

디지털 보존을 위한 PREMIS 3.0 기반 메타데이터에 관한 연구

A Study of Metadata for Digital Preservation Based on PREMIS 3.0

박희진(Heejin Park)

E-mail: papermod@hansung.ac.kr

한성대학교 디지털인문정보학트랙 부교수



논문접수 2020-11-03

최초심사 2020-11-08

게재확정 2020-11-24

ORCID

Heejin Park <https://orcid.org/0000-0002-7217-9869>

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

• 이 논문에는 국립중앙도서관 정책연구과제<국립중앙도서관 디지털자원 보존전략 수립 연구> 최종보고서의 일부분을 수정·보완한 내용이 일부 포함되어 있음.

초 록

PREMIS는 OAIS의 개념적 프레임워크의 한계를 넘어 보존 실무에 적용할 수 있도록 개발되었지만, 각 메타데이터 요소들을 실제 데이터와 어떻게 연결시킬 것인지에 대한 사례와 지침이 명확하지 않아 디지털 아카이브에서 실제 보존 메타데이터를 설계하거나 구현하는데 어려움이 있다. 이에 본 연구는 다양한 디지털 아카이브의 환경에서 디지털 자원을 관리하고 보존의 목적으로 활용되고 있는 PREMIS 3.0 기반의 보존 메타데이터 실제 사례들을 조사하고 보존 메타데이터 구현 및 설계를 위한 시사점을 도출하고자 하였다. 분석 결과를 토대로 PREMIS의 응용프로파일 지침에 따라 디지털 아카이브의 보존 메타데이터 설계 시 고려할 수 있는 보존 메타데이터 요소들을 제안하였다.

ABSTRACT

The development of the Preservation Metadata: Implementation Strategies (PREMIS) aims to preserve practices beyond the limits of the Open Archival Information System (OAIS) conceptual framework; however, the examples and guidelines on connecting each metadata element to the actual data are unclear, making the design of preservation metadata in real digital archives difficult. Hence, this study investigates actual PREMIS 3.0-based preservation metadata management cases, preserves digital resources in various digital archive environments, and derives implications for the implementation and design of preservation metadata. Based on the analysis, the study proposes preservation metadata elements that can be considered when designing preservation metadata for digital archives according to PREMIS 3.0 application profile guidelines.

Keywords: PREMIS, 장기보존, OAIS, 보존 메타데이터, 디지털 아카이브

PREMIS, Long-term preservation, OAIS, Preservation metadata, Digital archives

1. 서론

디지털 아카이브가 증가함에 따라 신뢰할 수 있는 디지털 자원의 관리와 보존에 대한 관심이 높아지고 있다. 디지털 자원은 기술의 변화에 밀접하게 영향을 받으며, 매체를 이동하고 변환하는 과정에서 데이터 손상 및 훼손의 위험이 따른다. 시간의 경과와 기술의 변화 속에서도 디지털 자원이 지속적으로 접근하고 이해할 수 있도록 하는 디지털 보존은 모든 아카이브의 공통적인 목표로 디지털 자원의 신뢰성과 상호운용성 확보를 위한 다각적인 노력이 이루어지고 있다. 이 중 메타데이터는 디지털 객체의 보존과 관리, 검색을 지원할 수 있는 기본으로 디지털 보존활동의 가장 큰 핵심 부분이라 할 수 있다.

보존 메타데이터는 디지털 객체의 무결성을 유지해야 하며 수록된 내용이나 형식이 임의적으로 변경되는 것을 방지할 수 있어야 한다. 또한 디지털 자원의 장기보존과 접근을 위해 변화하는 환경에서의 디지털 정보를 계속 이용하는데 필요한 상세 정보를 보존 메타데이터를 통해 제공할 수 있어야 한다. 이에 디지털 보존 관련 메타데이터는 디지털 아카이브가 증가하고 보존의 중요성이 확대되면서 관리적 메타데이터 영역에 포함시키던 기존 경향에서 독립된 별도의 보존 메타데이터 영역을 설정하는 실제 사례들이 증가하고 있다(이지영, 김희정, 2009, p. 50).

또한, 보존 메타데이터는 디지털 객체에 대한 포괄적인 기술사항을 제공하고 있으므로, 디지털 객체의 생애주기에 적용되는 설명 메타데이터, 기술 메타데이터 등과 상호운용성을 유지하여야 한다(이승민, 2015, p. 24). 보존 메타데이터의 실행에서 상호운용성에 초점을 두고 개발한 것이 PREMIS(Preservation Metadata: Implementation Strategies)이다. PREMIS는 2005년 초안을 발표한 이후, 도서관, 기록관, 박물관 등 문화유산기관을 비롯한 여러 커뮤니티에서 이를 활용하여 다양한 보존 메타데이터를 구축하여 실행하고 있다. 하지만 PREMIS는 각 메타데이터 요소들을 실제 데이터와 어떻게 연결시킬 것인지에 대한 사례와 지침은 명확하지 않아 디지털 아카이브에서 보존 메타데이터를 설계하거나 구현하는 데에 여러 가지 한계가 있다.

PREMIS는 그동안 다양한 환경의 보존 실무에서의 경험과 피드백들을 기반으로 여러 차례 수정, 보완을 걸쳐 2015년 현재의 3번째 개정판을 발표하였다(Library of Congress, 2016). PREMIS 3.0에서는 데이터 모델이 수정되고, 변화된 하드웨어 및 소프트웨어 환경이나 다양해진 디지털 정보를 표현하기 위해 기능이 확장되었다. 이에 본 연구는 PREMIS 3.0의 개정 내용을 살펴보고, 이를 적용한 국외 다양한 디지털 아카이브의 보존 메타데이터 구축 사례들을 조사하여 보존 메타데이터 구현 및 설계를 위한 시사점을 도출하고자 하였다. 분석 결과를 토대로 PREMIS의 응용프로파일 지침에 따라 디지털 아카이브의 보존 메타데이터 설계 시 고려할 수 있는 보존 메타데이터 요소들을 제안하고자 한다. 본 연구에서 제안한 메타데이터 요소는 직접적인 메타데이터 개발에 활용하기에는 한계가 있으나, 다양한 디지털 아카이브에서 포괄적인 보존의 목적으로 필요한 요소 제안에 목적을 두고 향후 보존 메타데이터 구축에 기초자료로 활용될 수 있을 것이다.

2. 이론적 배경

2.1 OAIS와 PREMIS

OAIS 참조모델은 디지털 정보의 장기보존을 위한 상호운용적인 아카이브 시스템 구축을 위한 국제적 개념틀을 제시한다. OAIS 참조모델은 보존 메타데이터 자체를 규정하지는 않지만, 디지털 보존과 관련된 폭넓은 범위의 디지털 아카이브 실무 체제와 공통 언어를 제공함으로써 디지털 아카이브 분야에서 범용되는 표준으로 인정되고 있다. OAIS는 디지털 정보의 장기적인 유지 및 관리를 위한 핵심적인 기능으로 입수(Ingest), 아카이브 저장(Archival

Storage), 데이터 관리(Data Management), 운영(Administration), 보존계획(Preservation Planning) 및 접근(Access)의 6가지 영역으로 제시하였다. OAIS 환경에서 정보는 제출정보패키지(Submission Information Package, SIP), 보존정보패키지(Archival Information Package, AIP), 배포정보패키지(Dissemination Information Package, DIP)의 정보패키지로 입수, 관리, 제공된다. 이들 정보패키지는 디지털 객체의 장기보존을 위한 상위 개념의 구조와 요건을 제공해 준다. 정보패키지의 보존기술정보(Preservation Description Information, PDI)로 참조정보(Reference Information), 맥락정보(Context Information), 출처정보(Provenance Information), 무결성 정보(Fixity Information), 권한 정보(Access Rights Information) 등을 포함하도록 하고 있다. 이러한 OAIS 모델의 정보구조는 보존 메타데이터를 개발하는 많은 기관에서 구조모형으로 활용되어 왔으며, 보존과 관련된 필요한 정보가 모두 포함되었는가를 확인하기 위한 유용한 기준점으로 평가되고 있다.

그러나 OAIS는 디지털 보존의 개념적 프레임워크를 제시한 것이므로 실제 보존 메타데이터를 설계하거나 실질적인 보존 업무에 적용하기에는 한계가 있다. 이에 OCLC와 RLG는 각국의 전문가로 구성된 실무그룹을 조직하여, 보존 메타데이터 설계에 필요한 공통된 요소를 찾아 핵심 요소를 개발하고자 하였다. 디지털 보존을 지원하기 위해 보존 리포지터리가 알아야만 하는 요소를 밝히기 위해서 개발된 것이 PREMIS(Preservation Metadata: Implementation Strategies)이다. PREMIS는 디지털 객체의 장기보존에 필요한 메타데이터 요소를 정의하고, 디지털 보존 시스템이 보존 메타데이터를 구현하기 위한 포괄적이고 실용적인 정보를 제공하는 것을 목적으로 한다. 이를 위해 PREMIS 워킹그룹은 디지털 아카이브의 운영 현황과 유형, 정책, 메타데이터의 실행 등을 파악하고 이를 토대로 디지털 보존을 위한 데이터 사전을 2005년 5월 발행하였다. ‘보존 메타데이터를 위한 데이터 사전: PREMIS 워킹그룹 최종 보고서(Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group)’의 초안 발표 이후, 실행성과 상호운용성에 초점을 맞추어 연구가 진행되었고 2008년 ‘보존 메타데이터를 위한 PREMIS 데이터 사전 버전 2.0’이 발표되었다(PREMIS Editorial Committee, 2008). 2011년 1월 PREMIS 데이터 사전 버전 2.1과 2012년 7월 데이터 사전 버전 2.2가 발표되었다. PREMIS 2.2는 데이터 모델과 데이터 사전으로 구성된다. PREMIS 2.2의 데이터 모델은 다음의 <그림 1>과 같이 지적개체(Intellectual Entity), 객체(Object Entity), 이벤트(Events), 행위주체자(Agents), 권한(Rights)의 개체 유형으로 구성된다.

PREMIS는 실제 보존 메타데이터 스키마 설계와 구현을 위해서는 각 개체에 관한 기술사항이 가능한 한 상세하고 명확하게 정의되어야 한다고 명시하고 있다. 이와 동시에 각 개체에 관한 기술정보는 해당 개체(Entity)에 연계되어야 한다. PREMIS의 데이터 사전은 보존활동에 필요하다고 생각되는 다양한 정보를 정리하고 체계화하여, 용어의 의미가 갖는 범위와 용법을 정의한 말 그대로 메타데이터 요소 사전이다. PREMIS는 기관이 처한 다양한 보존 환경에 맞춰 보존 메타데이터를 설계하고 실행할 수 있도록, 메타데이터 요소(element)라는 표현 대신 의미 단위(semantic unit)를 사용한다. 데이터 사전은 개체 유형별로 각 요소에 대한 요소의 객체, 이벤트, 행위자, 권한에 대한 이름, 하위요소, 정의, 데이터 제약사항, 객체 범주, 적용가능성(예: 반복가능성, 강제성), 생성/유지 안내, 사용안내를 정의하고 있다. 데이터 사전은 디지털 보존과 관련된 요소들을 각 개체 유형 별로 상위요소(1단계) 아래 최소 2단계에서 최대 4단계까지의 하위요소들로 구조화되어 있다. PREMIS가 보존과 관련해서 가장 중요시하는 것은 객체의 이력과 관련된 정보로, 디지털 출처 정보로 기록하고 다른 객체와의 관계 또한 기록화 하도록 하고 있다. PREMIS가 규정하는 것은 메타데이터의 의미단위이며 그것을 기록하기 위한 방식은 다양한 보존 환경에 따라 자유롭게 결정하도록 허용하되, 데이터 사전에 대응할 수 있는 XML 스키마를 작성하여 공개하고 있다.

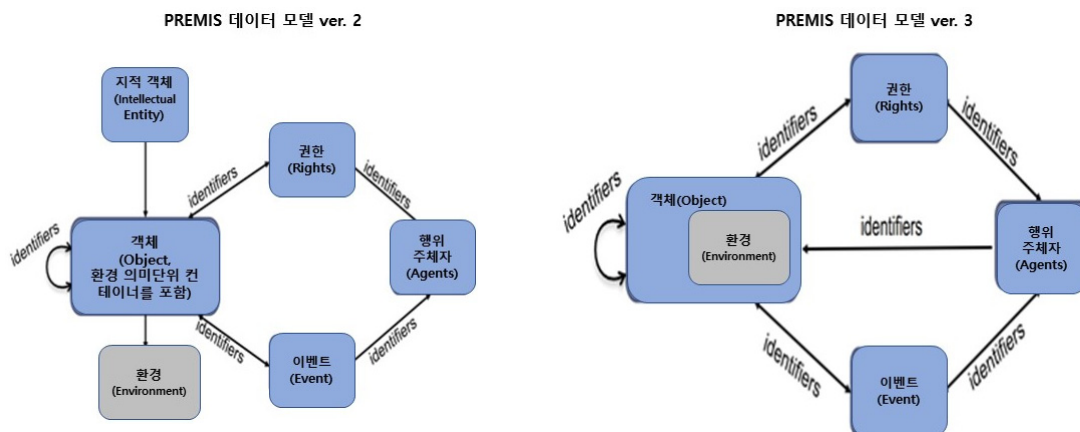
PREMIS의 특징은 디지털 정보의 구조와 특성을 고려하여, 디지털 객체와 관련하여 파일 뿐 아니라 표현물과 비트스트림 수준까지 기술 가능하다는 점이다. 또한, PREMIS는 다양한 보존 환경에서 디지털 객체의 장기 보존에 있어서 가장 필수적인 실행 가능성, 제공 가능성, 이해 가능성, 신뢰성을 지원한다. 보존 리포지터리에서 디지털 자료의 장기간 보존을 위해 알아야 할 핵심적인 정보를 제공하며, 체계적으로 정의된 데이터 모델과 데이터 사전을 활용하여 기관의 보존 업무에 적합한 메타데이터를 설계하고 실행할 수 있다. PREMIS는 보존 업무의 연속성을 지향하며, 보존 기술, 전략, 메타데이터 저장 및 관리에 대한 제약 없이 폭넓게 적용 가능하며 기술

중립적인 것이 특징이다. 이에 PREMIS는 디지털정보의 장기 보존을 위한 가장 종합적이고 포괄적인 메타데이터 표준 스키마로 평가되고 있다.

반면에 PREMIS가 가지는 한계점은 데이터 모델에서 제시하고 있는 지적 개체와 표현 개체를 실제 데이터에 적용하는데 그 개념의 구분이 명확하지 않다는 점이다. 데이터 모델의 구현을 위해 준수해야 할 구축상의 지침 및 최소한의 규정이 없으며, ‘의미단위’, ‘의미구성요소’ 또한 다양한 디지털 보존자원이 내포하고 있는 정보의 특성과 서지정보를 표현하는 데에는 한계가 있다.

2.2 PREMIS 3.0 개정과 특징

2015년 6월 10일, PREMIS 편집위원회는 “Data Dictionary for Preservation Metadata”(보존 메타데이터를 위한 데이터 사전)의 버전 3.0을 공개하였다. PREMIS 3.0은 PREMIS 2.2의 발표 이후 계속해서 실행성과 상호운용성에 초점을 맞추어 연구가 진행되었고 사용자들의 피드백을 받아 문제점을 보완하고 개선한 것이다. PREMIS 3.0의 데이터 모델은 객체 개체(Object Entity), 환경 개체(Environment Entity), 행위주체자 개체(Agent Entity), 권한 개체(Rights Entity), 이벤트 개체(Event Entity)로 구성되었으며 개체들 간의 관계는 다음의 <그림 1>과 같이 표현된다. PREMIS 3.0은 하드웨어 및 소프트웨어 환경이나 지적 개체 등의 정보를 표현하기 위한 기능을 확장하고 데이터 모델을 수정하였다(<그림 1> 참조). 이와 함께 XML 스키마와 OWL(Web Ontology Language) ontology도 함께 개정되었다.



<그림 1> PREMIS 데이터 모델 버전 2.2와 버전 3(Library Congress, 2016)

PREMIS 3.0의 주요 개정 내용은 크게 세 가지로 정리할 수 있다. 첫째, 지적 개체(Intellectual Entity)를 ‘객체’ 개체 범주로 포함시켰다. PREMIS 3.0의 데이터 모델 개정에서 가장 주목할 것은 지적 개체가 복잡한 디지털 객체의 속성을 표현할 수 있도록 ‘객체’ 개체의 일부로 수정된 것이다. 이는 그동안 실행성 측면에서 논리적으로만 존재하는 가상의 객체로 해석되었던 지적 개체의 개념을 객체의 개념으로 포괄시켜 디지털 객체의 복잡한 구조와 관계를 표현할 수 있도록 한 것이다. 이로써 실제 보존 업무에서 빈번히 발생하는 디지털 객체 수준에서 하나의 지적 개체가 여러 개의 다른 지적 개체로 구성될 수 있음을 표현할 수 있게 되었다. 예를 들어 하나의 지적 개체인 웹페이지에 또 다른 지적 개체인 사진이 포함된 경우, PREMIS의 수정된 ‘객체’ 개체를 적용하면 각각 웹페이지와 사진을 독립된 객체로 식별하고 두 객체를 ‘관계’ 개체를 통해 서로 연결하여 표현할 수 있게 되었다. 이러한 지적 개체에 대한 수정은 많은 보존 리포지터리에서 하나의 지적 개체가 여러 버전의 표현물로 생성되어 관리되

고 있는 것을 반영한 것이다. 하나의 디지털 이미지가 기관의 보존과 사용 환경에 따라 JPEG 이나 TIFF 등 다양한 포맷으로 저장, 관리될 수 있으며, 이러한 경우 개정된 PREMIS 3.0에서는 지적 개체와 관련된 표현물들로 식별하여 보존 관리에 관련된 특성을 기술하고, 서로 연계할 수 있도록 수정되었다.

둘째, ‘환경’ 개체를 ‘객체’ 개체에서 분리하여 새로운 독립된 개체로 추가하였다. ‘환경’ 개체는 디지털 객체를 재현하거나 이용할 때 필요한 운영시스템, 응용소프트웨어, 컴퓨팅 자원 등에 대한 기술 정보로, 기존의 객체 개체 데이터 모델과 분리하여 별도의 데이터 모델로 추가시킴으로써, 기술 변화에 따른 파일과 소프트웨어의 상세 정보를 지속적으로 확보할 수 있도록 하였다. PREMIS 3.0의 ‘환경’ 개체는 모듈화 되어 다른 객체들 간이나 리포지터리들 간에 재사용하고 공유하는데 필요한 보존 기술 정보를 제공할 수 있다. 또한 디지털 객체를 재사용하는데 있어서 관련된 행위자, 개체와의 관계 표현이 매우 중요함을 인지하고, ‘환경’ 개체에 대한 데이터 모델에 ‘related EnvironmentPurpose’, ‘relatedEnvironmentCharacteristic’ 등을 추가하였다.

셋째, ‘객체’ 개체에 물리적 객체(Physical Object)의 개념을 추가하였다. 물리적인 객체는 필사본, 인쇄 문서 등의 내용 정보만을 지칭하는 것이 아니라, 관련된 디지털 정보를 구현하기에 필요한 물리적인 하드웨어 기기 등의 ‘환경’ 개체, 디지털 객체의 인쇄형 표현물 등을 의미한다. 실제 보존 업무에서 이러한 물리적 객체가 여러 디지털 객체와 밀접하게 연관되어 있음을 파악하고, 데이터 모델에 물리적 객체를 추가하여 관련 정보를 제공할 수 있도록 하였다. 이러한 ‘객체’ 개체에 대한 수정으로, PREMIS 3.0은 객체의 형태가 디지털이든, 아날로그이든 관계없이 기술할 수 있게 되었으며, 물리적인 객체는 관련된 디지털 객체와 ‘관계’로 연계하여 기술하도록 하였다.

Dappert 등(2016)은 PREMIS 버전 3.0 튜토리얼을 통해 PREMIS 개정에서 가장 주목할 점은 다양한 디지털 리포지터리에서 수행되고 있는 보존 프로세스의 이해를 바탕으로 디지털 자원의 보존 생애주기를 반영하고 있다는 점을 들었다. Dublin Core나 다른 메타데이터 스킴에서 요소세트를 표준으로 제시하고 있는 반면, PREMIS는 다른 메타데이터 표준과 함께 보존 기관의 실무 환경에 따라 사용목적에 맞게 메타데이터 요소를 설계할 수 있는 확장성을 갖는다. PREMIS의 의미단위는 실제 보존 메타데이터 설계 시, 1:1, 1:多, 혹은 多:1의 관계로 보존환경에 따라 다양하게 구현이 가능하다. PREMIS의 데이터 사전은 디지털 리포지터리의 보존 정보패키지를 교환하는데 있어서 포괄적인 표준으로 활용될 수 있으며, 버전 3.0은 여러 디지털 리포지터리의 복잡한 보존 프로세스들을 염두에 두고 OAIS의 SIP(제출정보패키지), AIP(보존정보패키지), DIP(배포정보패키지)에 적용하여 실행하도록 제시하고 있다.

2.3 PREMIS 기반 보존 메타데이터 응용프로파일 설계지침

PREMIS를 기반으로 한 보존 메타데이터 실무 적용 구현 사례들을 수집, 분석한 보고서를 종합하여 OCLC는 2016년 “Digital Preservation Metadata for Practitioners: Implementing PREMIS”(Dappert, Guenther, & Peyrad, 2016)를 출판하였다. 이 보고서는 PREMIS의 특정 버전을 적용하는 방법보다는 현재 작업하고 있는 보존 실무에서 다루고 있는 자원의 특성과 기관의 요구에 따라 PREMIS를 응용하여 보존 메타데이터를 설계하는데 초점을 두고 있다. PREMIS는 디지털 보존 콘텐츠 유형, 기술적인 인프라구조, 조직의 요구에 따라 구현될 수 있다. 보존 메타데이터 실무가가 PREMIS를 기반으로 보존 메타데이터를 설계하거나 구현할 때 고려해야 할 사항들을 실무적인 관점에서 제시하고 있다.

PREMIS 데이터 모델을 기반으로 한 보존 메타데이터의 특징은 기관의 요구수준에 맞추어 필요에 따라 수정, 적용이 가능하다는 것이다. PREMIS가 발표된 이후, 도서관, 기록관, 박물관 등 문화유산 기관을 비롯하여 여러 디지털 아카이브에서 기관의 목적과 웹 정보자원, 동영상, 전자책 등 디지털 객체의 유형에 따라 PREMIS의 데이터 모델과 데이터 사전들을 적용하여 보존 메타데이터를 설계, 실행하는 사례가 많아졌다. PREMIS 편집위원회는 이용자의 요구와 디지털 아카이브 컬렉션의 목적과 검색, 브라우징, 접근, 데이터 교환, 관리 및 보존 활동 등을

종합적으로 검토하여 메타데이터 기능과 수준을 결정하고 적절한 PREMIS 데이터 사전을 선정하여 응용할 것을 권고하고 있다.

또한 PREMIS 데이터 모델과 데이터 사전을 활용하여 보존 메타데이터를 구현 시 고려해야 할 이슈들을 검토하여 PREMIS 기반 보존 메타데이터 응용 프로파일 설계 절차를 제시하고 있다(Peyrad, Dappert, & Guenther, 2016). 우선 보존 메타데이터 설계 및 실행 시 메타데이터의 핵심 요구사항이 무엇인지 파악하고, 디지털 보존 대상 객체나 자료로부터 메타데이터를 수동 또는 자동으로 추출할 수 있는지 검토한다. 그 다음 여러 출처에서 입수되는 다양한 포맷을 포괄적으로 수용할 수 있는 적합한 메타데이터 표준을 선택하고 마지막으로 메타데이터 표준을 기관의 보존 환경 및 디지털 자원 포맷에 맞추어 적용하기를 권고하고 있다.

보존 메타데이터는 보존을 수행하는 기관의 목적이나 상황에 따라 그 적용 범위가 다양하므로 표준화된 메타데이터 모델을 제안하는 것은 현실적으로 무리가 있다. 디지털 아카이브에서는 PREMIS 데이터 모델을 메타데이터 구조의 개념 모델로 삼아 특정한 보존 환경 시나리오에 적합한 맞춤형 보존 메타데이터 구현이 가능하다. 보존 대상과 범위가 결정되면 PREMIS 기반의 보존 메타데이터 응용프로파일 설계를 위한 구체적인 전략을 수립하여야 하는데 먼저 디지털 아카이브의 보존 메타데이터 설계를 위한 체계적인 요구 분석이 필요하다. 보존 대상이 되는 디지털 객체의 유형과 속성, 구조적 관계 등을 파악하고 보존 위험과 기능 분석의 종합적인 검토를 통해 구체적인 보존 시나리오를 파악하여야 한다. 위험 분석은 조직의 정책, 재정 및 법적, 기술적인 위험 등을 고려하여, 현재의 메타데이터가 실질적으로 디지털 자원의 장기보존에 적합한지, 포맷 변환과 같은 보존 활동의 수준과 방법을 평가하는 것이다. 요구 분석을 위해 OASIS 참고모형과 위험 분석을 위해서는 SPOT, DRAMBORA 등을 프레임워크로 활용할 수 있다(Peyrad, Dappert, & Guenther, 2016, p. 12).

3. PREMIS 3.0 기반 보존 메타데이터 사례 분석

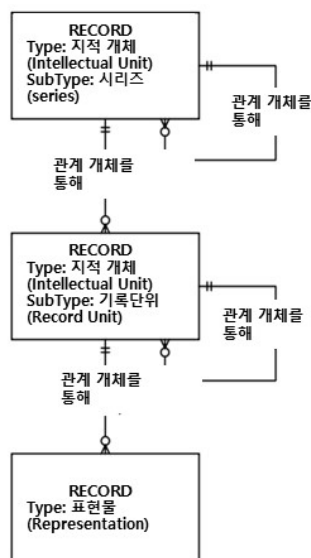
3.1 호주 국가기록원의 기록물시스템(Commonwealth Record Series)

호주의 기록관리 메타데이터 국가표준은 1999년 ‘Record keeping metadata standard for Commonwealth agencies (version 1.0)’으로 제정되었다. 기록물시스템(CRS) 메타데이터는 연방정부 기록물에서 영구보존의 가치를 지닌 기록물을 관리하고 검색하는데 활용할 수 있도록 개발되었다. 처음 개발 시에는 기록을 중심으로 한 단일 개체 모형을 기반으로 한 메타데이터로 설계되었다가 2008년 AS ISO 23081을 반영하여, 기록(Record), 행위자(Agent), 업무(Business), 법규(Mandate), 관계(Relationship)의 5개 개체를 기술하는 다중객체 모형의 메타데이터로 개정되었다.

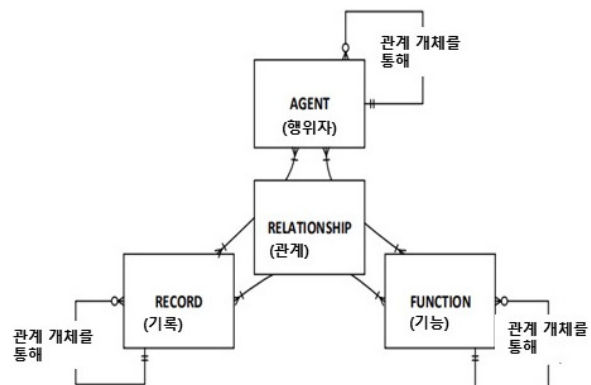
2018년 호주 국가기록원은 기관의 일상적인 디지털 업무를 지원하는 전략, 지침 및 기준을 제공하고자 ‘디지털 연속성 2020 정책(Digital Continuity 2020 Policy)’을 수립하고, 이에 따라 2011년 발표했던 디지털 보존 정책을 개정하였다. 보존 정책은 전자적으로 생산된 기록물(Born-digital Records)과 디지털화된 기록물에 모두 적용된다. 디지털 보존 정책에 정의된 전자기록물은 데이터 파일과 같은 객체로 된 내용정보(Content)와 그 내용을 설명하는 메타데이터로 구성된다. 보존 정책은 내용정보와 관련 메타데이터를 보존하고 그 둘이 서로 분리되지 않도록 영속적으로 연결(Link)하는 것을 원칙으로 한다. 또한 디지털 기록의 장기보존과 접근을 위해서는 기록물에 취해진 모든 보존 행위를 기록하는 메타데이터가 필수적임을 강조하고 있다. 보존 프로세스를 통해 전자기록물에 어떤 변경이 발생하는 경우 메타데이터를 통해 기록되고 관리되어야 하며, 디지털 객체와 메타데이터는 영속적으로 그 관계를 유지할 수 있어야 한다는 것이다(이지영, 2018, p. 31).

‘디지털 연속성 2020 정책’에 기반하여 CRS의 보존 메타데이터도 개정되었다. 기존의 CRS의 보존 메타데이터는 본문과 첨부파일로 구성되어 있는 이메일 기록을 비롯하여 주로 파일, 문서나 첨부파일 등으로 구성되어 있는

디지털 기록물의 복잡한 구조와 시스템을 반영하기에는 적절하지 않았다. 이에 최근 발표된 CRS의 새로운 보존 메타데이터는 디지털 기록의 속성과 구조 모델을 반영하고자 PREMIS 3.0의 데이터 모델을 기록 계층에 적용하였다(Garvie, 2020, p. 104). 전통적인 기록 조직에 사용되었던 기록물계열(series)과 기록물건(item)을 디지털 기록의 특성에 맞추어 PREMIS의 3.0의 객체 구조인 지적 개체와 표현물로 재정의하였다. 이를 통해 다양한 포맷으로 생산, 관리되는 디지털 기록들을 포괄하여 기술하고, 정보기술의 변화에 따른 마이그레이션 등의 보존활동 등으로 여러 버전으로 파생되는 기록물들을 효과적으로 관리하고자 하였다. 개정된 CRS 보존 메타데이터 구조는 다음의 <그림 2>, <그림 3>과 같다.



<그림 2> CRS 보존 메타데이터 구조(Garvie, 2020)



<그림 3> 호주 국가기록원의 기록통제 모델(Garvie, 2020)

CRS 보존 메타데이터는 ‘기록’ 개체에 PREMIS 3.0 데이터 모델의 ‘지적 객체’와 ‘표현물’ 개체의 개념을 적용하고, ‘행위자’ 개체는 외부기관과 연계를 고려하여 조직, 개인에 기타(other)를 추가, 확장하였다. ‘기능’ 개체는 기존의 CRS 시소러스의 내용을 행위자 개체와 연계할 수 있도록 새롭게 독립된 개체 수준으로 추가하였다. ‘관계’ 개체는 기록의 집합체 개념에서 그 범위를 확장하여 관련된 기록 객체 뿐 아니라 기술적인 측면에서 포맷변환이나 버전이 다른 관련 정보까지 표현할 수 있도록 수정한 것이다.

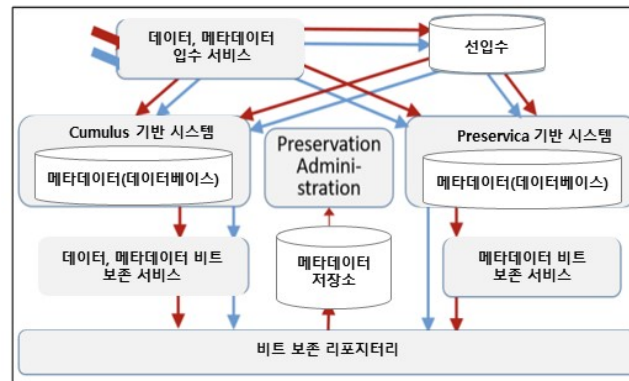
CRS의 보존 메타데이터는 기존의 보존 메타데이터가 영구기록물관리기관으로 이관된 기록을 대상으로 기록의 집합적, 계층적 정리와 기술을 목적으로 설계, 실행되었다면 현재의 보편화된 디지털 기록 환경에서는 기록의 전 생애에 걸쳐 메타데이터가 획득, 관리될 수 있는 유연한 구조가 필요함을 보여주고 있다.

3.2 덴마크 왕립도서관의 디지털 아카이브

덴마크 왕립도서관(Royal Danish Library)은 라디오 및 TV 방송 아카이브와 텍스트, 컴퓨터 게임, 웹 아카이브 등을 관리하였던 두 국립도서관의 아카이브를 통합하면서 각 아카이브에서 관리되고 있었던 다양한 디지털 자원을 통합 관리하고 장기 보존에 필요한 메타데이터가 필요하게 되었다(Zierau, 2019). 이에 PREMIS 3.0의 데이터 모델과 데이터 사전을 기반으로 한 보존 메타데이터를 개발하였다.

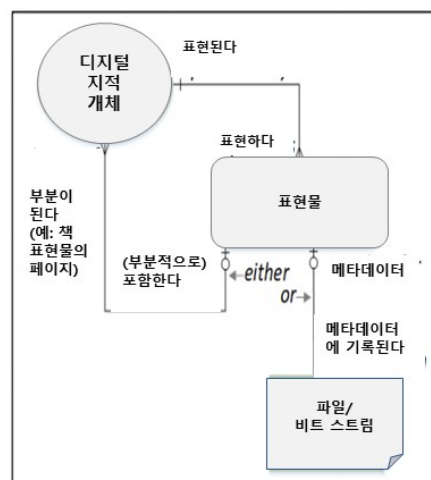
덴마크 왕립도서관은 디지털 자원의 생산된 그대로의 포맷 비트스트림 보존을 원칙으로 하며, 보존 자원의

데이터 무결성, 지속가능성과 이용을 보장할 수 있는 디지털 보존 기술 인프라를 구축하고 시스템 간 상호운용성을 지원할 수 있는 보존 메타데이터를 설계하고자 하였다. 이에 METS와 E-Ark 틀 내에서, 관리 메타데이터 영역에 기술(technical)정보와 디지털 출처에 PREMIS 3.0 데이터 모델과 데이터 사전을 적용하였다(<그림 6> 참조). 다음의 <그림 4>와 같이, 덴마크 왕립도서관 아카이브는 전체적인 디지털 자원의 생애주기를 기반으로 한 보존 프로세스를 반영할 수 있는 보존 인프라를 구축하고, Preservica, Culmmus 등 다양한 디지털자산관리시스템(Digital Asset Management System, DAMS)과 연계할 수 있는 메타데이터를 발표하였다.



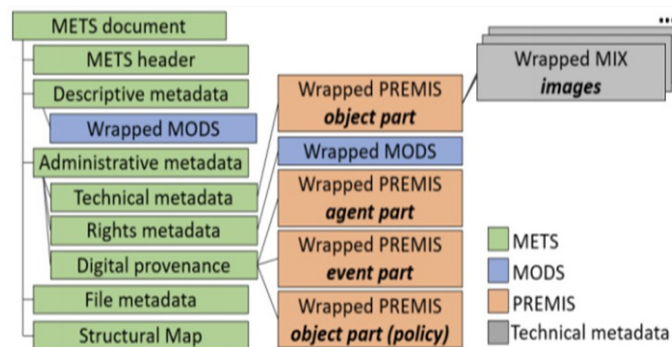
<그림 4> 덴마크 왕립도서관 보존 인프라 구조(Zierau, 2019)

덴마크 왕립도서관의 보존 메타데이터 모델은 PREMIS 3.0과 Planets 프로젝트 모델 등을 참조하여 단순화한 것으로, 디지털 지적 개체(digital intellectual entities), 표현물 개체, 파일/비트 스트림 객체로 구성된다(<그림 5> 참조). 디지털 지적 개체는 특정 디지털 자원을 식별하는 디지털 객체의 최상위 수준을 나타낸다. 디지털 지적 개체는 보존 활동, 조치와 상관없이 본래의 객체를 식별하기 위한 것으로, 여러 판(edition), 버전을 나타내는 복수의 표현물로 구성된다. 표현물은 디지털 지적 객체의 특정 버전, 판을 나타낸 것으로 파일과 해당파일의 메타데이터를 포함할 수 있다. 파일/비트스트림은 데이터 개체에 보존된 파일을 의미한다. 이러한 데이터 모델은 디지털 보존에서 발생할 수 있는 다양한 변화, 즉 자원의 디지털화, 초기 스캐닝 불량으로 인한 재디지털화, 포맷변환, 관련 메타데이터 수정 등 하나의 디지털 자원에 행해진 다양한 보존 행위를 기록할 수 있도록 한 것이다.



<그림 5> 덴마크 왕립도서관 보존메타데이터 모델(Zierau, 2019)

덴마크 왕립도서관의 보존 메타데이터는 디지털 객체와 메타데이터의 변경, 업데이트를 감사 추적하고 기록할 수 있도록 METS 구조에 PREMIS 3.0의 ‘객체’ 개체 메타데이터로 디지털 객체 관리에 필요한 기술(파일 형식 및 디지털화 정보) 정보와 ‘행위자’, ‘이벤트’, ‘객체’ 개체 메타데이터에 디지털 출처 정보를 기술하도록 하였다(<그림 6> 참조). 이로써 디지털 객체에 대한 데이터 변환과 메타데이터 업데이트 등으로 인해 시스템으로 이관하는 것이 지체되는 과정에서 발생할 수 있는 정보손실의 위험을 최소화할 수 있게 되었다. 디지털 지적 재산과 표현물, 파일 간의 관계는 덴마크 왕립도서관의 보존 메타데이터의 METS 내 structMap이나 fileSec 섹션에 기술되며, PREMIS 3.0 데이터 사전에 기반으로 한 관계 메타데이터(relationshipType, relationshipSubType) 등에 명시된다(Royal Danish Library, 2019).

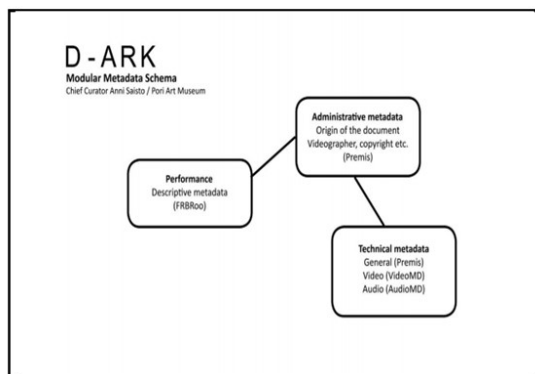


<그림 6> 덴마크 왕립도서관 METS 구조(Zierau, 2019)

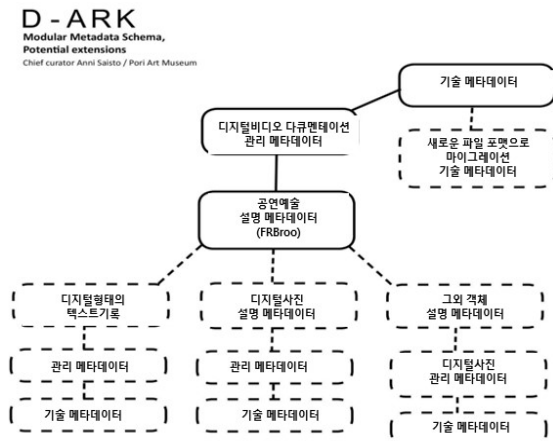
3.3 핀란드 공연예술 D-ark 아카이브

핀란드의 공연예술 디지털 아카이브인 D-ark(digital performance art archive)는 핀란드 예술 커뮤니티인 T.E.H.D.A.S가 주최한 공연 예술 행사의 다큐멘타리를 보존하기 위해 구축되었다. 이 아카이브는 2005년부터 클럽, 갤러리, 스트리트에서 이루어졌던 300 회 이상의 개별 공연들을 촬영한 비디오 자료를 중심으로, 핀란드의 현대 공연 예술 자원을 보존하기 위한 것이다. D-ark는 핀란드 포리 미술관(Pori Art Museum)과 협력하여 공연예술 아카이브의 체계적인 관리를 위해, CIDOC CRM 및 FRBRoo, PREMIS 3.0을 기반으로 하는 메타데이터를 설계하였다(Saisto, 2019).

D-Ark의 보존 메타데이터는 <그림 7>과 같이 공연예술에 대한 설명 메타데이터는 FRBRoo를 적용하여 기술하고, 디지털 자원 관리와 보존을 위한 기술 메타데이터와 관리 메타데이터에 PREMIS 3.0의 데이터 사전에 적용하였다. 공연예술자원의 관리 메타데이터 요소에는 PREMIS 3.0 데이터 사전의 documentation event, copyright status, date of copyright, the determination date of the copyright status 등이 적용되었다. 또한, 디지털 객체에 취해진 보존조치를 기록하기 위해서 PREMIS 2.2의 통제어휘를 사용하여 migrate, modify, use, disseminate, delete 등의 행위를 기록하도록 하였다. 기술 메타데이터로는 공연예술 아카이브 자원의 특성상 비디오나 오디오의 포맷표준인 VideoMD, AudioMD를 사용하거나 PREMIS의 ‘객체 특성(object characteristic)’을 확장하여 사용하도록 하였다(<그림 8> 참조).



〈그림 7〉 D-Ark 보존 메타데이터 모델(Saisto, 2019)



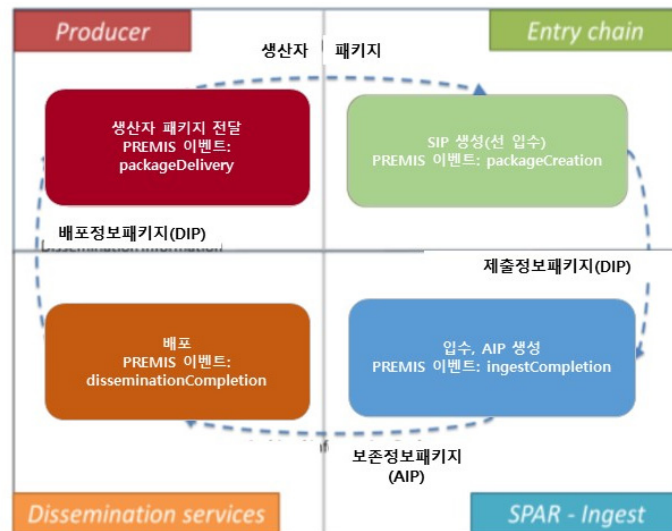
〈그림 8〉 D-Ark 보존 메타데이터 구조(Saisto, 2019)

D-Ark의 보존 메타데이터는 확장성을 염두에 둔 유연한 구조를 갖는 것이 가장 큰 특징이다. 보존 메타데이터는 아카이브 자원이 향후에 여러 시스템에서 학술 연구 등 다양한 목적으로 재사용할 수 있도록 디지털 자원이 생산되거나 획득되는 기술 환경을 식별하고, 어떤 환경이나 플랫폼으로부터 다른 환경이나 다른 보존 전략으로 효율적이며 성공적으로 마이그레이션 할 수 있도록 설계되었다. 이러한 메타데이터는 실행환경 속에 정의된 스키마의 범위를 넘어서는 다른 요소들의 추가를 허용하는 확장성을 갖고 있다. 최소한의 디지털 자원의 보존을 위한 핵심 요소들이 적절하게 처리되고 그 의미의 일관성이 유지될 수 있다면, 부가적인 메타데이터 요소들은 추후 목적에 맞추어 추가될 수 있다. D-Ark의 보존 메타데이터 모듈화는 디지털 아카이브의 기능성을 보장하면서 부가적인 요소들을 필요에 따라 추가, 확장하는 데 유용하다 할 것이다.

3.4 프랑스 국립도서관 SPAR(Système de Préservation et d'Archivage Réparti) 시스템

프랑스 국립도서관인 BnF는 프랑스 내의 모든 출판물을 수집, 보존하고 대중에게 제공하는 것을 목표로, 납본을 1537년 인쇄물에서, 1992년 전자문서, 2006년 인터넷에 이르기까지 다양한 형태의 출판물로 확대하였다. 디지털 문화유산을 보존하기 위해서 BnF은 확장 가능한 보존 및 아카이빙 시스템(SPAR)을 설계했다. 이 시스템은 OAIS 표준을 준수하고, 오픈 소스 기술 적용을 기반으로 모듈화할 수 있도록 설계되었다.

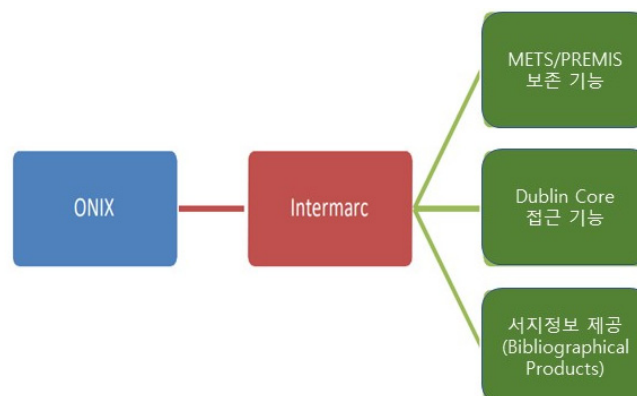
SPAR에서 가장 중요한 부분은 디지털 콘텐츠의 보존으로, 자원의 무결성과 장기적 이용가능성을 지원하는 것이 매우 중요하다. 이를 위해서 FTP 플랫폼으로 납본 제출 시, 자동으로 자원의 무결성과 포맷을 체크하도록 한다. SPAR를 통해 온라인 납본으로 제출되는 디지털 콘텐츠에 대한 보존 메타데이터는 자동으로 추출, 생성되는 것을 원칙으로 한다. 출판사로부터 납본 제출 시 ONIX 메타데이터에서 데이터 무결성과 포맷 관련 정보를 추출한 InterMarc는 보존 목적으로 METS와 PREMIS 기반의 메타데이터를 부여한다(〈그림 9〉 참조). 정보검색 서비스를 위해서는 Dublin Core와 효율적인 관리를 위한 서지 데이터를 부여한다. 보존정보패키지(AIP)는 보존될 디지털 파일에 따라 데이터 객체, 데이터 객체 이해에 필요한 표현정보와 보존기술정보(PDI)를 제공할 수 있는 메타데이터를 포함해 미디어에 패키지로 저장된 다양한 컴포넌트 간의 실질적, 논리적 링크를 설명하는 패키징 정보를 포함한다(Derrot, S. & Oury, C., 2014).



〈그림 9〉 SPAR 보존 메타데이터 구조(Derrot, S. & Oury, C. 2014, p. 7)

SPAR의 보존 메타데이터는 BnF의 디지털 콘텐츠의 장기적인 보존 관점에서 입수부터 관리, 제공까지의 디지털 콘텐츠의 생애주기 전반에 걸쳐 효율적인 메타데이터의 실행과 결정 과정에서 필요한 세부 내용들을 제공할 수 있도록 설계되었다. SPAR 보존 프로세스에서 입수부터 관리, 제공 과정에서 다양한 유형의 정보 패키지를 적용되며, 패키지는 고정된 형태로 존재하는 것이 아니며 보존 활동의 어떤 단계에 적용되는지에 따라 그 구성이 다르게 나타남을 강조하였다. SPAR 메타데이터는 SPAR의 보존 활동 전반에 필요한 기술과 보존기능을 지원하여야 하므로, 디지털 콘텐츠의 생애주기 전반에 걸쳐 각각의 단계에 맞게 PREMIS 기반의 메타데이터를 통해 관련 기술사항을 제공할 것을 제안한다(Caron, B., Houssaye, J., Ledous, T., & Reeht, S., 2017).

다음의 <그림 10>과 같이 PREMIS 기반의 보존 메타데이터는 생산자로부터 패키지가 생성되는 단계에서부터 선입수 시 SIP와 입수 후 AIP, DIP에 이벤트 개체를 통해 구현된다. SPAR의 보존 메타데이터는 장기적 보존 관점에서 디지털 콘텐츠의 생애주기 전반에 적용되는 PREMIS 3.0의 ‘이벤트’ 개체와 데이터 사전을 활용하여 보존 활동을 기술하는데 필요한 핵심적인 요소들을 제안하고 정보패키지 간의 상호운용성을 지원해주는 환경 구축을 제안하고 있다. 이로써 SPAR에서 행해지는 디지털 객체의 변화를 관련된 사건을 순서대로 기록하고 향후 객체의 내용이나 외관에 있어서의 변화를 추적하고 용이하게 관리할 수 있게 되었다.



〈그림 10〉 SPAR 정보패키지와 메타데이터(Caron, B., Houssaye, J., Ledous, T., & Reeht, S., 2017, p. 7)

3.5 PREMIS 3.0 적용 사례 분석의 시사점

본 연구에서 PREMIS 3.0을 적용한 보존 메타데이터 사례들을 조사, 분석한 내용을 종합해보면 기관의 보존 전략과 환경에 따라 그 적용 수준과 내용이 매우 다양하였다. PREMIS 3.0 데이터 모델을 기반으로 하여 보존 메타데이터의 개념 모형을 구축한 포괄적 수준(Garvie, 2020)에서부터 역동적인 디지털 보존활동을 반영할 수 있도록 PREMIS 3.0 데이터 사전에서 필요한 요소들을 추가, 확장 시킨 응용프로파일의 수준(Saisto, 2019; Zierau, 2019) 까지 다양하였다. 사례 기관에서 적용하고 있는 PREMIS 3.0의 데이터 모델과 데이터 사전들의 내용은 다음의 <표 1>과 같다. PREMIS 3.0의 데이터 모델의 개체를 중심으로 비교하면, 사례 보존 메타데이터는 크게 ‘객체’ 개체 중심의 보존 메타데이터와 ‘이벤트’ 개체 중심의 보존 메타데이터로 구분할 수 있다.

<표 1> PREMIS 3.0 적용사례 비교

	CRS (Garvie, 2020)	덴마크 왕립도서관 디지털 아카이브 (Zierau, 2019)	D-Ark (Saisto, 2019)	BnF의 SPAR (Caron et al., 2017)
PREMIS 개체	• 객체: 지적개체-표현물 • 행위자	• 객체: 디지털 지적 개체 - 표현물 - 파일/비트스트림 • 행위자 • 이벤트	• 권한 • 이벤트	• 이벤트
PREMIS 의미단위		• 관계, 관계유형(relationship Type, relationship SubType)	• 객체(object characteristic) • 이벤트, 권한: 디지털 객체의 생산, 접근 범위, 저작권 등(documentation event copyright status date of copyright the determination date of copyright status)	• 이벤트: 생애주기 전 반에 걸쳐 보존활동 이력을 기술
연계 메타데이터	• METS • DC	• METS • MODS • DC • MIX	• METS	• ONIX • METS • DC

디지털 객체의 특성을 반영하도록 하는 객체 중심의 보존 메타데이터는 호주 국립기록원의 CRS와 덴마크 왕립 도서관 디지털 아카이브 보존 메타데이터라 할 수 있다. 자원 중심의 메타데이터 모형에서 가장 중요한 영역은 ‘객체’이다. CRS와 덴마크 왕립도서관 디지털 아카이브는 (디지털) 지적개체, 표현물, 파일/비트스트림으로 구성 되는 PREMIS 3.0의 ‘객체’ 개체를 적용하여 하나의 디지털 자원에서 파생되는 관련 정보들을 포괄하여 기술하고 있다. 특히, 텍스트 뿐 아니라 동영상, 그래픽, 애니메이션 등 다양한 컴포넌트로 구성되어 있는 디지털 콘텐츠를 대상으로 하는 보존 아카이브의 경우, 각 객체들을 독립적인 개체로 식별하고 개체들 간에 존재하는 다양한 관계 들은 ‘관계’ 개체로 기술하고 있다. 이러한 접근은 디지털 자원의 구조와 관계를 체계적으로 기술하여 이용자의 구조적 접근과 발견을 돕는 데 목적이 있다. 디지털 자원의 양이 급증하고 그 매체의 성격과 유형이 다양해지면서 하나의 객체를 시작으로 해서 다양한 유형의 다른 객체, 관련된 정보들을 찾고 발견할 수 있는 기능이 중요해졌다. 디지털 자원의 생애주기 동안 파생되는 다양한 정보자원들을 ‘객체’와 ‘관계’의 개체를 통해 논리적으로 연결하여 이용자의 검색과 네비게이션을 지원하고 있다. 또한, 기존의 PREMIS 2.2 데이터 사전의 범위 밖에 있었던 지적 개체를 PREMIS 3.0에서는 객체 수준으로 포괄시켜, 객체 수준에 적용되었던 데이터 사전의 의미단위를 지적 개체에도 모두 적용할 수 있게 되었다. 따라서 하나의 지적 개체와 관련된 모든 자원, 즉 업데이트되거나 포맷이 변환된 자원들을 이용자가 파악하고 자신의 목적에 맞는 자원을 선택할 수 있게 되었다. CRS는 ‘객체’를 지적 개체, 표현물의 구조로, 덴마크 왕립도서관 디지털 아카이브는 디지털 지적 개체, 표현물로 자원의 다양한 관계를

효과적으로 표현할 수 있도록 하고 있다.

이벤트 중심의 보존 메타데이터는 디지털 자원의 생애주기에서 단계마다 발생하는 사건과 변화 내용을 기술하여 진본성과 무결성을 보장하는데 초점을 두고 있다. PREMIS는 ‘이벤트’ 개체에 디지털 자원의 출처와 관리 이력에 대해서 다양한 기술과 업무 환경에서 생산된 디지털 자원을 획득하고 그 이후 구조적 변화나 관련된 정보를 기술하도록 한다. 특히 PREMIS 3.0은 ‘이벤트’ 개체와 함께 ‘권한’ 개체를 ‘객체’와 관련된 다른 설명 메타데이터, 관리 메타데이터와 연계하여 디지털 자원의 생애주기에 걸쳐 중요한 보존과 출처 정보를 기술할 수 있도록 하고 있다. D-Ark의 경우, 공연예술정보의 특성상 저작권과 관련된 정보가 매우 중요하며 ‘권한’ 개체와 ‘이벤트’ 개체를 통해 저작에 대한 권리와 행위가 발생하는 과정에서 관련된 행위자와 저작권 상태, 생성날짜, 변화 내용 등을 기록화 하도록 하였다. BnF의 SPAR의 경우 PREMIS의 ‘이벤트’를 통해 디지털 자원의 전 생애주기 동안 발생하는 변화들을 상세히 기술하여 어떤 환경이나 플랫폼에서도 성공적으로 마이그레이션 할 수 있도록 지원하고 있다.

이러한 객체와 이벤트 중심의 보존 메타데이터는 모두 디지털 자원이 장기간에 걸쳐 이용, 관리할 수 있도록 기술적 정보 제공을 최우선으로 하고 있다. 기술적 정보는 디지털 자원이 생산되거나 획득되는 기술 환경에 대한 정보와 그 자원이 재생산되는데 필요한 관리적 정보로 PREMIS 3.0의 ‘객체’ 개체 내의 중요속성 정보와 ‘환경’ 개체의 정보로 제공된다. 이는 여러 시스템을 통해 획득된 디지털 자원을 지속적으로 이용 가능할 수 있을 뿐 아니라 여러 시스템에서 다른 목적으로도 재사용 가능할 수 있도록 하는 것으로 상호운용성 관점에서 매우 중요하다. PREMIS 3.0의 ‘환경’ 개체는 ‘객체’ 개체에 포함되는 객체의 중요한 속성, 소프트웨어, 하드웨어 등의 상세 정보들은 포괄한 것으로, 관리에 초점을 맞춘 다른 메타데이터에서 모듈처럼 재사용할 수 있다. 본 연구에서 조사한 보존 메타데이터 사례들은 모두 ‘환경’ 개체를 통해 관련된 디지털 자원의 기술적 정보들을 최소한의 노력으로 확보할 수 있도록 더블링크어를 비롯한 다른 외부 메타데이터와 연계하여 자동으로 추출하며 이를 METS 구조에서 재사용하고 있었다.

4. 디지털 아카이브의 보존 메타데이터 설계를 위한 고려사항

4.1 디지털 보존 위험 분석과 요구 분석

본 연구는 PREMIS 3.0의 개정 내용을 중심으로 보존 메타데이터 설계에 적용한 다양한 사례들을 살펴보고, 보존 메타데이터 설계에 고려해야 할 사항들을 파악하였다. 이를 토대로 신뢰할 수 있는 디지털 아카이브의 보존 메타데이터 설계 시, 고려해야 할 메타데이터 핵심 요소들을 제안하고자 한다. PREMIS의 응용프로파일 설계 지침에 따라 PREMIS 3.0 데이터 모델과 데이터 사전을 참고하여 디지털 아카이브의 보존 메타데이터에 필요한 요소들을 중심으로 구성하였다(Peyrad, Dappert, & Guenther, 2016).

Peyrad 등(2016)이 제시한 응용프로파일 설계 방법론에 따르면, 디지털 아카이브의 PREMIS 기반 보존 메타데이터 설계를 위한 SPOT(Simple Property Oriented Threat)과 OAIS 디지털 보존 요구 분석이 우선되어야 한다. SPOT은 뉴질랜드 통계청, OCLC, 그리고 플로리다 대학에서 디지털자료의 보존을 위해 개발된 것으로, 위험평가를 위하여 사용성, 정체성, 지속성, 렌더링 가능성, 이해가능성, 진본성이라는 6가지 필수 속성을 설정하고 속성별로 구체적인 위험요소를 제시하고 있다(<표 2> 참조).

〈표 2〉 디지털 보존 위험 평가요소 (Peyrad, Dappert, & Guenther(2016, p. 15), 국립기록원(2016, p. 74) 재구성)

보존 목표*	방법	위험요소
사용성 (Availability)	획득, 선별 (Capture, Selection)	• 시간 경과에 따른 디지털객체의 가치와 사용성 악화 • 디지털객체의 소멸 가능성 • 소유자의 접근 거부
정체성 (Identity)	기술 (Description)	• 불충분한 메타데이터 • 메타데이터와 실체 객체와의 연관성 부족 • 메타데이터의 사용성 부족
이해가능성 (Understandability)	기록화 (Documentation)	• 사용자 관심 미 고려 • 사용자에게 대한 정보제공 부족 • 정보의 전체 표현성 미확보에 의한 향후 이해불가능성
지속성 (Persistence)	저장, 매체관리 (Secure storage, Media management)	• 부정확한 제어 • 저장소 수명 종료 • 매체에 대한 독해 도구 사용불가 • 매체와 비트스트림에 대한 악의적 손상 • 매체와 비트스트림의 부주의에 의한 손상
렌더링 가능성 (Renderability)	포맷변환 (Format strategies)	• 하드웨어와 소프트웨어 연계가 부적절함 • 하드웨어와 소프트웨어가 운영, 관리되지 않음 • 정확한 렌더링 환경이 파악되지 않음 • 렌더링 대상의 정확한 특성이 파악되지 않음
진본성 (Authenticity)	진본확인 (Authentication)	• 객체의 메타데이터와 기록화 부족 • 메타데이터의 기술 오류 • 변화한 디지털객체에 대한 기록 부재

* 보존 목표는 보존 피라미드 계층 구조로 사용성을 가장 기본 구조로 하여, 최상위 단계의 진본성으로 구성된다(Peyrad, Dappert, & Guenther, 2016, p. 15)

OAIS 표준과 SPOT 표준에 기반한 보존 위험 분석을 통해 도출된 보존 메타데이터의 핵심 요소를 정리하면 <표 3>과 같다. 객체식별자, 객체특성, 관계, 행위자식별자, 이벤트식별자, 이벤트유형, 이벤트일시, 권한선언은 PREMIS 3.0 기반의 보존 메타데이터 사례(Saisto, 2019; Zierau, 2019)와 관련 연구(Dappert, Guenther, & Peyrad, 2016)에서 공통적으로 제안하고 있는 요소들로 보존 전반에 걸쳐 반드시 필수적으로 적용되어야 하는 요소로 고려되어야 한다.

〈표 3〉 요구 분석에 따른 보존 메타데이터의 핵심 요소

SPOT 속성	OAIS 정보	PREMIS 개체 / 의미단위	보존 메타데이터 요소(안)
사용성	참조 정보 맥락 정보 접근권한 정보	권한 개체, 권한 관련 행위자 객체 개체: ObjectIdentifier, storage, preservationLevel	객체식별자 권한선언 행위자식별자
정체성	참조 정보	객체 개체: ObjectIdentifier, originalName	객체식별자
지속성	무결성 정보	객체 형태를 변화시키거나 무결성을 체크하는 이벤트 (예: 무결성 체크, 포맷과 매체 마이그레이션, 위 이벤트 관련 행위자)	이벤트 식별자 이벤트유형 이벤트 일시
	참조 정보 출처 정보	객체 개체: preservationLevel, storage, fixity, size, format 환경 정보	객체특성
렌더링 가능성	구조 또는 표현정보 출처 정보	객체 개체: preservationLevel, format, environment, creatingApplication, structural relationships, objectCharacteristics 객체의 포맷을 변화시키는 이벤트 보존 행위를 수행하는 권한	객체특성 이벤트 식별자 이벤트유형 이벤트 일시
이해 가능성	의미 표현정보 변환 정보 속성	객체: 관련 지적 개체에 연결하는 relatedObjectIdentifier, significantProperties 지적 개체의 외부 설명메타데이터*	관계 객체특성 *설명메타데이터와 연계
진본성	무결성 정보	객체 개체: objectCharacteristics (예: size, fixity, signatureInformation)	객체특성

사용성과 정체성 측면에서 시간경과에 따른 디지털 객체의 가치를 판단하고 지속적으로 사용가능할 수 있도록 디지털 객체의 생성에서부터 물리적 형태의 변화에 이르기까지 보존의 대상이 되는 디지털 객체에 대한 포괄적인 기술이 필요하다. 또한 진본성 측면에서 보존된 디지털 객체의 무결성을 확인할 수 있도록 이와 관련된 정보 또한 충분하게 기술해 주어야 한다. 이를 위해 객체 개체 메타데이터 요소와 관련하여 객체식별자와 무결성, 포맷 등의 객체 특성이 필수적으로 제공되어야 한다.

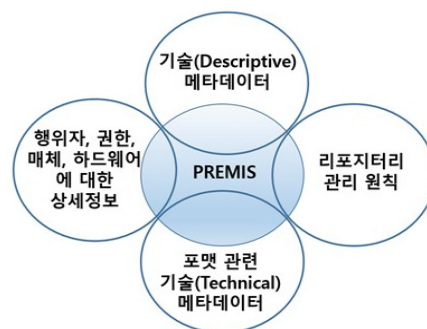
사용성 측면에서 보존된 디지털 객체 뿐 아니라 향후에도 지속적으로 이용가능한데 영향을 끼치는 사항과 해당 객체의 저작권 관련 사항을 포괄적으로 기술하여야 한다. 이는 PREMIS 3.0 데이터 모델의 권한 개체와 관련한 메타데이터 요소로 연결행위자 식별자, 연결객체 식별자와 같은 권한선언이 필요하다.

지속성과 렌더링 가능성 측면에서 디지털 객체의 보존과 관련된 제반 활동을 포괄적으로 기술해 주며 보존활동의 결과로 발생한 사항에 대해서도 상세하게 기술해 주어야 한다. 따라서 디지털 보존 메타데이터는 이벤트 개체 메타데이터 요소와 관련하여 객체 형태를 변화시키거나 무결성을 체크하는 무결성정보 등 보존행위에 대한 이벤트식별자, 이벤트유형, 이벤트일시 정보가 필수적으로 제공되어야 한다.

이해가능성 측면에서 보존의 대상이 되는 디지털 객체와 관련된 다른 객체에 대한 관계를 기술할 수 있어야 하며, 이를 위해 관계가 필수적으로 제공되어야 한다.

4.2 디지털 자원의 생애주기와 메타데이터 상호운용성

PREMIS 3.0은 PREMIS의 실행성과 상호운용성 관점에서 보존 생애주기를 반영하고 자원에 대한 다른 설명, 기술 메타데이터와 연계할 수 있도록 데이터 모델의 ‘객체’의 주요 속성과 연관된 ‘이벤트’, ‘행위자’, ‘권한’ 정보를 상호 연결하도록 하고 있다. PREMIS 보존 메타데이터는 디지털 객체에 관한 포괄적인 기술사항을 기반으로 하고 있기 때문에, 기존의 디지털 아카이브에 디지털 자원이 입수, 관리되는 과정에 생성되는 기술 메타데이터, 관리 메타데이터와의 상호운용성을 지원하며 효율적으로 보존 메타데이터를 생성, 관리할 수 있다(<그림 11> 참조). 본 연구의 보존 메타데이터 사례 분석을 통해서, 디지털 아카이브에서 납본, 수집, 구독, 연계 등 다양한 방법으로 입수되는 디지털 자원의 장기적 보존을 위해서 입수, 보존, 접근에 걸친 아카이빙의 전 단계에 걸쳐 적용할 수 있는 메타데이터 요소를 확인할 수 있었다(<표 4> 참조). PREMIS 3.0은 디지털 자원의 보존 생애주기 동안 관련된 서지 메타데이터나 포맷관련 기술 메타데이터와 연계하여 필요한 보존 메타데이터 요소를 확보하도록 하고 있다. 이는 PREMIS 데이터 모델의 객체 개체와 관련한 식별기호, 콘텐츠 유형, 형태 정보, 관련 자료를 비롯하여 이벤트 개체의 이벤트 시점, 행위자 개체와 관련한 행위자 정보, 권한 개체와 관련된 정보들로 보존 메타데이터의 필수 항목으로 고려되어야 한다.



〈그림 11〉 PREMIS와 설명 메타데이터, 기술 메타데이터와 연계

〈표 4〉 디지털 자원의 생애주기를 고려한 보존 메타데이터 요소

PREMIS 3.0		자동추출/ 연계 메타데이터 요소			보존 메타데이터 요소(안)
개체	의미단위	Dublin Core 요소	MODS 요소	WARC 기술 요소	
객체 Object	1.1 objectIdentifier	Identifier	identifier		객체식별자유형 객체식별자값
	1.2 objectCategory	Identifier	identifier		객체범주
	1.5 objectCharacteristics	Type, Format	typeOfResource physicalDescription	Format	객체특성
	1.14 linkingObjectIdentifier	Relationship	relatedItem		연결객체식별
이벤트 Event	2.1 eventIdentifier	Date		Date	이벤트일시
	2.2 eventType				
	2.3 eventDateTime				
행위자 Agent	3.1 agentIdentifier	Creator Contributor Publisher		Operator	행위자식별자유형 행위자식별자 값 행위자명
권한 Rights	4.1.1 RightsStatement Identifier	Right	accessCondition		권한문식별자값 권한근거

디지털 아카이브에 웹 아카이빙, 온라인 납본, 수집, 디지털화 등의 다양한 경로를 통해 디지털 자원이 입수되면 보존에 필요한 메타데이터 정보들은 자동적으로 추출될 필요가 있다. 본 연구에서 조사한 보존 메타데이터 사례 모두 디지털 자원이 디지털 아카이브에 최초 입수되는 시점에서 디지털 자원의 데이터 무결성과 포맷 관련 정보는 더블링크어와 연계하여 자동으로 추출하도록 하였다(Caron et al., 2017; Garvie, 2020; Zierau, 2019). 그 외, BnF의 SPAR는 ONIX(Caron et al., 2017)와 덴마크 왕립도서관 디지털 아카이브는 MODS와 웹 자원 저장 포맷의 표준인 WARC(Zierau, 2019)에 기술되어 있는 기술 요소들을 연계하여 관련 메타데이터 요소를 자동으로 입력되도록 하였다. WARC 파일 포맷에 정의된 필드(Named-field)와 코드 형식(Record Types)으로 저장되어 있는 정보는 보존 메타데이터의 객체 특성정보로 연계되어 자동, 추출될 수 있다. 콘텐츠 형식(Content-Type)과 관련해서 WARC의 기본적으로 표준화된 Application/warc-fields가 저장되며, 해당 WARC에 저장된 콘텐츠 크기(Content-Length)가 저장된다.

이처럼 디지털 아카이브의 성격에 따라, 아날로그 자원을 디지털화하여 입수하거나 외부 기관에서 수집, 수확되는 디지털 자원은 더블링크어나 MODS 등 해당 기관의 설명 메타데이터로부터 보존 메타데이터의 객체 정보로 연계되어 식별기호, 콘텐츠유형, 형태정보 등이 메타데이터 요소들로 자동, 추출될 수 있다. PREMIS 3.0 기반 보존 메타데이터 요소는 객체식별자유형, 객체식별자값, 객체범주, 객체특성, 연결객체식별, 이벤트일시, 행위자식별자유형, 행위식별자값, 행위자명, 권한식별자값, 권한근거로 관련된 다른 메타데이터와 연계하여 자동으로 입력되는 것을 제안하였다. 이로써 디지털 아카이브의 보존 자원 생애주기 전반에 걸쳐 불필요한 중복 업무를 줄이고 다른 메타데이터와의 상호운용성을 확보할 수 있을 것이다.

한편, 디지털 자원이 아카이브에 입수된 이후에는 디지털 자원의 무결성, 이용가능성 보장을 위해서 자원 상태에 영향을 미치는 모든 관리, 보존처리에 따른 이력 정보가 메타데이터에 기록되어야 한다. 장기보존 관점에서 디지털 자원의 생애주기 동안, 매체변환, 접근 및 이용권한의 변경 등과 같은 보존 관련 행위가 발생할 때 무슨 일이 일어났는지, 언제 발생했는지, 왜 발생했는지, 그리고 누가 수행했는지 등 발생한 모든 행위 이력에 대한 정보가 보존 메타데이터에 저장되어야 한다. 이를 위해 제안하는 디지털 아카이브의 보존 메타데이터에서는 디지털 자원의 입수 후, 보존관리 이력과 관련하여 이벤트 개체(이벤트 식별자, 이벤트유형)와 함께 객체(관계), 행위자(연결

이벤트 식별자, 연결권한선언 식별자), 권한(권한문식별자 유형)요소를 필수값으로, 보존행위를 수행한 행위주체와 행위유형, 행위일자를 명확하게 기술할 수 있도록 하였다. 또한, 디지털 자원의 장기보존을 위해서 PREMIS 3.0 데이터 사전에서 관계유형, 관계하위유형, 연결이벤트식별자, 연결권한식별자, 이벤트식별자, 이벤트유형, 이벤트상세사항, 이벤트결과정보, 연결행위자식별자, 연결객체식별자, 연결이벤트식별자, 연결권한선언식별자, 권한확장 등의 메타데이터 요소를 추가, 입력할 것을 제안하였다(<표 4> 참조).

5. 결론 및 제언

보존 메타데이터는 원 데이터에 대한 접근과 검색을 제공하고, 효율적으로 정보자원을 관리할 수 있도록 지원하는 것을 목적으로 한다. 이를 위해 보존 메타데이터는 디지털 자원의 장기간 유지가능성, 재연가능성, 이해가능성, 진본성, 무결성을 제공할 수 있어야 한다. PREMIS는 2005년 발표된 이후 디지털 자원의 장기보존을 위한 가장 종합적이고 포괄적인 메타데이터 요소 스키마로 받아들여지고 있으며, 2015년 보존 생애주기를 반영할 수 있는 PREMIS 3.0이 발표되었다. PREMIS 3.0은 다양한 디지털 리포지터리에서 수행되고 있는 보존 프로세스를 바탕으로 디지털 자원의 보존 생애주기를 반영하고 있다. PREMIS를 기반으로 하는 보존 메타데이터는 디지털 자원의 입수 과정에서 기술 메타데이터 레코드와 연계를 통해 최소한의 메타데이터 요소가 자동으로 생성될 수 있는데, 장기적인 이용가능성 측면에서는 객체의 중요한 속성, 관계, 이벤트, 환경 정보 기술이 반드시 보존 메타데이터의 핵심 요소들로 구현되어야 한다.

본 연구는 PREMIS 3.0을 기반으로 보존 메타데이터를 설계한 국외의 실제 사례들을 중심으로 PREMIS 기반 보존 메타데이터 설계 시 고려해야 할 사항들을 검토하고, 이를 토대로 PREMIS 기반의 보존 메타데이터 설계를 위한 메타데이터 요소를 제안하였다. 제안한 보존 메타데이터 요소들은 SPOT과 OAIS 기반의 보존 위험과 기능 분석을 통해 디지털 자원의 장기보존에 꼭 필요한 메타데이터 핵심요소로 구성된다. 또한 디지털 자원의 보존 생애주기에 적용된 기술, 관리 메타데이터 등과의 연계를 통해 자동으로 추출할 수 있는 요소들을 검토하고 디지털 아카이브나 보존 리포지터리에서 적용할 수 있는 PREMIS 기반의 보존 메타데이터 요소를 제안하였다.

디지털 보존 생애주기 전반에 걸쳐 수집, 관리, 보존, 이용 등 단계에서 정보자원의 무결성을 확보하고 지속적인 접근을 위해서는 각 단계의 작업이 유기적으로 이루어질 수 있도록 메타데이터의 통합적 연계가 필요하다. 디지털 아카이브의 입수 단계에서는 디지털 자원의 하드웨어, 소프트웨어 요소 등 보존에 있어서 중요한 정보가 자동으로 추출하여 보존 메타데이터로 연계하고 추후 보존 단계에서 정보패키징 변환을 지원하고 관련 기술 메타데이터를 추출하여 보존 메타데이터에 등록, 관리할 수 있다.

추후 디지털 아카이브의 국제적 수준의 상호운용성을 확보하기 위해서는 METS와 PREMIS 기반의 연계형 메타데이터 구현 방안을 모색할 필요가 있다. 제안하는 PREMIS 3.0 기반의 보존 메타데이터의 요소는 직접적인 메타데이터 요소 개발에는 한계가 있으나, 디지털 아카이브의 보존 활동을 기술하는데 필요한 핵심적인 요소들을 중심으로 최소한의 디지털 정보의 진본성, 이용가능성을 보장하는 보존업무를 수행하고 상호운용성을 확보하는 방안으로 고려할 수 있을 것이다. PREMIS는 다른 메타데이터 표준과 함께 보존 환경에 따라 사용목적에 맞게 설계할 수 있는 확장성을 갖는다. 추후 실제 디지털 아카이브의 사례를 중심으로 아카이브의 보존 업무와 환경에 따라 세부적인 기술요소들을 추가한 보존 메타데이터 요소들을 개발하는 기초 연구로 활용될 수 있을 것이다.

참고문헌

- 국가기록원 (2010). 기록관리 메타데이터 국가표준화 연구용역 최종보고서. 대전: 국가기록원.
- 국가기록원 (2016). 전자기록 기술정보 기반 장기보존 의사결정 지원체계에 관한 연구. 대전: 국가기록원.
- 김명옥, 리상용 (2010). 전자기록물의 장기보존을 위한 기능요소 연구. 한국기록관리학회지, 10(2), 101-126.
<https://doi.org/10.14404/JKSARM.2010.10.2.101>
- 박옥남 (2012). PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구: 국립중앙도서관 디지털 자료를 중심으로. 한국문헌정보학회지, 46(2), 83-113. <https://doi.org/10.4275/KSLIS.2012.46.2.083>
- 이승민 (2015). 디지털 아카이빙을 위한 보존 메타데이터 패키지 구축. 한국정보관리학회, 32(3), 21-47.
<https://doi.org/10.3743/kosim.2015.32.3.021>
- 이지영 (2018). 호주(NAA)의 전자기록 보존과 포맷 정책 동향. 기록인, 45, 30-33.
- 이지영, 김희정 (2009). 디지털이미지 기록관리를 위한 메타데이터 요소 연구. 정보관리연구, 40(4), 49-71.
<https://doi.org/10.1633/jim.2009.40.4.049>
- 한국기록관리협회 (2009). 전자기록물의 이해. 서울: 조은글터.
- Caron, B., Houssaye, J., Ledoux, T., & Reeht, S. (2017). Life and Death of an Information Package: Implementing the Lifecycle in a Multi-Purpose Preservation System. iPRES 2017 14th International Conference on Digital Preservation, Sep 2017, Kyoto, Japan. Retrieved November 10, 2020, from <https://hal-bnf.archives-ouvertes.fr/hal-01617645/document>
- Dappert, A., Guenther, R. S., Peyrard (2016). An Introduction to Implementing Digital Preservation Metadata. In Dappert, A., Guenther, R.S., Peyrard, S. (Eds.), Digital Preservation Metadata for Practitioners – Implementing PREMIS (pp. 11-22). Cham, Switzerland: Springer.
- Derrot, S. & Oury, C. (2014). Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France. Retrieved November 10, 2020, from <http://library.ifla.org/830/1/087-derrot-en.pdf>
- Garvie, C. Reimagining the Archival Control Model of the National Archives of Australia for the digital age. Archiving Conference, Archiving 2020 Final Program and Proceedings, pp. 101-105(5)
<https://doi.org/10.2352/issn.2168-3204.2020.1.0.101>
- Library of Congress (2016). PREMIS: Preservation Metadata Maintenance Activity. Retrieved November 10, 2020, from <https://www.loc.gov/standards/premis/ma.html>
- National Archives of Australia (2004). The CRS Manual: Registration and description standards for the Commonwealth Record Series (CRS) System. Retrieved November 10, 2020, from <https://recordsearch.naa.gov.au/manual/index.htm>
- Peyrard, S., Dappert, A., & Guenther, R. S. (2016). How to Develop an Digital Preservation Metadata Profile: Risk and Requirements Analysis. In Dappert, A., Guenther, R.S., Peyrard, S. (Eds.), Digital Preservation Metadata for Practitioners – Implementing PREMIS (pp.11-22). Cham, Switzerland: Springer.
- PREMIS Editorial Committee (2008). Data Dictionary for Preservation Metadata version 2.0. Retrieved November 10, 2020, <https://www.loc.gov/standards/premis/v2/premis-2-0.pdf>
- PREMIS Editorial Committee (2014). Data Dictionary for Preservation Metadata version 3.0, revised November 2014. Retrieved November 10, 2020, from <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>
- Royal Danish Library (2019). Royal Danish Library's Digital Preservation Policy. Retrieved November 10, 2020, from <http://en.statsbiblioteket.dk/national-library-division/digital-preservation>.
- Saisto, A. (2019). D-ark: A Shared Digital Performance Art Archive with a Modular Metadata Schema. Heritage, 2(1), 976-987. DOI: 10.3390/heritage2010064
- Zierau, E. (2019). Preservation of Metadata: A case study of a strategy to ensure technology agnostic metadata preservation. DOI:10.17605/OSF.IO/NGY4U

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Kim, Myeong-Ok & Lee, Sang-Yong (2010). A Study on the Functional Elements for Long-term Preservation of Electronic Records. *Journal of Korean Society of Archives and Records Management*, 10(2), 101-126.
- Korean Association of Archival Management (2009). *Understanding of Electronic Record*. Seoul:JoeunGulteo.
- Lee, Jiyoung & Kim, Hee-Jung (2009). A Study on Metadata Elements for Digital Image Records Management, 40(4), 49-71.
- Lee, Jiyoung (2018). Electronic Records Preservation and Strategies of National Archives of Australia. *National Archives of Korea*, 45, 30-33.
- Lee, Seungmin (2015). Construction of Preservation Metadata Package for Digital Archiving. *Journal of the Korean Society for Information Management*, 32(3), 21-47.
- National Archives of Korea (2010). *Records and archives management metadata research project final report of national standardization*. Daejeon: National Archives of Korea.
- National Archives of Korea (2016). *A study on long-term Preservation Decision Making Support System based on Electronic Records Technology Information*. Daejeon: National Archives of Korea.
- Park, Ok Nam (2012). A Study on developing preservation metadata based on PREMIS focusing on digital data in National Library of Korea. *Journal of the Korean Society for Library and Information Science*, 46(2), 83-113.

