

연구정보를 위한 보존 메타데이터 요소 개발에 관한 연구: 경제·인문사회연구회 연구관리시스템을 중심으로

A Study on Preservation Metadata Elements for Research Information

김판준(Pan-Jun Kim)*

초 록

가치 있는 디지털 정보자원으로서 연구정보를 위한 보존 메타데이터 요소를 개발하였다. 특히 국가정책 지식 생산의 주역이라고 할 수 있는 경제·인문사회 분야 정부출연연구기관의 연구정보를 장기적으로 보존하여 활용할 수 있는 기반으로 보존 메타데이터 요소를 개발하였다. 다양한 부서와 기관에서 분산 관리되고 있는 연구정보의 상호운용성을 확보하기 위하여 OAI 참조모형을 기반으로 유럽표준인 CERIF와 PREMIS 데이터 사전의 요소들을 비교 분석한 다음, 양자의 특성을 반영하여 상호보완적인 보존 메타데이터 요소를 개발하였다. 그 결과로서 개념적 차원이 아닌 실제 구현이 가능하고 시스템 간의 호환성이 전제된 연구정보 보존 메타데이터 요소들과 적용사례를 제시하였다.

ABSTRACT

This study aimed at developing preservation metadata elements and its applications for research information which is considered as a valuable digital resource these days. Specifically, the developed preservation metadata intends to provide a basis for the research information of the government-funded research institutes in economic and social science fields which are major knowledge producers of national policy. To ensure the interoperability of the research information across various departments and organizations, this study compared the elements from the CERIF(European Standard) and those from the PREMIS Data Dictionary which is based on OAI reference model (ISO 14721). Based on this comparative analysis, this study developed complementary preservation metadata elements based on the two standards' characteristics. Consequently, this study suggested a new preservation metadata elements and its applications that are compatible between the two systems and can be implemented in practice.

키워드: 연구정보, 보존 메타데이터, 보존 메타데이터 요소, OAI 참조모형, PREMIS 데이터 사전
CERIF(common european research information format), research information,
preservation metadata, OAI reference model(ISO 14721), PREMIS data dictionary

* 경제·인문사회연구회 전문위원(pjkim@nrsc.re.kr)

■ 논문접수일자: 2010년 11월 15일 ■ 최초심사일자: 2010년 11월 22일 ■ 게재확정일자: 2010년 11월 30일
■ 정보관리학회지, 27(4): 169-191, 2010. [DOI:10.3743/KOSIM.2010.27.4.169]

1. 서론

정보기술의 발전이 급속히 진행될수록 네트워크상의 정보가 기하급수적으로 증가하는 반면에, 흔적도 없이 사라지거나 훼손되어지는 다양한 영역의 정보자원에 대한 지속적인 관리와 보존의 필요성이 커지고 있다. 이러한 맥락에서 가치 있는 디지털 객체의 장기 보존과 이용을 위한 디지털 보존의 중요성이 자연스럽게 부각되고 있다. 연구정보는 연구활동의 전 생애주기 동안 생산 및 관리되는 모든 정보자원이라고 할 수 있으며, 이러한 연구정보는 연구활동에 대한 신뢰성 있는 증거로서 체계적으로 수집, 관리, 활용되어야 한다. 그러나 현재 연구정보는 다양한 부서와 기관에서 장기적인 보존 및 활용에 대한 고려가 거의 없는 환경에서 분산 관리되고 있다.

국가정책연구는 정부 정책의 기획과 실행, 그리고 평가를 위한 근거와 방향성을 제시하기 위한 연구이다. 여기서 정책은 “정부가 하는 일이나 하지 않기로 결정한 모든 것”(Dye 1984)이라 할 수 있으며, “바람직한 사회상태를 이룩하려는 정책목표와 이를 달성하기 위해 필요한 정책수단에 대하여 권위 있는 정부기관이 공식적으로 결정한 기본방침”이다(정정길 1996). 따라서 정책의 연구과정 및 결과에 대한 정보는 정책결정에 쓸모가 있도록 구성되어 모든 목표 지향적인 조직의 생존을 위하여 필수불가결한 것이며(Rappaport 1982), “정책결정자가 필요로 하는 정보뿐만 아니라 정책과정의 전 과정을 통하여 항상 요구되는 정보”이다(최봉기 1981). 이처럼 국가정책연구정보는 정책의 입안, 심의, 집행, 평가의 전 과정에서 필수적으로 요구되는

정보로서 상당한 가치를 갖는다고 할 수 있다. 특히 국가정책연구의 과정 및 결과에서 생산·관리되는 모든 정보는 가치 있는 디지털 정보자원으로서 장기적인 보존과 활용이 반드시 고려되어야 하며, 이를 위한 초석이 되는 것으로 보존 메타데이터 요소의 개발은 기관 간, 시스템 간의 호환성을 전제로 하여야 한다. 본 연구는 국가정책연구 관련 연구정보의 장기적인 보존과 접근을 위한 기반으로, 상호운용성에 바탕을 둔 보존 메타데이터 요소를 개발 및 적용해 보고자 한다. 특히 국가정책연구의 한 축을 담당하고 있는 경제·인문사회연구회의 연구관리 차원에서, 정부출연연구기관들의 연구활동을 통하여 생산 및 관리되는 연구정보를 위한 핵심적인 보존 메타데이터 요소를 개발 및 적용할 것이다.

2. 연구정보와 디지털 보존

2.1 연구정보의 개념과 범위

연구정보는 연구활동의 전 생애주기 동안 생산 및 관리되는 정보라고 할 수 있다. 즉 연구정보는 연구의 기획에서 시작하여 공모, 심의, 선정, 수행, 결과, 평가 등 연구과정 전반을 통해 창출되어 관리되는 모든 정보인 것이다. 이러한 연구정보는 연구사업의 수행 주체에 따라 정부와 민간에서 수행되는 연구정보로 구분할 수 있으나, 우리나라의 경우 대부분의 연구사업이 정부자금에 의존하고 있어 정부기관에서 관리되고 있는 경우가 많다. 따라서 우리나라에서는 이들 정부기관에서 관리되는 연구정보

의 중요성이 그 규모나 내용 측면에서 상당히 크다고 할 수 있다. 본 연구에서는 우리나라에서 경제·인문·사회 분야 국가정책연구정보의 생산 및 관리를 위한 핵심적인 주체인 경제·인문사회연구회의 연구관리 과정에서 생산 및 활용되는 연구정보에 초점을 맞추어 논의를 진행하고자 한다.

국무총리 산하의 경제·인문사회연구회는 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 근거하여 '정부출연연구기관의 설립·지원·육성'을 목적으로 설립된 '연구기관을 지도·관리'하는 기관이다(한국. 법률 제8852호, 제1조; 제19조). 여기서 연구기관은 경제·인문사회 분야 23개 정부출연연구기관¹⁾을 말하며²⁾, 경제·인문사회연구회(이후 연구회로 표기함)는 동법 제21조에 따라 연구기획, 연구기관의 연구실적 및 경영내용에 대한 평가, 연구기관간의 협동연구를 위한 지원 등의 사업을 수행한다. 연구회 소관 23개 정부출연연구기관들은 국가정책연구와 관련하여 다음과 같이 크게 세 가지 유형의 연구사업을 수행한다.

첫째, 기본연구사업: 연구기관의 기관고유사업으로서 정부출연금으로 추진되는 연구사업.

둘째, 수탁연구사업: 연구기관의 기관고유사업 이외에 정부 및 민간으로부터 제공되는 사업비로 추진되는 연구사업.

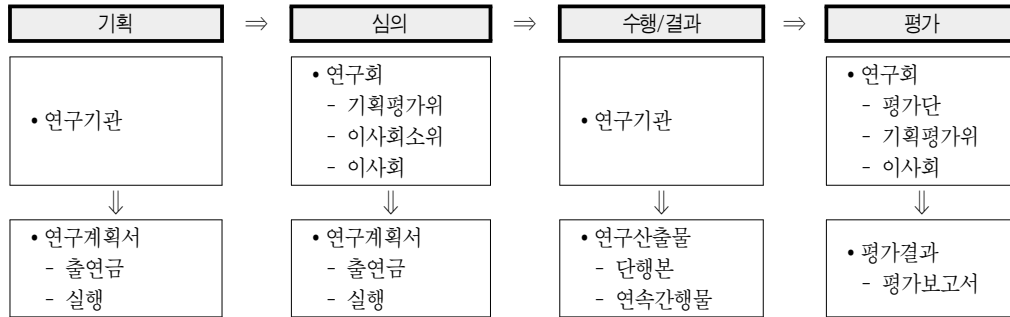
셋째, 협동연구사업: 연구회 또는 연구기관

이 주관하고 2개 이상의 국내의 연구기관 및 전문가가 공동 참여하여 정부출연금으로 추진되는 연구사업.

본 연구는 이러한 세 가지 유형의 연구사업 중에서 '기본연구사업'을 중심으로 연구정보를 다룬다. 그 이유는 첫째, 기본연구사업은 정부출연연구기관이 수행하고 있는 국가정책연구 중에서 가장 핵심적인 연구사업이다. 둘째, 수탁연구사업은 연구회에서 직접적으로 주관하지 않으므로 연구정보의 생산 및 관리과정에 관여하기 어렵다. 셋째, 협동연구사업의 진행과정이 기본연구사업과는 차이가 있어 연구정보의 생산 및 관리과정에서 다른 접근이 요구된다. 특히 연구회가 주관하는 협동연구사업의 경우 연구과제의 외부공모, 선정 등 기본연구사업의 수행과정에는 없는 단계가 존재한다.

경제·인문사회 분야 23개 정부출연연구기관이 수행하는 기본연구사업의 연구과정은 <그림 1>과 같다. 기본연구사업의 연구과정은 연구의 기획, 심의, 수행/결과, 평가로 이루어진다. 여기서 기본연구사업의 기획 및 수행/결과 단계는 연구기관에서 수행하며, 심의와 평가 단계는 연구회에서 수행한다. 연구사업에 대한 심의는 크게 출연금사업 심의와 실행사업 심의로 구분할 수 있으며, 각각 기획평가위원회, 이사회소위원회, 이사회의 심의 단계를 거친다. 또한 연구결과로서 연구산출물에 대한 평가는

1) “정부출연연구기관”이라 함은 정부가 출연하고 연구를 주된 목적으로 하는 기관을 말한다(동법 제2조).
 2) 연구회 소관 23개 정부출연연구기관: 과학기술정책연구원, 국토연구원, 대외경제정책연구원, 산업연구원, 에너지경제연구원, 정보통신정책연구원, 통일연구원, 한국개발연구원, 한국교육개발원, 한국교육과정평가원, 한국교통연구원, 한국노동연구원, 한국농촌경제연구원, 한국법제연구원, 한국보건사회연구원, 한국여성정책연구원, 한국조세연구원, 한국직업능력개발원, 한국청소년정책연구원, 한국해양수산개발원, 한국행정연구원, 한국형사정책연구원, 한국환경정책·평가연구원.



〈그림 1〉 경제·인문사회 분야 정책연구를 위한 기본연구사업의 연구과정

평가단에 의해 이루어지며, 평가결과에 대한 검토를 기획평가위원회와 이사회에서 수행한다. 이러한 연구과정의 각 단계에서 생산 및 관리되는 주요 연구정보로는 기획 및 심의 단계의 연구계획서(출연금, 실행)와 수행/결과 단계의 연구산출물(단행본, 연속간행물), 그리고 평가 단계의 평가결과(평가보고서)가 있다.

본 연구는 기본연구사업의 각 단계에서 생산 및 관리되는 연구정보를 중심으로 보존 메타데이터 요소를 개발 및 적용한다. 특히 이중에서도 심의 단계의 산출물로서 연구계획서(출연금, 실행)와 결과 단계의 산출물로서 단행본(단행본, 연속간행물)을 중심으로 한다. 그 이유는 첫째, 수행 단계의 중간산출물은 연구기관 자체적으로 관리되고 있으며, 둘째, 연구회 내부적으로 평가 단계에 대한 개선작업이 진행되고 있으므로 향후 그 결과를 반영하여 보존 메타데이터 요소를 개발할 필요가 있기 때문이다.

2.2 연구정보의 디지털 보존

우리나라에서 국가정책연구와 관련하여 연구사업을 수행하는 주체는 대학과 정부출연 연구기관, 그리고 민간기업의 연구기관으로 구분할

수 있다. 국가정책연구의 상당 부분을 주도적으로 수행하고 있는 정부출연연구기관의 경우, 연구관리의 주체는 해당 기관 내 연구관리 부서와 외부의 연구관리기관이 되며, 이들 부서 및 기관은 연구사업의 수행 과정과 성과, 그리고 예산에 대한 관리를 효율적으로 수행하기 위하여 정확하고 신뢰성 있는 연구정보를 필요로 한다. 여기서 연구정보는 연구의 수행과정과 연구사업비 관리로부터 발생하는 연구자(researchers), 과제(project), 결과(output) 및 예산(funding)에 관한 관리정보를 말한다(JISC 2010). 그러나 이러한 정보는 현재 다양한 기관 및 부서에 분산되어 있고 서로 호환성이 없는 형식으로 저장 및 활용되고 있다. 특히 본 연구의 적용 대상인 연구회 소관 23개 정부출연연구기관의 연구관리시스템에 대한 사전조사에서도, 각 기관의 연구정보가 여러 부서에서 분산 관리되고 있으며 연구정보 메타데이터 간에 호환성이 거의 없는 것으로 나타났다. 따라서 이들 기관 간에 연구정보의 교환·공유는 물론, 중장기적으로 연구정보의 효율적인 보존 및 접근을 위한 표준화된 체계가 필요하다.

연구기관 내외부의 연구관리 주체들 간에 연구정보를 보다 효율적으로 교환함으로써, 보다

효율적인 데이터 공유가 가능하고, 중복된 노력이 감소하고 보다 정확한 정보의 생산 및 활용이 가능할 수 있다. 또한 이를 통해 연구기관 및 연구관리기관은 연구 현황에 대한 정확한 이해를 바탕으로 전략적 기획의 기반이 되는 신뢰성 있는 정보를 제공할 수 있음은 물론, 보다 효율적으로 연구 결과 또는 성과에 대한 활용 및 평가를 수행할 수 있다.

디지털 보존은 디지털 정보자원에 대한 장기적인 접근과 저장을 보장하기 위한 과정과 활동을 말한다. 이러한 디지털 보존은 처음부터 디지털 형식으로 생산된 자료(born-digital)와 아날로그 형식의 자료를 디지털화한 자료(digitalized)에 대하여 모두 적용되며, 디지털 정보자원의 형태와 구조는 물론 내용을 함께 온전히 보존하는 것으로 디지털 아카이브 관리의 일부이다(Hodge 2000). 또한 디지털 보존이란 디지털 형태로 수록된 데이터의 유용성, 내구성, 및 지적 내용의 무결성(integrity)을 보장할 수 있도록 보존하는 과정(Graham 1994)을 의미하는 것으로, 원 정보의 손실 없이 디지털 정보자원을 장기간에 걸쳐 저장하고, 관리하고 접근하기 위한 활동이다. 구체적으로 OCLC/RLG Working Group(2002)은 디지털 보존을 0과 1로 이루어진 원 데이터(bit stream)를 계속적으로 유지하여 원형 그대로의 정보가 저장 매체로부터 읽혀질 수 있도록 하는 원형유지성(viability), 내용에 접근하여 이용자나 관리자가 처리 및 표기할 수 있도록 하는 재생가능성(renderability), 이용자가 디지털정보의 내용을 파악할 수 있게 해주는 이해가능성(understandability)의 세 가지 기능을 수행하는 활동으로 정의하였다. 요약하면, 디지털 보존은 디지털

원 데이터의 무손실 유지와 허가되지 않은 방법에 의한 변경 방지, 디지털 정보의 재생, 그리고 내용에 대한 이해라는 세 가지 기능을 갖는다.

디지털 보존은 시간의 흐름과 주변 환경의 변화에 상관없이 디지털 정보자원을 지속적으로 보존하고 활용할 수 있도록 하는 것으로, 이를 효율적으로 수행하기 위해서는 구조적, 체계적인 보존 메타데이터가 필요하다(OCLC/RLG Working Group 2002, 1). 따라서 가치 있는 디지털 정보자원으로서 국가정책연구 과정에서 생산 및 관리되는 연구정보의 장기적인 보존 및 접근을 보장하기 위한 보존 메타데이터 요소가 필요할 것이다. 보존 메타데이터의 개발에 있어 우선적으로 고려해야 할 사항은 상호운용성의 확보이다. 이를 위해서는 여러 디지털 아카이브에 적용 가능하고 디지털 아카이브 간에 서로 교환되는 표준의 개발이 필요하며, 이러한 표준의 개발은 기존에 연구되었던 현행 표준들을 토대로 개발이 가능하다(Day 2003, 3-4). 또한 연구정보의 장기 보존 및 접근과 관련한 전 과정에서 진본성(authenticity) 및 무결성이 보장되어야 하고, 중장기적으로 환경의 변화에 효과적으로 대응할 수 있도록 요소의 추가 또는 보완이 가능하여야 한다.

3. 연구정보와 보존 메타데이터 표준

3.1 연구정보 표준(CERIF)

CERIF(Common European Research Information Format)는 유럽국가연합(EU)의 회원

국들 간에 연구자, 연구기관, 연구비지원기관(funders), 연구과제, 산출물과 같은 주요 개체(entity)와 이들 간의 관계를 기술하는데 사용될 수 있는 표준으로 제안되었다(euroCRIS 2010). 1991년 유럽 위원회(European Commission)의 후원 아래 처음 개발된 CERIF는 2000년 이후에는 euroCRIS(euro Current Research Information Systems)의 주도 하에 지속적으로 개발이 진행되고 있다. CERIF 초안(CERIF91)은 유럽연합 국가들 간에 연구과제 정보에 대한 데이터 교환을 목적으로 데이터베이스 간 네트워킹을 위한 기본 포맷(format)으로 사용되었다. 그러나 CERIF91은 너무 엄격한 포맷과 반복되는 정보의 처리가 불가능하고 복수언어와 문자를 지원하지 않으며, 풍부한 표현을 제공하지 못한다는 한계를 보여주었다. 이후 이용자 요구사항을 반영하여 개정된 CERIF97은 2000년 euroCRIS에 이양되기 전까지 유럽 위원회에 의해 관리되었다.

2000년 euroCRIS가 주도한 CERIF2000은 연구자(person), 연구기관(OrganisationUnit), 산출물(Result) 등 연구정보의 관리를 위한 주요 개체들을 설정하고, 의미전달을 위한 유형(type)과 관계(relationship)을 도입한 데이터베이스 모형으로 개발되었다. 이후의 CERIF2006은 일부 개체의 추가 및 확장과 함께 의미전달을 위한 유형과 관계를 의미계층(semantic layer)으로 재조직하였다. 또한 XML과의 연계를 위한 기반으로 CERIF XML 교환 형식(CERIF XML interchange format)을 제안하였다.

가장 최근의 CERIF2008(euroCRIS 2010, 이후 CERIF2008로 표기)은 발간물 개체의 범위를 더욱 확장 및 개선하였고, 발간물 유형

(types)과 역할(roles)을 갖는 새로운 CERIF 의미체계(semantics)를 도입하였다. CERIF2008의 개념적 모형은 다양한 개체 유형과 속성들로 구조화되어 있다. CERIF2008의 주요 개체 유형으로는 핵심 개체(Core Entities), 산출물 개체(Result Entities), 2차 개체(2nd Entities), 링크 개체(Link Entities)의 4가지가 있으며, 이들은 다시 여러 개의 하위 개체로 세분된다.

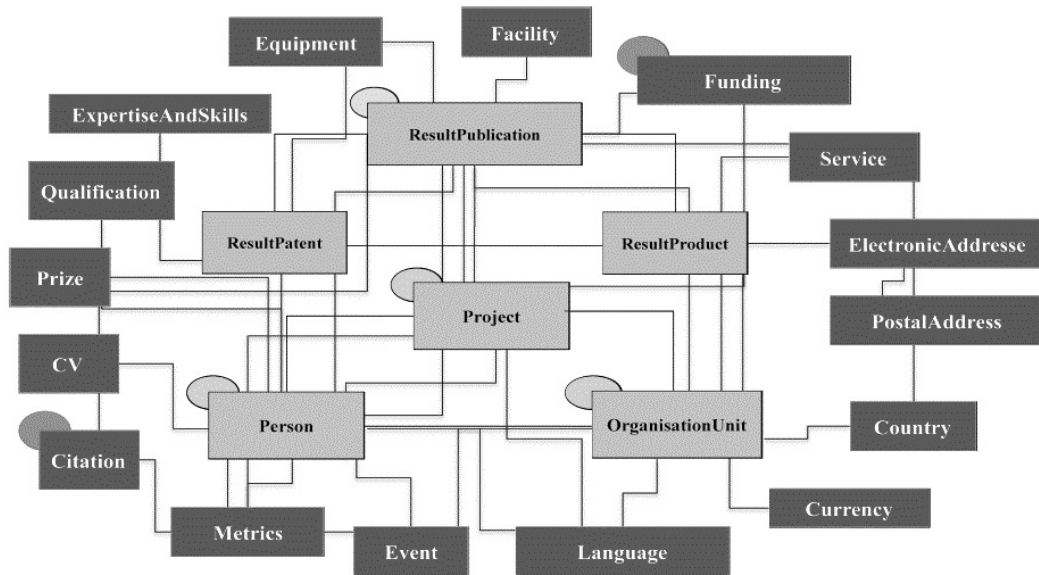
첫째, 핵심 개체: 과제(Project), 기관(OrganisationUnit), 인명(Person).

둘째, 산출물 개체: 발간물(ResultPublication), 특허(ResultPatent), 시제품(ResultProduct).

셋째, 2차 개체: 이벤트(Event), 언어(Language), 예산(Funding), 장비(Equipment), 서비스(Service), 화폐(Currency), 인용(Citation), 우편주소(PostalAddress) 등.

넷째, 링크 개체: 2개의 개체를 연결하여 양자 간의 관계를 설정.

〈그림 2〉는 주요 개체와 관계로 구성된 CERIF의 개념적 구조를 나타낸 것으로, 중앙의 핵심 개체와 산출물 개체를 바깥쪽의 2차 개체들이 감싸고 있으며, 이들 개체 간을 연결하는 링크는 실선으로 표시하였다. CERIF2008은 이러한 개념적 모형 이외에 실제적 구현을 위하여 논리적, 물리적 수준으로 상세한 구조를 제시하고 있다. 즉, CERIF2008은 연구 개체와 이들의 관계에 대한 개념적 명세(conceptual specifications)와 논리적 수준의 기술 모형(descriptive model), 그리고 물리적 수준의 형식화(database scripts)를 모두 제공하는 구현 중심의 실제적 표준이라 할 수 있다.



〈그림 2〉 CERIF의 주요 개체와 관계(euroCRIS 2010, 6)

CERIF는 연구정보의 교환 및 공유를 위한 국제적인 표준으로서, 특히 연구정보의 생산 및 활용 측면에서 세 가지 특성을 갖는다. 첫째, 연구정보를 위한 개념적, 논리적, 물리적 데이터 모형을 제공하고, 둘째, 연구 개체들과 이들의 활동에 대한 메타데이터 표현을 제공하며, 셋째, 형식적·의미적 관계를 갖는 고수준의 유연성(high flexibility)을 제공한다. 이러한 특성들에 기초하여 CERIF2008은 연구정보의 품질 유지, 아카이빙, 접근, 상호교환을 가능하게 하고 다양한 유형의 이해관계자를 위한 표준화된 연구정보를 제공할 수 있다. 따라서 유럽표준으로서 CERIF2008에 기반한 보존 메타데이터의 개발은 연구정보의 주체로서 관련기관(연구기관, 연구관리기관, 정부부처, 대학 등)의 의사결정 및 업무수행을 효과적으로 지원하기 위한 기획 및 심의, 연구산출물, 연구동향 관련 정보를 표준화된 방식으로 제공할 수 있는 기

반을 구축할 수 있다.

3.2 보존 메타데이터 표준(OAIS 참조모형, PREMIS 데이터 사전)

디지털 정보의 장기 보존을 위해서는 생산 시점부터 관리의 전 과정에 대한 정보를 제공할 수 있는 보존 메타데이터가 필요하다. 대부분의 디지털 보존 전략은 기본적인 구성요소로서 메타데이터의 획득과 생산 그리고 유지에 의존한다(Day 2003, 2). 보존 메타데이터는 디지털 정보를 장기간 보존하기 위한 과정을 기술하는 구조적 정보로 디지털 정보의 보존과 접근을 지원하여야 하고, 이를 위해 다른 메타데이터에 비하여 보다 더 다양하고 세부적인 항목들을 포괄하고 있다. 변환과 허가를 포함하는 관리 메타데이터, 기술 메타데이터, 구조 메타데이터 등의 다양한 범주가 보존 메타데이

터에 포함된다(서은경 2005).

보존 메타데이터 개발을 위한 시작점으로 세계적으로 가장 폭넓게 적용되고 있는 것은 국제 표준으로 제정된 OAIS 참조모형(ISO 14721)이다. OAIS 참조모형의 핵심 개념은 디지털 아카이브의 기능에 따라 전환되는 세 가지 범주의 정보패키지라고 할 수 있는데(OCLC/RLG Working Group 2002, 9), 이 중에서 보존정보 패키지(AIP)는 보존 메타데이터 개발을 위한 상위 차원의 개념적 표준이 될 수 있다. OAIS 참조모형의 정보패키지를 구성하는 4가지 정보 유형 중에서 보존 메타데이터와 관련된 핵심 정보는 내용정보(CI: Content Information)와 보존기술정보(PDI: Preservation Description Information)이다. 내용정보(CI)는 보존 대상인 원 데이터(Data Object, 즉 Bit Stream)에 대한 정보와 이를 재생 및 이해하기 위한 표현 정보(RI: Representation)로 구성된다. 원 데이터 정보에는 원 데이터의 내용이나 특징을 기술하는 것으로 주석, 크기, 유형에 관한 정보가 포함될 수 있다. 또한 표현정보는 원 데이터를 재생하고 의미를 전달할 수 있도록 하는 것으로, 재생을 위한 기술적 환경(Hardware, Software, Format 등)과 의미를 전달하기 위한 코드화 체계(ASCII, UTF-8 등) 등이 포함될 수 있다.

보존기술정보(PDI)는 내용정보(CI)를 보존하기 위한 활동에 관한 정보로서 다음과 같은 4가지 정보 유형을 포함한다.

첫째, 출처정보(Provenance Information): 내용정보(CI)의 기원 및 보존 내력에 관한 정보, 즉 내용정보의 기원과 출처, 그리고 생산 이후 발생한 변화, 보관권(custody)의 변화에 관한 정보이다. 예를 들면, 데이터의 생산 및 관리

자, 내용정보의 저장, 관리, 마이그레이션에 관한 정보, 디지털 원본에 대한 포인터, 수집 항목의 이전 버전에 대한 포인터 등이 포함된다.

둘째, 참조정보(Reference Information): 내용정보(CI)에 부여된 식별자정보, 즉 내용정보(CI)에 부여된 식별자 시스템을 확인하고 설명하는 정보이다. 예를 들면, 고유식별번호, URI, ISBN, ISSN 등이 이에 해당한다.

셋째, 인증정보(Fixity Information): 내용정보(CI)에 대한 인증정보, 즉 내용정보(CI)를 허가되지 않은 방식으로 변경할 수 없도록 보호하는 정보이다. 예를 들면, 체크섬(checksum)이나 디지털 서명, CRC(Cyclical Redundancy Check) 코드 등이 이에 해당한다.

넷째, 맥락정보(Context Information): 내용정보(CI)의 관계에 대한 정보, 즉 내용정보(CI)의 생산 이유와 다른 디지털 객체와의 관계에 대한 정보이다. 예를 들면 사업비 지원 내역, 도움말 파일, 이용자 지침서, 보정기록 자동 저장, 관련 데이터 집합(data sets), 출판 시점의 관련 문헌들에 대한 포인터 등이 이에 해당한다.

2005년 발표된 PREMIS 데이터 사전 1판은 구현가능하고 폭넓게 적용 가능한 핵심 메타데이터 요소들을 정의하고, 디지털 아카이빙 시스템의 보존 메타데이터를 코드화(encoding), 저장, 관리, 교환하기 위한 전략을 제시하였다(PREMIS Working Group 2005). 이는 디지털 아카이빙 시스템에서 보존 메타데이터를 구현하기 위한 망라적이고 실용적인 자원으로서, 보존 메타데이터의 관리와 사용을 위한 지침과 함께 핵심적인 보존 메타데이터들을 정의하였다.

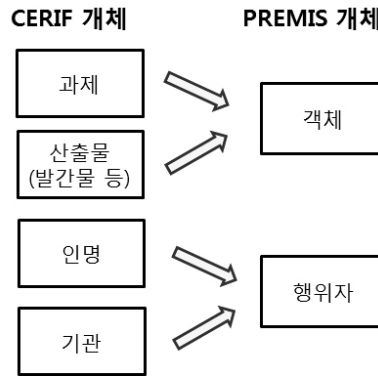
최근 발표된 PREMIS 데이터 사전 2판 (PREMIS Editorial Committee 2008)은 OAIS 개념구조를 그대로 따르기 보다는 실제 시스템의 구현에 초점을 맞추어 특정 기관의 환경에 적합한 것으로 개별 기관에서 실제적으로 적용할 수 있는 보존 메타데이터 요소들을 개발하기 위한 것이다. 보존 메타데이터 데이터 사전 2판 (이후 PREMIS 데이터 사전으로 표기함)은 디지털 보존을 위한 주요 개체 5개로 구성된 데이터 모델을 제시하고 있다(PREMIS Working Group 2005; PREMIS Editorial Committee 2008). 첫째, 지적 개체(Intellectual Entity)는 하나의 단위로 기술된 콘텐츠 집합, 예를 들면 단행본 한 권이 이에 해당한다. 둘째, 객체 개체(Object Entity)는 디지털 형식의 정보단위로서 예를 들면 PDF 파일이 이에 해당한다. 셋째, 이벤트 개체(Event Entity)는 보존 행위, 즉 리포지토리의 PDF 파일 입수(ingest)가 이에 해당한다. 넷째, 행위자 개체(Agent Entity)는 특정 보존 행위와 관련된 개인, 기관, 소프트웨어 프로그램, 즉 리포지토리에 저장된 PDF 파일의 생산자를 말한다. 다섯째, 권한 개체(Rights Entity)는 객체에 포함된 하나 이상의 이용허락, 즉 보존 목적으로 PDF 파일의 복사본을 생산할 수 있는 권한을 말한다. 이와 같이 PREMIS 데이터 사전은 핵심 보존 메타데이터의 구현과 이용을 위한 지침과 함께 개체들에 관련된 메타데이터에 대한 상세한 기술(description)을 제공하고 있다. 그러나 지적 개체에 관한 메타데이터는 제외하고 있으며, 이에 대해서는 기존에 개발되어 사용되고 있는 기술 메타데이터(Descriptive Metadata)를 이용할 것을 권고하고 있다.

보존 메타데이터의 상대적으로 짧은 역사를 되짚어보면 OAIS의 개념적 프레임워크에서 시작하여 실제적인 구현을 위한 PREMIS 데이터 사전에 이르기까지, 개념에서 구현으로의 진전을 이룬 것이다. 이처럼 보존 메타데이터는 이론(개념적/추상적 모형)에서 실제(논리적/구조적 데이터 모형)로 성장하는 단계에 있음을 알 수 있다. 보존 메타데이터와 관련하여 이 둘 두 가지 표준이 추구하는 최종 목표는 동일한 것이며, OAIS참조모형이 이론적 측면에서의 틀을 제공한 것이라면, PREMIS 데이터 사전은 실제 구현 측면에서의 지침(Guideline)으로 사용될 수 있다.

4. 연구정보 보존 메타데이터 기본 요소 개발

4.1 메타데이터 요소의 비교 분석

CERIF의 주요 개체를 중심으로 PREMIS 데이터 사전의 주요 개체들의 특성을 살펴 본 결과, 양자 간의 유사성에 따른 대응 관계는 <그림 3>과 같다. 첫째, CERIF의 핵심 개체 중에서 '과제'는 PREMIS 데이터 사전의 '객체'에 대응할 수 있다. 둘째, CERIF의 핵심 개체 중에서 '인명'과 '기관'은 PREMIS 데이터 사전의 '행위자'에 대응할 수 있다. 셋째, CERIF의 산출물 개체에는 '발간물', '특허', '시제품'이 있다. 이들은 모두 PREMIS 데이터 사전의 '객체'에 대응할 수 있다. 본 논문에서는 이 중에서 특히 CERIF의 '과제'와 '산출물' 개체와 PREMIS 데이터 사전의 '객체' 개체에 대하여 양자의 메



〈그림 3〉 CERIF 개체와 PREMIS 데이터 사전 개체 간의 대응관계

타데이터 요소들을 비교 분석하였다. 또한 이들 두 가지 표준의 요소들이 개념적 틀로서 OAIS 참조모형의 보존정보패키지(OAIS AIP)를 구성하는 내용정보와 보존기술정보(참조, 출처, 맥락, 인증) 중에서 어디에 해당하는지를 제시하였다.

CERIF의 핵심 개체인 '과제(Project)'를 중심으로 PREMIS 데이터 사전 '객체(Object)'의 요소들을 비교한 결과는 〈표 1〉과 같다. 여기서 메타데이터 요소는 CERIF 개체들의 속성들과 PREMIS 데이터 사전 개체들의 의미단위(Semantic Unit)를 말한다. 실제로 이들 속성과 의미단위 사이에는 일부 개념상의 차이가 있을 수 있으나, 본 연구에서는 넓은 의미에서 이들을 디지털 정보자원으로서 연구정보의 장기보존과 접근을 위한 메타데이터 요소로 간주하였다.

CERIF 산출물 개체 중 '발간물(Result Publication)'을 중심으로 PREMIS 데이터 사전 '객체(Object)'의 요소들을 비교한 결과는 〈표 2〉와 같다.

결과적으로, 연구정보 표준인 CERIF의 개체별 메타데이터 요소를 기준으로 PREMIS의 보존 메타데이터 요소를 비교 분석한 결과를

요약하면 다음과 같다.

첫째, CERIF의 요소들은 연구정보의 교환 및 기술(description)이 주된 목적이므로 필수적인 보존 메타데이터 요소가 거의 존재하지 않는다. 특히 연구정보의 보존을 위한 내용정보는 물론 보존기술정보에 해당하는 요소가 전혀 제공되지 않고 있다.

둘째, 이외는 대조적으로 PREMIS 데이터 사건의 요소들은 디지털 객체의 보존이 주된 목적이므로, 연구정보의 기술(description)을 위한 요소가 상대적으로 부족하다. 이는 PREMIS 데이터 사전에서 제시한 지적 개체 관련 요소를 다른 기술정보 중심의 표준을 사용하라는 권고와 맥락을 같이 한다.

셋째, CERIF와 PREMIS 데이터 사전에서 제시한 관계(relationships)는 적용 범위와 목적에서 서로 차이가 있다. 즉, CERIF는 모형을 구성하는 기술(description) 측면에서 모든 개체들 간의 링크를 중심으로 관계를 설정하고 있으나, PREMIS 데이터 사전은 보존 측면에서 여러 개체 중에서 객체 개체들 간의 관계만을 제시하였다.

〈표 1〉 CERIF 개체 '과제(Project)'와 PREMIS 개체 '객체(Object)' 간의 비교

CERIF 속성/한글명	PREMIS 의미단위	OAIS AIP	비교
cfProjectId/과제 ID	objectIdentifier - objectIdentifierType - objectIdentifierValue	참조정보	과제식별자/객체식별자
cfURI/과제 URI	objectIdentifier - objectIdentifierType - objectIdentifierValue	출처정보	과제식별자/객체식별자
cfAcro/과제약어명	-	기술정보	과제약어명/해당없음
cfStartDate/과제시작일	-	기술정보	과제시작일/해당없음
cfEndDate/과제종료일	-	기술정보	과제종료일/해당없음
cfTitle/과제명	-	기술정보	과제명(복수언어)/해당없음
cfKeyw/과제키워드	-	기술정보	과제키워드(복수언어)/해당없음
cfAbstr/과제초록	-	기술정보	과제초록(복수언어)/해당없음
cfClassId/과제분류	objectCategory	기술정보	과제분류/객체 유형분류체계
cfProjID/과제ID(링크)	relationship - relationshipType - relationshipSubType - relatedObjectIdentification - relatedEventIdentification	맥락정보	개체링크(과제와 과제)/ 객체관계(객체와 객체)
cfResPubId/발간물ID(링크)	relationship - relationshipType - relationshipSubType - relatedObjectIdentification - relatedEventIdentification	맥락정보	개체링크(과제와 발간물)/ 객체관계(객체와 객체)
cfOrgUnitId/기관ID(링크)	-	맥락정보	개체링크(과제와 기관)/해당없음
cfPersID/인명ID(링크)	-	맥락정보	개체 링크(과제와 인명)/ 해당없음
-	signatureInformation - signature	인증정보	해당없음/객체권한
-	objectCharacteristics - fixity - size - format - ...	내용정보	해당없음/객체특성 (인증, 크기, 포맷 등)
-	environment - software - hardware	내용정보	해당없음/객체환경
-	objectCategory	내용정보	해당없음/객체범주 (bit stream, file, representation)

〈표 2〉 CERIF 개체 ‘발간물(ResultPublication)’과 PREMIS 개체 ‘객체(Object)’ 간의 비교

CERIF 요소/한글명	PREMIS 의미단위	OAIS AIP	비교
cfResPubID/발간물 ID	objectIdentifier	참조정보	발간물식별자/객체식별자
cfISBN/ISBN	- objectIdentifierType	참조정보	발간물ISBN/객체식별자
cfISBN/ISSN	- objectIdentifierValue	참조정보	발간물ISSN/객체식별자
cfURI/URI		출처정보	발간물URI/객체URI
cfResPubDate/발간일	-	기술정보	발간일/해당없음
cfStartPage/시작쪽	-	기술정보	발간물시작쪽/해당없음
cfEndPage/끝쪽	-	기술정보	발간물끝쪽/해당없음
cfVol/권	-	기술정보	발간물권/해당없음
cfNum/호	-	기술정보	발간물호/해당없음
cfTitle/서명	-	기술정보	발간물서명/해당없음
cfcfSubtitle/부서명	-	기술정보	발간물부서명/해당없음
cfKeyw/키워드	-	기술정보	발간물키워드/해당없음
cfAbstr/초록	-	기술정보	발간물초록/해당없음
cfClassId/유형분류	objectCategory	기술정보	발간물유형분류/객체유형분류
cfResPubId/발간물ID(링크)	relationship - relationshipType - relationshipSubType - relatedObjectIdentification - relatedEventIdentification	맥락정보	개체링크(발간물과 발간물)/ 객체관계(객체와 객체)
cfResPubId/과제ID(링크)	relationship - relationshipType - relationshipSubType - relatedObjectIdentification - relatedEventIdentification	맥락정보	개체링크(발간물과 과제)/ 객체관계(객체와 객체)
cfOrgUnitID/기관ID(링크)	-	맥락정보	개체링크(발간물과 기관)/ 해당없음
cfPersID/인명ID(링크)	-	맥락정보	개체링크(발간물과 인명)/ 해당없음
-	signatureInformation - signature	인증정보	해당없음/객체권한
-	objectCharacteristics - fixity - size - format - ...	내용정보	해당없음/객체특성 (인증, 크기, 포맷 등)
-	environment - software - hardware	내용정보	해당없음/객체환경
-	objectCategory	내용정보	해당없음/객체범주 (bit stream, file, representation)

4.2 연구정보 보존 메타데이터 요소와 적용사례

CERIF와 PREMIS 데이터 사전의 요소를 비교 분석한 결과에 따라 주요 개체별로 연구정보를 위한 핵심적인 보존 메타데이터 요소들을 개발하고 실제적인 적용사례를 제시하였다. 본 연구에서 연구정보를 위한 보존 메타데이터 요소 개발의 기본원칙은 다음과 같다.

첫째, 연구정보의 적절한 기술(description)과 보존을 위한 2가지 유형의 표준으로서 CERIF의 기술정보 관련 요소와 PREMIS 데이터 사전의 보존정보 관련 요소를 상호보완적으로 적용한다.

둘째, CERIF와 PREMIS 데이터 사전의 기존 요소들을 가능한 한 있는 그대로 적용하여 연구정보의 교환 및 공유를 위한 기본적인 호환성을 확보할 수 있도록 한다.

셋째, 연구정보의 장기 보존 및 접근을 위한 진본성 및 무결성을 보장하면서, 이용자 및 관리자에게 최대한의 입력 및 관리상의 편의를 제공할 수 있도록 한다.

넷째, 연구정보 관리 및 이용의 편의와 품질 관리 측면에서 기본 요소들의 정보를 가능한 한 입력단계에서 자동 생성 및 획득할 수 있도록 한다.

다섯째, 증장기적 측면에서 연구정보의 내용에 대한 분석이 가능한 요소를 포함하고, 기존 요소들에 대한 확장성을 고려한다.

이러한 원칙에 따라 본 연구에서는 CERIF의 과제 개체와 산출물 개체를 중심으로 보존 메타데이터 요소를 개발 및 적용하였다. 그 이유는

경제·인문사회연구회에서 구축중인 연구관리시스템이 정부출연연구사업의 과제를 중심으로 하고 있으며, 이러한 과제의 수행 과정에서 다양한 산출물들이 발생하기 때문이다. 구체적으로 과제의 기획 및 심의 단계에서는 사업계획서(출연금, 실행), 과제의 종료 시점에는 단행본(연구보고서, 자료 등)과 연속간행물이 대표적인 산출물이 된다. 따라서 CERIF와 PREMIS 데이터 사전의 요소들을 상호보완적으로 적용하여, 정부출연연구사업과 관련된 핵심 개체 중에서 과제와 주요 산출물로서 사업계획서와 발간물(단행본/연속간행물) 개체를 위한 보존 메타데이터 요소와 구체적인 적용사례를 제시하였다. 이들 요소는 출처 측면에서 크게 세 가지 유형으로 구분할 수 있는데 기술(description) 관련 요소들은 CERIF에서, 보존 관련 요소들은 PREMIS 데이터 사전의 요소들을 주로 적용하였고, 이외에 일부 연구회 환경에서 필요한 요소(Local)를 추가하였다.

보존 메타데이터 요소와 적용사례를 제시한 이하의 표들에서 '필수/반복/자동' 항목은 해당요소의 필수(Mandatory)와 선택(Optional), 반복(Repeatable)과 반복불가(NonRepeatable), 그리고 자동(automatic)과 수동(manual)을 구분하기 위한 것이다. 특히 자동('a')으로 표시한 요소들은 연구회의 연구관리시스템에서 자동으로 생성 및 획득될 수 있다. 연구회에서 2010년 현재 구축하고 있는 연구관리시스템은 연구정보의 접수 및 입력을 위한 표준양식(한글 파일)을 사용하여 대부분의 메타데이터 요소들을 자동으로 데이터베이스에 저장한 다음, 입력 및 관리자가 저장된 내용을 확인 및 보정하는 기능을 제공한다. 즉 입력자가 보존 메타

데이터 요소들로 구성된 한글 형식의 파일을 작성하여 시스템에 업로드하는 동시에, 데이터베이스에 대부분의 보존 메타데이터 요소들의 값이 자동으로 입력되는 방식이다. 또한, 입력된 내용을 확인하고 보정하는 과정에서 특정 요소의 내용을 변경하고자 할 경우에는, 데이터베이스에 저장된 관련 정보(표준분류체계, 인명DB, 사업유형코드 등)를 검색한 결과에 대하여 자동 완성 및 채우기 기능을 활용할 수 있다. 이를 통해 연구정보 보존 메타데이터의 입력 및 관리상의 편의를 제공할 수 있음은 물론, 데이터 품질 관리 측면에서 신뢰성 및 무결성을 확보할 수 있다. 이외에도 '보정(Revision)', '마이그레이션(Migration)', '객체범주(ObjectCategory)', '기술환경(Environment)' 요소는 시스템이 자동으로 식별할 수 있거나 사전에 규정된(default)

정보이므로 자동 저장이 가능하다. 그러나 이탤릭체로 표시한 '링크(Proj_ResPubl)', '권한(Rights)', '인증(fixity)' 요소는 시스템에서 자동으로 획득할 수 없는 정보이므로 수작업이 필요하다.

연구정보를 위한 핵심 개체 중에서 '과제'의 보존 메타데이터 기본 요소와 적용사례는 <표 3>과 같다. 여기서 각 요소들은 출처 측면에서 CERIF, PREMIS, Local(자체개발)로 구분하여 제시하였다.

연구정보를 위한 산출물 개체 중에서 '사업계획서'의 보존 메타데이터 기본 요소와 적용사례는 <표 4>와 같다.

연구정보를 위한 산출물 개체 중에서 '발간물(단행본, 연속간행물)'의 보존 메타데이터 기본 요소와 적용사례는 <표 5>와 같다.

<표 3> 핵심 개체 '과제'의 보존 메타데이터 기본 요소와 적용사례

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 적용사례
ProjId(Local) - ProjIDType - ProjIDValue	M/NR/a	참조정보	과제식별자 ¹⁾ - 유형: 'UniqueID' - 값: 'NRCS-KDI-U-2009--003'
StartDate(CERIF) - StartDateType - StartDateValue	M/NR/a	기술정보 (Descriptive Information)	시작일 - 유형: 'ProjStartDate' - 값: '20090101'
EndDate(CERIF) - EndDateType - EndDateValue	M/NR/a		종료일 - 유형: 'ProjEndDate' - 값: '20091231'
ProjCont(Local)	M/NR/a		계속과제여부 ²⁾ - 값: 'no'
TotYear(Local)	M/NR/a		전체과제년 - 값: '1'
ProjYear(Local)	M/NR/a		과제년차 - 값: '1/1'
Title(CERIF) - TitleType - TitleValue	M/R/a		과제명 ³⁾ - 유형: 'ProjTitle_ko' - 값: '외국인 투자자 ...'

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 적용사례
Budget(Local) - CurrType - CurrValue	M/NR/a	기술정보 (Descriptive Information)	과제예산 - 유형: '₩' - 값: '810...'
Subject(Local) - SubjectScheme - SubjectValue	M/NR/a		과제분류 ⁴⁾ - 스킴: 'NRCS-CS' (경제·인문사회연구회 표준분류체계) - 값: 'B4(금융)'
Proj_ResPubl(CERIF) - Proj_ResPublType - Proj_ResPublValue	M/R/m	맥락정보	과제_산출물 ⁵⁾ - 유형: 'Prop'(사업계획서) - 값: 'NRCS-KDI-U-EMC-EUBB-P-2009--003', ...
Proj_ResPubl(CERIF) - Proj_ResPublType - Proj_ResPublValue	M/R/m	맥락정보	과제_산출물 ⁶⁾ - 유형: 'ResPubl'(발간물) - 값: 'NRCS-KDI-RB-2009--001', ...
Proj_OrgUnit(CERIF) - Proj_OrgUnitType - Proj_OrgUnitValue	M/NR/a	맥락정보	과제_기관 ⁷⁾ - 유형: 'SupVise'(주관기관) - 값: 'KDI'(한국개발연구원)
Proj_Proj(CERIF) - Proj_ProjType - Proj_ProjValue	M/NR/m	맥락정보	과제_과제 ⁸⁾ - 유형: 'ContProj'(계속과제) - 값: 'none'
ProjClass(CERIF) - ProjClassType - ProjClassValue	M/NR/a	맥락정보	과제범주 ⁹⁾ - 유형: 'GovGrant'(정부출연금사업) - 값: 'EUBB'(실행사업/기본연구사업)
Rights(PREMIS) - RightsType - RightsValue	M/R/m	출처정보	권한 ¹⁰⁾ - 유형: 'Creator_Record'(입력자) - 값: 'KDI_M_00017'
Rights(PREMIS) - RightsType - RightsValue	M/R/m	출처정보	권한 ¹¹⁾ - 유형: 'Manager_Record'(관리자) - 값: 'NRCS_M_00011'
Revision(PREMIS) - RevisionElement - RevisionValue - RevisionPersID - RevisionDate	M/R/a	출처정보	보정 ¹²⁾ - 보정요소: 'none' - 보정값: 'none' - 보정자식별자: 'none' - 보정일시: 'none'
Migration(PREMIS) - MigrationType - MigrationDate - MigrationPersID	M/R/a	출처정보	마이그레이션 ¹³⁾ - 유형: 'none' - 일시: 'none' - 마이그레이션담당자식별자: 'none'
Fixity(PREMIS) - Fixitytype - FixityValue	M/R/m	인증정보	인증 ¹⁴⁾ - 유형: 'NPKI'(공인인증서) - 인증코드: '*****'
ObjectCategory(PREMIS)	M/NR/a	내용정보	객체범주 ¹⁵⁾ - 값: 'bit stream'

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 적용사례
Environment (PREMIS) - hardware - hardwareType - hardwareValue - software - softwareType - softwareValue - softwareVersion	M/R/a	내용정보	기술환경 ¹⁶⁾ - 하드웨어 - 유형: 'Server_DB' - 값: 'HP' - 소프트웨어 - 유형: 'DBMS' - 값: 'ORACLE' - 버전: '11g'

- 1) 과제식별자는 경제·인문사회 연구회 연구관리시스템 구축 과정에서 개발된 것으로 입력시 자동 생성됨.
- 2) 계속과제는 다년간 수행되는 과제를 말하며, 계속과제인 경우 '전체과제년'과 연계하여 해당 과제의 '과제년차'를 입력함.
예) 3년 계속과제의 2년차: '2/3'
- 3) 과제명은 'TitleType'을 이용하여 다양한 유형의 표제 및 언어로 확장이 가능함.
- 4) 개체에 대한 주제분류는 연구회에서 개발한 '경제·인문사회연구회 표준분류체계(NRCS-CS)'를 적용함.
- 5)~8) 개체 간의 관계는 링크로 연결됨. 즉, 각 개체는 관계유형과 함께 관련 개체의 식별자로 연결함.
- 9) 과제범주는 정부출연급사업 아래 연구회에서 사용하고 있는 사업유형코드로 구분함.
- 10)~11) 해당 레코드에 대한 권한정보로 유형으로 권한을 구분하고, 값으로는 권한 주체의 식별자를 부여함.
- 12) 해당 레코드에 변경이 발생한 경우 변경된 요소, 값, 변경자식별자, 일시를 자동저장함.
- 13) 해당 레코드에 대한 마이그레이션을 수행한 경우 마이그레이션 유형, 담당자식별자, 일시를 자동저장함.
- 14) 개체 및 객체의 내용에 접근 및 관리 행위를 수행할 경우, 관련된 권한에 대하여 '공인인증서(NPKI)'로 인증함.
- 15) 디지털 객체의 범주를 bit stream, file, representation으로 구분하여 저장함.
- 16) 기술환경은 연구회에서 규정한 환경으로 통일하여 자동 저장함.

<표 4> 산출물 개체 '사업계획서'의 보존 메타데이터 기본 요소와 적용사례

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
ResPubId (Local) - ResPubIDType - ResPubIDValue	M/NR/a	참조정보	사업계획서식별자 ¹⁾ - 유형: 'uniqueID' - 값: 'NRCS-KDI-U-EMC-EUBB-P-2009--003'
URL (PREMIS) - URLType - URLValue	M/NR/m	출처정보	파일링크 ²⁾ - 유형: 'Metadata' - 값: 'http://.../NRCS-KDI-U-EMC-EUBB-P-2009--003.hwp'
ResPubClass (CERIF) - ResPubClassType - ResPubClassValue	M/NR/a	기술정보 (Descriptive Information)	산출물범주 ³⁾ - 유형: 'Review'(심의단계) - 값: 'Proposal'(사업계획서)
Keyw (CERIF) - KeywType - KeywValue	M/R/a		키워드 ⁴⁾ - 유형: 'ProjKeyw_ko' - 값: '외국인 투자', '주식', '거시경제'
OverView (Local) - OverViewType - OverViewValue	M/NR/a		과제의 필요성 및 목적 ⁵⁾ - 유형: 'ProjOverView_ko' - 값: '최근 원/달러 환율 급등의 중요한 요인으로...'
Method (Local) - MethodType - MethodValue	M/NR/a		과제수행방법 ⁶⁾ - 유형: 'ProjMethod_ko' - 값: '문헌조사', '실증분석', '시계열회귀분석'

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
Contents(Local) - ContentsType - ContentsValue	M/NR/a	기술정보 (Descriptive Information)	과제내용 ⁷⁾ - 유형: 'ProjContents_ko' - 값: '금융위기 이후 다시 관심의 대상이 되고 있는 외국인 투자자의...'
Benefits(Local) - BenefitsType - BenefitsValue	M/NR/a		과제기대효과 ⁸⁾ - 유형: 'ProjBenefits_ko' - 값: '향후 외국인 투자자금 환수가...'
Budget(CERIF) - CurrType - CurrValue	M/NR/a		과제예산 ⁹⁾ - 유형: '₩' - 값: '810...'
Subject(Local) - SubjectScheme - SubjectType - SubjectValue	M/NR/a		과제분류 ¹⁰⁾ - 스킴: 'NRCS-CS' (경제·인문사회연구회 표준분류체계) - 유형: ProjSubject - 값: 'B4(금융)'
Proj_Pers(CERIF) - Proj_PersType - Proj_PersValue	M/NR/a	맥락정보	과제책임자 ¹¹⁾ - 유형: 'PriInvest'(연구책임자) - 값: 'KDI_R_00033(신**)'
Proj_Pers(CERIF) - Proj_PersType - Proj_PersValue	M/R/a	맥락정보	과제참여자 ¹²⁾ - 유형: 'Participants'(연구참여자) - 값: 'KDI_R_00037(입**)'
Proj_OrgUnit(CERIF) - Proj_OrgUnitType - Proj_OrgUnitValue	M/NR/a	맥락정보	과제기관 ¹³⁾ - 유형: 'SupVised'(주관기관) - 값: 'KDI'(한국개발연구원)
Proj_OrgUnit(CERIF) - Proj_OrgUnitType - Proj_OrgUnitValue	M/R/a	맥락정보	과제기관 ¹⁴⁾ - 유형: 'Participated'(참여기관) - 값: 'none'
Relation(PREMIS) - RelationType - RelationObjectidentifier	M/NR/m	출처정보	관계 ¹⁵⁾ - 유형: 'FollowedBy'(선행/후행관계) - 값: 'NRCS-KDI-U-ESC-EUBB-P-2009--003'
Relation(PREMIS) - RelationType - RelationObjectidentifier	M/R/m	출처정보	관계 ¹⁶⁾ - 유형: 'CausedBy'(인과관계) - 값: 'NRCS-KDI-U-2009--003'
Rights(PREMIS) - RightsType - RightsValue	M/NR/m	출처정보	권한 ¹⁷⁾ - 유형: 'Creator_Record'(입력자) - 값: 'KDI_M_00017'
Rights(PREMIS) - RightsType - RightsValue	M/NR/m	출처정보	권한 ¹⁸⁾ - 유형: 'Manager_Record'(관리자) - 값: 'NRCS_M_00011'
Revision(PREMIS) - RevisionElement - RevisionValue - RevisionPersID - RevisionDate	M/R/a	출처정보	보정 ¹⁹⁾ - 보정요소: 'none' - 보정값: 'none' - 보정자식별자: 'none' - 보정일시: 'none'
Migration(PREMIS) - MigrationType - MigrationDate - MigrationPersID	M/R/a	출처정보	마이그레이션 ²⁰⁾ - 유형: 'none' - 일시: 'none' - 마이그레이션담당자식별자: 'none'

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
Fixity(PREMIS) - Fixitytype - FixityValue	M/NR/m	인증정보	인증 ²¹⁾ - 유형: 'NPKI'(공인인증서) - 인증코드: '*****'
ObjectCategory(PREMIS)	M/NR/a	내용정보	객체유형 ²²⁾ - 값: 'File'
Environment(PREMIS) - hardware - hardwareType - hardwareValue - software - softwareType - softwareValue - softwareVersion - Format - FormatType - FormatVersion	M/NR/a	내용정보	전산환경 ²³⁾ - 하드웨어 - 유형: 'DeskTop' - 값: 'All' - 소프트웨어 - 유형: 'Application' - 값: '한글' - 버전: '2007이상' - 파일포맷 - 유형: 'HWP' - 값: '실행_과제별_사업계획서_1.가.기본연구사업_v1.0.hwp'

- 1) 사업계획서식별자는 경제·인문사회 연구회 연구관리시스템 구축 과정에서 개발된 것으로 입력시 자동 생성됨.
- 2) 보존 메타데이터 입력파일(한글파일)에 대한 링크정보를 입력함.
- 3) 산출물범주는 'Type'으로 산출단계(기획/심의/수행/결과)를 구분하고, 'Value'로는 산출물의 자료유형(사업계획서, 발간물(단행본, 연속간행물))을 저장함.
- 4) ~9) 사업계획서의 내용에 대한 분석에 사용하기 위한 요소들임.
- 10) ~14), 17) ~22) 과제 개체와 동일하게 적용함.
- 12) 공동연구 또는 협동연구일 경우, 참여 연구자와 소속기관을 입력함.
- 15) ~16) 산출물 개체들 간의 관계는 'Relation' 요소로 연결함. 구체적으로 사업계획서 간의 관계는 심의 단계별로 선행/후행, 사업계획서와 산출물 간의 관계는 원인/결과로 입력함.
- 23) 기술환경은 연구회에서 규정한 환경으로 자동 저장함. 또한 파일포맷의 값으로는 사업계획서 관련 메타데이터 요소를 자동입력하기 위해 업로드 되는 한글파일명을 자동 저장함.

<표 5> 산출물 개체 '발간물(단행본, 연속간행물)'의 보존 메타데이터 기본 요소와 적용사례

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
ResPublId(Local) - ResPublIDType - ResPublIDValue	M/NR/a	참조정보	발간물식별자 ¹⁾ - 유형: 'uniqueID' - 값: 'NRCS-KDI-RB-2009--001'
URL(PREMIS) - URLType - URLValue	M/NR/m	참조정보	파일링크 ²⁾ - 유형: 'Metadata' - 값: 'http://.../NRCS-KDI-RB-2009--001.hwp'
ResPublClass(CERIF) - ResPublClassType - ResPublClassValue	M/NR/a	기술정보 (Descriptive Information)	산출물범주 ³⁾ - 유형: 'Result'(결과단계) - 값: 'Report'(연구보고서)
ResPublDate(CERIF)	M/NR/a		발간일 ⁴⁾ - '20091231'
Vol(CERIF)	M/NR/a		권 ⁵⁾ - 'none'
Num(CERIF)	M/NR/a		호 ⁶⁾ - 'none'

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
StartPage(CERIF)	M/NR/a	기술정보 (Descriptive Information)	시작쪽 ⁷⁾ - '1'
EndPage(CERIF)	M/NR/a		끝쪽 ⁸⁾ - '174'
TotPage(Local)	M/NR/a		총페이지수 ⁹⁾ - '174'
Title(CERIF) - TitleType - TitleValue	M/R/a		발간물명 ¹⁰⁾ - 유형: 'BookTitle_ko' - 값: '외국인 투자자 ...'
Keyw(CERIF) - KeywType - KeywValue	M/R/a		키워드 ¹¹⁾ - 유형: 'AuthorKeyw_ko' - 값: '외국인 투자', '주식', '거시경제', '매매'
OverView(Local) - OverViewType - OverViewValue	M/NR/a		연구의 필요성 및 목적 ¹²⁾ - 유형: 'ResPublOverView_ko' - 값: '자본시장 개방 이후 외국인 투자자...'
Method(Local) - MethodType - MethodValue	M/NR/a		연구방법 ¹³⁾ - 유형: 'ResPublMehtod_ko' - 값: '주성분 분석기법 등'
Contents(Local) - ContentsType - ContentsValue	M/NR/a		연구내용 및 결과 ¹⁴⁾ - 유형: 'ResPublContents_ko' - 값: '본 연구는 금융위기 이후 다시 관심의 대상이 되고 있는 외국인 투자자의 주식매매...'
Benefits(Local) - BenefitsType - BenefitsValue	M/NR/a		기대효과 ¹⁵⁾ - 유형: 'ResPublBenefits_ko' - 값: '외국인 투자자의 주식 매매 자금흐름을 살펴봄으로써...'
Abstr(CERIF) - AbstrType - AbstrValue	M/NR/a		초록(또는 목차) ¹⁶⁾ - 유형: 'ResPublAbstr_ko' - 값: '본 연구는 금융위기 이후 다시 관심의 대상이 되고 있는 외국인 투자자의 주식매매...'
Subject(Local) - SubjectScheme - SubjectType - SubjectValue	M/NR/m		발간물분류 ¹⁷⁾ - 스킴: 'NRCS-CS' (경제·인문사회연구회 표준분류체계) - 유형: ResPublSubject - 값: 'B4'(금융)
ResPubl_Pers(CERIF) - ResPubl_PersType - ResPubl_PersValue	M/NR/a		연구책임자 ¹⁸⁾ - 유형: 'PriInvest'(연구책임자) - 값: 'KDI_R_00033'(신**)
ResPubl_Pers(CERIF) - ResPubl_PersType - ResPubl_PersValue	M/NR/a		연구참여자 ¹⁹⁾ - 유형: 'Participants' - 값: 'KDI_R_00037'(임**)
ResPubl_OrgUnit(CERIF) - ResPubl_OrgUnitType - ResPubl_OrgUnitValue	M/NR/a		발간기관 ²⁰⁾ - 유형: 'Survised'(주관기관) - 값: 'KDI'(한국개발연구원)

메타데이터 요소	필수/반복/자동	OAIS AIP	한글명 및 사례
Relation (PREMIS) - RelationType - RelationObjectIdentifier	M/R/m	출처정보	관계 ²¹⁾ - 유형: 'CausedBy' (인과관계) - 값: 'NRCS-KDI-U-2009--003'
Rights (PREMIS) - RightsPersType - RightsPersIDValue	M/NR/m	출처정보	권한 ²²⁾ - 유형: 'Creator_Record' (입력자) - 값: 'KDI_M_00033'
Rights (PREMIS) - RightsPersType - RightsPersValue	M/R/m	출처정보	권한 ²³⁾ - 유형: 'Manager_Record' (관리자) - 값: 'NRCS_M_00011'
Revision (PREMIS) - RevisionElement - RevisionValue - RevisionPersID - RevisionDate	M/R/a	출처정보	보정 ²⁴⁾ - 보정요소: 'none' - 보정값: 'none' - 보정자식별자: 'none' - 보정일시: 'none'
Migration (PREMIS) - MigrationType - MigrationDate - MigrationPersID	M/R/a	출처정보	마이그레이션 ²⁵⁾ - 유형: 'none' - 일시: 'none' - 마이그레이션담당자식별자: 'none'
Fixity (PREMIS) - Fixitytype - FixityValue	M/NR/m	인증정보	인증 ²⁶⁾ - 유형: 'NPKI' (공인인증서) - 인증코드: '*****'
ObjectCategory (PREMIS)	M/NR/a	내용정보	객체범주 ²⁷⁾ - 값: 'File'
Environment (PREMIS) - hardware - hardwareType - hardwareValue - software - softwareType - softwareValue - softwareVersion - Format - FormatType - FormatVersion	M/NR/a	내용정보	전산환경 ²⁸⁾ - 하드웨어 - 유형: 'DeskTop' - 값: 'All' - 소프트웨어 - 유형: 'Application' - 값: '한글' - 버전: '2007이상' - 파일포맷 - 유형: 'HWP' - 값: '발간물_단행본_연구보고서_기본연구사업_v1.0.hwp'

- 1) 발간물식별자는 경제·인문사회 연구회 연구관리시스템 구축 과정에서 개발된 것으로 입력시 자동 생성됨.
- 2) 보존 메타데이터 입력파일과 발간물 원문에 대한 링크정보를 입력함. 사업계획서와 달리 발간물은 보존 메타데이터 입력파일과 원문파일(hwp, pdf)이 별도로 존재하므로, 'Type'에서 이를 구분하고, 'Value'에 해당 파일의 URL을 저장함.
- 4)~9) 발간물에 대한 기본적인 기술정보임.
- 10) 과제명과 다를 경우에 입력함.
- 11)~15) 발간물의 내용에 대한 분석에 사용함. 최종 확정된 사업계획서의 내용과 상호비교할 수 있음.
- 16) 14)의 '연구내용 및 결과' 요소와는 별도로 발간물의 내용에 대한 기술정보로 추가함. 한글과 영어로 입력함.
- 3), 17)~27) 과제 및 사업계획서 개체와 동일하게 적용함.
- 28) 기술환경은 연구회에서 규정한 환경으로 자동 저장함. 또한 파일포맷의 값으로는 발간물(단행본, 연속간행물) 관련 메타데이터 요소를 자동입력하기 위해 업로드되는 한글파일명을 자동 저장함.

5. 결론

가치 있는 디지털 정보자원으로서 국가정책 연구를 위한 연구정보의 장기적 보존과 접근을 보장하기 위한 보존 메타데이터 요소를 개발하여 실제로 적용한 사례를 제시하였다. 특히 경제·인문사회연구회의 연구관리 차원에서 국가정책연구정보의 장기적인 접근과 보존을 위한 기반으로 보존 메타데이터 요소를 개발 및 적용하였다.

본 연구에서 개발 및 적용한 보존 메타데이터는 첫째, 연구정보를 위한 유럽표준인 CERIF의 요소와 보존 메타데이터의 구현을 위한 대표적인 표준으로서 PREMIS 데이터 사전의 요소를 비교 분석한 결과에 따라 양자의 요소를 상호보완적으로 적용하였다. 둘째, 연구정보의 생산 및 보존의 전 과정에서 핵심적인 보존 메타데이터의 자동 추출 및 생성이 가능하도록 하였다. 셋째, 중장기적 측면에서 환경의 변화에 효과적으로 대응할 수 있도록 요소의 추가 및 보완을 통한 확장이 가능하도록 하였다.

본 연구에서 개발한 보존 메타데이터 요소들을 연구회 소관 23개 정부출연연구기관에서 지난 3년간(2007년~2009년) 수행된 기본연구사업 관련 연구정보에 적용하였다. 그 결과로 사업 계획서가 1340건, 산출물이 2589건으로 총 3929건에 대하여 대부분의 기본 요소들에 대한 실제 데이터를 입력하였다. 이 과정에서 일부 요소들에 대한 데이터가 누락되는 경우가 있었는데, 특히 기술정보 중에서 키워드 요소와 맥락정보, 출

처정보, 인증정보 관련 요소들이 그러하였다. 이는 이들 요소에 대한 기존 데이터가 전혀 생산되지 않았거나 생산 시점에 해당 데이터를 관리 및 보존할 수 있는 시스템이 존재하지 않은 관계로 소급입력 과정에서 발생하는 문제이며, 이를 보완하기 위해서는 기존 연구정보의 디지털화 및 변환에 대한 연구가 추가적으로 필요하다. 그러나 신규 입력되는 연구정보는 본 연구에서 제안한 한글파일로 작성된 표준 양식을 통하여, 이러한 보존 메타데이터 요소들이 자동입력 및 관리되므로 이러한 문제가 발생하지 않는다.

연구정보 보존 메타데이터 요소를 통하여 체계적으로 축적된 연구정보는 다음과 같이 다양한 목적으로 활용될 수 있다. 첫째, 가치 있는 디지털 정보자원으로서 연구정보의 장기 보존과 접근, 둘째, 연구활동에 대한 신뢰성 있는 기록의 제시 및 유지, 셋째, 전략적 연구기획과 연구성과 분석, 넷째, 연구성과의 효율적 관리 및 확산.

본 연구의 제한점은 연구정보의 범위를 경제·인문사회 연구회에서 주관하는 국가정책 연구 중에서 정부출연연구기관의 기본연구사업 수행과정에서 생산 및 관리되는 정보로 한정하여 보존 메타데이터 요소를 개발한 것이다. 따라서 이 연구에서 개발 및 적용된 보존 메타데이터 요소를 확장하여 경제·인문사회 분야 정부출연연구기관의 전체 연구사업을 포괄하여 적용될 수 있는 보다 범용적인 보존 메타데이터 요소의 개발을 위한 후속 연구가 필요할 것이다.

참 고 문 헌

- 교육과학기술부. 2009. 『교육과학기술부 연구개발사업 성과관리업무 매뉴얼』. 서울: 교육과학기술부.
- 서은경. 2005. 디지털자원의 보존 메타데이터 요소셋 개발에 관한 연구. 『정보관리학회지』, 22(3): 233-260.
- 이지수, 정영미. 2010. 국가 정책지식의 구조와 특성에 관한 연구. 『정보관리연구』, 41(2): 1-30.
- 정정길. 1996. 『정책결정론』. 서울: 대명출판사.
- 최봉기. 1981. 政策情報의 機能과 價値評價. 『한국행정학보』, (15): 157-172.
- 최상미, 리상용. 2007. 현용전자기록물의 메타데이터 요소에 관한 연구. 『한국기록관리학회지』, 7(1): 39-60.
- 홍현진, 노영희. 2008. 정책정보통합서비스시스템 구축 모형에 관한 연구. 『한국문헌정보학회지』, 42(1): 95-125.
- Caplan, P. 2009. Understanding PREMIS. [cited 2010.10.11]. <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>.
- CCSDS. 2002. Reference Model for an Open Archival Information System(OAIS). CCSDS 650.0-B-1, BLUE BOOK. [cited 2010.9.20]. <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>.
- Day, Michael. 2003. Preservation Metadata Initiatives: Practicality, Sustainability, and Interoperability. [cited. 2010.10.3]. <<http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf>>.
- Dye, T. R. 1981. *Understanding Public Policy*, 5th ed. Englewood Cliffs, N.J.: Prentice-Hall.
- euroCRIS. 2010. CERIF 2008 - 1.1 FDM: Model Introduction and Specification. [cited 2010.9.7]. <http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/CERIF2008_1.1_FDM.pdf>.
- Graham, P. S. 1994. Intellectual Preservation: Electronic Preservation of the Third Kind. [cited 2010.9.11]. <<http://www.clir.org/pubs/reports/graham/intpres.html>>.
- Hodge, G. M. 2000. "Best Practices for Digital Archiving: An Information Life Cycle Approach." *D-Lib Magazine*, 6(1). [cited 2010.9.7]. <<http://www.dlib.org/dlib/january00/01hodge.html>>.
- JISC. 2010. Research Information Management: Towards a common standard for exchanging research information. [cited. 2010.10.8]. <<http://www.jisc.ac.uk/publications/briefingpapers/2010/bpexriv1.aspx#d>>

- downloads>.
- Lupovici, C. and J. Masanès 2000. Metadata for the Long Term Preservation. [cited 2010.10.1]. <<http://nedlib.kb.nl/results/preservationmetadata.pdf>>.
- OCLC/RLG Working Group. 2002. Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects. [cited 2010.10.11]. <http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf>.
- PREMIS Working Group. 2005. Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group. [cited 2010.9.11]. <<http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf>>.
- PREMIS Editorial Committee. 2008. *Data Dictionary for Preservation Metadata version 2.0*. [cited 2010.9.11]. <<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>>.
- Rappaport, A. 1982. *Information for Decision-making: Quantitative and Behavioral Dimensions*, 3rd ed. Englewood Cliffs, N.J.: Prentice-Hall.
- Rogers, N., Lesly Huxley, and Nicky Ferguson. 2009. Exchanging Research Information in the UK. [cited 2010.10.15]. <http://ie-repository.jisc.ac.uk/448/1/exri_final_v2.pdf>.
- UKOLN. 2010. Research Information Management in the UK: CERIF and metadata alignment. [cited 2010.10.10]. <<http://www.ukoln.ac.uk/rim/dissemination/2010/rim-cerif.pdf>>.