PREMIS 데이터모델 적용을 위한 사무문서 컨텐츠모형 설계 연구*

A Study on Contents Model for Business Records by the Application of the PREMIS Data Model

문주영(Juyoung Moon)**
김태수(Taesoo Kim)***

초 록

본 연구에서는 OAIS 참조 모형을 구체적으로 발전시킨, 사실상의 보존 메타데이터 표준인 PREMIS 데이터모델과 데이터사전을 사무문서에 적용하기 위한 사무문서 컨텐츠모형을 개발하였다. 대상 문서는 'A사 B국 해외 석유사업 및 유전개발 문서'로 국가 차원 이상의 영구 보존 가치를 지니는 문서 컬렉션이다. PREMIS 데이터모델을 사무문서에 구체적으로 적용하기 위하여 PREMIS 모델 내의 지적개체에 대한 문서 차원의 개념 정립과 이해를 시도하였다. 즉, 문서 컨텐츠의 계층을 구분하는 원칙과 구조를 설계하였고 그에 맞추어 사무문서 컨텐츠를 대상으로 한 계층 모형을 만들어 사무문서 컨텐츠모형을 도출하였다. 이 과정에서 기록물 기술 규칙을 준수하였다.

ABSTRACT

This study presents a contents model designed for business records that require long-term preservation. The contents model is based on the PREMIS(Preservation Metadata: Implementation Strategies) data model and the ISAD(G)(General International Standard Archival Description). The study selected the record collection of "the records of the overseas petroleum business and oil field development of A company located in B country." This collection requires permanent preservations by the nation and even beyond. It was attempted to establish the concepts of intellectual objects in the PREMIS data model to apply the PREMIS data model to the business records specifically. In other words, the study established the principles for differentiation of the classes in the record contents and the hierarchy structure, and the hierarchy model was developed for business records contents to derive the business records model based on those principles.

키워드: 보존메타데이터, 데이터모델, 사무문서, 기업문서, 문서관리
PREMIS preservation metadata, business records, archive, archiving, records management

---

* 이 논문은 연세대학교 박사학위 논문의 일부를 요약한 것임.

** 숭의여자대학 비서행정과 전임강사(jymoonglory@hanmail.net) (제1저자)
*** 연세대학교 문헌정보학과 교수(btree@yonsei.ac.kr) (공동저자)

■ 논문접수일자 : 2011년 2월 15일 ■ 최초심사일자 : 2011년 2월 23일 ■ 게재확정일자 : 2011년 3월 10일
■ 정보관리학회지, 28(1): 43-68, 2011. 〔DOI:10.3743/KOSIM.2011.28.1.043〕

1. 서 론

문서의 중요성은 문서가 담고 있는 내용이 중요하기 때문이기도 하지만 특히 사무문서(business records)의 경우 조직이나 기관의 업무수행의 산물로서 활동의 결과이며 법적인 증빙자료가 된다는 점에서 더욱 중요하다고 할 수 있다. 따라서 문서는 조직의 과거와 현재와 미래를 담고 있는 것이며 종류에 따라서 그 조직이 유지되는 한, 또는 그 사회가 유지되는 한 영구적으로 보존해야하는 가치를 지니기도 한다. 더욱이, 사무문서는 그 사회의 구성원들의 삶과 경제·사회적으로 더 밀접히 관련된 경우가 많으므로 장기 보존 가치가 있는 사무문서의 보존에 실패하였을 때 오히려 그 사회의 구성원들에게 더 직접적이고 막대한 피해를 미칠 수도 있을 것이다.

공공부문에서는 2005년 이후 기록관리 혁신 정책이 빠르게 추진되면서 표준화의 중요성이 부각되었다. 특별히 디지털 문서의 보존에 대한 문제의식과 연구가 활발하여 OAIS 참조모형의 준수, 보존메타데이터의 활용 등에 대한 연구가 계속되고 있다. 그러나 이러한 연구는 모두 공공부문의 정부 문서나 대학의 학술연구논문, 연구기관의 데이터세트, 역사적 가치가 있는 사진 이미지 등을 주 대상으로 연구되고 있으며 기업의 사무문서 자체를 대상으로 한 연구는 거의 찾아볼 수 없다. 지금까지 기업 기록관리가 기록관리 학문의 한 연구 영역으로 연구되고 있긴 하였지만 주로 기업의 역사를 중심으로 하는 기업사 연구 측면에서 사료관리 체계 구축이나 관리 방법론 위주로 몇몇 연구가 진행되었을 뿐이다.

전자문서 보존은 미래를 담보로 하는 모험일 수 있다. 문서의 장기적 관리와 보존이란 미래에도 문서의 재구성이 가능하고 활용이 가능하도록 보존하는 것을 의미하는데 아무도 미래에 가보지 않는 이상 보존이 성공적으로 수행되었는지의 여부를 미리 검증할 길은 없다. 따라서 오랜 기간이 지난 후에도 그 목적을 성취할 수 있음을 가능한 한 보장받을 수 있는 기제가 필요한데 바로 이점이 전자문서 보존에 있어서 표준의 정립과 준수가 그 무엇보다도 중요한 이유이다.

전자문서 아카이빙 관련 표준으로 대표적인 것은 OAIS 참조모형과 보존메타데이터이다. OAIS 참조 모형은 디지털 보존에 관련된 모든 활동 및 프로젝트의 개념적 기반으로 디지털 보존 분야의 기반 표준의 위상을 차지해 왔다. 즉 2002년 OAIS 참조모형이 ISO의 장기보존을 위한 디지털 아카이빙 시스템 표준으로 채택되면서 분야와 대상 자료 형태는 다르지만 'OAIS 준수'를 표방하는 많은 아카이빙 시스템들이 연구 되었다(Van de Sompel et al. 2005; Caplan 2004; Sawyer et al 2002; Pinna et al. 2005). 그러나 최근의 연구는 추상적으로 개념적인 틀을 제시하는 OAIS 참조모형에 대해서 'OAIS 참조모형 준수'(OAIS compliance)가 과연 무엇을 의미하는 가에 대한 논란과 함께 OAIS 준수에 따른 다양한 견해가 존재한다는 보고서도 보고되었다(PREMIS Working Group 2004). 이에 따라 신뢰할 수 있는 디지털 아카이브 인증기준에 대한 연구와 함께 OAIS를 준수하는 PREMIS와 같은 실질적인 표준을 중심으로 디지털 아카이빙을 시도하는 연구 흐름을 보이고 있다.

PREMIS 데이터모델과 사전은 OAIS 기능모델과 정보모델을 모두 만족하도록 개발되어(OCLC/RLG 2002) 사실상의 보존메타데이터 표준으로 사용되고 있는 보존 메타데이터이다. OAIS 참조모형이 국제표준으로서 가장 기본적이고도 개념적인 틀을 제시했다면 PREMIS는 이를 실제로 적용할 수 있도록 구체적으로 발전시킨 표준이라고 할 수 있다.

사무문서는 정부기관이든 일반기업이든 전 분야, 전 조직에 걸쳐 업무과정에서 필연적으로 발생되는 모든 종류의 문서 또는 서류를 통칭한다. 이러한 사무문서 중에서 본 연구에서는 기업 사무문서를 중심으로 연구를 진행하였으며 구체적인 대상 문서는 특정 기업이 해외에서 개발 중인 유전과 관련된 문서로 상대방 국가와의 유전 계약서 등이 포함된, 한 기업체의 범위를 넘어서 전 국가 차원의 보존 가치를 지닌 중요한 문서 컬렉션이다.

PREMIS 데이터모델을 사무문서에 적용하기 위해서는 PREMIS 모델 내의 지적개체에 대한 문서차원의 개념 정립과 이해를 필요로 한다. 이는 PREMIS가 '문서'만을 위한 데이터모델로 개발된 것이 아니라 보존을 필요로 하는 모든 디지털 개체를 위한 데이터 모델이기 때문이다. 또한 PREMIS는 데이터 모델의 구현을 위해 따라야 하는 어떤 실질적인 구축상의 지침이나 최소한의 규정도 제시하지 않고 있어 지적개체를 문서에서 어떻게 적용할 것인가를 먼저 고려해야 한다.

PREMIS 데이터모델의 5가지 개체 중에서 문서 개체와 직접적으로 관련 있는 개체는 지적개체와 객체개체(object entity)이다. 이 중 객체 개체는 file, bitstream, representation의 3가지 하위 유형을 가지는 개체로 문서에 직접 적용할 수 있다. 그러나 PREMIS는 지적개체를 관리(management)와 기술(description)을 위한 하나의 지적단위(a single intellectual unit)로 정의하며, 그 예로 지도(map), 사진, 데이터베이스, 웹사이트, 웹페이지 등을 들고 있는데(PREMIS Editorial Committee 2008), 이를 문서에 적용하기 위해서는 문서 컨텐츠에 대한 이해와 분석이 선행되어야 한다. 즉, 문서에 있어서 무엇을 '관리와 기술을 위한 하나의 지적단위'로 볼 것인가의 문제와 문서의 집합적인 성격을 어떻게 감안할 것인지에 대한 논의가 필요하다. 또한 문서관리 메타데이터 체계에서는 문서 계층에 따라 적용되는 메타데이터 요소가 다르다는 점도 고려하여야 한다. 이러한 문제를 해결하기 위해 본 연구에서는 기록물 기술규칙을 적용하여 사무문서 컨텐츠모형을 설계 하였다.

문서의 계층을 구분하는 원칙과 계층 구조를 먼저 정립하고 그에 맞추어 사무문서 컨텐츠를 대상으로 한 계층 모형을 만들었다. 이를 '사무문서 컨텐츠모형 설계' 작업으로 정의하였으며, 먼저 기술할 계층 구조를 정하고 대상 문서 컬렉션을 대상으로 기술 계층 구조대로 기술할 개체를 분석하고 구조를 모델링하였다. 이 과정에서 기록물 기술 규칙이 준수되도록 하였다.

2. 디지털 사무문서의 특성과 보존 필요성

사무문서라 함은 고유 업무 및 일반적인 업무에서 발생하는 문서, 예를 들면 대외문서, 보

고서, 결재서류, 각종 대장 등의 행정 업무상 발생하는 일반 서류들을 말한다(심재균 1998). 이들은 공기관이든 사기업이든 전 분야, 전 조직에 걸쳐 거의 필연적으로 발생되는 문서이며, 대규모 조직인 경우에는 문서 관리 규정을 두어 발생에서부터 유통(주로 결재), 배포, 보관, 이관, 폐기 등에 이르는 문서의 생명주기(life cycle) 및 보관 관리 체계를 정의하고 있다. 어떤 조직에서건 발생하는 문서들이 중요하지 않은 곳이 없으며, 특히 결재 과정을 거치는 문서들은 여러 사람의 지적 행위와 의지가 들어 있는 주요 문서이고 필수 관리 대상이 된다. 보통 문서를 작성 시기나 이용 빈도 등을 고려하여 현용문서(current record), 준현용문서(semi-current record), 비현용문서(non-current record) 또는 활용문서, 준활용문서, 비활용문서 등으로 나누어 구분 하는데 본 연구의 대상은 준현용문서 이상, 특별히 보존 필요성이 높은 문서로 한정하였다.

사무문서의 특성은 첫째, 기업체나 조직체, 정부 등의 어떤 명백한 행정적인, 법적인, 업무상의 또는 기타의 사회적인 목적을 달성하기 위해서, 조직의 기능적인 활동의 과정과 직접적인 연관성 속에서 생산된다. 둘째, 업무과정의 진행에 따라 이와 관계된 문서는 자연적이고 연속적이며 유기적으로 생산되고 축적된다. 따라서 문서는 내적 외적으로 어떤 조직적인 체계가 존재하게 된다. 셋째, 사무문서는 조직이나 기관의 업무수행의 산물로서 법적인 효과를 지닌다. 넷째, 사무문서는 대량으로 인쇄되어 생산되는 도서와는 달리 전문 업무나 일반 행정업무 수행과정에서 생산 접수된 유일한 원본이다(이원영 2000). 따라서 사무문서는 조직의 과거와 현재와 미래를 담고 있는 것이며 종류에 따라서 그 조직이 유지되는 한, 또는 그 사회가 유지되는 한 영구적으로 보존해야 하는 가치를 지니기도 한다.

불과 십 몇 년 전에 전 세계는 정보의 디지털화라는 큰 변화의 시기를 맞이하였다. 기업 환경에서 중요한 화두 중 하나는 페이퍼리스 오피스(paperless office)였고, 종이문서 시대로부터 전자문서 시대로 발 빠르게 전환되었다. 종이 문서의 경우 생성과 동시에 물리적으로 현실에 존재하는 것이기에 원본성과 표현성(look & feel)을 가지며 다른 복사본 혹은 유사한 내용을 담은 문서와 차별화 된다. 하지만 전자 문서는 문서를 생성하였을지라도 물리적으로 존재하는 것이 아니라 전자적 조합으로만 존재하기 때문에 보존성과 표현성을 가지고 있다고 볼 수 없다. 다만 전자문서의 디지털 코드를 보관할 수 있는 매체와 이를 표현할 수 있는 소프트웨어가 결합되었을 때에 사람이 인식하고 확인할 수 있는 형태가 되는 것이다. 이에 따라 문서에 포함되어 있는 정보가 특정 회사 소프트웨어에 종속되며 그에 따라 소프트웨어 개발사의 부도, 제품의 단종 등의 이유로 그동안 생산된 문서를 활용하지 못하는 경우도 흔하다. 또한 폐쇄된 문서 포맷은 문서의 유통 및 활용에 제한을 가져오게 되고 특정한 폰트나 특정 표현 방식도 시대가 지나면서 유지되지 못하게 된다(장윤기 2007). 또한 사무문서는 거래처, 고객, 사업파트너, 정부, 내부문서 등 다양한 입수경로와 형태를 지닌 다양한 문서로 구성되고 관리해야 할 문서의 양이 엄청남에도 불구하고 세계적으로 인정된 표준시스템이나 관리방식에 대한 제시가 없어 혼란스러운 상태

이다(Phillips 2009). 특히 사무용 소프트웨어의 경우 특정 업체의 기술 의존성이 매우 높으며 시장이 광범위하여 프로그램의 등장과 쇠퇴 속도도 매우 빠르다. 이러한 사무용 소프트웨어로 작성된 디지털 사무문서의 양은 기하급수적으로 늘고 있고 소프트웨어 기술의 한세대라고 볼 수 있는 기간의 단축 속도는 점점 빨라지고 있다. 따라서 지금 현재 작성되고 있는 문서를 10년 후에도 그대로 읽을 수 있다고 아무도 장담할 수 없다.

사무문서 영역에서 재난·위험관리 차원으로 실질적인 경각심이 들기 시작한 시점은 2001년 미국 세계무역센터 테러사건 이후이다. 이러한 위험인식의 기류와 함께 디지털문서 보존에 관해서도 우려의 시각이 현실적으로 인식되기 시작하였다. 그러나 기업의 생존과 발전에 제1전략을 둘 수밖에 없는 기업에서는 문서의 장기적 보존이나 관리 보다는 현용 문서의 검색과 문서 지식 활용 중심의 시스템적 접근에 집중하고 있는 실정이다(Andolsen 2008). 대부분은 현용문서의 관리 차원으로 대기업 중심의 개별 기업들이 문서관리 시스템(EDMS)을 도입하여 운영하기도 하지만 전사적 차원에서 체계적인 문서관리 비전을 수립하고 표준을 준수하여 추진한다기 보다는 기술적인 관점과 프로그램 차원의 접근을 하는 경우가 많다.

국가적으로는 이미 '기록관리의 과학화' 바람이 불었고 정부기관을 대상으로 기록관리 혁신 정책이 빠르게 추진되었다. 특히 전자문서의 관리와 보존의 필요성이 역설되고 여러 표준의 수용과 보존 방안이 수립되었다. 그러나 이러한 움직임은 일차적으로 국가기관의 업무결과로 생산된 공적문서, 정부문서 등을 중심으로 체계가 잡혀가고 있으며 학계의 주요 연구대상도 공문서 관리 및 보존이나 역사문서 보존 등에 집중되고 있는 실정이다. 이에 따라 OAIS 참조모형 및 기록관리 표준의 준수, 보존메타데이터의 활용 등에 대한 연구 대상에서도 사무문서 영역은 제외되어 왔다. 그러나 이제 사무문서 관리 및 보존 영역에도 장기적 관점의 투자와 연구 노력이 필요한 시기가 도래하였다. 행정영역 뿐만 아니라 기업 영역의 문서관리에도 과학화가 이루어져야 진정한 의미의 기록 선진국이 될 수 있을 것이다.

3. PREMIS 보존 메타데이터

보존 메타데이터는 아카이브의 디지털 보존 과정을 지원하는 정보로, 보존 환경에서 디지털 자원의 장기간의 영속성(viability), 재현성(renderability), 이해가능성(understandability), 진본성(authenticity), 일치성(identity)을 유지하는 핵심적인 기능을 한다(OCLC/RLG 2005). 이러한 보존 메타데이터 요소 세트는 1990년대 후반 이후 여러 기록 선진국들의 노력으로 각국의 국가 표준들로 나오게 되다가 2002년에 개발된 OCLC/RLG 보존메타데이터가 이전에 나온 여러 보존메타데이터 표준들을 분석하고 아우르는 요소 세트를 개발함으로써 사실상의 표준으로 사용되게 되었다. 이후 OCLC/RLG 보존메타데이터는 PREMIS 데이터모델과 데이터사전으로 이어져 현재는 PREMIS가 보존메타데이터의 표준으로 확실한 자리매김을 하고 있다고 할 수 있다. 2005년 버전1.0이 발표된 PREMIS 데이터모델과 사전은 OAIS 참조

모형의 기능모델과 정보모델을 모두 만족하도록 개발되었다(OCLC/RLG 2005). 이어 2008년 3월 PREMIS 데이터 사전 version 2.0과 수정된 데이터모델이 발표되었다. OAIS 참조모형이 국제표준으로서 가장 기본적이고도 개념적인 틀을 제시했다면 PREMIS는 이를 실제로 적용할 수 있도록 구체적으로 발전시킨 표준이라고 할 수 있다.

OAIS 참조모형은 디지털 보존에 관련된 모든 활동 및 프로젝트의 개념적 기반으로 디지털 보존 분야의 기반 표준의 위상을 차지해 왔다. 그러나 이는 순수한 참조모형일 뿐 구현을 위한 실질적 안내자로서의 역할은 할 수 없다. 그러므로 OAIS 참조모형에서 제공하는 보존 메타데이터 요구사항에 따른 메타데이터를 구현하기 위해서 적용할 수 있는, 좀 더 구체화된 표준이 필요하다. 따라서 본 연구에서는 사무문서 아카이빙 구축을 위한 보존 메타데이터 설계를 위해 PREMIS 데이터모델을 사용하였다. 문서 생산에 관여한 개체가 무엇인지, 개별 개체의 내부가 어떻게 구성되는지, 그리고 개체 간에 어떠한 관계가 형성되는지를 나타내는 데이터모델은 문서관리 메타데이터 설계 시 중요하고 필수적인 도구라 할 수 있다.

3.1 데이터모델

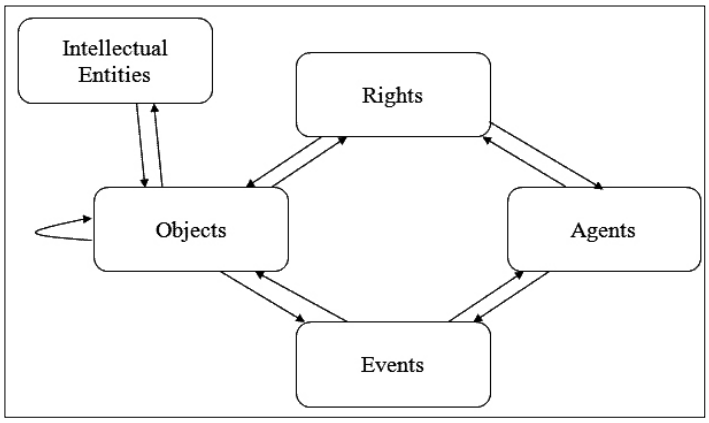
PREMIS는 OAIS 참조 모형의 정보 모형을 검토하고 OAIS의 개념 구조와 보존 메타데이터 요소와의 매핑을 통하여 보존 메타데이터 요소를 개발하였다. 특이할 만한 것은 PREMIS가 기존에 일반적으로 사용되던 '메타데이터 요소'(metadata element)라는 용어 대신에 '의미단위'라는 용어를 사용하고 있다는 점인데 이는 PREMIS의 '구현에 있어서는 독립적인'(implementation independent)입장 때문이다. 즉, 의미단위의 의미는 '정보나 지식의 작은 조각'인 것에 반해 메타데이터 요소는 '메타데이터 스키마나 데이터베이스 등의 방법으로 어떤 정보를 표현하도록 정의된 방식'이라 정의할 수 있다. 따라서 PREMIS는 이와 같이 메타데이터의 구현에 있어서 어떤 특정 방식을 특정화하지 않으며 단지 리포지터리가 보존을 위해 알아야 할 요소들을 정의할 뿐(PREMIS Editorial Committee 2008)으로 자신들의 임무를 한정하고 있기 때문에 메타데이터 요소라는 단어를 사용하지 않았다.

PREMIS는 메타데이터 요소의 논리적 조직화를 쉽게 하기 위해 〈그림 1〉과 같은 5개의 개체(entity)로 이루어진 데이터모델을 제시하였다.

이 5개의 개체는 지적개체, 객체개체, 이벤트개체, 행위주체개체, 권한개체로 구성되며 각각의 개체들은 서로 관계를 갖게 된다. 이 관계를 이어주기 위해 각 개체들은 식별자(identifier)를 가지게 되며 〈그림 1〉 상의 화살표와 같은 관계 방향을 갖게 된다.

3.1.1 지적개체

5개 개체 유형 중에서 지적개체는 개념적인 개체(conceptual entity)로 FRBR(Functional Requirements for Bibliographic Records) 모형에서의 저작(work)이나 표현형(expression)에 해당하는 개념이다(Wendler 2006). FRBR 모형에서 저작은 지적으로나 예술적으로 독창성을 지닌 창작물이며 표현형은 하나의 저작을



〈그림 1〉 PREMIS 데이터모델

(출처: PREMIS Editorial Committee 2008)

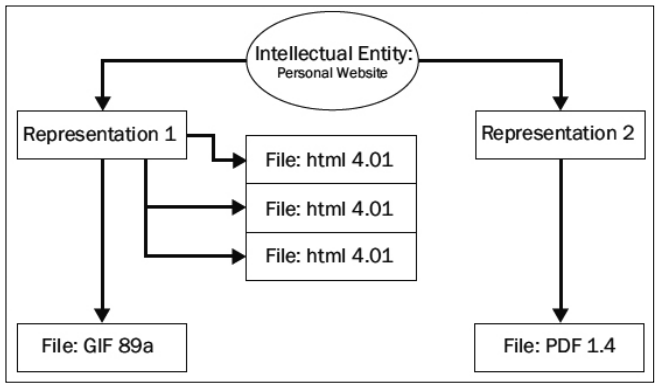
지적, 예술적으로 실현한 것을 말하는데(IFLA Study Group on the FRBR 2003) 이 두 개체는 모두 추상적인 것이다. 이처럼 PREMIS 데이터사전의 지적개체는 디지털 개체에 있어서 의미를 지닌 하나의 지적단위 또는 그것을 실현한 추상적이고 개념적인 단위가 된다. 즉, PREMIS는 지적개체를 '관리(management)와 기술(description)을 위한 하나의 지적단위(a single intellectual unit)'로 정의하며 그 예로 지도, 사진, 데이터베이스, 웹사이트, 웹페이지 등을 들고 있다(PREMIS Editorial Committee 2008). 지적 개체는 또 다른 지적개체를 포함할 수 있으며(예: 웹페이지 내에는 여러 개의 웹사이트들을 가질 수 있다), 여러 개의 디지털 표현(representation)을 가질 수 있다(하나의 웹 페이지 내에는 JPEG 이미지가 포함된 하나의 html파일과 그것을 pdf화한 PDF파일이 있을 수 있다). 지적개체는 기술 메타데이터(descriptive metadata)로 기술하도록 되어 있으므로 보존 메타데이터용 PREMIS 데이터 사전에서는 설명이 제외되었다. 이는 대상 자료의 성격에 따라 기술 메타데이터가 상이하며 각각의 해당 메타데이터에서 용법을 상세하게 다루고 있을 것이기 때문이다(PREMIS Editorial Committee 2008).

3.1.2 객체개체

객체개체는 '디지털 형태로된 정보의 개별 단위'(a discrete unit of information in digital form)로 정의되며 표현(Representation), 파일, 비트스트림의 3개의 서브타입 유형을 가진다. 본 연구의 설계에서는 이 3가지 유형 중에서 표현과 파일 단계의 객체에 대해서만 다루었으며 비트스트림 단계는 제외하였다.

1) 표현(Representation)

표현은 '특정한 지적개체를 완벽하게 잘 재현하기 위한 구조적 메타데이터를 포함한 파일들의 집합'으로 정의된다(PREMIS Editorial Committee 2008). 〈그림 2〉는 Paradigm Project



〈그림 2〉 지적개체와 표현, 파일의 관계

(출처: Paradigm Project 2005)

(2005)에서 설명한 표현의 예이다. 지적개체인 하나의 개인 웹사이트는 html 4.01 버전으로 작성된 3개의 html파일과, GIF 89a 버전으로 저장된 하나의 gif 파일로 이루어진 표현1과 장기 보존 등의 이유로 동일한 웹페이지가 PDF 버전 1.4로 저장된 표현2 두 개로 구성되어 질 수 있다. 이 두 가지 표현은 서로 다른 보임과 느낌(look and feel)을 주겠지만 근본적으로는 같은 정보와 같은 내용을 다루고 있다.

이런 방식으로 하나의 지적개체를 표현과 파일로 나누어서 생각할 때의 장점은 리포지터리가 관련 있는 메타데이터의 중복을 최소화하고 한꺼번에 모아서 관리 및 저장할 수 있다는 점에 있다(Paradigm Project 2005).

2) 파일(File)

컴퓨터 운영 체제의 파일 시스템에 의해 하나의 파일로 인식되는 것을 말하는데 그 예를 다음과 같이 들 수 있다.

- Portable Document Format 1.4 파일
- Windows용 WordPerfect 5.1 파일
- Macintosh용 WordPerfect 파일
- GIF(Graphics Interchange Format) 1989a 파일

3) 비트스트림(Bitstream)

비트스트림은 파일 내에 존재하는 연속적이거나 비연속적 데이터를 의미한다. 비트스트림은 단독적인 파일로는 전환될 수 없는데 예를 들면 다음과 같다.

- 하나의 Zip 파일 내에 압축된 형식으로 존재하는 PDF 1.4 파일
- TIFF 파일포맷의 문서 내에 삽입된 이미지 파일
- Microsoft Word 파일 내에 삽입된 이미지 파일

3.1.3 행위주체 개체

행위주체 개체는 데이터 객체의 생애주기에서 보존과 관련된 활동과 권한에 관계된 행위를 하는 주체에 관한 정보로, 개인뿐만 아니라 조직, 컴퓨터 소프트웨어 등도 행위주체에 포함된다.

3.1.4 이벤트 개체

이벤트 개체는 하나 이상의 객체나 행위주체를 포함하는 활동에 관한 정보이다. 이벤트에 관한 메타데이터는 디지털 객체와 분리되어 기록되고 저장되며, 객체를 수정하는 추가 활동 등도 이벤트로 기록된다. 이벤트 개체의 요소는 이벤트가 발생하는 시기마다 그 활동에 관한 정보인 메타데이터로 획득되어야 한다.

3.1.5 권한 개체

권한 개체는 저작권이나 지적 재산권법 또는 협약에 의해 행위주체에 부여된 권한에 대한 정보이다. 권한 정보는 아카이브에서 소장하고 있는 객체와 관련하여 아카이브가 수행할 수 있는 권한과 허가를 받은 것이 무엇인가에 대한 정보라 할 수 있다.

3.2 데이터 사전

PREMIS 데이터 사전은 '디지털 보존을 지원하기 위해 보존 리포지터리가 알아야만 하는 요소를 밝히기 위해서'라는 목표로 2005년 2월 버전 1.0이 출시되었고 2008년 3월 버전 2.0이 출시되었으며 최상위 13개의 요소와 하위요소를 가진다.

PREMIS 데이터 사전은 '기술메타데이터'(descriptive metadata) 요소와 '특정 포맷 중심의 기술적인 메타데이터'(technical metadata), '행위주체와 권한, 미디어나 하드웨어 등과 관련된 세부적인 요소', '리포지터리를 위한 운영 규칙' 등이 제외되고 발표되었다(Caplan 2009). 따라서 PREMIS 데이터 사전은 거의 대부분이 객체 개체에 대한 의미단위의 기술과 설명에 지면을 할당하고 있다. PREMIS 데이터 사전에 포함된 보존메타데이터 요소들은 다음과 같다(Caplan 2009).

- 객체를 위한 유일한 식별자 정보
- 체크섬이나 알고리듬과 같은 고정 정보 (fixity information)
- 객체의 크기(size) 정보
- 객체의 포맷 정보
- 객체의 원래 이름
- 작성과 관련된 정보
- 제한자(inhibitors)에 관한 정보
- 중요 속성에 관한 정보
- 환경에 관한 정보
- 어떤 미디어에 저장되었는지에 관한 정보
- 디지털 서명관련 정보
- 다른 객체나 다른 타입의 개체와의 관련성(relationship) 정보

이들 요소를 구체화한 13개의 상위요소는 〈표 1〉과 같다.

3.3 PREMIS 보존 메타데이터의 장단점

Paradigm Project(2005)에서는 PREMIS의 장점을 다음과 같이 제시하고 있다.

1) 표준의 유지 관리에 미국 국회도서관이 지원하고 있다.
2) 전체적으로 이해가 쉽다.
3) 포맷 레지스트리 등의 외부 참조가 쉽다.
4) XML 스키마가 가능하다.
5) 여러 디지털 리포지터리들에서 PREMIS를 구현하고 있다.

〈표 1〉 PREMIS 데이터 사전의 13개 상위 요소

PREMIS 상위요소	
1.1 objectIdentifier	객체 식별자
1.2 objectCategory	객체 구분
1.3 preservationLevel	보존 단계
1.4 significantProperties	중요 속성
1.5 objectCharacteristics	객체 특성
1.6 originalName	원본명
1.7 storage	저장
1.8 environment	환경
1.9 signatureInformation	서명 정보
1.10 relationship	관계
1.11 linkingEventIdentifier	연결된 이벤트 식별자
1.12 linkingIntellectualEntityIdentifier	연결된 지적객체 식별자
1.13 linkingRightsStatementIdentifier	연결된 허가사항 식별자

6) 보존 계층(preservation level)을 책정하기 위해서 리포지터리 내의 정책 수립이 우선적이다.
7) 많은 기관에서 PREMIS를 이용하여 아카이브를 구축하고 있으며 구현상의 실재적 경험들이 쌓이고 있다.
8) 정보의 저장에 있어서 어떤 방법도 강제하지 않는다.
9) 모든 형태 디지털 자료, 즉 디지털 아카이브로부터 E-Book에 이르기까지 적용가능하다.
10) OAIS 참조 모형 적용 시, METS 파일 등과 결합하여 AIP 생성이 가능하다.
11) 개체간 연결 능력이 우수하고 1:1관계, 1:다 관계, 다:1 관계 등 관계 설정이 자유롭다.
12) 개체 타입에 상관없이 적용가능하다.
13) 필요한 만큼 의미단위의 반복사용이 가능하다.
14) 의미단위 중 필수 요소가 매우 적다.

반면, 단점을 다음과 같이 지적하고 있다(Paradigm Project 2005).

1) '지적개체', '표현'(representation)의 개념 이해가 쉽지 않다.
2) 운영을 위한 리포지터리의 운영 규칙에 있어서 합의가 필요하다.
3) 제어 어휘집 개발 등 데이터 컨텐츠 표준에 대해 내부 합의를 필요로 한다.
4) 표준 없는 제어 어휘집 사용은 리포지터리 간의 상호운영성을 훼손할 수 있다.
5) 특정 타입의 개체나 미디어, 하드웨어 등의 기술적(technical) 속성에 관한 메타데이터가 보완되어야 한다.

4. 사무문서 컨텐츠모형 설계

PREMIS 데이터모델의 단점은 지적개체와 표현(representation) 개념의 적용 어려움이다(Paradigm Project 2005). 따라서 PREMIS

데이터모델을 사무문서에 적용하기 위하여 PREMIS 모델 내의 지적개체에 대한 문서차원의 개념정립과 이해가 선행되어야 한다. 이는 PREMIS가 '문서'만을 위한 데이터모델로 작성된 것이 아니라 보존을 필요로 하는 모든 디지털 개체를 위한 데이터모델로 개발된 것이기 때문이다. 또한 PREMIS는 데이터모델의 구현을 위해 따라야 하는 어떤 실질적인 구축상의 지침이나 최소한의 규정도 제시하지 않고 있다.

PREMIS의 5가지 개체 중에서 문서 개체와 직접적으로 관련 있는 개체는 지적개체와 객체개체이다. 이중 객체개체는 representation, file, bit stream의 3가지 서브타입을 가지는 개체로 문서에 직접적으로 적용가능하다. 그러나 PREMIS는 지적개체를 관리(management)와 기술(description)을 위한 하나의 지적단위로 정의하며 그 예로 지도, 사진, 데이터베이스, 웹사이트, 웹페이지 등을 들고 있는데(PREMIS Editorial Committee 2008), 이를 문서에 적용하기 위해서는 문서 컨텐츠에 대한 이해와 분석이 선행되어야 한다. 즉, 문서에서 무엇을 관리와 기술을 위한 하나의 지적 단위로 볼 것인가의 문제와 문서의 집합적인 성격을 어떻게 감안할 것인지를 논의할 필요가 있다. 문서관리 메타데이터 체계에서는 문서 계층에 따라서 적용되는 메타데이터 요소가 다르다는 점도 계층구조에 대한 분석이 선행되어야 하는 이유가 된다.

이를 해결하기 위해 기록물 기술 규칙을 적용하였다. 이어 문서의 계층을 구분하는 원칙과 계층 구조를 먼저 정립하고 그에 맞추어 사무문서 컨텐츠를 대상으로 한 계층 모형을 만드는 작업을 수행하였다. 이를 사무문서 컨텐츠모형 설계 작업으로 정의하였으며 먼저 기술할 계층 구조를 정하고 대상 문서 컬렉션을 대상으로 기술 계층 구조대로 기술할 개체를 분석하고 구조를 모델링하였다.

4.1 설계 원칙

PREMIS의 지적개체의 개념을 사무문서에 적용하기 위해, 문서는 개별적으로 존재한다기보다는 집합적인 성격을 가지며 기술에 있어서도 집합기술의 원칙을 가진다는 문서의 중요한 특징을 고려하였다. 따라서 사무문서 컨텐츠모형 설계에 앞서, 본 연구에서는 다음과 같은 두 가지의 큰 설계 원칙을 세우고 이를 준수하였다.

- PREMIS 보존 메타데이터 데이터모델을 기본 모델로 한다.
- 기록물 기술 규칙을 준수한다.

국제적 차원의 기록물 기술 규칙은 ISAD(G)(General International Standard Archival Description)이다. 이 규칙은 ICA의 기술표준위원회(Commission on Descriptive Standard; ICS/DDS)가 개발하여 ICA가 공인한 보존기록 기술에 대한 국제표준이다. 1994년에 초판이 공표되었고, 2000년에 제2판이 발표되었으며 매 5년마다 개정토록 되어 있다. 이 표준에서는 기록의 계층에 따라 다계층 기술을 하도록 정하고 있고, 26개의 기술요소를 제시하고 있다(한국기록관리학회 2008).

우리나라에서는 국가기록원이 2006년 2월 「국가기록원 기록물 기록규칙(안)」을 원내 표준으로 제정하였다. 이 규칙은 ISAD(G)의 원칙 및 요소를 반영하였고, 다계층 기술 및 집합적 기술

원칙에 따라 기록계층 기술요소를 제시하고 있으며 모든 기록 매체와 형식을 포괄할 수 있도록 설계되었다(국가기록원 2006).

이들 기술 규칙의 가장 중요한 핵심은 집합적 기술의 원칙과 다계층 기술의 원칙이다. 이러한 기술원칙은 ISAD(G) 기록물 다층기술의 4가지 원칙(ICA 2000)과 이를 바탕으로 마련된 한국 국가기록원 기록물 기술원칙(국가기록원 2006)에서도 잘 보여진다.

ISAD(G)에서 말하는 기록물 다층기술의 4가지 원칙(ICA 2000)은 다음과 같다.

- 포괄적 계층에서 구체적 계층으로 기술한다.
- 기술계층에 적절한 정보를 기술한다.
- 각 기술을 상위계층의 기술단위에 연결한다.
- 정보를 반복 기술하지 않는다.

기록물 기술 제1원칙은 '다계층 기술'(multi-level description) 원칙이다. 기록의 계층분석에 입각한 기술 및 관리를 하도록 하고 있다. 이는 상위계층의 기술을 먼저 한 후 하위계층의 기술을 하라는 원칙으로 가장 넓은 범위를 기술한 후 아래 계층으로 내려가면서 기술하되 각 계층이 연계되어야한다.

둘째, 기술되는 기록계층에 적절한 정보만을 기술해야한다. 기술계층이 기록군이라면 기록철에 대한 상세한 내용정보를 제공할 필요가 없으며, 해당 기술단위의 생산자가 '과'나 '지부'인 경우, '부'나 '국' 전체 조직의 연혁을 제공할 필요가 없다.

셋째, 각 기술을 연결해야 한다. 기술 계층의 위치를 명확히 구현해야 하며, 각각의 기술은 상위의 기술 계층과 연결되어야 한다.

넷째, 정보를 반복하여 기술하지 말아야 한다. 계층적으로 관련된 기술 간에 정보가 중복되지 않도록 하기 위한 원칙으로 최상위 계층의 기술에서는 전체 계층에 적용되는 공통 정보를 제공하고, 하위 계층 기술에서는 상위계층 기술에서 이미 제공된 정보를 중복 기술하지 않아야 한다(한국기록관리학회 2008).

한편 한국 국가기록원 기록물 기술규칙을 정리하면 다음과 같다(국가기록원 2006).

- 출처존중의 원칙
- 원질서 존중의 원칙
- 집합기술의 원칙
- 다층기술의 원칙

출처존중의 원칙은 기록물의 소장자보다 생산자가 더 중요하며, 기록물과 생산자와의 관계가 중요하다는 것을 의미한다. 이는 기록물 자체뿐만 아니라 기록물 생산자에 대한 정보를 함께 보존하고자 하는 것이며 출처주의는 자료가 만들어진 배경에 대한 정보를 집적하고자 하는 것이다.

둘째, 원질서 존중의 원칙은 생산기관이 공식 활동을 수행하는 과정에서 기록을 축적한 순서와 질서를 그대로 유지해야 한다는 원칙으로 이 원칙은 특정한 기록물 계열들 사이의 상관관계를 포함하여 전체 기록물철의 구조를 보존하기 위한 것이다. 보존기록의 경우, 생산환경을 전혀 알지 못하는 먼 미래의 이용자까지 고려해 주어야 하므로 출처정보와 원질서를 존중하여 기술하는 것이 미래 이용자들에게는 생산 환경에 대한 정보를 함께 제시해 주는 중요한 정보가 될 것이다.

셋째, 집합기술의 원칙은 출처와 원질서 존중의 원칙에 입각한 기술은 개별 건이 아닌 집합

적 수준에서 이루어져야 한다는 원칙이다. 이 원칙은 개별 기록의 내용보다는 전체적인 내용의 윤곽을 중요시 하는 것이다.

넷째, 계층통제의 원칙은 출처와 원질서를 존중하면서 기록물을 처리하는 과정이 유지되어야 한다는 원칙으로 이 원칙은 가장 포괄적인 상위계층으로부터 가장 구체적인 하위계층의 연속선을 따라 기록물을 구분하여 기술하라는 원칙이다.

따라서 본 연구에서는 전술한 ISAD(G)와 한국 국가기록원 기록물 기술규칙을 종합하여 다음과 같은 기술 원칙을 세우고 이를 준수하였다. 특히, 다층기술의 원칙에는 ISAD(G)의 다층기술의 4원칙을 모두 포함하였다.

- 출처와 원질서 존중의 원칙
- 다층기술의 원칙
  - 포괄적 계층에서 구체적 계층으로 기술한다.
  - 기술계층에 적절한 정보를 기술한다.
  - 각 기술을 상위계층의 기술단위에 연결한다.
  - 정보를 반복 기술하지 않는다.

4.2 문서 선정

본 논문에서 적용한 사무문서는 일반 기업체의 사무문서 중 특히 보존가치가 높은 문서를 우선 선정 대상으로 고려하였다. 따라서 가능한 여러 업종의 기업체를 탐색하였으며 그 중 사무문서의 장기 보존 가치가 가장 높다고 판단되는 A기업 B국 석유사업 및 유전개발 문서를 적용대상으로 선정하였다. 문서 포맷으로는 doc, hwp, xls, pdf 등의 텍스트 문서와 gif, jpg 등의 그림파일 포맷으로 구성되어 있으며 업무 구조대로 폴더를 중심으로 관리되고 전체적인 문서관리시스템은 구축되어 있지 않다.

적용대상 사무문서의 종류는 다음과 같다.

- 계약서: 해당국의 정부와 맺은 광구 개발 계약서로 20년 이상 보관중인 문서도 있다. 계약서에는 지분양도 계약서(farmout agreement)와 협력운영협의서(joint operation agreement) 등 매우 중요한 문서가 존재한다.
- 광구 참여 제안서: 광구 계약 전 개발에 어떤 조건으로 참여하고자 하는지 제안한 문서이다.
- 기술문서: 탄성파 자료(디지털 자료, 종이 출력 자료, 마그네틱테이프 형태로 존재), 각종 지도, 각종 시추 자료, 시기별로 다양한 기술평가를 한 광구 평가 보고서 등이 존재한다. 또한 설계도면, 운영 매뉴얼, 관리지침 등의 시설·설비 관련 자료, 물품 창고 관련 목록을 적은 문서 등이 포함된다.
- 현장 운영 문서: 일일 생산량, 일일 작업 인력, 시설점검사항, 기타 안전관리 상황에 관한 일일 보고서로 현장의 기록을 담은 매우 중요한 문서이다.
- 일반 행정 관리문서: 일반 행정업무상 발생한 문서로 인사규정, 내무규정, 본사와의 서신 등이 포함된다.
- 단기 프로젝트 관련 문서: 프로젝트 관련 문서로 프로젝트 기안서, 프로젝트 승인과 관련된 내부승인문서, 본사 승인문서, 공동 운영사 승인 문서, 해당 정부 승인문서 등이 포함되며 입찰준비와 관련된 각종

문서, 낙찰 처리 관련 문서, 용역 계약서 등이 여기에 해당한다. 또한 프로젝트 관련 안전문제 관련 계약서, 작업 용역 관련 문서, 작업 진행에 따른 추가 계약서 등이 있다.

A기업 B국 사무소 전체 문서의 량은 수백 기가 byte에 해당하는 량으로 본 연구를 위해서 사무소 전체 문서에 대한 모델링은 불가능하다. 따라서 보존 필요성이 가장 높은 문서 및 그와 업무 관련성이 높은 문서들의 한 축을 포착하여 대상 문서를 선정하였다. 이를 위하여 42.65 Gbyte에 해당하는 895폴더 총 15,000여 파일을 일주일에 걸쳐 일일이 탐색하는 문서 선정 작업을 거쳤으며 그 결과로 대상 사무소의 업무 특성 및 기능을 가장 잘 반영하고 문서들 간의 관계구조가 최대한 다양하게 반영될 만한 문서군을 〈표 2〉와 같이 최종 선정하였다. 대상 문서의 종류, 수, 및 크기(byte)는 다음과 같으며 전체적인 구조는 〈그림 3〉과 같다. 선정된 문서는 '경제성평가및입찰, 계약서, 기술문서, 시설물사진, 행정관리문서'의 총 5개의 폴더로 구성되며 각 폴더 안에 총 27개의 디지털 파일이 존재한다.

〈그림 3〉의 구조 중에서 특징적인 구조를 가진 몇 가지를 서술하면 다음과 같다.

- 상단의 큰 글자는 각 폴더명을 의미하는데 「계약서」 폴더내의 PSA_2007.doc과 PSA_2007.pdf 파일은 같은 문서가 두 가지 포맷의 표현(representation)으로 존재하는 것이다.
- 「기술문서」 폴더에는 2개의 하위폴더가 있는데 〈EvalDevelPoten_Rep〉로 표시된 하위폴더는 EvalDevelPoten_Rep.zip으로 압축되어 압축 파일이 함께 보관되고 있다.
- 「시설물사진」 폴더의 DSC01017.jpg, DSC01017.gif, DSC01017.tif 파일은 하나의 사진이 3개의 포맷으로 함께 존재한다.
- 「행정관리」 폴더에 있는 첨부1_국외여행계획서.hwp는 해외출장품의서.hwp의 첨부 문서이다.

4.3 구조 설계

전자문서의 계층구조를 정한 예로는 영국 TNA(The National Archives)의 파일플랜과

〈표 2〉 대상 문서의 종류와 수 및 크기(byte)

폴더명	데이터종류	하위폴더	파일수	크기(byte)	설명
경제성평가및입찰	xls, doc	·	2	212,992	유전경제성평가 및 광구입찰제안서
계약서	pdf, doc	·	5	10,410,413	광구계약서, 생산물분배계약서
기술문서	zip, jpg, hwp, pdf	2	7	160,991,630	유전, 광구개발과 관련된 기술보고서
시설물사진	jpg, gif, tif	·	5	1,320,343	유전, 광구 및 시설물 사진
행정관리문서	hwp, jpg, xls, doc	·	8	688,128	사무소의 행정관리 사무문서
총합	hwp, jpg, doc, xls, pdf, jpg, gif, tif	·	27	173,623,506	·

〈그림 3〉 적용대상 문서의 전체구조

Paradigm Project(2005)의 개인 전자기록물 계층수준이 있다. 우선, TNA는 전자문서의 계층 구조를 '클래스-폴더-파트(하위 폴더의 개념)-레코드-컴포넌트'로 5단계 구분하였다. 클래스 자체에는 기록이 포함되지 않고 메타데이터로 구성되며 일부 메타데이터는 하위계층으로 상속되기도 한다(Public Records Office 2002). 개인용 보존 디지털 아카이브 개발을 위한 Paradigm project(2005)에서는 전자문서의 계층구조를 구조화하여 '컬렉션-액세션-시리즈-하위시리즈-개별기록물'의 5단계로 기술 단위를 삼고 있다. 예를 들면, 컬렉션 계층으로는 X정치가 개인 소유 기록물 전체를 액세션 계층으로는 특정일에 만들어진 X정치가의 기록물을 예로 들 수 있다. 또한 그의 1999년부터 2008년 사이의 이메일 기록물은 시리즈 계층으로, 그 중 특히 2007-2008년 특정주제와 관련된 이메일 폴더는 하위시리즈 계층으로 설명하고 있다. 마지막으로 개별기록물 계층은 특정일에 행해진 연설문 파일이나 특정한 사진 한 장을 말한다.

PREMIS는 지적개체에 대해서 개념만 소개하고 있을 뿐 실제 적용을 위한 어떤 지침도 제공하지 않고 있다. 따라서 본 연구에서는 기록의 계층적 기술을 위한 사전 작업으로 PREMIS 데이터모델의 지적개체를 'Collection 계층, Accession 계층, SubjectFolder 계층, Item 계층'으로 개념적으로 단계를 나누었다. 특히 본 연구에서는 기록의 계층적 기술을 위해 원질서 원칙에 입각하여 추출된 각 문서 폴더를 SubjectFolder 계층으로 명명하여 지적개체의 제3 계층에 적용하였다. 폴더의 개념은 영국 국립기록청(National Archive: TNA)에서 제시한 전자문서의 계층구조에서도 나온 개념으로 이는 문서

기술의 기본단위를 정의하고 있는 ISAD(G)나 EAD(Encoded Archival Description)의 계층 구분상에서는 'Series'단계에 해당하는 것이다. 그러나 ISAD(G)나 EAD가 디지털 문서가 아닌 물리적 기록을 주 대상으로 정의하고 있고 'Series'의 개념은 상당 부분 공문서의 생산방식에 기인된 것으로도 볼 수 있으므로 디지털 문서에는 Series라는 개념 보다는 Folder를 문서를 모으는 기본 단위 개념으로 정의하였다.

지적개체가 '관리와 기술을 위한 하나의 지적단위'로 동질 속성을 공유하는 집합체를 의미하므로 문서의 내용과 업무기능을 중심으로 하나의 지적 단위체의 성격을 갖는 문서 폴더는 지적개체 내 하나의 계층으로 충분히 적용될 수 있다. 또한 PREMIS는 하나의 지적개체는 하위의 또 다른 지적개체를 포함할 수 있다고 밝혔으므로(예: 하나의 웹사이트는 한 웹페이지를 포함할 수 있다) 폴더 안에 또 다른 폴더를 포함해야 하는 문서구조에 잘 적용될 수 있다. 본 연구에서 제시한 기술의 계층 수준과 각 표준의 비교는 〈표 3〉에서 볼 수 있다.

전자문서의 계층 구조는 전통적인 종이 문서의 계층 구분과는 다른 특징을 가지며 물리적 계층 보다는 논리적 계층이 강조되어야 한다. 하나의 전자 문서건은 다시 여러 개의 표현(representation)의 형태를 띌 수 있다.

Collection, Accession 계층 자체에는 기록이 포함되지 않으며 기술 메타데이터로만 구성되고, 하위 계층으로 메타데이터가 상속된다. SubjectFolder는 기록관리의 기본 단위(primary unit)이며, 메타데이터로 구성되는 이른바 기록을 담는 단위용기로 볼 수 있다. SubjectFolder가 속한 상위계층 즉, Collection 계층, Accession 계층으로부터 메타데이터를 상속받고, Subject Folder의 메타데이터도 그것에 속한 Item에 상속된다.

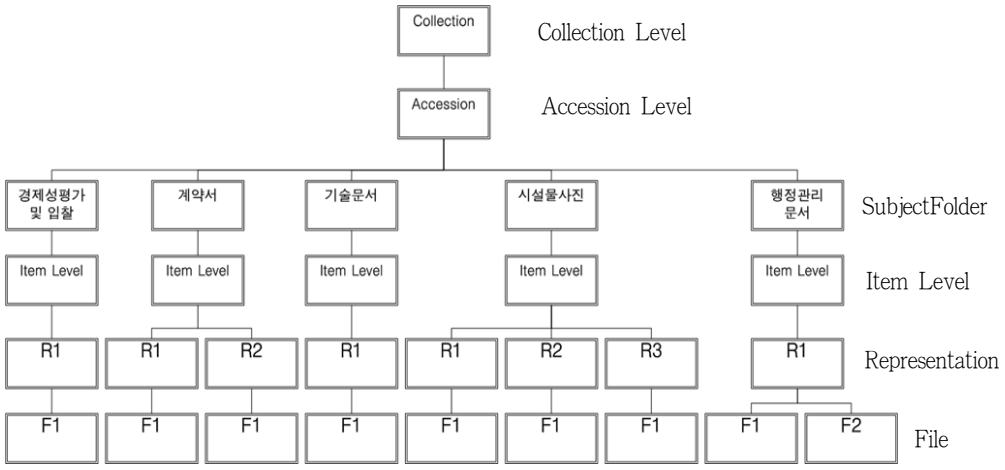
〈그림 4〉에서 보듯이 하나의 아이템은 하나 이상의 표현으로 구현된다. PREMIS의 '표현'(representation)은 지적개체의 하나의 구현 사례(a single digital instance of an intellectual entity)로 정의할 수 있다(PREMIS Editorial Committee 2008). 이는 FRBR의 '구현형'(manifestation)과 혼동을 피하기 위해 representation으로 명명 되었다. 하나의 Item은 하나 이상의 representation으로 구현되며 이것은 물리적 파일 개체를 하위에 가지게 된다. 〈표 4〉는 본 연구에서 제시한 계층 수준에 대한 실제 문서 예를 설명하고 있다.

〈표 3〉 ISAD(G), EAD, TNA, Paradigm Project 기술 계층과의 비교

계층 수준		한국	ISAD(G)	EAD	TNA	Paradigm Project
지적개체	Collection	기록물군	Fonds	Collection	Class	Collection
	Accession	기록물계열	Sub-fonds	Subgroup	(Subclass)	Accession
	SubjectFolder	하위기록물계열	Series	Series	Folder	Series
	subSubjectFolder	기록물철	Subseries	Subseries	Part	Subseries
	Item	기록물건	Item	Item	Record	Item
객체개체	Representation/File				Component	

〈표 4〉 계층 수준에 대한 예

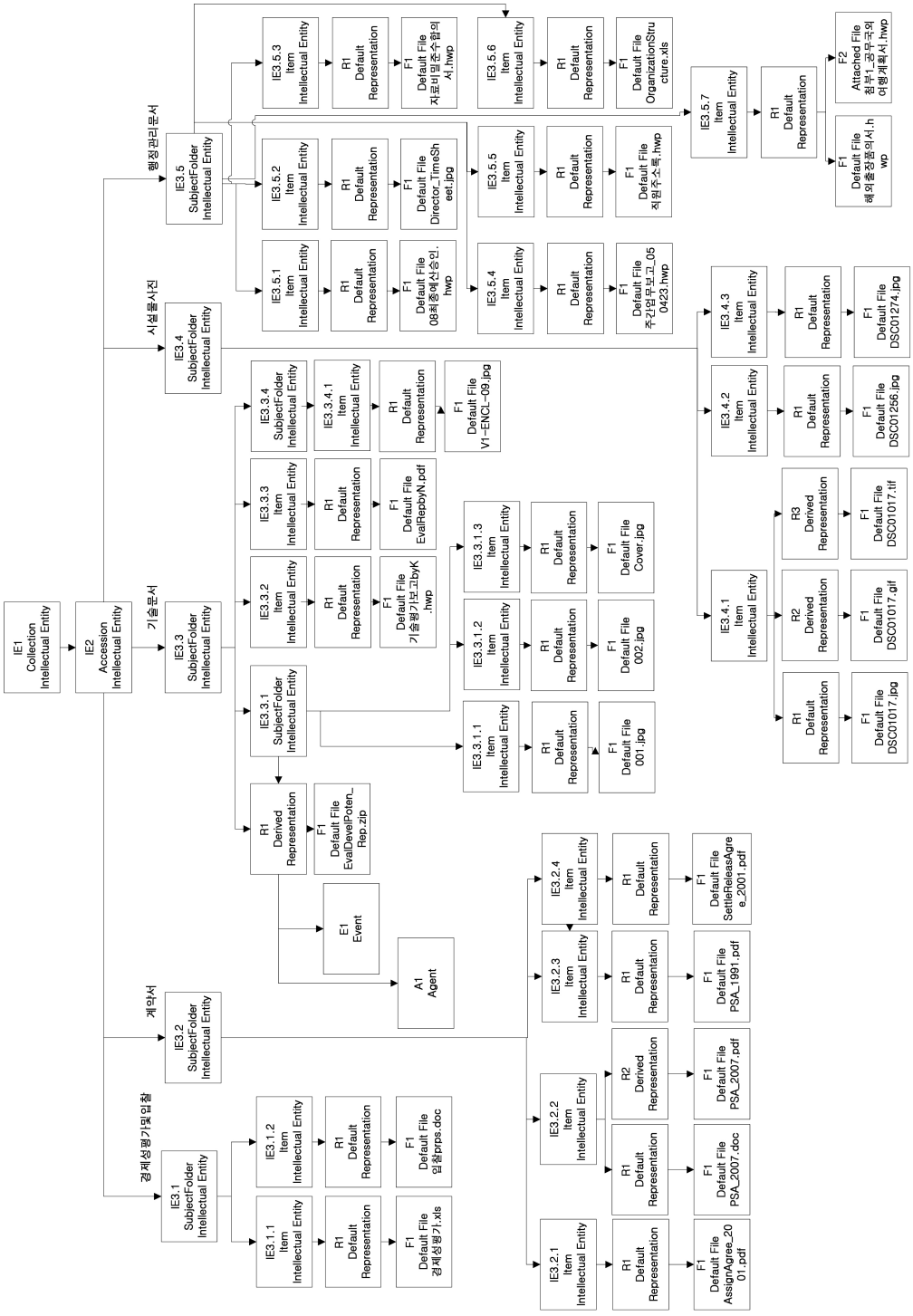
계층 수준	예시	PREMIS 데이터 모델 개체
Collection 계층	'A기업 B국 사무소 문서' 전체	지적개체
Accession 계층	'C광구 관련 문서 중 본 연구를 위해 09년 7월 선별된 문서파일 아카이브'	
SubjectFolder 계층	Ex: '계약서' 폴더	
Item 계층	'2007년 양국합의 유전개발계약 및 생산물 분배합의서'	
Representation	'2007년 양국합의 유전개발계약 및 생산물 분배합의서'의 워드파일 또는 pdf 파일	객체개체
File	'2007년 양국합의 유전개발계약 및 분배합의서 파일인 PSA_2007.doc'	



〈그림 4〉 전체 대상문서의 계층 구조 설계

4.4 모형 설계

본 연구에서는 기록의 계층적 기술을 위한 사전 작업으로 PREMIS 데이터모델의 지적개체를 〈그림 4〉와 같이 'Collection계층, Accession 계층, SubjectFolder 계층, Item 계층'으로 개념적으로 단계를 나누었다. Item 계층 밑으로는 개념적 개체가 아닌 물리적 개체에 해당하는 Representation과 File이 존재한다. 맨 위의 'Collection' 계층은 A사 B국 사무소의 모든 문서 집합을 의미한다. 그 다음 단계인 'Accession 계층'은 본 연구를 위해서 모은 문서 집합 전체를 의미한다. 세 번째 단계부터 사무문서 컨텐츠모형 설계의 기본 단위인 'SubjectFolder 계층'이며 '경제성평가및입찰, 계약서, 기술문서, 시설물사진, 행정관리문서'의 총 5개의 폴더로 구성되었다. 5개의 폴더 안에는 하위 폴더가 포함되기도 하며 총 27개의 디지털 파일이 존재한다. 이것을 대상 문서에 적용하면 〈그림 5〉와 같다. 이것은 모든 폴더들의 컨텐츠 모형도를 종합한 것이며 각각의 폴더에 대한 모형은 다음 절에서 각각 설명하였다.


IE1 Collection Intellectual Entity
IE2 Accession Intellectual Entity
IE3 SubjectFolder Intellectual Entity
기술로서
IE3.1 SubjectFolder Intellectual Entity
경제성평가및입찰
IE3.2 SubjectFolder Intellectual Entity
계약서
IE3.4 SubjectFolder Intellectual Entity
시설물사진
IE3.5 SubjectFolder Intellectual Entity
행정관리문서
IE3.1.1 Item Intellectual Entity
R1 Default Representation
F1 Default File 경제성평가.xls
IE3.1.2 Item Intellectual Entity
R1 Default Representation
F1 Default File 입찰prps.doc
IE3.2.1 Item Intellectual Entity
R1 Default Representation
F1 Default File AssignAgree_2001.pdf
IE3.2.2 Item Intellectual Entity
R1 Default Representation
F1 Default File PSA_2007.doc
R2 Derived Representation
F1 Default File PSA_2007.pdf
IE3.2.3 Item Intellectual Entity
R1 Default Representation
F1 Default File PSA_1991.pdf
IE3.2.4 Item Intellectual Entity
R1 Default Representation
F1 Default File SettleReleasAgree_2001.pdf
E1 Event
A1 Agent
R1 Derived Representation
F1 Default File EvalDevelPoten_Rep.zip
IE3.3.1 SubjectFolder Intellectual Entity
IE3.3.1.1 Item Intellectual Entity
R1 Default Representation
F1 Default File 001.jpg
IE3.3.1.2 Item Intellectual Entity
R1 Default Representation
F1 Default File 002.jpg
IE3.3.1.3 Item Intellectual Entity
R1 Default Representation
F1 Default File Cover.jpg
IE3.3.2 Item Intellectual Entity
R1 Default Representation
F1 Default File 기술평가보고by K.hwp
IE3.3.3 Item Intellectual Entity
R1 Default Representation
F1 Default File EvalRepbyN.pdf
IE3.3.4 SubjectFolder Intellectual Entity
IE3.3.4.1 Item Intellectual Entity
R1 Default Representation
F1 Default File V1-ENCL-09.jpg
IE3.4.1 Item Intellectual Entity
R1 Default Representation
F1 Default File DSC01017.jpg
R2 Derived Representation
F1 Default File DSC01017.gif
R3 Derived Representation
F1 Default File DSC01017.tif
IE3.4.2 Item Intellectual Entity
R1 Default Representation
F1 Default File DSC01266.jpg
IE3.4.3 Item Intellectual Entity
R1 Default Representation
F1 Default File DSC01274.jpg
IE3.5.1 Item Intellectual Entity
R1 Default Representation
F1 Default File 08최종예산승인.hwp
IE3.5.2 Item Intellectual Entity
R1 Default Representation
F1 Default File Director_TimeSheet.jpg
IE3.5.3 Item Intellectual Entity
R1 Default Representation
F1 Default File 자료비일준수협의서.hwp
IE3.5.4 Item Intellectual Entity
R1 Default Representation
F1 Default File 추가업무보고_0423.hwp
IE3.5.5 Item Intellectual Entity
R1 Default Representation
F1 Default File 직원주소록.hwp
IE3.5.6 Item Intellectual Entity
R1 Default Representation
F1 Default File OrganizationStructure.xls
IE3.5.7 Item Intellectual Entity
R1 Default Representation
F1 Default File 해외출장품의서.hwp
F2 Attached File 첨부1_공동주최여행계획서.hwp

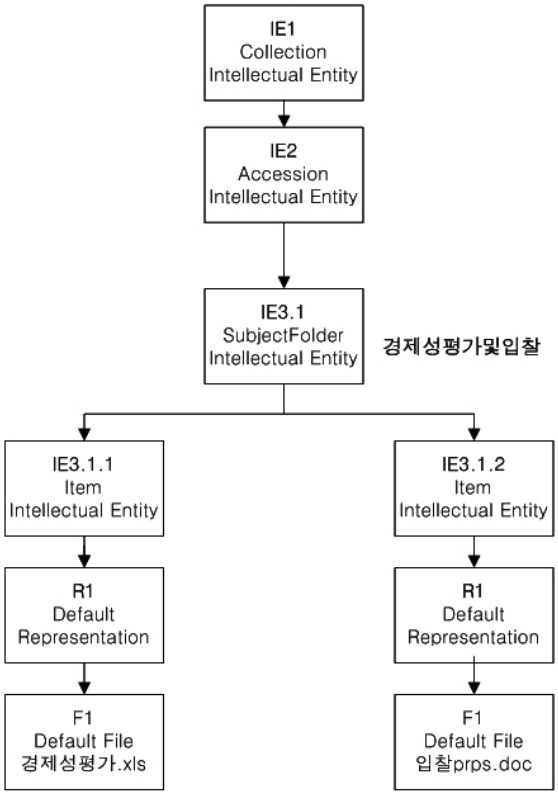
〈그림 5〉 전체 컨텐츠모형

4.4.1 「경제성평가 및 입찰」 폴더의 컨텐츠 모델링

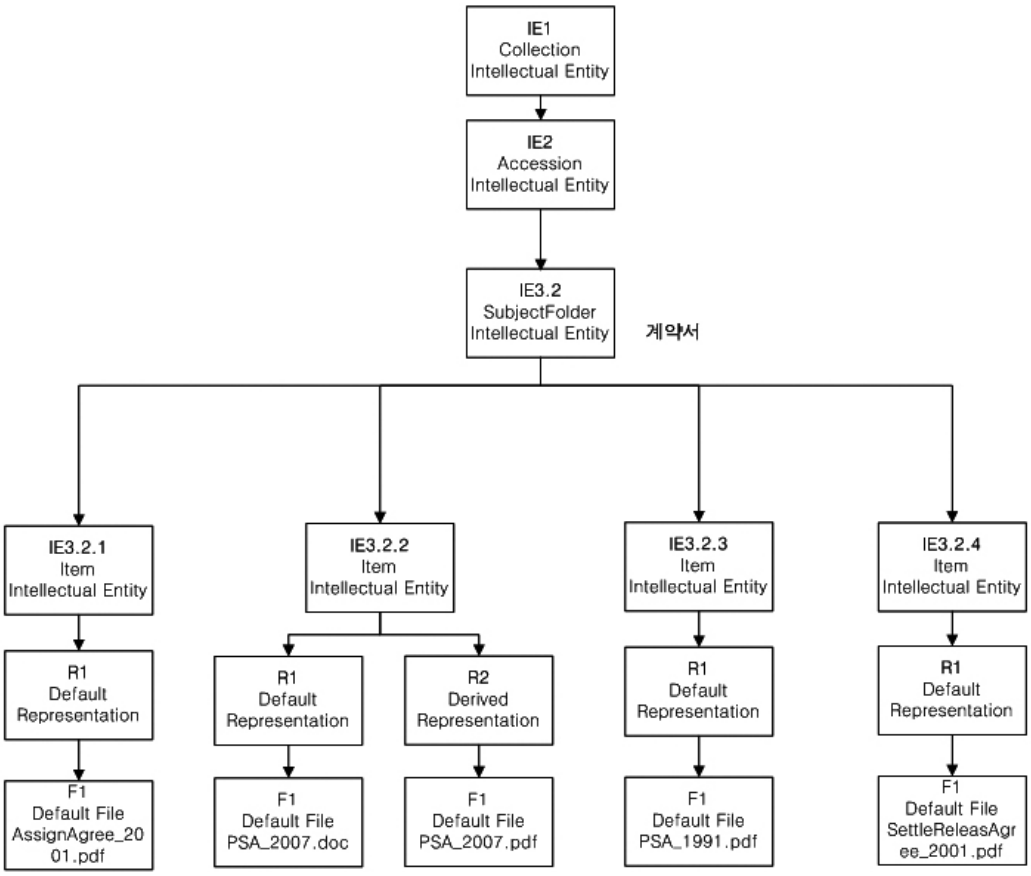
〈그림 6〉은 첫 번째 폴더인 「경제성평가 및 입찰」 폴더에 대한 컨텐츠 모델링 결과를 보여준다. 맨 위의 사각형 IE1은 A국 B사무소의 문서 전체 컬렉션을 의미하며 두 번째 사각형 IE2 본 연구를 위해 모은 문서 집합을 나타낸다. 세 번째 사각형 IE3.1이 SubjectFolder 계층인 「경제성평가 및 입찰」 폴더를 말한다. 여기서, IE3.1의 '3'은 해당 객체가 본 모형의 계층 단계에서 3번째 단계인 것을 의미하며 '1'은 3번째 계층의 첫 번째 개체임을 나타내는 것이다. 9 이상의 숫자 확장을 위해 숫자 사이에 '.'을 찍어 각 계층을 구분하였다. 이 폴더 내에는 두 개의 물리적 파일 '경제성평가.xls'와 '입찰prps.doc'가 존재하는데 이 두 파일을 표현해 주기 위해 그 위로 각각의 R1(representation1)과, Item 계층인 IE3.1.1과 IE3.1.2가 존재한다.

4.4.2 「계약서」 폴더의 컨텐츠 모델링

〈그림 7〉은 「계약서」 폴더에 대한 컨텐츠 모델링 결과를 보여준다. 다음 〈그림 6〉과 마찬가지로 맨 위의 사각형 IE1은 A국 B사무소의 문서 전체 컬렉션을 의미하며 두 번째 사각형 IE2 본 연구를 위해 모은 문서 집합을 나타낸다. 세 번째 사각형 IE3.2가 본 SubjectFolder 계층인 「계약



〈그림 6〉 「경제성평가및입찰」 폴더 컨텐츠모형



〈그림 7〉「계약서」 폴더 컨텐츠모형

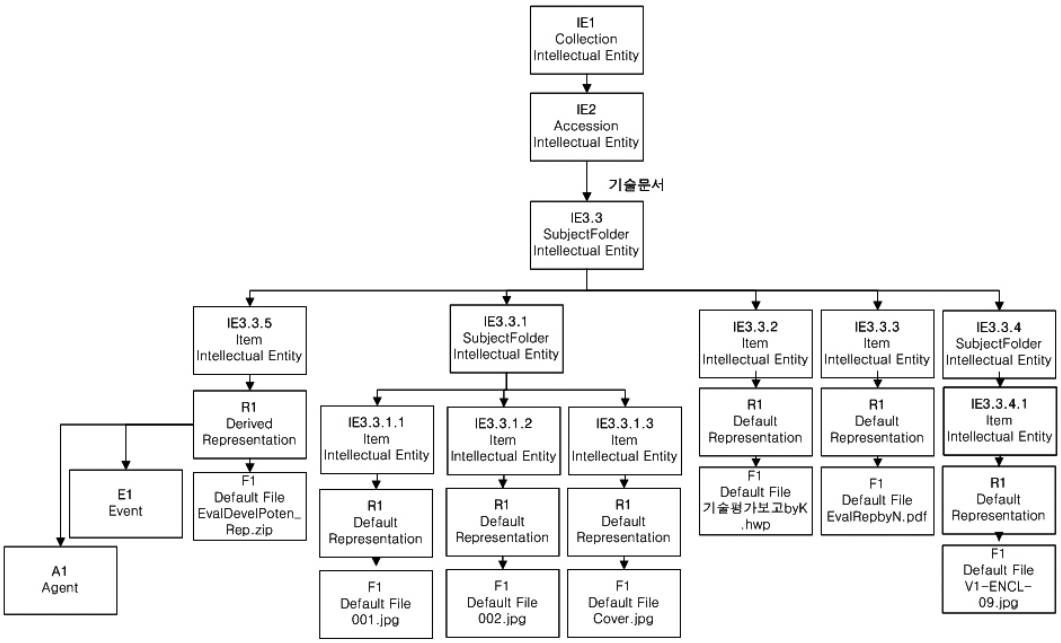
서」 폴더를 말한다. 여기서의 특이사항은 「계약서」 폴더내의 PSA_2007.doc과 PSA_2007.pdf 파일은 같은 문서가 두 가지 포맷의 표현으로 존재하는 것이다.

4.4.3 「기술문서」 폴더의 컨텐츠 모델링

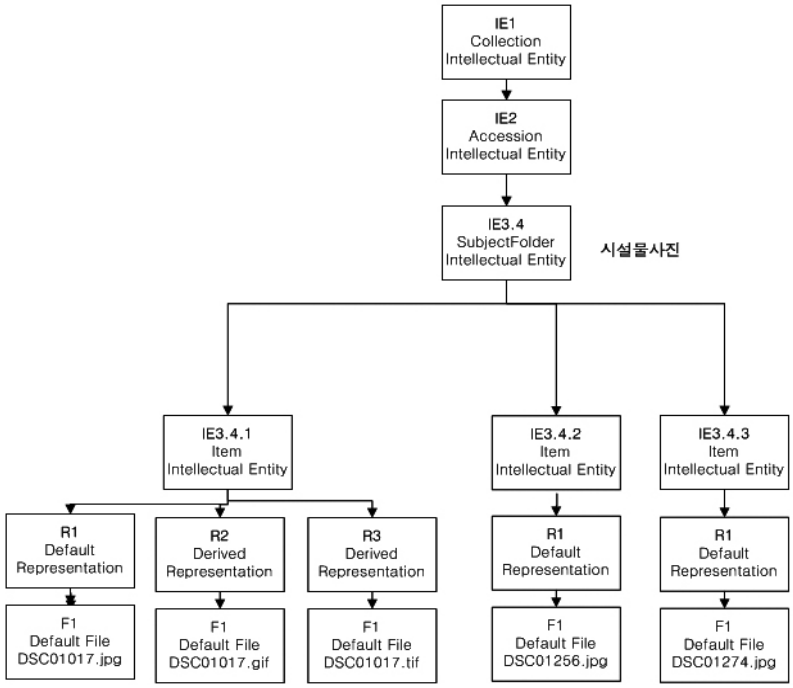
〈그림 8〉은 「기술문서」 폴더에 대한 컨텐츠 모델링 결과를 보여준다. 역시 위의 〈그림 6, 그림 7〉과 마찬가지로 위의 사각형 IE1은 A국 B사무소의 문서 전체 컬렉션을 의미하며 두 번째 사각형 IE2는 본 연구를 위해 모은 문서 집합을 나타낸다. 세 번째 사각형 IE3.3이 본 SubjectFolder 계층인 「기술문서」 폴더를 말한다. 이 폴더에서의 특이점은 「기술문서」 폴더에는 2개의 하위폴더가 있는데 그 중 〈EvalDevelPoten_Rep〉 폴더가 EvalDevelPoten_Rep.zip으로 압축되어 압축 파일이 폴더 내에 함께 존재한다는 점이다.

4.4.4 「시설물사진」 폴더의 컨텐츠 모델링

〈그림 9〉는 「시설물사진」 폴더에 대한 컨텐츠 모델링 결과를 보여준다. IE1과 IE2는 전술한 폴더들과 함께 공유하는 지적개체이며 세



〈그림 8〉「기술문서」폴더 컨텐츠모형



〈그림 9〉「시설물사진」폴더 컨텐츠모형

번째 사각형 IE3.4가 본 SubjectFolder 계층인 「시설물사진」 폴더를 말한다. 이 폴더 내에 DSC01017.jpg, DSC01017.gif, DSC01017.tif 파일은 내용이 동일한 사진이 서로 다른 3개의 포맷으로 존재하는 것이다.

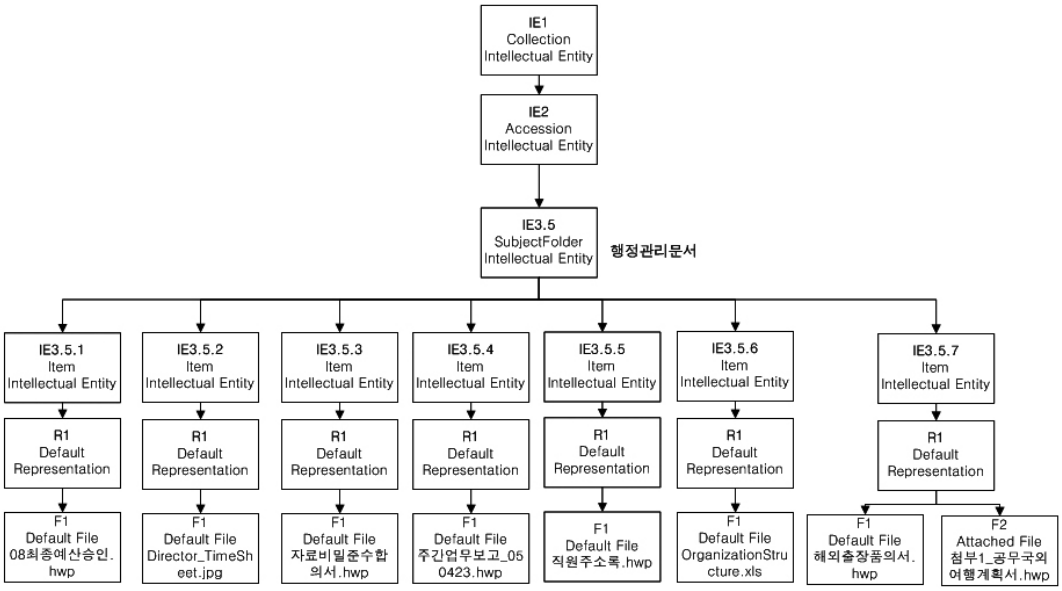
4.4.5 「행정관리문서」 폴더의 컨텐츠 모델링

〈그림 10〉은 「행정관리문서」 폴더에 대한 컨텐츠 모델링 결과를 보여준다. 전술한 것과 마찬가지로, IE1과 IE2는 위의 폴더와 함께 공유하는 지적개체이며 세 번째 사각형 IE3.5가 이번 SubjectFolder 계층인 「행정관리문서」 폴더를 말한다. 이 폴더에 있는 첨부1_국외여행계획서.hwp는 해외출장품의서.hwp의 첨부 문서이다. 이 두 파일이 함께 Representation 계층인 'IE3.5.7_R1'을 구성하게 된다.

4.4.6 사무문서 컨텐츠모형의 확장 적용

본 연구에서 PREMIS 데이터모델을 사무문서에 적용하기 위해 설계된 컨텐츠모형은 PREMIS 메타데이터 요소와 결합하여 여러 유형의 사무문서에 확장 적용될 수 있다.

문서는 동일한 활동에서 산출된 문서들 간의 상호 관계를 의미하는 문서 결합관계(archival bond)를 가지며 이 결합관계를 통해 특정 업무와 관련된 문서들 간의 관계뿐만 아니라 선행처리 업무와 후속 처리 업무에서 산출된 문서들 간의 관계도 파악될 수 있다(한국기록학회 2008). 이러한 문서의 결합관계를 표현해 낼 수 있는 도구로 본 연구에서는 2가지 방법을 제안한다. 즉, 컨텐츠모형을 통하여 문서간의 집합적이고 계층적인 관계를 1차적으로 표현해 낼 수 있으며 문서간의 업무흐름 및 시간 흐름에 따른 2차적인 수평적 관계를 PREMIS 메타데



〈그림 10〉 「행정관리문서」 폴더 컨텐츠모형

이터 요소의 기술을 통하여 표현할 수 있다.

문서의 메타데이터를 기술한다는 것은 문서의 내용, 맥락, 구조가 잘 들어날 수 있도록 기술하는 것을 말하며 특별히, 업무 활동을 통하여 생산된 사무문서에 있어서는 업무에 따른 문서작성 맥락과 문서간의 구조적 관계가 들어날 수 있도록 기술되어야 한다.

PREMIS에서는 '관계'(relationship)요소를 이용하여 문서간의 구조적, 파생적 관계를 기술하도록 규정하고 있는데 관계를 기술할 수 있는 하위요소로 관계유형(relationshipType) 및 관계하위유형(relationshipSubType)요소가 있다. 관계 요소는 아카이브의 재량에 따라 제어어휘값을 두어 필요한 요소를 설계하여 사용할 수 있으며 아카이브는 기술할 문서의 유형에 맞게 제어어휘값을 미리 선정해 둘 수 있다. PREMIS 데이터사전에서는 관계요소의 값으로 '구조적'(structural)관계, '파생'(derivation)관계 등의 값을 제안하였으나 아카이브 정책에 따라 변경 및 추가가 가능하다. 관계하위유형의 요소에도 제어어휘값이 필요하며 PREMIS가 제안한 값으로는 'has sibling', 'is part of', 'has part', 'is source of', 'has source', 'has root', 'includes', 'is included in' 등이 있다. 이외에 연결된지적개체식별자, 연결된이벤트식별자, 연결된허가사항식별자 요소를 이용하여 지적개체 및 객체 개체와의 연결관계를 표현할 수 있다.

5. 결 론

전자문서의 보존 문제에 있어서는 오랜 기간이 지난 후에도 그 목적을 성취할 수 있음을 가능한 한 보장받을 수 있도록 표준의 정립과 준수가 중요하다. 따라서 본 연구에서는 OAIS 참조 모형을 구체적으로 발전시킨 사실상의 보존 메타데이터 표준인 PREMIS 데이터모델과 데이터사전을 사무문서에 적용하기 위한 사무문서 컨텐츠모형을 개발하였다. 대상 문서는 'A사 B국 해외 석유사업 및 유전개발 문서'로 국가 차원 이상의 영구 보존 가치를 지니는 문서 컬렉션을 선정하였다. 이론적 배경으로 디지털 사무문서(business records)의 특성과 보존 필요성에 관하여 서술하고 PREMIS 보존 메타데이터에 관해 자세히 살펴보았다. 이후, PREMIS 데이터모델을 사무문서에 구체적으로 적용하기 위하여 PREMIS 모델 내의 지적개체에 대한 문서 차원의 개념 정립과 이해를 시도하였다. 즉, 문서 컨텐츠의 계층을 구분하는 원칙과 구조를 설계하였고 그에 맞추어 사무문서 컨텐츠를 대상으로 한 계층 모형을 만드는 작업을 실시하여 사무문서 컨텐츠모형을 도출해 내었다. 이 과정에서 기록물 기술 규칙을 준수하였다.

본 연구의 의의 첫째, PREMIS 보존메타데이터를 사무문서에 적용하였다는 점이다. 보존 메타데이터는 리포지터리나 아카이브의 디지털 보존 과정을 지원하는 정보로, 보존 환경에서 디지털 자원의 장기간의 영속성, 재현성, 이해가능성, 진본성, 일치성을 유지하는 핵심적인 기능을 하게 된다. 또한 기록관리에 표준 메타데이터를 이용하면 기록의 4대 속성인 진본성, 신뢰성, 무결성, 이용가능성을 보장할 수 있다. 본 연구에서 사무문서 보존에 국제적 표준 보존 메타데이터를 이용함으로써 이러한 요건들을 충족시키고 지원할 수 있게 되었다. 둘째, 주로

비전자문서인 종이문서에 적용되는 기록물 기술 규칙을 전자문서에 접목시킴으로써 비구조화된 데이터인 전자문서를 구조화시키는 방안을 제시하였다. 많은 디지털 아카이빙 연구 대상이 학술논문, 사진, 과학기술 데이터세트 등에 머무르는 것은 이들 대상 데이터가 이미 구조화되어 나온 데이터이거나 구조화시키기에 어렵지 않은 데이터들이기 때문이다. 지금까지 문서의 경우는 구조화시킬 수 있는 방법이 기록물 기술규칙에 한정되었고 이것도 주로 종이문서 적용에 한정되어 연구가 진행되어 왔다. 이러한 점에서 본 연구는 전자문서 보존과 관련하여 구조화된 모형의 적용가능성을 제시하였다. 셋째, PREMIS 데이터모델의 적용을 위해 사무문서 컨텐츠모형을 설계함으로써 PREMIS의 개체에 대한 문서차원의 개념 정립과 이해를 시도하였다. PREMIS 데이터모델은 전체적으로 이해가 쉽고 개체 간 연결 능력이 우수하며 모든 형태의 디지털 자료에 적용 가능하다는 장점이 있으나 '지적개체'나 '표현' 개체의 직접적인 적용이 쉽지 않다는 단점이 있다. 그러나 본 연구는 사무문서 컨텐츠모형 설계를 통해 PREMIS의 개체에 대해 문서차원의 구조화된 개념을 제시하였다. 결과적으로 개념적이고 추상적 계층인 지적개체를 'Collection, Accession, SubjectFolder, Item'으로 계층화시키고 정의하였다.

본 연구를 통해 제시될 수 있는 향후 연구과제는 다음과 같다. 첫째, 본 연구는 장기보존용 아카이브를 구축함에 있어서 표준의 적용을 중심으로 컨텐츠모형 개발에 집중하여 연구를 진행하였으며 향후 이를 기반으로 한 메타데이터 요소 적용 및 시스템 구축이 이루어 져야 할 것이다. 둘째, 본 연구에서 PREMIS 데이터모델을 적용하기 위해 개발한 사무문서 컨텐츠모형을 더욱 폭넓게 적용하기 위해서 사무문서 이외의 다양한 문서 유형에 대한 정의와 문서 관계에 대한 분석이 필요하며 그에 맞는 메타데이터 요소 값도 개발되어야 할 것이다.

참 고 문 헌

심재균. 1998. EDMS구축 사례 및 기대효과. 『EDMS 기술백서』. 서울: 한국경제신문사.

이원영. 2000. 기록물분류의 원리: 문헌분류와의 비교. 『기록학연구』, 2: 103-128.

장윤기. 2007. 소리 없는 전쟁, 문서포맷. 『온더넷』, 132: 124-129.

한국 국가기록원. 2006. 『국가기록원 기록물 기술규칙(안)』. [대전]: 국가기록원.

한국기록학회. 2008. 『기록학 용어 사전』. 서울: 역사와비평사.

Andolsen, Alan A. 2008. "The Pillars of Vital Records Management." *Information Management Journal*, April: 47-52.

Caplan. 2009. *Understanding PREMIS*. [online]. [cited 2009.7.15].

〈http://www.loc.gov/standards/premis/understanding-premis.pdf〉.

Caplan, Priscilla. 2004. *DAITSS overview.* [online]. [cited 2007.5.20]. 〈http://www.fcla.edu/digitalArchive/pdfs/DAITSS.pdf〉.

IFLA Study Group on the Functional Requirements for Bibliographic Records. 2003.『서지레코드의 기능상의 요건』. 김태수 역. 서울: 국립중앙도서관.

International Council on Archives. 2000. *ISAD(G): General International Standard-Archival Description. Second Edition.* Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999. Ottawa. [online]. [cited 2009.6.3].

OCLC/RLG. 2002. *Preservation Metata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects.* [online]. [cited 2007.9.3]. 〈http://www.oclc.org/research/pmwg〉.

OCLC/RLG. 2005. *Data Dictionary for Preservation Metadata version 1.0. Final Report of the PREMIS Working Group.* [online]. [cited 2006.9.3]. 〈http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf〉.

Paradigm Project. 2005. *Workbook on Digital Private Papers.* [online]. [cited 2009.5.3]. 〈http://www.paradigm.ac.uk/workbook/index.html〉.

Phillips, John T. 2009. "Selecting Software for Managing Physical & Electronic Records." *Information Management Journal,* June: 25-39.

Pinna, Gian Maria, Vincenzo Beruti, Stephane Mbaye, Mathias Moucha, Valter Spaventa, and Davide Castellazzi. 2005. "From HARM to SAFE: the ESA's proposal for a standard archive format for Europe." *In: PV 2005: Ensuring Long-Term Preservation and Adding Value toScientific and Technical Data. Edinburgh.* 21-23 Nov. [online]. [cited 2009.8.13]. 〈http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/018.pdf〉.

PREMIS Editorial Committee. 2008. *PREMIS Data Dictionary for Preservation Metadata Version 2.0.* [online]. [cited 2008.4.1]. 〈http://www.loc.gov/standards/premis/v2/premis-2-0.pdf〉.

PREMIS Working Group. 2004. *Implementing preservation repositories for digital materials: current practice and emerging trends in the cultural heritage community.* [online]. [cited 2009.7.3]. 〈http://www.oclc.org/research/projects/pmwg〉.

Public Record Office. 2002. *Requirements for Electronic Records Management System*

2: MetadataStandard 2002 revision: final version. [online]. [cited 2009.7.3]. 〈http://www.nationalarchives.gov.uk/documents/metadatafinal.pdf〉.

Sawyer, Donald, Lou Reich, David Giaretta, Patrick Mazal, Claude Huc, Michel Nonon-Latapie, and Nestor Peccia. 2002. *The Open Archival formation System (OAIS) Reference Model and its usage.* [online]. [cited 2006.1.31]. 〈http://www.ccsds.org/documents/so2002/spaceops02_p_t5_39.pdf〉.

Van de Sompel, Herbert, Jeroen Bekaert, Xiaoming Liu, Luda Balakireva, and Thorsten Schwander. 2005. *aDORe: A modular, standards-based digital object repository.* [online]. [cited 2006. 1.31]. 〈http://arxiv.org/abs/cs/0502028〉.

Wendler, Robin. 2006. "The status of preservation metadata in the digital library community." *DIGITAl Preservation.* 60-77. London: Facet publishing.