PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구*

- 국립중앙도서관 디지털 자료를 중심으로 -

A Study on Developing Preservation Metadata Based on PREMIS Focusing on Digital Data in National Library of Korea

박 옥 남(Ok Nam Park)**

목 차

1. 서 론
2. PREMIS
3. 선행연구
4. 방법론
5. PREMIS 주요요소 분석
6. 보존 메타데이터 및 활용방안 제시
7. 결 론

초 록

본 연구는 국립중앙도서관이 수집하는 디지털정보자원의 효율적인 관리 및 보존을 위하여 장서관리 시스템에서 보존용 메타데이터 입력을 위한 PREMIS 기반 목록규칙을 개발하는 것은 물론 METS를 기반으로 한 목록레코드에 PREMIS를 적용하는 방안을 제시하고자 하였다. 이를 위해 PREMIS 기반 구현 구현사례를 분석함은 물론, 국립중앙도서관 디지털 자원 수집 현황 및 보존현황, 담당자와의 면담 등이 이루어졌다. 본 연구는 객체(object), 이벤트(event), 행위자(agent), 권한(rights)의 총 4개 분야로 나누어 메타데이터 요소를 도출하였다. 또한 현재 국립중앙도서관 디지털 장서관리에서 인코딩 표준으로 사용되고 있는 METS를 기반으로 한 PREMIS분석을 통해 도출된 메타데이터를 적용한 후 이에 따른 목록의 예를 제시하여 실무의 적용가능성을 높였다.

ABSTRACT

This research seeks to develop preservation metadata and to provide guidelines to apply PREMIS-METS decisions in a METS profile in the National Library of Korea. The process of developing metadata is divided into three steps, that is, analysis of metadata schema in other digital preservation initiatives, analysis of resources and preservation policies, and interviews with librarians in the National Library of Korea. A metadata schema consists of four main entities - object, event, agent, and rights. The study also provides a cataloging example as well as principles using PREMIS-METS for archiving digital resources. The schema has values to manage digital resources effectively and efficiently in the National Library of Korea.

키워드: PREMIS, 보존메타데이터, METS, 국립중앙도서관, 디지털 도서관
PREMIS, Preservation Metadata, METS, National Library of Korea, Digital Library

---

* 본 연구는 2011년 국립중앙도서관 'PREMIS 기반 보존 메타데이터 목록규칙 연구'의 일부분을 수정・보완한 것임.
** 한남대학교 문헌정보학과 조교수(ponda@hnu.ac.kr)
논문접수일자: 2012년 4월 15일 최초심사일자: 2012년 4월 16일 게재확정일자: 2012년 5월 8일
한국문헌정보학회지, 46(2): 83-113, 2012. 〔http://dx.doi.org/10.4275/KSLIS.2012.46.2.083〕

www.kci.go.kr

1. 서 론

정보통신 환경의 급속한 발달로 많은 지적 창작물이 디지털 형태로 생산되고 유통되고 있으며 이러한 형태의 생산물은 짧은 수명을 가지고 일정기간이 지나면 사라져버리는 소멸성을 지니고 있다. 그럼에도 불구하고 인터넷을 통해 생산되고 유통되는 디지털 자원의 정보자원으로서의 가치는 현재 더 커지고 있으며 이러한 이유로 세계 각국의 주요 도서관은 90년대 중반부터 이미 인터넷 지적 자원을 수집하고 보존하는 프로젝트를 추진하였고 현재 다양한 연구 성과들이 발표되어 디지털 자원의 보존이 중요하게 인식되었다.

Gomes, Miranda와 Costa(2011)의 보고에 따르면 2003년 이후 웹 아카이빙은 급속한 성장을 보였으며, 특히 선진국에서 이러한 성장은 두드러지게 나타났다. 1996년에는 4개의 아카이브 이니셔티브에 그치던 것이 2011년에는 42개로 급증하였다. 또한 42개의 웹 아카이빙 이니셔티브는 26개국에서 수행되고 있으며, 유럽(23개), 북미(10개), 아시아(6개), 오세아니아(1개)에 걸쳐 수행되고 있다. 이 가운데 절반은 OECD국가이며, OECD 국가 중 21개(62%)는 적어도 하나 이상의 웹 아카이빙 이니셔티브를 수행하고 있다. 이렇듯, 매년 웹 아카이빙 이니셔티브의 수가 증가하고 있으며 OECD 국가 가운데 절반 이상이 하나 이상의 웹 아카이빙 이니셔티브를 수행하고 있는 것은 웹 아카이빙의 중요성을 반증하는 결과로 해석할 수 있다.

웹 아카이빙의 성공적인 구축을 위해서 디지털 아카이빙을 위한 기반 환경이 마련되어야 한다. 디지털 정보의 관리와 검색을 조력하는 표준 메타데이터를 선정하고 표준 메타데이터의 적용방안을 마련하는 것이 이에 해당할 것이다. 이를 통해 기존 데이터와 보존 데이터 간 무리없이 소통이 가능한 상호운용성을 확보하게 될 것이다.

기존 연구들을 살펴보면 다양한 영역에서 자원의 장기보존 목적으로 PREMIS를 활용하려는 노력이 있었음을 알 수 있다(서은경 2008; 이경남 2006; 오삼균, 권기성 2007; 김판준 2010; 문주영, 김태수 2011). 그러나 PREMIS를 개발한 뒤 직접 데이터에 적용해봄으로써 적용가능성을 논의하고 있는 부분이 많지 않았으며 특히 PREMIS에서 지적개체(intellectual entity), 즉 목록데이터와 어떻게 연결해야 하는지에 대한 연구가 많이 이루어지지 않았다. 이러한 부분은 실제 도출된 PREMIS 요소를 기존 목록요소에 적용하고자 할 경우에 한계로 지적될 수 있다. 그러므로 보존 메타데이터 스키마를 개발하는 것에 그치지 않고 기존 목록 데이터에 보존 데이터를 어떻게 적용할 것인지에 대한 고민이 요구된다.

이에 본 연구는 사실상 보존규칙으로 받아들여지는 PREMIS를 기반으로 보존 메타데이터 스키마를 개발하는 것을 목적으로 하는 것은 물론 기존 데이터에 보존 메타데이터를 활용할 수 있는 가이드라인을 제시하고자 한다. 이를 위하여 국립중앙도서관 디지털 자원을 사례로 채택하였다.

국립중앙도서관은 국가 대표 도서관으로서 가치 있는 인터넷 자료를 국가적인 차원에서 수집·축적하여 미래 세대에 연구 자료 제공과 온라인 디지털자원 수집 보존의 표준모델 연구 등에 그 목적을 두고(2010년도 국립중앙도서관

www.kci.go.kr

연보) 공개용 온라인 디지털자원 수집・보존 사업인 OASIS(Online Archiving & searching Internet Sources) 프로젝트를 추진하는 것은 물론 판매용 디지털 자원의 보존 프로젝트를 통해 디지털 자원의 보존을 수행하고 있다.

이렇듯 디지털 자원의 보존이 수행되고 있음에도 불구하고 자원을 보존하기 위한 입력규칙이 마련되어 있지 않고 있어, 현재 시스템에서 보존 데이터를 기록할 수 있는 입력 틀이 부재한 상황이다. 보존은 자원의 보존은 물론 자원의 보존 활동과 관련한 메타데이터를 함께 관리하는 것임(한국기록관리협회 2009, 35)을 고려할 때 현재 국립중앙도서관이 수행하고 있는 디지털 자원의 보존을 보다 효율적으로 수행하기 위해서 디지털 자원의 보존데이터를 입력하기 위한 틀이 제시되어야 할 것이다. 이와 함께 기존의 목록 레코드에 보존 메타데이터를 어떻게 취합할 것인지에 대한 가이드라인이 제시되어야 한다.

이에 본 연구는 PREMIS를 기반으로 보존 메타데이터 스키마를 개발하는 것은 물론 디지털 도서관 영역의 범용 서지정보 표준 메타데이터 체계인 MODS(Metadata Object Description Schema)와 디지털 자원의 메타데이터 인코딩 표준인 METS(Metadata Encoding and Transmission Standard)를 기반으로 한 국립중앙도서관의 기존 목록 레코드에 보존 메타데이터를 활용할 수 있는 가이드라인을 제시하고자 한다. 이를 위하여 실제 아카이브 이니셔티브에 PREMIS를 적용하였던 사례 및 그 각 요소의 활용을 조사한 뒤, 국립중앙도서관 디지털 자원 및 보존정책에 대한 분석을 바탕으로 요소의 적용가능성 및 적용 값을 도출하였다 또한 PREMIS를 기존 METS에 적용하였던 사례 및 미국회도서관에서 제시하고 있는 지침을 기반으로 국립중앙도서관 디지털 자원에 적용방안을 제시하였다.

이는 추후 다른 국내외 아카이브 이니셔티브들이 스키마를 개발하고 운용하는데 초석이 될 것이며 기존 데이터의 입수 및 반출을 위해 기존 데이터에 PREMIS 기반의 보존 메타데이터를 첨가하는데 기초 가이드라인으로 활용될 수 있을 것이다.

2. PREMIS

PREMIS는 OAIS(Open Archival Information System)(ISO 14721) 참조모델에 기반을 둔다. OAIS 참조모델은 Open Archival Information System의 약자로 디지털 정보를 장기보존 하는데 요구되는 시스템, 즉 아카이브를 위한 개념적 구조 틀을 마련한 ISO 표준이다.

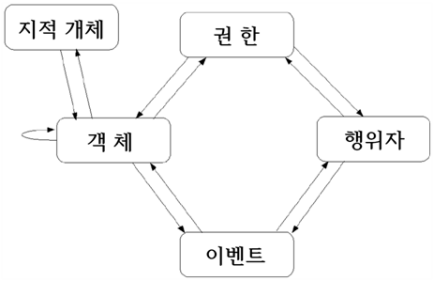
OAIS 참조모델의 의미는 수년에 걸쳐 서로 다르게 사용되던 디지털 장기보존에 관한 용어 및 기능모델에 대한 합의를 도출하였다는 점이다. 그러나 OAIS를 기반으로 실제 보존 시스템을 구축하기에는 한계를 지닌다. OAIS는 개념적 프레임워크를 제시한 것이므로 실제 적용 시에는 변환의 과정을 거쳐야 하기 때문이다. 이에 OCLC/RLG는 2003년 각국의 전문가로 구성된 실무그룹을 조직하여, 다양한 실무에서 공통된 요소를 찾아 "핵심(Core)" 요소를 개발하고자 하였다. 이러한 노력의 결과로 2005년 디지털 보존을 위한 데이터 사전을 발행하였으며, 데이터 사전이 발행된 후 장기보존을 위한

www.kci.go.kr

메타데이터 및 데이터 모델이 보완되어, 2008년 3월 데이터사전 제2판(Data Dictionary for Preservation Metadata Version 2.0)이 발행되었으며, 2011년 1월 데이터사전 V2.1이 발행되었다(PREMIS Data Dictionary for Preservation Metadata 2.1).

PREMIS 데이터 사전 V2.1은 데이터 모델과 데이터 사전으로 구성된다. 데이터 모델은 PREMIS 데이터 사전 내에 정의되어 있는 요소들을 논리적으로 조직하기 위한 것으로 지적 개체(Intellectual Entity), 객체(Object), 이벤트(Event), 행위주체(Agents), 권한(Rights)의 개체유형으로 구성된다(〈그림 1〉 참조). PREMIS 데이터 모델에서 정의하고 있는 개체 중 지적 개체는 관리와 기술 목적에서 단일한 지적 단위로 여겨지는 일련의 콘텐츠로 기술 메타데이터로 설명되므로 보존 메타데이터 요소에서는 제외되었다. 객체는 디지털 형태에 관한 정보이며 객체의 관리를 위하여 관련된 특성을 기술하는 정보이다. 이벤트 개체는 하나 이상의 객체나 행위자를 포함하는 활동에 관한 정보이다. 행위주체 개체는 객체의 생애주기에서 발생하는 보존과 관련된 활동과 권한과 관련이 있는 사람, 기관 또는 시스템에 대한 특성에 관한 정보이다. 마지막으로 권한 개체는 저작권이나 지적 재산권법에 따라 객체 또는 행위주체에게 적용되는 권한에 관한 정보이다(국가기록원 2007).

PREMIS 데이터 사전은 각 요소에 대하여 요소의 객체, 이벤트, 행위자, 권한에 대해 이름, 하위요소, 정의, 데이터 제약사항, 객체 범주, 적용 가능성, 예, 반복 가능성, 강제성, 생성/유지 안내, 사용안내를 사용하여 정의하고 있다.



〈그림 1〉 PREMIS 데이터 모델

3. 선행연구

서은경(2005)은 보존 메타데이터는 사용자가 보다 쉽게 진본데이터에 접근하고 검색 할 수 있으며 관리자는 보다 효율적으로 정보자원을 관리하고 보존해야 함을 논하였다. 또한 디지털자원의 보존을 위해 각국의 국립도서관이나 대표기관의 프로젝트에서 제시한 보존메타데이터를 중심으로 각 요소를 비교 분석하였다. PREMIS를 비롯하여 CEDARS Project, OCLC/RLG Working Group, NEDLIB Project, Natural Library of Australia(NLA) 표준안, National Library of New Zealand(NLNZ) 표준안을 비교 분석함으로서 우리나라의 실정에 맞는 보존 메타데이터 요소세트를 제안하기 위해 대학도서관을 대상으로 보존 메타데이터의 구축 현황과 각 요소의 적합성을 설문조사 및 보고하였다. 연구자는 설문조사 결과와 외국 사례조사를 바탕으로 7가지 범주(서지정보, 관리정보, 구조정보, 기술정보, 통제정보, 맥락정보, 출처정보)아래 36개 항목으로 구성된 보존 메타데이터 요소세트를 제안하였다.

이경남(2006)은 디지털 정보의 장기적 보존

www.kci.go.kr

에서 부각되는 보존 메타데이터의 국제적인 연구 성과들을 검토하고 그 가운데 기존의 메타데이터와는 개념적인 적에서 보다 실무적인 것으로 진전한 형태인 호주 빅토리아주 기록보존소의 VERS(Victorian Electronic Records Strategy)와 PREMIS 데이터사전을 중심으로 비교 분석하여 전자기록의 장기적 보존을 위한 메타데이터 요소를 제안하였다.

오삼균, 권기성(2007)은 전통예술분야의 아카이빙을 위한 보존 메타데이터 스키마를 설계하기 위해 국립국악원 국악자료를 바탕으로 전통예술자료의 특성을 분석하고 보존과 활용에 필요한 최적의 보존 메타데이터 스키마를 제안하였다. 먼저 PREMIS를 비롯한 각종 보존 메타데이터와 멀티미디어 자원에 관련된 메타데이터를 분석하고 실제 시스템을 구현하는데 있어 실제적인 가이드라인을 제시하고 있는 PREMIS 데이터사전을 기본 틀로 채택하였다. 또한 PREMIS 데이터 사전은 그 규모가 방대하기 때문에 선행연구를 비교 분석하여 공통되는 요소를 기본 요소로 제안하고, 객체에 대한 기술적 메타데이터는 추가적 요소로 제안하였다. 제안된 메타데이터세트는 크게 4개의 블록(Objects, Events, Agents, Rights)으로 나누어 구성하였다.

김판준(2010)은 국가정책연구 과정에서 생산 및 관리되는 연구정보의 장기적인 보존 및 접근을 보장하기 위한 보존 메타데이터 요소를 위해 유럽의 표준인 CERIF(Common European Research Information Format)와 PREMIS 데이터사전의 요소들을 비교 분석하고 양자의 특성을 반영한 상호보완적 보존 메타데이터 요소를 도출하였다.

문주영, 김태수(2011)는 PREMIS 데이터사전을 사무문서에 적용하기 위한 사무문서 콘텐츠 모형을 설계하였다. 일반 기업체의 사무문서 중에서도 특히 보존가치가 높은 문서를 대상으로 디지털 사무문서의 특성과 보존 필요성에 대해 서술하고 PREMIS 데이터모델을 적용하기 위해 PREMIS 모델 내의 지적개체에 대한 문서 차원의 개념 정립과 이해로 문서를 구조화하고 기록의 계층적 기술을 위해 계층 수준을 컬렉션, 접근, 폴더, 아이템, 표현, 파일로 구분하였다. 즉, 계층 수준으로 개념적 단계(아이템 계층 위)와 물리적 단계(표현, 파일)로 나누었고 본 연구에서는 사무문서의 콘텐츠를 구분하기 위한 폴더와 디지털 파일을 구분하였다. 이는 PREMIS의 지적개체와 객체개체를 조금 더 세분화하여 사무문서에 적용하였다는 점에서 그 의의가 있다고 할 수 있다.

기존 연구들을 종합해볼 때 다양한 영역에서 자원의 장기보존 목적으로 PREMIS 활용하려는 노력이 있었음을 알 수 있다. 그러나 PREMIS를 개발한 뒤 직접 데이터에 적용해봄으로써 적용가능성을 논의하고 있는 부분이 많지 않았으며 특히 PREMIS에서 지적개체(intellectual entity)와 연결을 어떻게 구현해야 하는지에 대한 연구가 많이 이루어지지 않았다. 이러한 부분은 실제 도출된 PREMIS 요소를 기존 목록요소에 적용하고자 할 경우에 한계로 지적될 수 있다. 그러므로 보존 메타데이터 스키마를 개발하는 것에 그치지 않고 기존 목록 데이터에 보존 데이터를 어떻게 적용할 것인지에 대한 고민이 요구된다.

www.kci.go.kr

4. 방법론

본 연구는 보존 메타데이터 설계를 위해 국립중앙도서관 디지털정보자원 현황 분석, 국립중앙도서관 디지털자원 목록규칙 및 보존정책분석, 해외 아카이브 PREMIS 사용현황 조사, PREMIS 기반 메타데이터 요소 도출, 검수, METS 인코딩 표준 기반 PREMIS 적용방안을 마련하는 단계를 거쳤다. 최종 메타데이터를 개발하였으며 이를 위해 기존 데이터의 대한 분석은 물론, 국립중앙도서관 디지털정보자원 보존 담당자와의 면담이 이루어졌다. 본 연구에서 사용된 메타데이터 설계원칙 및 분석 사례는 다음과 같다.

4.1 메타데이터 설계원칙

메타데이터 설계원칙으로 첫째, OAIS 참조모형을 준수하며 보존 메타데이터를 설계하기 위하여 제시된 표준 PREMIS 스키마를 준용하였다. 이는 PREMIS를 사용하는 시스템과의 호환성을 보장하기 위함이다. 둘째, 경제성의 원칙을 준수한다. 이는 이미 MODS 요소를 통해 지적개체가 기술되고 있으며 국립중앙도서관의 경우, 개방성 및 공공성의 원칙아래 접근금지 조항 및 세부 사항에 대한 정책 등 무결성을 보장하기 위한 정책이 엄격하지 않은 것을 고려하여, 최소한의 요소만을 제시하도록 하였다. 셋째, 더블린 코어 추상모델에서 제안하는 덤-다운(Dumb-down)원칙을 준수한다. 일반적으로 '간략 DC(Simple Dublin Core)'라는 말의 의미는 각 문장이 값 문자열만을 수록하고 있고, 부호화 체계와 요소 세목을 사용하지 않은 DC 메타데이터를 지칭할 때 사용되며, 반대로 '확장 DC(Qualified Dublin Core)'라는 말은 여기서 설명한 추상 모델의 모든 자질을 사용한 메타데이터를 지칭하는 데 사용된다. 확장 DC를 간략 DC로 변환하는 과정을 일반적으로 '더밍-다운'(dumbing-down)이라고 한다.

4.2 PREMIS 사례연구

PREMIS를 사용하는 프로젝트 중 기존에 이미 METS를 기반으로 목록레코드를 구축해오고 있으며 보존 메타데이터 입력을 위해 PREMIS를 추가하고 있는 사례를 조사하였다. 이는 국립중앙도서관과 같이 이미 기 구축된 목록데이터, 즉 PREMIS 데이터 모델에서 제시하는 지적개체에 대한 부분이 이미 제시된 프로젝트를 의미한다. 이러한 선정 기준에 따라 핀란드 국가도서관, HATHI TRUST, 플로리다 디지털 아카이브(이하 FDA), MathArc 프로젝트, SHERPA 디지털 리포지터리의 PREMIS 요소 사용 현황을 조사하였다.

핀란드 국가도서관(이하 NLF)은 단행본 디지털화 사업을 계획하고 METS 프로파일을 사용하여 목록레코드를 구축해오다가 보존정보를 기록하기 위하여 PREMIS를 첨가하고 있다. PREMIS가 포함된 METS프로파일과 PDF 파일, 각각의 디지털화된 페이지 등이 METS 패키지를 구성할 계획이다. 현재 프로파일은 아직 구축된 상태는 아니지만 2009년 샌프란시스코에서 열린 PREMIS 구현박람회에서 PREMIS 정보를 METS에 어떻게 결합시킬지에 대한 아이디어, 가능한 이벤트, 행위자, 권한 정보 값 등에 대한 아이디어를 공유한 바 있다.

HATHI TRUST는 2008년을 시작으로 미국

www.kci.go.kr

미시간 주립대학과 캘리포니아 주립대학이 인터넷 아카이브 콘텐츠의 입수 프로세스를 개발하기 위하여 공동 진행하고 있는 프로젝트이다. 이 양 기관은 디지털자원의 공유를 목적으로 하며 현재 30개가 넘는 기관이 파트너로 참여하고 있다. 이 프로젝트는 디지털화를 위해 METS를 사용하고 있으며 보존의 목적으로 PREMIS를 포함할 것을 제시하고 있다. 또한 프로젝트 홈페이지를 통해 METS 프로파일을 공개하여 METS 프로파일에서 PREMIS가 어떻게 활용되고 있는지를 공유하고 있다.

플로리다 디지털 아카이브는 플로리다 주의 교육, 학습, 연구를 지원하고 디지털 자원의 장기보존 리포지터리를 제공하는 것을 목적으로 하고 있으며, 오픈 소스 보존 리포지터리 소프트웨어 어플리케이션인 DAITSS(Dark Archive in the Sunshine State)를 사용하여 개발하고 있다. FDA DAITSS 사이트를 통해 DAITSS 보존과정 및 보존 서비스는 물론 AIP단계의 디스트립터를 살펴볼 수 있다. 플로리다 디지털 아카이브 역시 METS기반이며, DAITSS 1.0버전에서는 PREMIS를 부분적으로 수용하였으나 DAITSS 2.0에서는 PREMIS를 전적으로 수용하고 있다.

MathArc 프로젝트는 코넬 대학과 SUB Gottingen가 공동으로 진행하고 있는 연구프로젝트로 미국과학재단으로부터 지원을 받아 진행하고 있는 연구프로젝트이다. 본 연구는 수학 및 통계분야의 디지털 연속간행물의 장기보존 프로토타입을 구현하는 것을 목적으로 한다. 기본 원칙은 OAIS 참조모델을 기반으로 하고 있으며, 이를 구현하기 위하여 METS를 사용하고 METS에서도 PREMIS를 수용하기 위한 시도를 하고 있다. 또한 홈페이지를 통해 AIP교환을 통한 메타데이터 스키마, PREMIS 요소가 수용된 샘플 문서를 살펴볼 수 있다.

SHERPA 디지털 리포지터리는 영국의 예술 및 인문분야 학술정보를 위한 장기 보존 프로젝트이다. 시스템은 공유할 수 있는 보존 저장소, 보존 계획, 보존 기능을 수행함과 동시에 접근을 제공할 수 있는 서비스를 제공하고 있다. 홈페이지를 통해 프로젝트의 상세사항을 살펴볼 수 있으며 보존 메타데이터 세트에 대한 문서 역시 살펴볼 수 있다. SHERPA 디지털 리포지토리 역시 METS를 기반으로 지적개체에 대한 목록을 기술하고 있으며 PREMIS를 수용하기 위한 시도를 하고 있다.

5. PREMIS 주요요소 분석

5.1 국립중앙도서관 디지털 정보자원 수집 및 보존현황

국립중앙도서관의 디지털 자원의 유형, 특성, 현황, 아카이빙 대상 자원을 분석하였다. 국립중앙도서관은 OASIS(Online Archiving and Searching Internet Sources) 프로젝트를 통해 수행중인 공개용 온라인 디지털 자원 및 디브러리를 통하여 서비스를 제공 중인 판매용 온라인 자료를 수집하고 있으며, 수집된 모든 디지털자원에 대하여 보존하고 있다.

공개용 온라인 디지털 자원에 대한 수집 및 보존을 살펴보면, 국립중앙도서관은 2001년 온라인저작물 수집・보존 시스템을 구축하고 2004년부터 웹 자료를 수집, 2006년 2월부터 OASIS

www.kci.go.kr

홈페이지를 통해 대국민 서비스를 실시하고 있다. 최근 2010년부터 OASIS를 위해 국제적으로 통용되고 있는 수집로봇인 Heritrix를 사용하여 온라인 자료를 수집하고, 수집된 자료를 기술하기 위하여 메타데이터 포맷 역시 더블린 코어에서 MODS로 변경하여 확장성을 확보하고 있다. 이 결과 2010년에는 웹문서 14만 4,452건, 웹사이트 7,365건을 수집하여 총 15만 1,817건의 공개용 온라인 자료를 수집하였다(2010년도 국립중앙도서관 연보 2010, 66-68).

현재 국립중앙도서관에서 수집하고 있는 공개용 온라인 자료는 전 분야의 웹사이트와 연구보고서, 간행물, 정책자료, 통계자료 등 주로 웹사이트 내 공개자료실에서 제공되는 웹 문서로 그 수집범위는 한국에 관한 것, 한국 도메인(~kr)상에 있는 것, 한국과 관련되며 중요한 주제이거나 한국저자에 의해 쓰여진 것, 해외 온라인 자료 중 한국 사람에 의해서 쓰여지거나 한국을 주제로 다룬 것 등을 포함한다. 자료 형태로는 HTML, PDF, PPT, ZIP JPEG, HWP, DOC 등 다양한 형태의 자료가 수집・보존되고 있다.

판매용 온라인 자료는 디브러리(http://www.dibrary.net)를 통하여 제공되고 있다. 2009년 도서관법이 개정된 이래 공개용 온라인 자료뿐만 아니라 전자책, 전자잡지와 같은 판매용 온라인 자료에 대해서도 본격적으로 수집하고 있다. 2010년 기준으로 전자책 30,410책, 학회지 570종 702,459건, 이미지 16종 52,452건, 방송자료 15종 455건, 오디오북 69종 143건, 이러닝 자료 28종 122건을 수집하였다.

판매용 온라인 자료는 한국전자출판협회 인증목록 및 유통업체 판매목록을 활용하여 수집대상 자료를 선정하며, 선정된 자료에 대하여 유통업체 및 제작사에 자료 제공을 요청한다. 국립중앙도서관은 제공받은 원본파일을 인터넷이 차단된 별도의 서버에 저장하여 영구 보존하고, MODS를 기반으로 메타데이터를 구축하여 장서관리시스템에 입력한다. 메타데이터 항목으로는 자료 유형마다 약간의 차이가 있지만 제목, 저자, 자료유형, 장르, 발행지, 발행자, 발행년, 언어, 자료 형태, 분류기호, 페이지 수, URL 정보, 파일명 등이 입력된다. 국립중앙도서관은 제공받은 자료에 대한 증명으로 온라인 자료 제공자에게 도서관자료 수집증명서를 발급하고 온라인 자료 제공자가 제출한 도서관자료 보상청구서를 토대로 수집 온라인 자료에 대한 보상금을 지급한다.

이상과 같이 국립중앙도서관의 디지털 자료의 분석결과 PDF, HTML페이지는 물론 JPEG, TIFF를 포함하는 다양한 형태를 포함하는 것은 물론 방송자료, 오디오북, 전자책, 웹사이트 등 다양한 유형의 자료를 보존할 수 있는 메타데이터 스키마가 필요하다는 것을 알 수 있으며 디지털 자원의 목록규칙을 분석한 결과, 〈표 1〉과 같은 상위요소가 사용되고 있는 것을 파악할 수 있었다.

5.2 기존 메타데이터 분석을 통한 요소 추출 및 확장

국립중앙도서관 디지털 자원을 보존하기 위한 메타데이터 요소를 기존 메타데이터 분석을 통하여 추출하게 되었는데, 〈표 1〉에서 나타난 바와 같이, 첫째 분석된 프로젝트에서 공통적으로 사용되는 요소를 추출하였다. 둘째 추출

www.kci.go.kr

〈표 1〉 MODS 사용요소

MODS Elements	요소명	비고	MODS Elements		
〈titleInfo〉	표제정보	표제, 대등표제, 부표제 등	〈note〉	서지적주기	
〈name〉	저자정보	저자, 역자, 삽화가, 공헌자 등	〈subject〉	주제	
〈typeOfResource〉	콘텐츠유형	텍스트, 전자저널 등	〈classification〉	분류기호	
〈genre〉	장르	단행본, 연속간행물, 보고서, 학술논문	〈relatedItem〉	연관정보	기타 형태 관계
〈originInfo〉	출처정보	출판사 정보 등	〈identifier〉	식별기호	
〈language〉	언어		〈location〉	소재정보	URL
〈physicalDescription〉	형태기술정보	페이지수, 자료형태(HWP, PDF, DOC 등)	〈accessCondition〉	접근제한정보	외부공개, 외부비공개 여부
〈abstract〉	요약정보(초록)		〈part〉	내용정보	
〈tableOfContents〉	요약정보(목차)		〈extension〉	부가정보	
〈targetAudience〉	이용대상자	일반이용자용	〈recordInfo〉	레코드정보	레코드 소스 및 생성일, 식별자 등

된 공통 요소 중에서 국립중앙도서관 보존정책 및 디지털 자원에 대한 분석을 바탕으로 핵심적으로 요구되는 요소를 선별하고 불필요한 요소를 배제하였다. 셋째 도출된 요소를 바탕으로 보존 담당자와의 면담 및 보존 메타데이터 적용을 통해 필요하다고 판단되는 요소를 선별하고 추가하여 확장하였다.

PREMIS 주요 요소의 사용에 대하여 분석한 결과, 다음 〈표 2〉와 같이 나타났다.

- 객체식별자(objectIdentifier)는 모든 프로젝트에서 구현하고 있는 요소이며 내부 식별자를 사용하기도 하며 경우에 따라서는 표준 형태의 식별자를 생성하기도 한다.
- 보존수준(preservationLevel)은 프로젝트에 따라 사용여부가 다르다. 또한 그 값 역시 객체 생성주기, 파일 포맷의 보존 또는 보존권한의 적절성, 포맷 파일 포맷의 기대 수명 등 시스템의 목적에 따라 달리할 수 있다. 프로젝트에 따라서는 0 또는 1로 값이 일정하게 정해진 경우도 있으며 비트레벨(Bit-Level), 콘텐츠 레벨(Content-Level), 전체(Full)로 구분하고 있다.
- 객체범주(objectCategory)는 대부분의 프로젝트에서 사용하는 있는 요소이며 표현, 파일, 비트스트림 등 OAIS에서 제시하고 있는 정보모델을 값으로 사용하고 있다.
- 중요속성(significantProperties)은 PREMIS에서도 정의가 명확하게 되어있지 않은 부분으로, 프로젝트에 따라 사용여부가 다르며 부여되는 값에 있어서도 차이가 났다. 예를 들어 SHERPA 프로젝트의 경우 이 요소를 포맷의 하위요소로 지정하고 있는 반면, HATHI TRUST의 경우는 파일 페이지 수를 표시하였다.
- 구성단계(compositionLevel)의 경우는 압축 또는 암호화된 데이터의 수집여부에 따라 그 값을 달리한다. 대부분 압축된 데이터를 보존하지는 않으므로 이 요소를 사용하는 경우도 “0” 또는 “00”으로 처리하였으며 그 외에 정책상 압축된 데이터를 보존하지 않는 경우 전혀 사용하지 않는 경우도 있었다.

www.kci.go.kr

〈표 2〉 메타데이터 요소 비교분석

메타데이터 요소/프로젝트	SHERPA	MathArc	FDA	NLF	HATHITRUST
Object Entity					
objectIdentifier	○	○	○	○	○
- preservationLevel	○	○			○
- objectCatetory	○	○	○	○	○
significantProperties	○				○
objectCharacteristics	○	○	○	○	○
- compositionLevel	○	○	○	○	○
- fixity	○		○	○	○
- messageDigestAlgorithm	○		○		
- messageDigestOriginator	○		○		
- messageDigest	○				
size	○		○		
format	○	○	○	○	○
- formatDesignation	○	○	○		
- formatName	○		○		
- formatVersion	○		○		
- formatNote			○		
- formatRegistry	○		○		
inhibitors	○				
- inhibitorType	○				
- inhibitorTarget	○				
- inhibitorKey	○				
creatingApplication	○	○	○		
- creatingApplicationName	○	○	○		
- creationApplicationVersion	○	○			
- dateCreatedByApplication		○	○		
originalName		○			
environment		○			
- environmentCharacteristics		○			
- dependency		○			
- software		○			
- hardware		○			
relationship	○	○	○	○	○
- relationshipType	○	○	○	○	○
- relationshipSubType	○	○	○	○	○
- relatedObjectIdentification	○	○	○	○	○
- relatedObjectSequence	○	○			
- relatedEventIdentification	○	○	○	○	○
- linkingAgentIdentifier			○		○
- linkingAgentRole					○
- linkingObjectIdentifier			○		
Event Entity					
eventIdentifier	○	○	○	○	○
eventType	○	○	○	○	○
eventDateTime	○	○	○		○

www.kci.go.kr

메타데이터 요소/프로젝트	SHER PA	MathArc	FDA	NLF	HATHI TRUST
eventDetail	○				
eventOutcomeInformation	○		○		
eventOutcome	○	○	○		
eventOutcomeDetail	○	○			○
Agent Entity			○	○	○
agentIdentifier			○		○
agentName			○		○
agentType			○		○
agentNote			○		○
Right Entity			○	○	

- 고정성(fixity)의 경우, "체크섬" 또는 "불변성"을 체크하기 위한 것이다. 체크섬은 제대로 보존 아카이브로 전송되었으므로 이후 고의적 또는 무의식적으로 변경되지 않았음을 인증하는 중요한 요소이다. 불변성 체크(fixity check)를 수행하기 위해서 이전날짜에 계산된 체크섬과 나중에 계산된 체크섬을 비교하게 된다. 또한 이를 확인한 날짜를 이벤트로 기록하며, 기록의 결과를 이벤트결과(eventOutcome)에 기록한다. 대부분의 저장소가 적어도 하나의 체크섬 알고리즘을 사용하고 있으며 가장 보편적인 형태가 MD5와 SHA-1이다.
- 사이즈(size)는 모든 저장소에서 사용되고 있으며 바이트(bytes)의 형태로 기록된다. 보통 파일 크기는 시스템 기능을 통해서 자동 기록이 가능하다.
- 포맷(format)의 경우 모든 저장소에서 사용되고 있는 요소로 PREMIS 데이터 사전은 포맷 지시자(formatDesignation)이나 포맷 레지스트리(formatRegistry)를 사용할 것을 요구하고 있다. 저장소에 따라 복잡한 포맷 식별 경로를 개발하기도 하고, 가장 간단하지만 정확성이 완전히 보장되지 않은 방법을 사용하기도 한다.
- 금지사항(inhibitors)은 특정사용을 금지하는 등 금지사항을 기록하기 위함이다. 이 요소를 사용하는 저장소는 많지 않으며 필요할 경우 중요한 속성(significant properties)에 정보를 저장하고 있다.
- 생성어플리케이션(creatingApplication)은 디지털 객체를 생성하는데 사용된 어플리케이션을 기록하기 위함이다. 대부분의 저장소에서 이 요소는 사용되고 있으나 데이터의 생성방법은 다소 차이가 있다. 이 정보는 파일 헤더에 저장되거나 메타데이터 추출 도구를 사용하여 자동 추출이 가능하기 때문에 자동으로 생성되는 경우가 많으나, 이 경우에는 생성 날짜 정보가 헤더의 정보와 PREMIS 데이터 사전이 권고하고 있는 ISO 표준이 일치할 때만 의미가 있다 할 수 있다. 이러한 이유로 저장소에서 수동으로 가능한 한 자세한 정보를 저장할 것을 제시하기도 한다.
- 원본명(originalName)은 대부분의 저장소에서 입수 단계에서 원본 파일 이름을 수

www.kci.go.kr

집하고 있으므로 필수 요소임에도 대부분의 저장소에서 사용하지 않는다.

- 저장소(storage) 정보는 비록 대부분의 저장소가 객체를 어떻게 저장하는지에 대해서 알고 있음에도 불구하고, 메타데이터를 사용하여 분명하게 명시하는 경우는 거의 없었다. 콘텐츠의 위치 값이나 파일 위치 값을 기록할 수 있다.
- 환경(environment)에 대해서는 객체와 상호작용하거나 렌더링하기 위해 요구되는 환경을 의미한다. PREMIS 데이터 사전에서도 분명하게 명시되어 있지 않으며 사용하고 있는 저장소도 많지 않다. 사용하는 경우는 대부분 객체의 유형정보, 소프트웨어, 하드웨어를 포함하는 기술정보를 연결하기 위하여 사용하고 있다.
- 서명정보(signatureInformation)는 거의 사용하고 있는 저장소가 없었다.
- 관계(relationship)는 가장 다양한 방법으로 저장소에 의해 구현되는 방법이다. MathArc는 현재 구체적인 관계를 기록하지 않고 있으며, 파일 사이의 구조적 관계는 METS structMap을 통해 XML에서 저장하고, 관계 요소는 출처정보를 표현하기 위하여 유도(derivation)관계를 기록하기 위해서만 사용된다. FDA관계는 구조적 관계(structural)인 "has part"와 "is part of"관계를 통해 연결된다. 그러므로 파일 간의 형제관계는 이들 둘 사이에서 추론될 수 있다. SHERPA 프로젝트가 관계요소를 비교적 상세하게 규정하고 있는데, 관계유형을 구조, 유도, 의존관계(structural, derivation, dependency)로 표현하고, 하위유형으로 "is child of", "is parent of", "is part of", "has part", "source of" 등으로 표현하고 있다.
- 이벤트 개체(event entity)는 저장소에 의해 일반적으로 기록되고 있으나 이벤트 유형이나 실제 구현에 대한 부분은 상당히 다양하다. 예를 들어, SHERPA 프로젝트의 경우에는 전체 이벤트 개체를 구현할 것을 제안하고 있으며 모든 요소들이 이벤트 과정 내에서 자동으로 기록될 것으로 기대하는 반면, FDA는 DAITSS 보존 저장소에 의해 수행되는 지역적인 이벤트에 한정하여 기록하고 있다.
- 이벤트 식별자(eventIdentifier)의 경우 대부분의 프로젝트에서 구현되고 있으며 이벤트 식별자는 관계와 이벤트를 연계하는 링크로 사용된다.
- 이벤트 유형(eventType)은 많은 프로젝트에서 구현되고 있으며, 이벤트 유형의 값 역시 PREMIS에서 제시하고 있는 리스트를 차용하는 것에서부터 시작하여 통제어휘로 확장하는 등 다양한 방법으로 구현되고 있다.
- 행위자(Agent)의 경우 대부분의 저장소가 행위자 개체 형태를 사용하고 있었으나 구현하는 방법에 있어서는 상당한 차이가 있었다. HATCHI 프로젝트는 행위자 식별자는 데이터베이스로부터 자동 추출하는 반면, FDA와 같은 프로젝트는 직접 기록하고 있다.
- 권한개체(Rights entity) 역시 행위자 개체와 마찬가지로 매우 다른 형태로 구현되고 있으며, PREMIS에서 지시하는 바와

www.kci.go.kr

일치되지 않는 경우도 있었다. FDA 경우는 권한 시맨틱 단위에 대응하는 모든 메타데이터 요소를 사용할 것을 권고하고 있으며, 이 개체를 표현단계와 연계할 것을 제시하고 있다. NDF 경우 역시 지적재산권 여부에서의 접근권한은 표시하고 있다.

다음 〈표 3〉은 앞서 밝힌 기본 요소 추출과정을 통해 선별된 요소이다.

〈표 3〉 공통된 기본 요소

개체	메타데이터 요소
Object Entity	objectIdentifier; objectCategory; objectCharacteristics(Composition Level); Format, Relationship(Relationship; relatedObjectIdentification, relatedEventIdentificatio) - 대소문자 구별
Event Entity	eventIdentifier, eventType

5.3 국립중앙도서관 보존방향 분석 및 요소 확장

종합적인 분석 결과 도출된 기본 요소를 전제로 담당자와의 면담을 통해 보존정책을 분석하였다. 이를 통해 국립중앙도서관에 필요 없는 요소는 제외하고 기본요소에서는 도출되지 않았으나 보존정책을 위하여 요구되는 요소를 확장하였다. 현재 국립중앙도서관은 확실한 보존정책을 세운 것은 아니며, 기본 원칙만을 정하고 있다. 그 기본원칙 및 메타데이터 요소 고려사항은 다음과 같다.

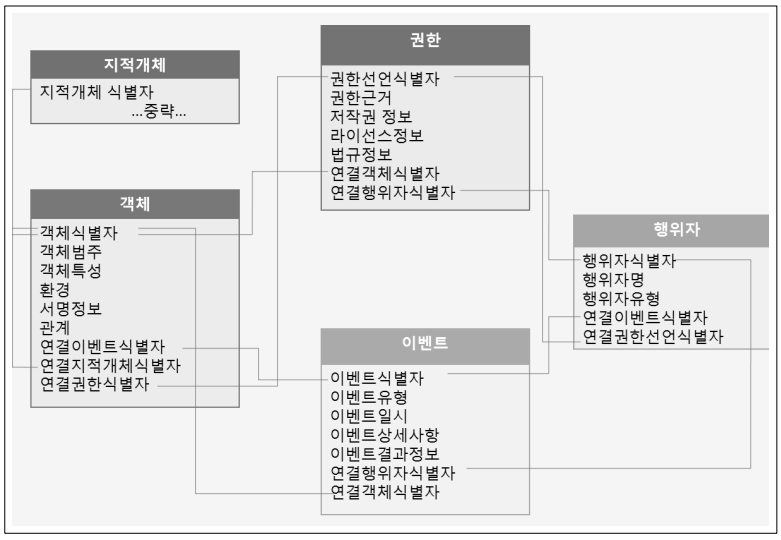
- 국가대표도서관으로서 수집하는 모든 디지털 자원은 유형에 상관없이 보존하는 것을 목적으로 한다.
- 국가대표도서관으로서 보존되는 모든 자료는 공공성을 가지므로 비밀번호 등을 표시하는 금지사항 요소는 요구되지 않는다.
- 저작권법, 장애인법에 따라 접근 제한되는 정보가 존재한다. 그러므로 권한개체를 사용하여 객체에 연관된 권한에 대해 정보를 제공한다.
- 행위주체에 대해서는 원칙적으로 이벤트와 관련한 행위주체가 국립중앙도서관 자체가 되므로 행위주체 정보가 요구되지 않을 수도 있으나, 무결성을 보장한다는 측면에서 행위주체정보는 요구된다.
- 대부분의 파일은 압축되지 않는다. 그러므로 기본적으로 구성단계는 압축되지 않는 파일을 표시하는 값을 기본 값으로 설정하도록 한다.
- 향후 마이그레이션 등 장기보존정책이 구현될 것을 고려하여, 생성어플리케이션, 환경정보, 불변성 등의 요소는 포함되어야 한다.

6. 보존 메타데이터 및 활용방안 제시

6.1 설계된 보존 메타데이터 요소

종합적인 분석 결과 설계된 메타데이터는 객체(object), 이벤트(event), 행위자(agent), 권한(rights)의 총 4개 분야로 나누어졌다. 각 분야의 메타데이터는 다양한 PREMIS 구현사례 및 국립중앙도서관 온라인 자료의 수집현황 및 정책분석을 바탕으로 디지털 자원의 기술메타데이터와의 연계, 객체 개체와의 유기적 연계가 가능하도록 설계되었다(〈그림 2〉 참조).

www.kci.go.kr



〈그림 2〉 메타데이터 구조

객체 개체에서는 PREMIS에서 필수요소로 규정하고 있는 객체 식별자, 객체범주, 객체특성(구성단계)를 포함하였으며 포맷, 생성어플리케이션, 환경, 서명정보, 관계의 요소를 포함하였다. 포맷정보는 MODS의 〈physicalDescription〉에서 형태기술정보를 기록할 수 있는 요소가 있으나 지적개체가 디지털 형태가 아니지만 보존되는 파일의 형태가 디지털인 경우를 고려하여 포함하였다. 환경정보는 현재와 같이 명료하며 단순한 디지털 정보자원만을 수집하는 상태에서는 필요성이 높지 않으나 향후 마이그레이션, 에뮬레이션, 인캡슐레이션과 같은 장기보존전략을 수립할 때 특히 에뮬레이터 개발에 필요한 정보이므로 포함하였다.

구성단계의 요소값으로 자료의 대부분이 압축되지 않은 파일 형태임을 반영하여 기본값을 "0"으로 제공하였으며, 포맷 레지스트리에 대한 정보 역시 국립중앙도서관이 절대적인 부분을 차지하므로 제외되었다. 또한 포맷노트는 PREMIS에서 선택요소이며 국립중앙도서관 디지털 자원의 포맷 정보가 명확하여 부가적인 정보를 기재할 필요가 없다고 판단되어 포함하지 않았다. 금지사항의 경우, 저작권의 이유로 원문사용이 금지되는 요소는 MODS의 〈accessCondition〉 요소에 기재되므로 포함하지 않았다.

이벤트 객체에서는 PREMIS에서 필수요소로 지정하고 있는 이벤트 식별자, 이벤트 유형, 이벤트 일시 등을 포함하였으며, 그 외에 이벤트 상세사항 및 이벤트 결과정보를 포함하였다. 현재 식별자 유형이 지정되지 않고 있음을 고려하여 이벤트를 지칭하기 위하여 "NDL Repository Event ID"를 초기 값으로 갖기로 하며 이벤트 값은 "e-yyyymmdd-0001"부터 자동 생성되는 일련번호 구조를 제시하였다. 이벤트 유형은 이벤트의 속성을 분류한 것으로 통제어휘를 추가하여 참조할 것을 제시하였으며, 이벤트 일시는 이벤트가 생성된 날짜와 시간을 기록하기 위한 것으로 가능한 한 가장 구체적인 시간을 기록

www.kci.go.kr

〈표 4〉 객체 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
객체식별자 (PREMIS:objectIdentifier)	객체식별자유형 (PREMIS:objectIdentifier Type)			1:1	DLC
	객체 식별자 값 (PREMIS:objectIdentifier Value)			1:1	http://nrs.harvard.edu/urn3:FHCL.Loeb:sa1
객체범주 (PREMIS:objectCategory)				1:1	file
객체특성 (PREMIS:object Characteristics)	구성단계 (PREMIS:compositionLevel)			1:1	0
	불변성 (PREMIS:fixity)	메시지요약 알고리즘 (PREMIS:messageDigest Algorithm)		1:1	SHA-1 MD5
		메시지요약 (PREMIS:messageDigest)		1:1	2e2107b67a1924419563
	사이즈(PREMIS:size)			0:1	2038937
	포맷 (PREMIS:format)	포맷지시자 (PREMIS:format Designation)	포맷명 (PREMIS: formatName)	1:1	PDF JPEG GIF HTML
			포맷버전 (PREMIS: format Version)	0:1	6.0
	생성 어플리케이션 (PREMIS:creating Application)	생성 어플리케이션명 (PREMIS:creating ApplicationName)		0:1	MSWord
		생성 어플리케이션버전 (PREMIS:creating ApplicationVersion)		0:1	2003
		생성일 (PREMIS:dateCreatedBy Application)		0:1	2003-03-19T12:25:14-05:00
환경 (PREMIS:environment)	소프트웨어 (PREMIS:software)	소프트웨어 명 (PREMIS:swName)		1:1	Adobe Acrobat Reader
		소프트웨어 버전 (PREMIS:swVersion)		0:M	6.0
		소프트웨어 유형 (PREMIS:swType)		1:1	
		소프트웨어 기타 정보 (PREMIS:swOther Information)		0:M	Install acroread(PREMIS: Adobe Acrobat)
		소프트웨어 종속성 (PREMIS:swDependency)		0:M	GNU gcc 〉= 2.7.2
	하드웨어 (PREMIS:hardware)	하드웨어명 (PREMIS:hwName)		1:1	Intel Pentium III 1 GB DRAM Windows XP-compatible joystick
		하드웨어 유형 (PREMIS:hwType)		1:1	
		하드웨어 기타 정보 (PREMIS:hwOther Information)		1:M	32MB minimum Required RAM for Apache is unknown

www.kci.go.kr

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
관계 (PREMIS:relationship)	관계유형 (PREMIS:relationship Type)			1:1	structural
	관계하위유형 (PREMIS:relationshipSub Type)			1:1	is part of
	연관객체식별 (PREMIS:relatedObject Identification)	연관객체 식별자유형 (PREMIS:relatedObject IdentifierType)		1:1	객체 식별자 예시 참조
		연관객체 식별자값 (PREMIS:relatedObject IdentifierValue)		1:1	객체 식별자 예시 참조
		연관객체 순서 (PREMIS:relatedObject Sequence)		0:1	객체 식별자 예시 참조
	연관이벤트식별 (PREMIS:relatedEvent Identification)	연관이벤트 식별자 유형 (PREMIS:relatedEvent IdentifierType)		1:1	이벤트 예시 참조
		연관이벤트 식별자값 (PREMIS:relatedEvent IdentifierValue)		1:1	이벤트 예시 참조
		연관이벤트 순서 (PREMIS:relatedEvent Sequence)		0:1	이벤트 예시 참조
연결이벤트 식별자 (PREMIS: inkingEventIdentifier)	연결이벤트 식별자유형 (PREMIS:linkingEvent IdentifierType)			1:1	이벤트 식별자 유형 예시 참조
	연결이벤트 식별자값 (PREMIS:linkingEvent IdentifierValue)			1:1	이벤트 식별자 값 예시 참조
연결지적개체 식별자 (PREMIS: inkingIntellectualEntity Identifier)	연결지적개체 식별자 유형 (PREMIS:linkingIntellectual IdentifierType)			1:1	지적개체 식별자 유형 예시 참조
	연결지적개체 식별자 값 (PREMIS:linkingIntellectual IdentifierValue)			1:1	지적개체 식별자 값 예시 참조
연결권한 식별자 (PREMIS: inkingRightsStatement Identifier)	연결권한선언 식별자 유형 (PREMIS:linkingRights StatementIdentifierType)			1:1	권한식별자 유형 예시와 동일
	연결권한선언 식별자 값 (PREMIS:linkingRights StatementIdentifierValue)			1:1	권한 식별자 값 예시 참조

하고 표준 시간대를 표기하는 것을 권장하므로 ISO 8601(yyyy-mm-ddThh:mm:ss)표준을 사용할 것을 제시하였다. 이벤트 결과정보 역시 PREMIS가 통제어휘를 사용할 것을 권고하고 있으므로 통제어휘를 제시하였다. 이상의 이벤트 개체 메타데이터 요소를 정리하면 다음 〈표 5〉와 같다.

www.kci.go.kr

〈표 5〉 이벤트 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
이벤트 식별자 (PREMIS:event Identifier)	이벤트식별자유형 (PREMIS:eventIdentifier Type)			1:1	Repository Event
	이벤트식별자값 (PREMIS:eventIdentifier Value)			1:1	e-yyyymmdd-0001
이벤트 유형 (PREMIS:event Type)				1:1	통제어휘(불변성 체크, 메시지 요약, 파일변환, 검증, 바이러스 체크 등)
이벤트일시 (PREMIS:event DateTime)				1:1	20050704T071530-0500
이벤트상세사항 (PREMIS:event Detail)				0:1	ISO 8601(2003-03-19T12:25:14-05:00)
이벤트 결과 정보 (PREMIS:event Outcome Information)	이벤트 결과 (PREMIS:eventOutcome)			0:1	통제어휘 (00, 01, 02 등)
	이벤트결과 상세사항 (PREMIS:eventOutcome Detail)	이벤트결과상세정보 (eventOutcome DetailNote)		0:M (상세사항) 0:NR (상세정보)	헤더에 비표준 태그 발견
연결 행위자 식별자 (PREMIS:linking AgentIdentifier)	연결 행위자 식별자 유형 (PREMIS:linkingAgent IdentifierType)			1:1	행위자 식별자 유형 예시 참조
	연결 행위자 식별자 값 (PREMIS:linkingAgent IdentifierValue)			1:1	행위자 식별자 값 예시 참조
연결 객체 식별자 (PREMIS:linking ObjectIdentifier)	연결 객체 식별자 유형 (PREMIS:linkingObject IdentifierType)			1:1	객체 식별자 유형 예시 참조
	연결 객체 식별자 값 (PREMIS:linkingObject IdentifierValue)			1:1	객체 식별자 값 예시 참조

행위자 개체는 PREMIS에서 필수요소로 지정하고 있는 행위자 식별자 외에 행위자 명을 포함하였다. 행위자명은 행위자의 텍스트 형태의 이름을 명시하는 것으로 객체 개체에 행해지는 모든 이벤트에 대한 행위자를 기록하기 위함이며 추후 객체에 행해지는 이벤트에 대한 책임

www.kci.go.kr

을 명확히 하여 기록의 무결정 및 진본성을 보장하는 것을 목적으로 한다. 그 외에 행위자유형 및 행위자 설명은 포함하지 않았다. 이는 모든 이벤트의 주체는 국립중앙도서관일 경우가 절대적이기 때문이다. 행위자 개체의 메타데이터 요소를 정리하면 다음 〈표 6〉과 같다.

권한 개체는 권한선언, 권한근거, 저작권 정보, 라이선스 정보, 법규 정보 등을 표시하였다. 권한근거는 PREMIS에서 통제어휘를 사용할 것을 권고하는 바, 통제어휘를 제시하였으며, 법규정보다 라이선스 정보는 많이 활용되는 요소는 아니나, 장애인법과 같이 법에 의해 권한이 부여되거나 통제되는 경우가 있는 것을 반영하기 위하여 포함하였다. 이상의 권한개체요소를 정리하면 다음 〈표 7〉과 같다.

〈표 6〉 행위자 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
행위자식별자 (PREMIS:agentIdentifier)	행위자식별자 유형 (PREMIS:agentIdentifier Type)			1:1	LCNAF SAN MARC 기관 코드 URI
	행위자식별자값 (PREMIS:agentIdentifier Value)			1:1	92-79971 Owens, Erik C. MH-CS /a000001 info:lccn/n78890351
행위자명 (PREMIS:agentName)				0:M	박성환 국립중앙도서관 FDA Virus Check Service
연결이벤트식별자 (PREMIS: inkingEventIdentifier)	연결이벤트식별자유형 (PREMIS:linkingEvent IdentifierType)			1:1	이벤트 식별자 유형 예시 참조
	연결이벤트식별자값 (PREMIS:linkingEvent IdentifierValue)			1:1	이벤트 식별자 값 예시 참조
연결권한선언식별자 (PREMIS: inkingRightsStatement Identifier)	연결권한선언식별자유형 (PREMIS:linkingRights Statement IdentifierType)			1:1	권한선언 식별자 유형 예시 참조
	연결권한선언식별자값 (PREMIS:linkingRights Statement IdentifierValue)			1:1	권한선언 식별자 값 예시 참조

www.kci.go.kr

〈표 7〉 권한 개체 메타데이터 요소

시맨틱 단위 (Semantic unit)	의미 요소 (Semantic components)			필수: 반복	예시
권한선언식별자 (PREMIS:rights StatementIdentifier)	권한문식별자 유형 (PREMIS:ightstatement IdentifierType)			1:1	
	권한문식별자 값 (PREMIS:ightsStatement IdentifierValue)			1:1	URI 로컬식별자
권한근거 (PREMIS:rightsBasic)				1:1	통제어휘(저작권, 라이선스, 법령)
저작권정보 (PREMIS:copyright Informaiton)	저작권상태 (PREMIS:copyrightStatus)			1:1	통제어휘(미상, 저작권, 저작권만료, 공개)
	저작권적용영역 (PREMIS:copyright Jurisdiction)			1:1	ISO 3166 (410 또는 KOR 또는 KR)
	저작권발효일 (PREMIS:copyrightStatus DeterminationDate)			0:1	20111011
	저작권설명 (PREMIS:copyrightNote)			0:M	저작원은 2010년에 만료될 예정 저작권에 대한 선언문은 파일 헤더에 포함되어 있음
라이선스정보 (PREMIS:license Information)	라이선스식별자 (PREMIS:licenseIdentifier)	라이선스식별자유형 (PREMIS:license IdentifierType)		1:1	URI 로컬식별자
		라이선스식별자값 (PREMIS:license IdentifierValue)		1:1	
	라이선스약관 (PREMIS:licenseTerms)			0:1	
	라이선스설명 (PREMIS:licenseNote)			0:M	라이선스는 파일헤더의 XMP 블록에 포함되어 있음
법규정보 (PREMIS:statue Information)	법규관할구역 (PREMIS:statueJurisdiction)			1:1	us de
	법규인용 (PREMIS:statueCitation)			1:1	도서관법 장애인법 National Library of New Zealand
	법규정보결정일 (statueInformation DeterminationDate)			0:1	2007-12-01 ISO 8601(2003-03-19T12:25:14-05:00)
	법규설명 (statuteNote)			0:M	
연결객체식별자 (PREMIS: inkingObjectIdentifier)	연결객체식별자유형 (PREMIS:linkingObject IdentifierType)			1:1	객체 식별자 유형 예시 참조
	연결객체식별자값 (PREMIS:linkingObject IdentifierValue)			1:1	객체 식별자 값 예시 참조
연결행위자 식별자 (PREMIS:linkingAgent Identifier)	연결행위자식별자 유형 (PREMIS:linkingAgent IdentifierType)			1:1	객체 행위자 유형 예시 참조
	연결행위자식별자 값 (PREMIS:linkingAgent IdentifierValue)			1:1	객체 행위자 값 예시 참조

www.kci.go.kr

6.2 PREMIS-METS 사용 가이드라인

현재 국립중앙도서관 디지털 자원의 목록은 서지정보 메타데이터 표준으로 MODS를, 인코딩 표준으로 METS를 기반으로 한다. 그러므로 PREMIS를 기 구축된 시스템에서 활용하는 방안이 제시되어야 할 것이다. MODS는 2002년 7월 미국 의회도서관의 Library of Congress Network Development and MARC Standard Office에서 발표한 것으로 MARC과 호환되는 XML 스키마 기반의 디지털 도서관 영역의 범용 서지정보 표준 메타데이터 체계이다. METS는 시스템간의 '복합디지털객체'의 전송 및 아카이빙을 위한 디지털 자원에 대한 허브(hub)문서의 인코딩 규칙을 규정하는 XML 스키마 기반의 명세(specification)로 미의회도서관에서 개발되었다.

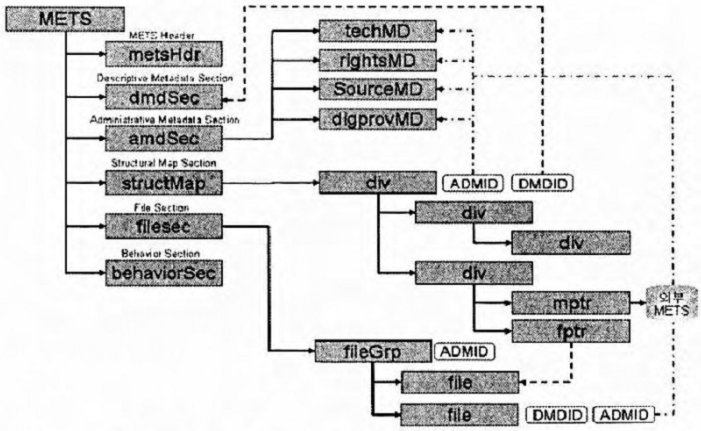
METS와 PREMIS를 동시에 사용하는 것은 가능하며 실제로 다양한 프로젝트에서 METS를 기반으로 목록데이터를 입력하고 PREMIS를 활용하여 보존 메타데이터의 입력을 시도하고 있다. 그러나 METS의 요소와 PREMIS 개체간의 이질성 및 중복성의 이유로 PREMIS 데이터를 METS와 같은 인코딩 포맷에 적용할 시에는 이러한 문제를 해결하는 방안이 필요하다. 이는 기관이 가지고 있는 자료의 유형 및 특성에 대한 이해를 바탕으로 해야 한다.

METS 스키마는 다음의 7개 영역으로 구성된다(〈그림 3〉, 〈표 8〉 참조).

PREMIS 시맨틱 단위를 METS 영역에 연결시키기 위한 지침(Guenther 2008) 및 미국 회도서관이 제시하고 있는 샘플레코드를 고려하여 볼 때 PREMIS와 METS 활용 시 다음의 네 가지를 고려하여야 한다.

6.2.1 PREMIS와 METS 매핑

- premis:object는 techMD나 digiProvMD에 기술하는데, 이는 객체기술내용이 대부분 기술정보(예: 사이즈, 포맷)를 포함하고 있으므로 이 경우는 techMD에 기술한다.
- premis:event는 행위에 대한 것이므로 digiProvMD에 기술하는 것이 적절하다.



〈그림 3〉 METS기본구조 다이어그램(서보동 2008, 27)

www.kci.go.kr

〈표 8〉 METS 스키마

상위요소	하위요소	요소명	정의
〈metsHdr〉		헤더(Header)	METS 문서가 인코딩하는 디지털 객체를 설명하는 메타데이터가 아닌 METS 문서자체에 대한 메타데이터
〈dmdSec〉		서지정보 메타데이터 (Descriptive Metadata)	디지털 객체의 모든 아이템(디지털 객체를 구성하는 하나 이상의 파일 아이템)에 대한 모든 서지정보 메타데이터
		외부 서지정보 메타데이터 (External Descriptive Metadata)	MARC 레코드, DC 레코드 등의 외부에 위치한 메타데이터를 URI를 통해 이용
		내부 서지정보 메타데이터 (Internal Descriptive Metadata)	METS문서내의 DC같은 외부 표준을 이용해 표현하거나 MARC 레코드와 같은 이진 레코드로 인코딩해서 사용하는 방식
〈amdSec〉		관리메타데이터 (Administrative Metadata)	디지털 객체의 모든 아이템들에 관한 관리정보
	〈techMD〉	기술정보메타데이터(Technical Metadata)	
	〈sourceMD〉	소스메타데이터(Source Metadata)	
	〈rightsMD〉	저작권메타데이터(Rights Metadata)	
	〈digprovMD〉	출처정보 메타데이터 (Digital Provenance Metadata)	
〈fileSec〉		파일섹션(File Section)	디지털 객체를 구성하는 모든 파일들에 대한 리스트정보
〈structMap〉		구조맵(Structure Map)	디지털 문서의 계층적 구조를 정의
〈structLink〉		구조링크(Structural Links)	구조맵에서도 표현된 계층구조의 각 노드들을 연결하는 하이퍼링크에 대한 정보
〈behaviorSec〉		행위영역(Behavior Section)	디지털 객체의 콘텐츠 정보를 사용하여 실행할 수 있는 모든 행위들에 대한 정보

- premis:Rights은 rightMD에 기술한다.
- premis:Agent는 권한이나 이벤트에 대한 행위자를 표현하기 위한 것이므로 rightMD나 digiProvMD에 기술하는 것이 바람직하다.

6.2.2 PREMIS와 METS의 중복성

요소에 따라 METS와 PREMIS에서 중복적으로 나타나는 경우가 있다. 예를 들어 METS의 〈file〉요소의 속성으로 체크섬, 체크섬 유형을 포함하고 있으며, PREMIS에서도 〈fixity〉요소의 하위요소로 〈messageDigest〉, 〈messageDigest-Algorithm〉을 가진다. 물론 두 경우는 차이가 있다. METS의 경우 체크섬 정보가 속성에 표현되므로 중복될 수 없으나, PREMIS에서는 하위요소이므로 중복이 가능하다. 또한 METS의 MINETYPE이나 PREMIS의 〈formatName〉 역시 유사한 정보를 표현한다.

이 경우 중복적으로 기술할지, 선택적으로 한 곳에서만 기술할지를 결정해야 한다. 이는 중복되는 요소를 반복해야 할 필요가 있는지, 추가적인 정보를 PREMIS에서 제공해야 하는지에

www.kci.go.kr

따라 결정해야 할 것이다. 예를 들어, 기존의 구축된 METS 파일에 이미 정보가 기록되어 있지만 추가적인 정보를 PREMIS를 통해 제공해야 할 필요가 있거나 속성이 아닌 반복 가능한 하위 요소로 표현해야 할 경우 PREMIS 사용하는 것이 바람직 할 것이며, 그렇지 않은 경우는 중복 기술은 지양해야 할 것이다.

국립중앙도서관 METS파일을 분석한 결과, 〈file〉요소의 속성으로 체크섬, 체크섬 유형을 포함하지 않고 있으며, MINEType 역시 기재하지 않고 있으므로, PREMIS의 〈fixity〉 요소 및 〈formatName〉을 사용하여 필요한 정보를 기재하는 것이 바람직하다.

6.2.3 구조 메타데이터의 기록

METS 문서의 〈strucMap〉요소를 통해 디지털 자원의 구조적 관계를 표현할 수 있다. 그러나 PREMIS에도 구조적 관계를 표현하는 요소를 표현하고 있으므로 중복표현할지, 한 곳에만 표현할지 결정해야 한다. 가이드라인에 따르면 METS의 〈strucMap〉요소가 관계를 표현하는데 유연하고, METS에 있어서 〈strucMap〉요소는 핵심 요소라고 할 수 있으므로 이 요소를 우선적으로 사용하고 그 외에 관계를 표현할 경우 PREMIS사용하는 방법이 있을 수 있다. 그러나 〈strucMap〉를 사용할 경우 관계의 의미(파생, 구조적, 의존관계)를 표현하는데 한계가 있으므로 요소와의 관계를 〈structMap〉으로 연결하고, 그 관계의 의미를 PREMIS를 사용하여 표현하는 방법이 바람직하다.

6.2.4 PREMIS와 METS요소간의 연결

구조 METS문서는 일반적으로 다수의 관리 메타데이터 영역, 기술 메타데이터 영역, 참조 파일을 가질 수 있으므로, 메타데이터와 파일을 연결시키는 것을 중요하다. METS는 METS 문서의 부분을 연결시키기 위하여 ID/IDREF를 사용한다. PREMIS 역시 연계된 개체를 연결시키는 요소를 정의하고 있다. 이를 위해 PREMIS 메타데이터가 METS문서에 쌓여있는 경우 METS ID/IDRef 구조를 사용해야하며, PREMIS 메타데이터가 분리된 경우에는 PREMIS linking elements를 사용해야 한다(〈그림 4〉, 〈그림 5〉 참조).

```
<fileSec> <fileGrp>
 <file ID="FID1" ADMID="TMD1PREMIS TMD1MIX DP1EVENT DP1AGENT">
</fileGrp></fileSec>

<techMD ID="TMD1PREMIS">
 <premis:object>
  <premis:linkingEventIdentifier>
   <premis:linkingEventIdentifierType>ECHODEP Hub Event
   </premis:linkingEventIdentifierType>
   <premis:linkingEventIdentifierValue>echo12345</premis:linkingEventIdentifierValue>
   </premis:linkingEventIdentifier>
<digiprovMD ID="DP1EVENT">
  <premis:event>
    <premis:eventIdentifier>
     <premis:eventIdentifierType>ECHODEP Hub Event</premis:eventIdentifierType>
     <premis:eventIdentifierValue>echo12345 </premis:eventIdentifierValue>
    </premis:eventIdentifier>
    <premis:eventType>ingestion</premis:eventType>
    <premis:eventDateTime>2006-05-02T15:12:53 </premis:eventDateTime>
</premis:event>
```

〈그림 4〉 METS ID 및 PREMIS 식별자 사용(Guenther 2008)

www.kci.go.kr

```
Defined in METS:
<structMap TYPE="physical">
 <div ORDER="1" TYPE="text">
 <fptr FILEID="FID9"/>
   <div ORDER="1" TYPE="page" LABEL=" Page [1]">
   <fptr FILEID="FID1"/></mets:div>
    <div ORDER="2" TYPE="page" LABEL=" Page [2]">
    <fptr FILEID="FID2"/></mets:div>
 </div>

Defined in PREMIS (in metadata for the file FID1):
<premis:relationship>
  <premis:relationshipType>structural</premis:relationshipType>
  <premis:relationshipSubType>is sibling of </premis:relationshipSubType>
 <premis:relatedObjectIdentification>
   <premis:relatedObjectIdentifierType>UCB</premis:relatedObjectIdentifierType>
   <premis:relatedObjectIdentifierValue>FID2</premis:relatedObjectIdentifierValue>
   <premis:relatedObjectSequence>1</premis:relatedObjectSequence>
```

〈그림 5〉 METS 〈structMap〉과 PREMIS 〈relationship〉요소(Guenther 2008)

6.3 국립중앙도서관 PREMIS-METS 적용

국립중앙도서관의 디지털형태 자료의 유형에 따라 PREMIS 적용사항이 구별될 수 있다. 첫째, 원시자료로부터 재생산된 파일인지, 원래 디지털 자료(Born in Digital)인지 여부, 둘째, 자료는 한 파일로 구성함과 동시에 각 페이지를 개별 파일로 구성하였는지 아니면 자료를 한 파일로만 구성했는지 여부, 셋째, 디지털 파일을 추후에 다른 형태의 파일로 마이그레이션 하였는지 여부이다. 각 경우에 따라 PREMIS적용방안은 상이할 수 있다.

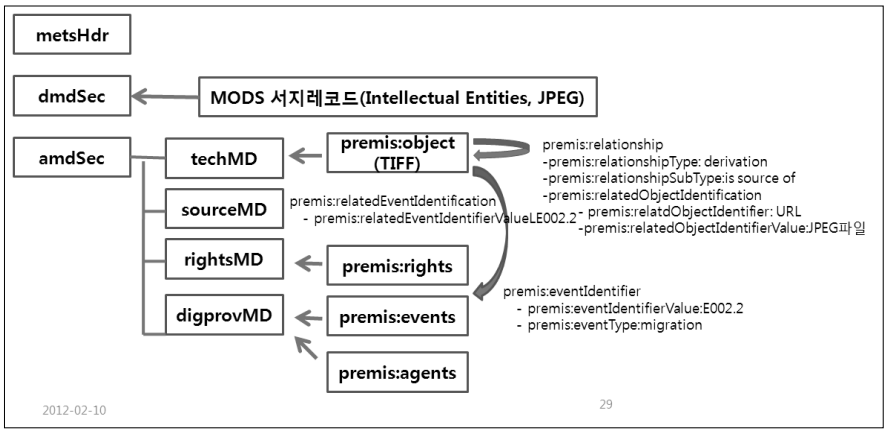
지적개체에서 기술된 디지털 파일의 보존정보를 기록하고자 하는 경우, MODS의 〈physical-Description〉요소를 사용하여 디지털 파일의 형태사항이 기록되었으므로 PREMIS의 〈format〉요소는 중복 기재할 필요가 없다. 단 〈format〉의 경우는 PREMIS에서는 필수요소이므로 빈 요소(empty element)를 사용하여 요소만 표기하기로 한다.

물리적 파일의 유형이 디지털 형태(예: JPEG 파일)로 생산된 후, 마이그레이션되어 다른 형태의 디지털 행태의 파일(예: Tiff 파일)이 있는 경우, 기존의 자원의 지적개체에 대한 기술은 MODS를 사용하여 기록하고, 기존 자원의 보존정보는 PREMIS를 사용하여 객체 개체에 기술하기로 한다. 또한 마이그레이션된 파일은 〈premis:relationship〉을 통해 기술하고, 마이그레이션 등의 이벤트 정보를 이벤트 개체를 사용하여 기술한다(〈그림 6〉 참조).

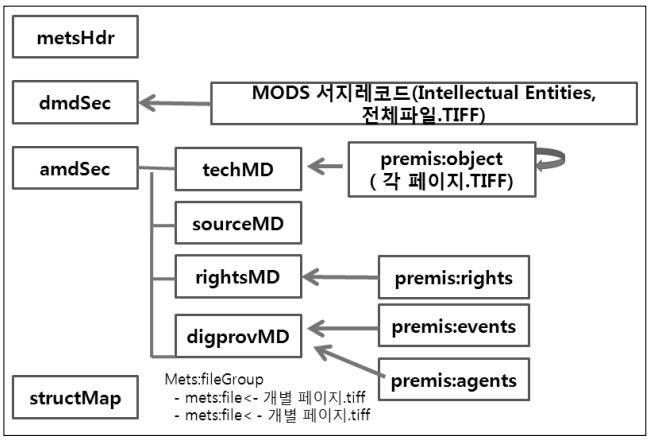
마지막으로 기존의 파일이 디지털 형태(예: Tiff 파일)이며 각 페이지가 역시 디지털 형태(예: Tiff 파일)를 구성하고 있는 경우, 전체 파일과 각 페이지 파일의 구조적 관계는 METS의 structMap을 사용하여 표현하며, PREMIS의 객체 개체의 〈premis:relationship〉을 사용하여 파일 간 관계의 의미를 밝힌다(〈그림 7〉 참조).

자세한 목록 예는 〈부록〉에 첨부하였다.

www.kci.go.kr



〈그림 6〉 마이그레이션 파일 가이드라인



〈그림 7〉 개별페이지 가이드라인

7. 결 론

본 연구는 디지털 정보자원의 효율적인 관리 및 보존을 위하여 PREMIS를 기반으로 보존 메타데이터 스키마를 개발하는 것은 물론 기존 METS를 기반으로 한 목록데이터에 보존 메타데이터를 적용하는 방안을 제시하고자 하였다. 이를 위하여 국립중앙도서관 디지털 자원을 사례로 채택하였다.

연구방법론으로 PREMIS 구현 방법론 및 전략이 논의되었던 선행 연구 및 구현사례를 분석함은 물론, 국립중앙도서관 디지털 자원 수집 현황 및 보존현황, 담당자와의 면담이 이루어졌다.

이를 통하여 보존 메타데이터의 요소를 추출하였다. 앞서 밝힌 메타데이터 선정 및 요소 도출 기준에 따라 객체(object), 이벤트(event), 행위자(agent), 권한(rights)의 총 4개 분야로 나누

www.kci.go.kr

어 도출하였다. 각 분야의 도출된 기본 요소는 각 메타데이터 포맷의 하위요소를 선별하여 확장한 뒤 기술하지 못하는 정보에 대한 하위요소의 보다 상세한 확장 과정을 거쳤다.

상세한 확장 과정을 거친 메타데이터 및 통제어휘는 콘텐츠 분석결과, 담당자와의 검증, 국립중앙도서관 디지털 자원의 색인테스트 결과가 반영되어 정제되는 과정을 거쳤다.

객체(object), 이벤트(event), 행위자(agent), 권한(rights)의 4개 개체를 PREMIS 데이터 모델에 입각하여 서로 연결되는 객체를 연결 식별자(linkingIdentifier)를 사용하여 메타데이터와 연결되도록 하여 자원이 객체, 이벤트, 행위자, 권한에 대한 정보들의 중심에서 유기적으로 정보를 제공할 수 있도록 하였다. 마지막으로, 현재 국립중앙도서관 디지털 장서관리에서 인코딩 표준으로 사용되고 있는 METS를 기반으로 PREMIS분석을 통해 도출된 메타데이터를 적용한 목록의 예를 제시하여 실무의 적용가능성을 높였다.

표준 메타데이터 PREMIS를 활용한 목록규칙의 설정 및 적용은 한중일 디지털도서관 프로젝트에 디지털 정보자원의 관리 및 공유를 위한 표준메타데이터 요소로 활용될 것이며, 이는 삼국도서관 자료의 통합 검색 기반 환경의 초석이 될 것이다. 이를 통해 국내 지식정보 콘텐츠의 재생산은 한층 활발하게 이루어질 뿐만 아니라 삼국의 정보자원은 새로운 아시아의 정보문화권을 형성하는데 기여할 것으로 기대된다.

참 고 문 헌

[1] 국가기록원. 2007.『전자기록의 관리와 보존을 위한 국제협력 아젠다 개발』. 서울: 국가기록원.

[2] 국립중앙도서관. 2010.『2010년도 국립중앙도서관연보』. 서울: 국립중앙도서관.

[3] 김판준. 2010. 연구정보를 위한 보존 메타데이터 요소 개발에 관한 연구.『정보관리학회지』, 27(4): 169-191.

[4] 문주영, 김태수. 2011. PREMIS 데이터모델 적용을 위한 사무문서 컨텐츠모형 설계 연구.『정보관리학회지』, 28(1): 43-68.

[5] 서보동. 2008.『MODS/METS를 적용한 전자도서관의 메타데이터 표준 연구』. 석사학위논문, 연세대학교 대학원, 컴퓨터공학과.

[6] 서은경. 2005. 디지털자원의 보존 메타데이터 요소세트 개발에 관한 연구.『정보관리학회지』, 22(3): 233-260.

[7] 오삼균, 권기성. 2007. 전통예술자료 아카이빙을 위한 보존 메타데이터 설계에 관한 연구.『한국문헌정보학회지』, 41(4): 387-412.

[8] 이경남. 2006. 전자기록의 장기적 보존을 위한 보존 메타데이터 요소 분석.『기록학연구』, 14:

www.kci.go.kr

191-240.

[9] 한국기록관리협회. 2009.『전자기록물의 이해』. 서울: 조은글터.

[10] Gomes, Daniel, Miranda, Joao, & Costa, Miguel. 2011. "A survey on web archiving initiatives." In *Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL'11): Research and Advanced Technology for Digital Libraries*, 408-420. Berlin: Springer-Verlag.

[11] Guenther, R. S. 2008. "Battle of the buzzwords." *D-Lib Magazine*, 14(7/8). [online]. [cited 2012.4.10]. 〈http://www.dlib.org/dlib/july08/guenther/07guenther.html〉.

[12] Hathi Trust. [online]. [cited 2012.4.10]. 〈http://www.hathitrust.org/〉.

[13] Florida Digital Archive. [online]. [cited 2012.4.10]. 〈http://daitss.fcla.edu/〉.

[14] MathArc. [online]. [cited 2012.4.8]. 〈http://www.library.cornell.edu/dlit/MathArc/web/index.html〉.

[15] National Library of Finland. [online]. [cited 2012.4.10]. 〈http://www.nationallibrary.fi/〉.

[16] PREMIS Data Dictionary for Preservation Metadata 2.1. [online]. [cited 2012.4.10]. 〈http://www.loc.gov/standards/premis/v2/premis-2-0.pdf〉.

[17] SHERPA. [online]. [cited 2012.4.6]. 〈http://www.sherpadp.org.uk/index.html〉.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

[1] National Archives of Korea. 2007. *International Cooperation Agenda for Management and Preservation of Electronic Records*. Seoul: National Archives of Korea.

[2] National Library of Korea. 2010. *Yearbook of 2010 National Library of Korea*. Seoul: National Library of Korea.

[3] Kim, Pan-Jun. 2010. "A study on preservation metadata elements for research information." *Journal of the Korean Society for Information Management*, 27(4): 169-191.

[4] Moon, Juyoung & Kim, Taesoo. 2011. "A study on contents models for business records by the application of the PREMIS data model." *Journal of the Korean Society for Information Management*, 28(1): 43-68.

[5] Seo, Bo-dong. 2008. *The Study on Metadata Standards Based on MODS/METS*. M.A. thesis, Yonsei University, Computer Science.

[6] Seo, Eun-Gyong. 2005. "A study on preservation metadata for digital resources." *Journal of the Korean Society for Information Management*, 22(3): 233-260.

www.kci.go.kr

[7] Oh, Sam-Gyun & Kwon, Ki-Seong. 2007. "Designing preservation metadata for archives of Korean traditional arts." *Journal of the Korean Society for Library and Information Science*, 41(4): 387-412.

[8] Lee, Kyung-Nam. 2006. "A study on preservation metadata for long term preservation of electronic records." *The Korean Journal of Archival Studies*, 14: 191-240.

[9] The Korea Association of Archives Management. 2009. *Understanding Electronic Records*. Seoul: Joeungeulteo.

www.kci.go.kr

〈부록 - 목록에〉


```
<?xml version="1.0" encoding="UTF-8"?>
- <mets:mets xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd"
  xmlns:premis="info:lc/xmlns/premis-v2" xmlns:lc="http://www.loc.gov/mets/profiles/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/TR/xlink"
  xmlns:mets="http://www.loc.gov/METS/">
  - <mets:metsHdr LASTMODDATE="2012-01-03T06:10:12" CREATEDATE="2012-01-03T06:10:12">
      - <mets:agent TYPE="ORGANIZATION" ROLE="CREATOR">
            <mets:name>NDL</mets:name>
        </mets:agent>
    </mets:metsHdr>
  - <mets:dmdSec CREATED="2012-01-03T06:10:12" ID="dmdid-00001">
      - <mets:mdWrap MDTYPE="MODS" MIMETYPE="text/xml">
          - <mets:xmlData>
              - <mods:mods xsi:schemaLocation="http://www.loc.gov/mods/v3
                http://www.loc.gov/standards/mods/v3/mods-3-3.xsd" xmlns:xlink="http://www.w3.org/1999/xlink"
                version="3.3" xmlns:ndlklds="http://nl.go.kr/schema/license/terms/"
                xmlns:mods="http://www.loc.gov/mods/v3">
                - <mods:titleInfo type="title">
                      <mods:title>학교 교육정보화의 효과적 지원방안 연구</mods:title>
                      <mods:subTitle>2010년 정책연구개발사업</mods:subTitle>
                      <mods:partNumber/>
                      <mods:partName/>
                      <mods:nonSort/>
                  </mods:titleInfo>
                - <mods:name type="corporate">
                      <mods:namePart>교육과학기술부</mods:namePart>
                    - <mods:role>
                          <mods:roleTerm/>
                      </mods:role>
```


〈이하 MODS 부분 중략〉


```
  - <mets:amdSec>
      - <mets:techMD ID="object1">
          - <mets:mdWrap MDTYPE="PREMIS:OBJECT">
              - <mets:xmlData>
                  - <premis:object xsi:schemaLocation="info:lc/xmlns/premis-v2
                    http://www.loc.gov/standards/premis/v2/premis-v2-0.xsd" xsi:type="premis:file">
                      - <premis:objectIdentifier>
                            <premis:objectIdentifierType>hdl</premis:objectIdentifierType>
                            <premis:objectIdentifierValue>ndl.education/13245</premis:objectIdentifierValue>
                        </premis:objectIdentifier>
                      - <premis:objectCharacteristics>
                            <premis:compositionLevel>0</premis:compositionLevel>
                          - <premis:fixity>
                                <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
                                <premis:messageDigest>36b03197ad066cd719906c55eb68ab8d</premis:messageDigest>
                            </premis:fixity>
                            <premis:size>20800896</premis:size>
                          - <premis:format>
                              - <premis:formatDesignation>
                                    <premis:formatName>PDF</premis:formatName>
                                    <premis:formatVersion>5.0</premis:formatVersion>
                                </premis:formatDesignation>
```


www.kci.go.kr

```
        </premis:format>
      - <premis:creatingApplication>
            <premis:creatingApplicationName>Adobe Acrobat</premis:creatingApplicationName>
            <premis:creatingApplicationVersion>5.0</premis:creatingApplicationVersion>
            <premis:dateCreatedByApplication>2010-09-20T08:29:02</premis:dateCreatedByApplication>
        </premis:creatingApplication>
    </premis:objectCharacteristics>
  - <premis:environment>
      - <premis:software>
            <premis:swName>Adobe Acrobat Reader</premis:swName>
            <premis:swVersion>5.0</premis:swVersion>
            <premis:swType>renderer</premis:swType>
        </premis:software>
  - <premis:software>
        <premis:swName>Windows</premis:swName>
        <premis:swVersion>XP</premis:swVersion>
        <premis:swType>operatingSystem</premis:swType>
    </premis:software>
  - <premis:hardware>
        <premis:hwName>Intel x86</premis:hwName>
        <premis:hwType>processor</premis:hwType>
        <premis:hwOtherInformation>60 mhz minimum</premis:hwOtherInformation>
    </premis:hardware>
  - <premis:hardware>
        <premis:hwName>64 MB RAM</premis:hwName>
        <premis:hwType>memory</premis:hwType>
        <premis:hwOtherInformation>32 MB minimum</premis:hwOtherInformation>
    </premis:hardware>
</premis:environment>
<premis:relationship>
    <premis:relationshipType>derivation</premis:relationshipType>
    <premis:relationshipSubType>is source of</premis:relationshipSubType>
  - <premis:relatedObjectIdentification>
        <premis:relatedObjectIdentifierType>URL</premis:relatedObjectIdentifierType>
        <premis:relatedObjectIdentifierValue>http://www.mest.go.kr/web/1106/ko/board/view.do?
            bbsId=148boardSeq=22376mode=view</premis:relatedObjectIdentifierValue>
        <premis:relatedObjectSequence>0</premis:relatedObjectSequence>
    </premis:relatedObjectIdentification>
  - <premis:relatedEventIdentification>
        <premis:relatedEventIdentifierType>NDL Repository Event</premis:relatedEventIdentifierType>
        <premis:relatedEventIdentifierValue>e-20110706-0021</premis:relatedEventIdentifierValue>
        <premis:relatedEventSequence>1</premis:relatedEventSequence>
    </premis:relatedEventIdentification>
</premis:relationship>
```


www.kci.go.kr

```
- <premis:linkingEventIdentifier>
    <premis:linkingEventIdentifierType>NDL Repository Event</premis:linkingEventIdentifierType>
    <premis:linkingEventIdentifierValue>e-20101006-0001</premis:linkingEventIdentifierValue>
  </premis:linkingEventIdentifier>
- <premis:linkingEventIdentifier>
    <premis:linkingEventIdentifierType>NDL Repository Event</premis:linkingEventIdentifierType>
    <premis:linkingEventIdentifierValue>e-20101006-0002</premis:linkingEventIdentifierValue>
  </premis:linkingEventIdentifier>
- <premis:linkingIntellectualEntityIdentifier>
    <premis:linkingIntellectualEntityIdentifierType>URI</premis:linkingIntellectualEntityIdentifierType>
    <premis:linkingIntellectualEntityIdentifierValue>http://policy.dibrary.net/posts/list/410/289398/0.do
      </premis:linkingIntellectualEntityIdentifierValue>
  </premis:linkingIntellectualEntityIdentifier>
</premis:object>
</mets:xmlData>
</mets:mdWrap>
</mets:techMD>
- <mets:digiprovMD ID="event1">
  - <mets:mdWrap MDTYPE="PREMIS:EVENT">
    - <mets:xmlData>
      - <premis:event>
        - <premis:eventIdentifier>
            <premis:eventIdentifierType>NDL Repository Event</premis:eventIdentifierType>
            <premis:eventIdentifierValue>e-20101006-0001</premis:eventIdentifierValue>
          </premis:eventIdentifier>
          <premis:eventType>validation</premis:eventType>
          <premis:eventDateTime>2010-10-06T00:00:00.001</premis:eventDateTime>
          <premis:eventDetail>jhove</premis:eventDetail>
        - <premis:eventOutcomeInformation>
            <premis:eventOutcome>00</premis:eventOutcome>
          - <premis:eventOutcomeDetail>
              <premis:eventOutcomeDetailNote>Well-formed and valid</premis:eventOutcomeDetailNote>
            </premis:eventOutcomeDetail>
          </premis:eventOutcomeInformation>
        - <premis:linkingAgentIdentifier>
            <premis:linkingAgentIdentifierType>AgentID</premis:linkingAgentIdentifierType>
            <premis:linkingAgentIdentifierValue>age0001</premis:linkingAgentIdentifierValue>
          </premis:linkingAgentIdentifier>
        - <premis:linkingObjectIdentifier>
            <premis:linkingObjectIdentifierType>hdl</premis:linkingObjectIdentifierType>
            <premis:linkingObjectIdentifierValue>ndl.education/13245</premis:linkingObjectIdentifierValue>
          </premis:linkingObjectIdentifier>
        </premis:event>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:digiprovMD>
- <mets:digiprovMD ID="event2">
  - <mets:mdWrap MDTYPE="PREMIS:EVENT">
    - <mets:xmlData>
      - <premis:event>
        - <premis:eventIdentifier>
            <premis:eventIdentifierType>NDL Repository Event</premis:eventIdentifierType>
            <premis:eventIdentifierValue>e-20101006-0002</premis:eventIdentifierValue>
          </premis:eventIdentifier>
          <premis:eventType>ingestion</premis:eventType>
          <premis:eventDateTime>2010-10-06T00:00:05.001</premis:eventDateTime>
          <premis:eventDetail>ingester1_0.exe</premis:eventDetail>
        - <premis:eventOutcomeInformation>
            <premis:eventOutcome>00</premis:eventOutcome>
          </premis:eventOutcomeInformation>
```


www.kci.go.kr

```
            - <premis:linkingAgentIdentifier>
                <premis:linkingAgentIdentifierType>AgentID</premis:linkingAgentIdentifierType>
                <premis:linkingAgentIdentifierValue>age0001</premis:linkingAgentIdentifierValue>
              </premis:linkingAgentIdentifier>
            - <premis:linkingObjectIdentifier>
                <premis:linkingObjectIdentifierType>hdl</premis:linkingObjectIdentifierType>
                <premis:linkingObjectIdentifierValue>ndl.education/132451</premis:linkingObjectIdentifierValue>
              </premis:linkingObjectIdentifier>
            </premis:event>
          </mets:xmlData>
        </mets:mdWrap>
      </mets:digiprovMD>
    - <mets:digiprovMD ID="event3">
        - <mets:mdWrap MDTYPE="PREMIS:EVENT">
            - <mets:xmlData>
                - <premis:event>
                    - <premis:eventIdentifier>
                        <premis:eventIdentifierType>NDL Repository Event</premis:eventIdentifierType>
                        <premis:eventIdentifierValue>e-20110706-0021</premis:eventIdentifierValue>
                      </premis:eventIdentifier>
                      <premis:eventType>migration</premis:eventType>
                      <premis:eventDateTime>2011-07-06T00:00:00.006</premis:eventDateTime>
                      <premis:eventDetail>Adobe Acrobat 6.0</premis:eventDetail>
                    - <premis:eventOutcomeInformation>
                        <premis:eventOutcome>00</premis:eventOutcome>
                      </premis:eventOutcomeInformation>
                    - <premis:linkingAgentIdentifier>
                        <premis:linkingAgentIdentifierType>AgentID</premis:linkingAgentIdentifierType>
                        <premis:linkingAgentIdentifierValue>age0001</premis:linkingAgentIdentifierValue>
                      </premis:linkingAgentIdentifier>
                    - <premis:linkingObjectIdentifier>
                        <premis:linkingObjectIdentifierType>hdl</premis:linkingObjectIdentifierType>
                        <premis:linkingObjectIdentifierValue>ndl.education/132451</premis:linkingObjectIdentifierValue>
                      </premis:linkingObjectIdentifier>
                  </premis:event>
              </mets:xmlData>
          </mets:mdWrap>
        </mets:digiprovMD>
    - <mets:digiprovMD ID="agent1">
        - <mets:mdWrap MDTYPE="PREMIS:AGENT">
            - <mets:xmlData>
                - <premis:agent>
                    - <premis:agentIdentifier>
                        <premis:agentIdentifierType>AgentID</premis:agentIdentifierType>
                        <premis:agentIdentifierValue>age0001</premis:agentIdentifierValue>
                      </premis:agentIdentifier>
                      <premis:agentName>NDL Repository</premis:agentName>
                      <premis:agentType>organization</premis:agentType>
                  </premis:agent>
              </mets:xmlData>
          </mets:mdWrap>
        </mets:digiprovMD>
  </mets:amdSec>
- <mets:fileSec>
    - <mets:fileGrp>
        - <mets:file ID="학교 교육정보화의 효과적 지원방안 연구" MIMETYPE="pdf" DMDID="dmdid-00001" ADMID="admid-
          00001">
            - <mets:FContent>
                <mets:binData>NjI2OQ==</mets:binData>
              </mets:FContent>
          </mets:file>
        </mets:fileGrp>
  </mets:fileSec>
- <mets:structMap>
    - <mets:div>
        - <mets:div DMDID="dmdid-00001" ADMID="admid-00001">
            <mets:fptr FILEID="학교 교육정보화의 효과적 지원방안 연구.pdf"/>
          </mets:div>
        </mets:div>
  </mets:structMap>
```


www.kci.go.kr

www.kci.go.kr