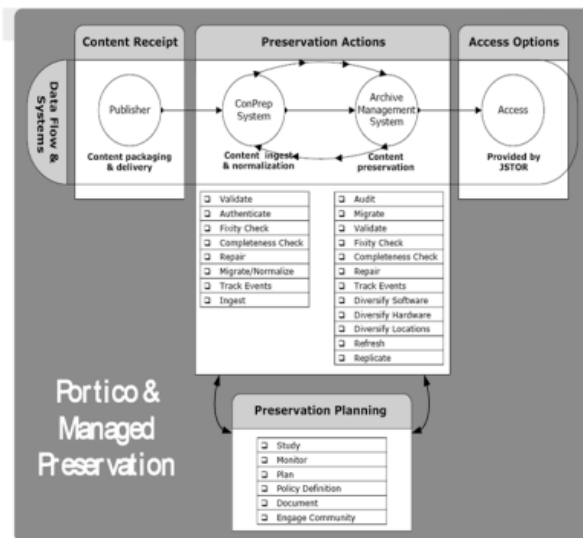
Portico의 보존 프로세스는 <Figure 8>과 같다. 전자기록물 생성자가 콘텐츠를 패키징하여 전달한 것을 수집하고 일반화 과정을 거친다. 이 과정에서 유효성 검사를 수행하여 콘텐츠의 무결성 및 진본성을 보장한다. 다음으로 보존 관리 시스템에서 콘텐츠에 사용된 소프트웨어, 하드웨어, 위치 정보 등을 추출

하여 마이그레이션하고, 다시 한 번 유효성 검사를 실시하여, 보존 프로세스를 거친 후에도 진본성을 유지한다. 이 과정은 모두 보존 계획에 기반하여 수행된다.

4. 복합전자기록물 장기보존 전략

4.1 복합전자기록물 보존 전략

복합전자기록물을 크게 이관대상 복합전자기록물과 수집대상 복합전자기록물로 나누고 각각에 대한 보존 전략을 수립할 필요가 있다. 이관대상 복합전자기록물은 보존 메타데이터가 이관시 함께 전달되게 되지만, 수집대상 복합전자기록물은 이관 시 보존 메타데이터를 생성해 주어야 한다. 이에 따른 장기보존 전략은 다음과 같다.



<Figure 5> The Archiving Process



〈Figure 6〉 The Lifecycle of Transferred Composite Electronic Records

4.1.1 이관대상 복합전자기록물 보존 전략

이관대상 복합전자기록물에는 CAD 도면류와 전자도서가 있다. 이들은 〈Figure 6〉과 같은 생애주기를 갖는다. 생산기관에서 생성된 전자기록물은 기록관으로 전달되어 활용되다가 보존 기관으로 이관되어 영구 보존되게 된다. 이러한 생애주기를 갖는 이관대상 복합전자기록물은 생산·활용·보존 단계별 메타데이터가 다르며, 보존 기관에 이관할 때 보존 메타데이터가 생성되게 된다. 따라서, 이관대상 복합전자기록물은 활용기관에서 보존 포맷으로 변환하여 이관하고, 이관 시 보존 메타데이터도 함께 생성하여 이관하는 형태가 바람직할 것이다.

4.1.2 수집대상 복합전자기록물 보존 전략

〈Figure 7〉과 같은 생애주기를 갖는 수집대상 복합전자기록물의 메타데이터 특성은 생산·활용 메타데이터가 같지만 보존 메타

데이터는 생성하지 않는다는 데 있다. 즉, 보존 기관이 보존에 필요한 메타데이터를 기록물 수집 시 생성해야 한다. 수집대상 복합전자기록물의 경우에는 보존 기관에서 수집된 복합전자기록물을 적당한 보존 포맷으로 변환하고, 필요한 보존 메타데이터를 추출하는 보존 전략이 필요하다. 수집대상 복합 전자기록물에는 소셜 네트워크 서비스 기록물이 해당된다.

4.2 복합전자기록물 보존 포맷

복합전자기록물은 여러 종류의 전자기록물로 구성된 것이므로, 단일 보존 포맷을 지정하기 어렵다. 따라서, 각 전자기록물 마다 적합한 보존 포맷 후보군을 제시한다.

- CAD 도면류 : KOSDIC, PDF/E, STEP
- SNS : 새로 정의
- 전자도서 : PDF, EPUB

CAD/CAM 도면류는 우리나라 관련 표준



〈Figure 7〉 The Lifecycle of Collected Composited Electronic Records

인 KOSDIC이나 PDF/E가 후보군이 될 수 있다. 또한 국제 표준이 STEP도 보존 포맷 후보가 된다. 그러나 KOSDIC도 STEP을 준수하고 있고, PDF/E는 아카이빙을 목적으로 제정된 표준이 아니므로 STEP가 가장 적합한 보존 포맷으로 판단된다. SNS는 관련하여 보존 포맷이 존재하지 않으므로 새로 정의해야 하며, 전자도서의 경우에는 전자문서 표준인 PDF나 국제 표준인 EPUB가 그 후보군이 될 수 있다.

4.3 복합전자기록물 보존 메타데이터

복합전자기록물을 위한 보존 메타데이터는 여러 종류의 전자기록물로 구성된 복합 전자 기록물의 구성을 잘 표현해 줄 수 있어야 하며, 전자기록물마다 다른 표준으로 작성된 메타데이터를 포함시켜 줄 수도 있어야 한다.

이와 관련하여 가장 적합한 표준은 METS (Metadata Encoding and Transmission Standard) 표준이다. METS는 디지털 저작물에 대한 구조적, 기술적, 관리적 메타데이터를 입력하기 위한 XML 언어 스키마로, 리포지토리 간의 디지털 저작물 교환을 표준화하기 위해 사용된다(문헌정보학용어사전). 이 표준은 디지털 도서관의 객체가 어떻게 조직되어 있는지 이해하기 위해서 구조 메타데이터가 필요하고, 디지털 형태의 자원이 원자원을 얼마나 정확하게 표현하는지를 이해하기 위해 기술(technical) 메타데이터가 필요하다는 요구에 대한 해결책으로 제시되었다. 나아가 오디오 또는 비디오 형태와 같이 시간에 기반을 둔 디지털 자원의 구조에 대한 기술(記述)과 외부 자원에 대한 링크 기능, 기술 및 관리

메타데이터에 더 많은 확장성을 제공하기 위해서 개발되게 되었다.

METS는 디지털도서관 연맹의 사업으로 개발되어 2001년 6월 베타버전이 발표되었고 현재는 미국 의회도서관 네트워크 개발 및 마크 표준국이 관리하고 있으며 2010년 버전 1.6이 발표되었다.

METS의 목적은 다음과 같다.

- 디지털 도서관 객체의 계층적 구조를 표현하는 XML 문서 생성
- 객체를 구성하는 파일들의 이름과 위치를 기록
- 관련된 메타데이터들을 기록

METS 스키마는 크게 7개 부분으로 구성되어 디지털 도서관이나 디지털 아카이브를 위한 설명, 관리 및 구조에 관한 메타데이터를 XML 스키마언어를 사용하여 표현하며 OASIS 프레임워크 위에 구축된다. METS 스키마의 7개 부분은 METS 헤더, 기술적 메타데이터(Descriptive Metadata Section), 관리적 메타데이터(Administrative Metadata Section), 파일(File Section), 구조적 지도(Structural Map Section), 구조적 연결(Structural Link Section), 행위(Behavior Metadata Section)이다<Figure 8>.

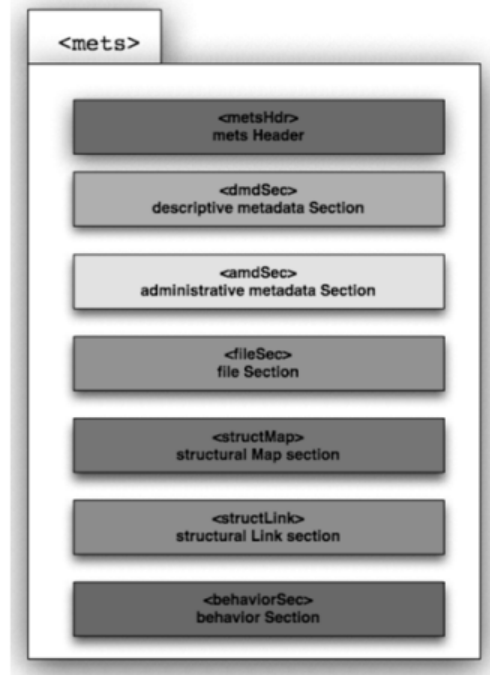
- 헤더섹션(mets Header) : METS 문서 자체에 대한 메타데이터를 저장하는 곳으로, METS 문서가 인코딩하는 디지털 객체 자체에 대한 메타데이터가 아님. 헤더섹션에서는 에이전트 이름, 역할, METS 문서에 대한 대체 레코드 ID, METS 문서 생성 및 갱신 시간, METS 문서 상

태를 기록할 수 있음

- 서술메타 데이터섹션(descriptive meta-data Section) : 디지털 객체와 관련된 서술 메타데이터를 기술
- 관리메타 데이터섹션(administrative metadata Section) : 디지털 객체를 관리하는데 필요한 메타데이터를 기술. 관리메타데이터에는 기술 메타데이터, 소스메타데이터, 저작권 메타데이터, 변경정보 메타데이터가 있으며, 이들 역시 기술 형식을 제공하지 않고, 다른 메타데이터 표준으로 기술된 관리 메타데이터를 METS 문서에 포함시킬 수 있도록 함
- 기술 메타데이터 : METS 문서가 표현하고자 하는 디지털 객체를 구성하는 파일이 어떻게 생성되고 저장되었는지를 기술
- 소스 메타데이터 : 디지털 객체가 디지털 형태로 생산되지 않고 아날로그 상태에서 디지털 형태로 변환된 경우 원자원에 대한 서술 및 관리 메타데이터를 기술할 수 있도록 함
- 저작권 메타 데이터 : 디지털 객체의 지적재산권 정보를 기술
- 변경정보 메타 데이터 : 디지털 객체를 구성하는 파일간의 출발지 및 목적지간의 관계를 기술
- 파일섹션(file Section) : 디지털 객체를 구성하는 디지털 파일 정보를 기술하는 곳으로 디지털 파일 자체를 포함하거나 다른 곳에 위치한 디지털 파일로의 연결정보를 기술. 파일섹션의 용도는 하나 이상의 관련된 파일을 하나로 묶어주는 역할을 함으로써, 용도 및 형태를 기준

으로 METS 문선에 포망되는 디지털 파일을 그룹핑하는 역할을 담당

- 구조맵섹션(structural map Section) : 디지털 객체를 네비게이션할 수 있도록 파일간의 계층적 구조를 정의
- 구조연결섹션(structural link Section)은 구조맵 내의 구성 단위인 div 간의 하이퍼링크를 표현
- 행위섹션(behavior Section)은 디지털 객체와 관련된 실행가능한 행위에 대한 정보를 기술



〈Figure 8〉 METS

METS는 각 요소를 구조화하기 위한 논리적 구조를 제공한다. 다양한 메타데이터는 METS에서 정의된 7개의 섹션에 논리적으로 매핑되어 상호교환을 위한 형태를 가질 수 있게

〈Table 5〉 The Mapping from KOSDIC to METS

KOSDIC		METS
Name		Descriptive Metadata Section
Time-stamp		Behavior Section
Organization		Descriptive Metadata Section
Originating_system		File Section
Authorization	Project_Information()	Administrative Metadata Section
	Project_Additional_Information()	Administrative Metadata Section
	Employer_Information()	Descriptive Metadata Section Administrative Metadata Section
	Contractor_Information()	Descriptive Metadata Section Administrative Metadata Section Behavior Section
	Drawing_Preview_Information()	File Section
	Drawing_Index_Information()	File Section
	Drawing_Additional_Information()	Descriptive Metadata Section
	Drawing_Change_Information()	Behavior Section
	Facility_Information()	Descriptive Metadata Section
	User_Defined_Information()	Descriptive Metadata Section

된다. 또한 METS는 OAIS 참조모형을 지원하여 SIP, DIP를 인코딩하고 전달할 수 있으며 AIP를 인코딩할 수 있는 프레임워크를 제공한다. 즉, 전달, 저장을 위한 구문을 제공하고 이를 통해 디지털 객체를 이용할 수 있는 기반을 제공한다. METS를 사용하여 분산된 환경에서 생성된 다양한 형식의 메타데이터를 변환하지 않고 사용할 수 있다.

〈Table 4〉는 KOSDIC 메타데이터를 METS로 표현한 것이다. 모든 항목이 일대일로 매핑되지는 않지만, 규칙을 통해 반자동화 방법으로 필요한 보존 메타데이터를 추출할 수 있을 것이다.

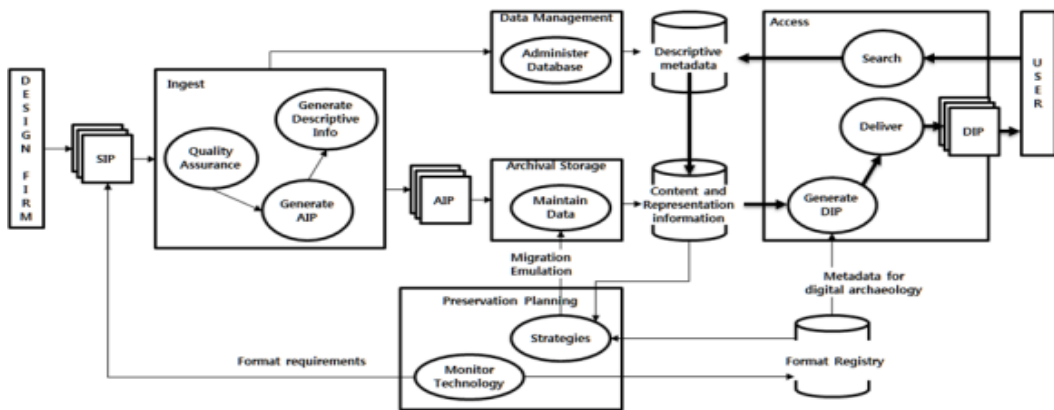
4.4 복합전자기록물 워크플로우

OAIS를 기반으로 이관대상 및 수집대상 복합전자기록물 워크플로우를 정의하였다.

4.4.1 이관대상 복합전자기록물

이관대상 복합전자기록물 워크플로우는 〈Figure 9〉와 같다. 메커니즘은 [12]를 국내 실정에 맞추어 정의하였다.

이관대상 복합전자기록물 워크플로우는 총 여섯 단계로 이루어진다. 생산기관에서 생산된 복합전자기록물은 SIP 형태로 입수 단계에 전달된다. 입수 단계에서는 영구보존에 필요한 정보



〈Figure 9〉 The Workflow of Transferred Composite Electronic Records

들을 추출하고 이관된 복합자기기록물을 AIP 형태로 변환한다. AIP는 아카이브 저장소로 전달되며, 관리자에 필요한 데이터베이스 관리 단계를 거치게 된다. 이러한 보존 프로세스는 보존 정책을 따르게 되며, 후에 접근 단계를 통해 사용자는 자신이 필요한 정보에 접근할 수 있다.

1) 입수

입수 단계에서는 SIP를 AIP로 변환하는데 필요한 품질보증과 메타데이터를 생성한다. 입수 단계는 다음과 같은 프로세스를 갖는다.

- 기본 프로젝트 정보로부터 초기 카탈로그 레코드를 생성
- 장기보존 무결성을 보장하기 위해 모든 데이터에 대해 자동적으로 체크섬을 생성하고, 그 정보와 디지털 문서를 함께 업로드 함
- 파일이 손상되지 않도록 품질 보존 체크 완료
- 모든 디지털 문서에 “보류” 상태를 지정
- 웹에서 사용가능하도록 TIFF에서 JPG

이미지 생성

- AIP 생성

디지털 설계 데이터 AIP는 다음과 같은 요소를 포함해야 한다.

- 디지털 콘텐츠 원본
- 포맷에 대한 표현 정보
- 데이터 무결성 및 진본성 검증 정보
- 접근 모듈에서 검색에 필요한 기술 메타데이터
- 보존 전략 및 출처 정보와 같은 관리 메타데이터
- 파일들 간의 관계성에 대한 구조 메타데이터 및 문서

2) 데이터 관리

데이터 관리 모듈은 관리자에 필요한 데이터베이스를 관리한다. 이를 위해 디지털 데이터에 관한 정보 즉, 메타데이터를 관리하는데, 메타데이터는 정보시스템을 구성하거나 특정 아이템을 검색하는데 사용된다. 만약 메타데이터가 효과적으로 사용될 필요가 있다면, 데이터를 적절하게 분류해야 한다.

3) 아카이브 저장소

아카이브 저장소는 디지털 콘텐츠 자체를 관리하는 레포지토리이다. 이 레포지토리는 다음과 같은 사항들을 반드시 따라야 한다.

- 디지털 콘텐츠 “집합소”
- 디지털 객체에 대한 메타데이터와 디지털 콘텐츠에 대해 ID로 관리
- 디지털 문서의 비트스트림 보존 보장
- 보존 기관에서 수립한 보존 전략에 따라 보존 절차 수행
- 파일 포맷, 버전, 관계된 보존 전략을 추적할 수 있도록 포맷 레지스트리와의 관계성 유지
- 적절한 데이터 백업 보장
- 오류 복구 기능 제공

4) 보존 계획

데이터 보존은 매우 복잡한 문제이다. 예전에는 종이 기반 보존 정책은 물리적 개체를 보존하는데 초점을 맞추었다. 그러나 디지털 데이터를 보존하기 위해서는 데이터를 저장하는 물리적 미디어를 보존하는 것부터 시작

된다. 디지털 보존은 장기간 후에도 저장된 데이터를 읽을 수 있도록 물리적 미디어를 복구시키는 것과 데이터를 손상시키거나 변경하지 않고 데이터에 접근하는 방법을 보장해야 한다.

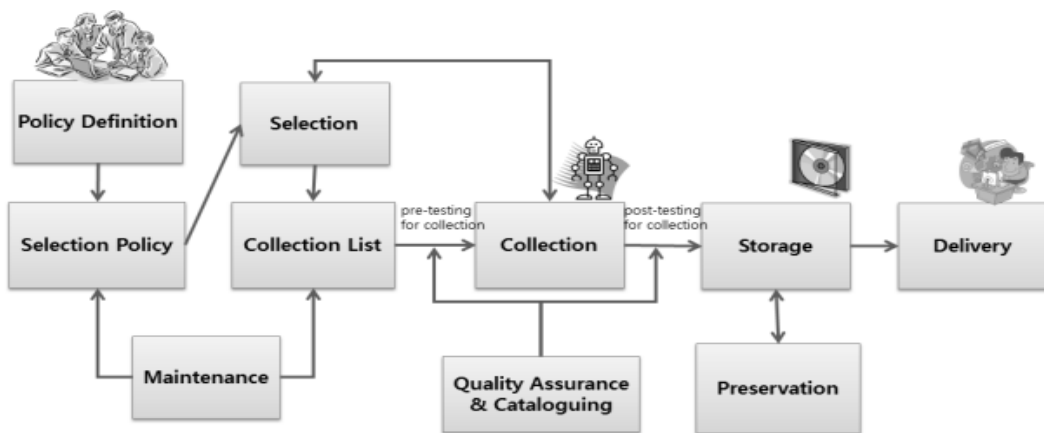
5) 접근

접근 모듈은 사용자가 기술 메타데이터를 이용하여 영구보존된 데이터를 검색할 수 있도록 하고 DIP로 전달한다. DIP는 사용자가 요구한 디지털 콘텐츠, 관련 메타데이터와 경우에 따라서 데이터와 상호작용하거나 보기 위한 방법도 제공한다.

4.4.2 수집대상 복합전자기록물

본 연구에서 개발한 복합자기기록물 아카이빙을 위해 대상되는 수집대상 복합 전자기록물들을 선별, 수집, 보존, 전달되는 워크플로우는 <Figure 10>과 같다.

매커니즘(워크플로우)은 국내외 아카이빙 사례조사 및 웹 기록물 아카이빙에 대해 조



<Figure10> The Workflow of the Collected Composite Electronic Records

사한 문서[8]를 바탕으로, 국내 실정에 맞추어 정의했다. 선별위원회에서 적절하게 정해진 선별 정책에 의해 선별된 사이트는 정해진 수집방법을 통해 수집된다. 수집시 품질 보증을 위해 전후로 사전 수집, 사후 수집 테스트를 시행한다. 이와 같은 품질 보증 및 목록에 의해 진본성을 유지하며, 일정한 매체에 저장되어 장기간의 보존된다. 저장되고 보존된 아카이빙은 전달 시스템을 통해 사용자가 아카이빙시의 모습 그대로를 확인할 수 있도록 제공한다.

6) 선별

복합전자기록물 아카이빙에 있어 적당한 선별정책을 결정하는 것은 반드시 선행되어야 한다. 이는 정보 자원이 구조적(Structural), 임시적(Temporal), 정보의 질적(Informational qualities)인 특징이 있기 때문이다. 선별정책의 결정 요소에는 수집 기관의 목표, 지적 재산권 관련, 기구 자원의 존재성 등이 있다.

선별 절차는 크게 선별정책(Selection policy), 선별(Selection), 유지(Maintenance)로 구분되어 진다. 선별정책에는 관련된 외부 선별 정책에 근거하여 정책을 확립되어야 하며 선별정책에서 잘 정의된 선별기준을 포함하여, 여러 개의 선별방법 중 적절한 선별방법을 정해야 한다. 선별은 복합전자기록물에 대하여 어느 부분까지 선별 할 것인지를 결정하는 선별영역의 정립이 필요하며, 언제 혹은 얼마나 자주 아카이빙 해야 하는 타이핑과 횟수에 대해 정해야 한다. 또한 선별과 수집이 계속 이루어질 경우 보편성을 보증하기 위해 선별정책과 수집리스트가 유지되어야 한다. 이는 사전 수집테스팅의 결과로부터 정제될

수 있다.

7) 수집방법 및 수집

수집방법(Collecting Methods)으로는 크게 콘텐츠-기반 수집(Content-driven collection)과 이벤트-기반 수집(Event-driven collection)으로 나눌 수 있다. 콘텐츠-기반 수집은 콘텐츠에 기반 한 즉, 웹 사이트의 기본적인 콘텐츠를 아카이빙하기 위한 방법이다. 이벤트-기반 수집(Event-driven collection)은 웹 서버와 브라우저 사이에 발생된 실제적인 트랜잭션(Transaction)을 처리하는 것이다.

8) 품질보증 및 목록화

품질보증(Quality assurance)은 복합전자기록물 아카이빙 절차에 필수적인 구성 요소이다. 품질보증의 본질 및 등급은 요구 사항과 수집하기 원하는 리소스에 많이 좌우된다. 품질보증은 크게 사전-수집 테스트(Pre-collection testing)과 사후-수집 테스트(Post-collection testing)이 있다. 사전-수집 테스트는 수집 전에 자료 분석과 테스트 수집을 통해 선별 수집방법이 잘 이루어지는지, 수집 인자에 수정할 사항이 있는지 확인하는 것이다. 사후-수집 테스트는 선별정책과 선별된 사이트의 리스트를 관리하는데 필수적인 단계이다. 두 개의 웹 브라우저 윈도우를 사용하여 현재 웹과 아카이브된 웹을 비교하는 테스트 스크립트를 바탕으로 테스트 한다. 다음의 기술들은 수집 기관의 능력을 크게 높여주는 역할을 한다.

목록화(Cataloguing)는 적절한 관리나 사용자들의 접근을 제공할 수 있게 하도록 아카이브된 컬렉션에 메타데이터와 같은 설명 정보를 포함시키는 것이다. 도서관이나 박물관

관의 경우 각자 고유한 분류방식을 가지고 있고 복합전자기록물에서 작성한 것들도 통합하고자하지만, 복합전자기록물의 특징에 맞지 않아 성공적으로 적용시키는 수 없다. 따라서 복합전자기록물 특징에 맞는 메타데이터의 설계가 필요하다.

9) 보존

보존(Preservation)의 목적은 대상의 가치가 유지되면서, 영속적인 접근을 보장하는 것이다. 즉, 성공적인 보존이란 사용자들의 접근이 가능하며 본래의 가치를 그래도 보존하여 전달해야 하는 것이다. 이러한 복합전자기록물의 보존은 디지털 대상의 보존과 유사하여 같은 기술이 적용된다. 하지만 이러한 보존은 아카이빙이 시작된 1996년에 이루어졌지만 여전히 초기 단계에 머물러 있다. 보존 전략으로는 원래 형식에서 보존된 대상에 접근할 수 있는 새로운 메소드를 개발하는 에뮬레이션(Emulation)과 기존의 대상을 현재 메소드를 통해 접근 가능하도록 새로운 형식으로 변환하는 마이그레이션(Migration)으로 정의된다.

특히 장기의 보존을 위해서는 설명정보의 메타데이터와 기술정보의 메타데이터의 저장에 기본적인 필수 사항이다. 이는 메타데이터가 복합전자기록물에 대해 이해하고 해석하는 방법을 제공하기 때문이다.

10) 전달

전달(Access)은 보존된 아카이빙된 복합전자기록물을 사용자가 사용할 수 있게 제공하는 것을 말한다. 아카이빙된 복합전자기록물에 대한 발견이나 확인에 있어 기본적인 메소드는 크게 검색 접근(Searching access)

과 열람 접근(Browsing access)이 있다.

검색 접근은 사용자에게 필수적인 툴을 제공하는 것으로 작가는 제목이나 URL의 검색부터 크게는 전체 컬렉션에 걸친 전문 검색(full-text search)을 제공한다.

열람 접근(Browse access)은 미리 정해진 계층구조로 자료들이 분류되어 있어 사용자들은 “drill-down” 식으로 접근 가능하다. 정해진 분류체계가 사용자의 예상과 같으면 매우 유용하지만, 사용자의 그룹이 많기 때문에 어려운 점이 있다.

5. 결 론

지금까지 이 논문에서는 복합전자기록물 영구보존과 관련하여 국내·외 관련 연구를 파악하였다. 국내에는 복합전자기록물 영구보존과 관련하여 표준 및 관련기술이 거의 전무한 실정이다. 이는 오래전부터 전자기록물과 관련하여 연구를 진행해온 호주, 미국, 캐나다, 영국의 경우도 같다. 이처럼 국내·외적으로 복합전자기록물 보존과 관련한 정책 및 표준, 관련기술은 시작 단계라고 볼 수 있다. 따라서 복합전자기록물 보존 정책 및 표준을 우리나라에서 먼저 제정함으로써 국제적 파급 효과를 가져 올 수 있을 것으로 판단된다.

이 논문에서 우리는 CAD 도면류 및 전자도서를 이관대상 복합전자기록물로, 소셜 네트워크 서비스 기록물은 수집대상 복합전자기록물로 분류한 뒤, 각각에 대한 보존 전략, 보존 포맷 및 보존 메타데이터를 제안하였다. 이관대상 복합전자기록물은 활용기관에서 보존 포맷으로 변환하여 이관하고, 이관 시 보

존 메타데이터도 함께 생성하여 이관하는 형태가 바람직하며, 수집대상 복합전자기록물의 메타데이터는 생산·활용 메타데이터가 같지만 보존 메타데이터는 생성하지 않으므로 보존 기관에서 수집된 복합전자기록물을 적당한 보존 포맷으로 변환하고, 필요한 보존 메타데이터를 추출하는 보존 전략이 필요하다.

각 복합 전자기록물의 보존 포맷 후보군은 아래와 같다.

- CAD/CAM 도면류 : KOSDIC, PDF/E, STEP
- SNS : 새로 정의
- 전자도서 : PDF, EPUB

복합전자기록물을 구성하는 각 전자기록물은 서로 다른 보존 포맷 및 메타데이터를 가지는데, 이들 각각을 특성을 모두 포함할 수 있는 보존 메타데이터가 필요하다. 이를 통해 이렇게 함으로써 전자기록물의 종류와 상관없이 동일한 형태의 보존 메타데이터를 갖게 되어 아카이빙 및 활용에 일관성을 기할 수 있다. 이를 만족하는 메타데이터 표준으로는 METS가 있다.

우리는 보존 워크플로우를 크게 이관대상 및 수집대상으로 나누어 살펴보았다. 이관대상이 되는 복합전자기록물에 대한 보존 워크플로우는 인수정책에 따라 이관된다. 이관대상 복합전자기록물은 생산 단계부터 보존과 관계된 메타데이터를 생산/관리가 용이하나, 수집대상 복합전자기록물은 그렇지 않다. 따라서, 이에 대한 고려사항이 필요하다. 수집대상 복합전자기록물의 경우에는 웹 아카이빙 프로세스와 유사할 것이다. 이들은 국가기록원의 선별정책에 따라 각 페이지 및 리소스를 수집하고, 수집목록을 작성한다. 수집된

리소스들은 보존 포맷으로 규격화하여 저장되고 보존되는 일련의 과정을 거친다. 이 두 가지 복합전자기록물에서 모든 메타데이터를 자동 추출하는 것은 어렵다. 이는 SNS 서비스가 가지는 모든 콘텐츠 소유권이 동일하지 않고, 그 리소스의 출처 또한 다양하기 때문이다. 따라서 자동 추출 가능한 메타데이터는 보존 포맷 생성일자, 생성자, 원 리소스 생성자 정보 등이 있을 것으로 판단된다.

디지털 보존 환경은 다음과 같이 정의하였다. 생성 단계는 새로운 정보가 전달되는 초기 단계에 해당하며, 조합 단계는 아카이빙 관련 객체를 평가하는 단계로써, 향후 보내게 될 정보를 완전하게 컴파일하기 위한 모든 프로세스를 처리한다. 이는 관련 커뮤니티의 심도 있는 회의에 의해 진행되는데, 영구 보존에 필요한 객체를 추출하고 후에 재사용될 수 있는 객체를 선별해 내기 위해서는 관련 지식이 풍부한 전문가들이 회의를 진행해야 한다. 채택 단계는 재사용을 위해 보존 단계에서 제공하는 정보들을 표시, 검사, 채택, 통합하는 모든 과정을 수행한다. 이 단계에서는 데이터 재사용을 위해 필요한 모든 프로세스를 수행하는데, 전달, 조합, 맥락화 등이 이에 해당한다. 재사용 단계는 소비자나 다른 프로세스에서 필요한 정보를 사용하는 단계다.

References

- [1] A Collaborative Initiative of The Library of Congress, Preserving Your Digital Memories.

- [2] Antonio, G., Alessio S., "The Index- able Web is more than 11.5 billion pages" [cited 2008-11-17> 2005.
<<http://www.cs.uiowa.edu/~asignori/web-size/>>.
- [3] ArchiveFacebook, <https://addons.mozilla.org/ko/firefox/addon/archivefacebook/>
- [4] Archive-It, <http://www.archive-it.org/>.
- [5] Archivist Desktop, <http://visitmix.com/work/archivist-desktop/>.
- [6] Brocks, H., Kranstedt A., Jaschke, G., Hemmje, M., "Modeling Context for Di- gital Preservation," Smart Information and Knowledge Management : Advances, Challenges, and Critical Issues(Springer Studies in Computational Intelligence), 2010.
- [7] Center for Research Libraries, Trust- worthy Repositories Audit and Certifi- cation(TRAC) Criteria and Checklist. Chicago, 2008.
- [8] Cha, S. J., Choi, Y. J., Lee, K. C., "Metadata Design for Archiving Public Deep Web Records," The Journal of Society for e-Business Studies, Vol. 14, No. 4, pp. 181-193, 2009.
- [9] Consultative Committee on Space Data Systems, Reference model for an open ar- chival information system, ISO 14721 : 2003.
- [10] Digital Library Federation, METS, Ver- sion 1.6 revised, 2010.
- [11] FacebookArchive, <http://www.whois mind.com/whois/facebookarchive.com.html>.
- [12] Goncalo, A., Jose Bar, Jose Bor, "A Re- ference Architecture for Digital Preser- vation," Austrian Computer Society, 2010.
- [13] International Standard Organization. ISO 15489-1 : 2001(E) Information and doc- umentation : Records Management. Part 1 : General, Geneva : ISO, 2001.
- [14] Jane, H., Sharmin C., "PANIC-An Inte- grated Approach to the Preservation of Composite Digital Objects using Semantic Web Services," International Journal on Digital Libraries, Vol. 6, No. 2, pp. 174-183, 2006.
- [15] Johan, S., "The NEDLIB Guidelines," Bureau D'Arts, Amsterdam, 2000.
- [16] Kevin, Bradley, Deborah Wood, "Preser- vation Metadata for Digital Collection," Metadata for long term preservation in NEDLIB, 2000.
- [17] Korean Institute of Construction Tech- nology, Korea Standard of Drawing Infor- mation in Construction Vol. 1, No. 2, 2011.
- [18] Kim, K. O., Kim, S. H., Lim, S. B., Choy, Y. C., "A Comparison of Electronic book metadata formats and Development of Electronic Book of Korea Standard meta- data," A Comparison of Electronic book metadata formats and Development of Electronic Book of Korea Standard meta- data, International Conference CALS/EC KOREA, pp. 511-521, 2001.
- [19] Laura, E. C., Beth D., "Digital Preservation : The Twitter Archives and NDIIPP."

- [20] LOTAR, Project, <http://www.lotar-international.org/>.
- [21] Ministry of Security and Public Administration, National Information Society Agency, 2003~2007 e-Government White Paper, 2008.
- [22] Ministry of Security and Public Administration, Act on the Management of Public Archives, 2012.
- [23] National Archives of Korea, Research and Development for Web Archiving Technology and Experimentation, 2008.
- [24] National Archives of Korea, NAK/TS 3 : 2008(v1.0) Standard of Archival Information Package, 2008.
- [25] National Archives of Korea, NAK/TS 2. : 2008(v1.0) Digital Document File Format for Long-term Preservation, 2008.
- [26] Park, B. J., Cha, S. J., Lee, K. C., "Extension of the Longterm Archival Information Package for Electronic Records to Accommodate Web Records," The Journal of Society for e-Business Studies, Vol. 15, No. 4, pp. 33-47, 2010.
- [27] Policy, <http://www.facebook.com/hipolicy>.
- [28] Portico, Portico Content Type Action Plan : E-Book Content, Vol. 1, No. 1, 2009.
- [29] Portico, Portico Content Type Action Plan : E-Journals, Vol. 1, No. 1, 2009.
- [30] Public Record Office Victoria, Export of Electronic Records to PROV PROS 99/007(Version2) Specification 5, 2006.
- [31] Kaitlin, L., Jason P., "Archiving Scholars' Tweets," Society of American Archivists, 2011.
- [32] SocialSafe, <http://socialsafe.net/>.
- [33] The Archival Community Digitization Program, <http://www.cdncouncilarchives.ca/ACDP.html>.
- [34] The Art Institute of Chicago Department of Architecture, "Collecting, Archiving and Exhibiting Digital Design Data Section2 : Archiving Digital Design Data : Practices and Technology."
- [35] The National Archives, Digital Preservation Guidance Note : 4 Graphics File Formats, 2008.
- [36] TwapperKeeper, <http://twapperkeeper.com/index.html>.
- [37] TweetDoc, <http://www.tweetdoc.org/>.
- [38] TweetScan, <http://tweetscan.com/>.
- [39] Wikipedia, Comparison of e-book formats, http://en.wikipedia.org/wiki/Comparison_of_e-book_formats.

저 자 소 개



황윤영
 2002년
 2004년
 2011년
 2009년~2011년
 2011년~2012년
 2012년~현재
 관심분야

(E-mail : yyhwang@kisti.re.kr)
 충남대학교 컴퓨터공학 (학사)
 충남대학교 컴퓨터공학 (석사)
 충남대학교 컴퓨터공학 (박사)
 충남대학교 소프트웨어연구소 전임연구원
 충남대학교 BK21 박사후연구원
 한국과학기술정보연구원 연구원
 아카이빙, 유비쿼터스, 소셜 서비스 네트워크, 데이터베이스
 정제



이규철
 1984년
 1986년
 1990년
 1989년~현재
 2000년~현재
 2003년~현재
 2005년~현재
 2006년~2012년
 2009년~현재
 2009년~현재
 관심분야

(E-mail : klee@cnu.ac.kr)
 서울대학교 컴퓨터공학 (학사)
 서울대학교 컴퓨터공학 (석사)
 서울대학교 컴퓨터공학 (박사)
 충남대학교 컴퓨터공학과 교수
 한국전자거래학회 이사
 한국기록관리학회 이사
 충남대학교 소프트웨어연구소 소장
 대통령자문 정부혁신 지방분권위원회 의원
 웹코리아 포럼 의장
 KISTI 정보화사업 평가위원회 위원
 데이터베이스, XML, 정보통합, 아카이빙, 사물인터넷(IoT)