

과학기술분야의 연구성과물정보 표준 메타데이터 개발에 관한 연구

박 동 진

공주대학교 산업시스템공학과

A Study on the Development of Metadata Standard for Research Outcomes Information in Science & Technology

Dong-Jin Park

Dept. of Industrial and Systems Engineering, Kongju National University

It has lately been difficult to share the research outcomes information due to the information system not being compatible with one another. To solve this problem, it is believed that a metadata standard needs to be developed for all the institutions to follow. This paper shows how to make the research outcome of Science and Technology R&D project a standard metadata. We first classified all the research outcome into the following seven groups: Report, Journal article, Presentation material, Book, Dissertation, Industrial rights, and Technology development. We then identified three components that constitute a standard metadata. We employed the Dublin Core Metadata Set and CERIF(Common European Research Information Format) for worldwide interoperability. This Standard can be applied in the following areas : 1) Developing a standard architecture to manage research result within the organization. 2) Designing a conceptual/logical database to measure and evaluate the research outcome. 3) Developing a metadata system for research outcome as a subsystem of an Industrial Technology MDR (Metadata Registry) System

Keywords : Metadata, Standards, Research Outcomes, Dublin Core, CERIF

1. 서 론

연구개발 성과정보들의 수집, 교환 및 평가를 가능하게 하는 성과관리 정보시스템의 구축에 앞서 필수적으로 선행되어야 할 것이 연구성과 정보에 대한 저장 체계의 표준화이다. 특히 연구성과물 공유 차원에서 각 연구기관별로 생산된 연구성과물을 상호연계하고 유통을 하기 위해서는 상호운용성이 확보되어야 한다.

이는 각 연구기관에서 연구성과정보를 일관성있고 체계적으로 관리할 때 가능하다. 그러나 현재 연구성과물에 대한 정보는 연구기관별로 서로 상이한 체계로 관리

하고 있어서 연구성과물의 호환성 및 상호운용성이 문제가 된다.

본 연구의 목표는 과학기술 분야에 있어서 연구개발 성과물의 상호운용성 제고를 위하여, 연구성과로 인정되는 각종 산출물에 대한 표준 메타데이터를 개발하는 것이다.

제2장 연구의 배경에서는 연구성과정보시스템 실태, 연구성과물 범위 및 메타데이터 개발에 대하여 알아보고, 제3장에서는 메타데이터 표준 개발의 절차에 따른 결과물을 제시한다. 제4장에서는 본 연구를 통하여 개발된 메타데이터 표준의 활용방안을 제시하고, 마지막으로 제5장에서 결론과 연구의 한계를 논의한다.

2. 연구의 배경

2.1 연구성과정보 검색의 실태

과학기술부가 국가연구개발성과의 활용에 관한 애로 사항을 파악하여 과학기술혁신정책에 반영코자 실시한 설문에서 보면 연구성과활용이 미진한 이유로서 “성과물의 축적과 관리 미흡”이 13.1%에 달한 것으로 파악하였다[2].

2006년 현재 과학기술분야의 연구성과물은 한국과학기술정보연구원의 “과학기술통합검색”을 통하여 각 연구기관의 연구보고서 및 출원특허 등에 관한 정보를 개략적으로 파악할 수 있으나, 해당 연구성과물이 어떤 연구사업 및 과제와 관련되는지를 파악하기가 어렵다[3]. 또한 한국과학기술기획평가원에서 운영중인 “국가연구개발사업종합관리시스템”에서는 각 사업과 관련과제들을 검색할 수는 있으나, 해당 연구사업과 과제의 연구업적을 평가하기 위한 연구성과물을 파악하기가 어렵다. 또한 각 연구기관별로는 각 기관의 자체연구성과물인 보고서, 산업재산권 및 특허 등에 대한 검색과 조회가 가능하나, 각 기관별 데이터베이스 스키마에 차이가 있어서 상호호환성이 없으며 관리 대상물의 범위도 서로 달라서 종합적인 연구성과물의 파악이 어렵다.

즉 이상에서처럼 현재의 연구성과정보시스템은 첫째, 정보검색 측면에서 연구성과물 중 연구보고서와 같이 공식적으로 발표된 것들에 대한 정보만을 검색 가능하고 각종 회의 및 세미나 자료, 논문, 단행본 등을 포함하는 기타의 연구성과물에 대한 정보의 검색이 불가능하다. 둘째, 연구성과측정 측면에서 특정 과제와 관련하여 산출된 모든 성과물에 대한 추적 및 관리가 시스템적으로 불가능하다. 셋째, 연구기관별 연구성과물 유통과 정보교환이 불가능하다.

2.2 메타데이터를 통한 상호운용성 확보

현재 거의 모든 연구기관에서는 제한적이지만 웹환경에서 나름대로의 방법으로 연구성과정보를 관리하고 있다. 이러한 상황에서 단일의 데이터 구조형식으로 통합하여 전체를 관리하는 것은 현실적으로 한계가 있다. 따라서 각 기관의 연구성과정보관리의 다양성을 인정하면서 상호운용성(interoperability)을 제고할 수 있는 방안이 필요하다. 상호운용성이란 두개 또는 그 이상의 기술적인(technical) 시스템이 시스템 사용자가 만족할 수 있는 방법으로 정보를 직접 교환할 수 있는 상태를 말한다[15]. 이러한 상호운용성을 확보하기 위한 기반이 해당영역의 메타데이터이다.

메타데이터는 데이터에 관한 구조화된 데이터로서 자원과는 독립적으로 존재하면서 자원에 대한 다양한 접근점과 네트워크 주소를 포함하는 데이터이다[15]. 메타데이터는 다음과 같이 활용된다. 첫째, 자원의 의미를 요약해 준다. 둘째, 자원의 검색을 가능하게 한다. 셋째, 자원의 필요성에 대한 판단을 가능하게 해 준다. 넷째, 다른 자원들과의 연관성을 말해준다[17].

2.3 메타데이터 표준의 개발

메타데이터의 표준을 개발하고 활용하는 것은 디지털 자원을 검색하고, 공유하고, 관리하는데 있어서 매우 중요하다. 잘 구조화된(well-structured) 메타데이터는 자료를 검색하여 표현하는데 있어서 거의 무한가지의 방법을 제공한다[12]. 메타데이터의 개발시 기존의 표준을 채택하는 것은 메타데이터간의 상호호환성을 확보하기 위하여 매우 중요하다. 대표적인 웹자원을 기술하는 표준 메타데이터가 Dublin Core이다[10]. 그리고 과학기술분야의 연구개발 성과물을 위해서 CRIS에서 CERIF-2004 메타데이터를 제시하였다[9].

메타데이터의 개발에 있어서 조직에서 유연성, 확장성, 응용성, 가독성이 있는 어휘를 결정하는 것은 매우 어려운 일이기 때문에 메타데이터의 개발 과정은 복잡하고 다양한 사람들이 관여되어야 한다[11]. 그러나 일반적인 절차는 다음과 같다. 먼저 적절한 메타데이터 요소의 집합(element set)을 선택하고, 각 데이터 요소에 대하여 어휘(vocabulary)를 정의하고 그리고 값의 표현 방법인 스킴(scheme)을 결정한다. 많은 경우 메타데이터 스킴은 기존의 표준이나 온톨로지(ontology)들로부터 도입한다.

2.4 메타데이터 기반 정보시스템 개발단계

메타데이터 개발단계는 개발의 목적, 범위 등에 따라서 많이 다르므로 일반적으로 통용되는 절차 및 방법은 없다. 대안에서 디지털 도서관 및 박물관을 위한 메타데이터 개발에서 채택한 절차와 내용은 <표 1>과 같다[8]. 이들을 중심으로 한 메타데이터 개발팀은 국내 및 국제적 표준포맷, 메타데이터 간의 관계, 기존 정보와의 호환성, 그리고 데이터베이스 구현 등이 중요한 이슈로서 검토하였다.

Xie와 Shibasaki(2005)에 의해서 지구상의 Water Cycle 과 관련된 각종 자료 관리 및 서비스를 위한 메타데이터 시스템이 개발되었다[18]. 이 시스템은 크게 3개의 과정을 거쳐서 개발되었다. 첫째는, 메타데이터개념 모델의 구축이다. UML으로 기술한 이 모델은 메타데이터

의 요소를 핵심(공통)요소, 분야별 특성요소, 그리고 조직관련 요소 등으로 구성하였다. 둘째는, 메타데이터 구현 모델로서 메타데이터개념 모델을 XML Schema, XSL로 구현을 하였다. 마지막으로 웹서비스 응용시스템을 개발하였다. 이는 메타데이터의 저장 및 관리와 웹을 통하여 이용자에게 각종서비스를 제공할 수 있는 기능을 포함한다.

<표 1> 메타데이터 시스템 개발 절차

순서	단 계	내 용
1	Analysing the attributes of collections	정보 대상물의 내용 및 특성을 반영하는 속성을 파악
2	Needs assessment of metadata users	메타데이터의 잠재적 사용자에게 대한 정보요구와 정보획득 행위를 분석
3	Interoperability consideration	국내외 관련 표준과의 호환성 및 상호운용성을 고려
4	Semantic design of metadata	의미를 반영한 메타데이터의 체계설계
5	Developing metadata management system	메타데이터의 전자적 표현체계 및 관리를 위한 틀 개발
6	Developing tagging guide and user manual	사용자가 직접 메타데이터를 입력, 수정, 및 조회할 수 있는 가이드와 매뉴얼 작성
7	Providing training courses	메타데이터 관련자 및 사용자에게 대한 교육 실시

2.5 연구성과물의 범위

과학기술분야에 있어서 연구 성과물은 일반적으로 첫째, 저널이나 기타 문헌들을 통한 연구성과물의 발표(publication)나 실험과 관찰을 통하여 획득한 데이터세트 등을 말한다[14]. 또한 시제품 및 제조기술도 중요한 연구성과물로 볼 수 있다.

문헌을 통한 연구성과물은 학술적 검토 과정을 거치며 공식적인 채널을 통하여 입수가 가능한 저널 논문(journal articles)과 생산, 유통, 자료의 이용자 및 서지 기술적 특성이 독특하여 공식적인 출판경로를 통해서 입수하기 어려운 회색문헌(grey literature)으로 구분된다[4]. R&D 관점에서의 회색문헌은 출판전 배포문, 연구 계획 제안서, 연구보고서, 일반적으로 컨퍼런스 발표논문, 기술보고서, 학위논문, 회의/세미나/워크샵 자료, 기술계약에 관한 보고서, 그리고 특허 등이다[6]. Jeffery (1999)은 R&D에서 생성된 대부분의 연구결과물은 회색 문헌이다 라고 주장하였다[10]. 국내 연구소에 대한 조사의 결과에 따르면 회색문헌은 최신의 연구결과이므로

과학기술의 발전을 위해서 회색문헌의 학술적 가치 및 활용상 가치가 매우 크며, 각 기관은 물론 국가적 차원에서 이를 체계적으로 관리해야 할 필요가 있음을 주장하였다[5].

또한 과학기술 R&D 프로젝트에서 실시한 각종 측정(measurement)데이터, 실험(experiment)데이터, 및 시뮬레이션데이터도 중요한 연구 결과물로 간주된다[14]. 따라서 과학 연구 데이터세트에 대한 메타데이터 표준을 제정하고, 각 연구기관에서 이를 기반으로 메타데이터를 작성하고 관리함으로써 귀중한 연구결과물의 활용을 촉진할 수 있다.

3. 연구성과물 메타데이터 표준개발

본 연구를 통하여 개발될 표준은 과학기술분야의 연구개발 성과물인 연구보고서, 각종 논문, 지식(산업)재산권 그리고 기술개발 등의 정보들을 일관성 있고 체계적으로 구성한 통합 메타데이터 표준이다.

연구성과물 메타데이터 표준개발의 절차는 다음과 같다. 먼저 메타데이터의 핵심(core) 모델을 개발하고 기타 관련 데이터요소들을 그룹화 한다. 다음으로 전체적 메타데이터 구조 및 상세정도를 결정한 후 최종적으로 각 범주별 메타데이터를 기술하는 것이다.

3.1 표준 개발의 목표

본 표준은 과학기술분야의 연구개발(R&D) 성과물에 관한 통합 메타데이터 표준으로서 각 연구성과물에 관한 정보를 일관성 있고 체계적으로 관리하고 또한 이를 통하여 연구성과의 측정 및 평가, 연구성과물 검색의 효율성 및 상호운용성을 제고하는 것을 목적으로 하며 구체적으로 다음과 같은 세부 목표가 있다. 첫째, 연구성과물 정보관리 차원의 “본원적 모델(The Generic Metadata Model)”로서 각 기관별 그리고 학문분야별 적용 가능하여야 한다. 둘째, 국내외 학문 및 기관간의 연구성과물 메타데이터 상호운용성을 확보하여야 한다. 셋째, 정보시스템(메타데이터 카탈로그/검색) 개발의 인프라 역할을 해야 하며 XML 기반의 메타데이터 스키마, 관계형 데이터베이스 모델 구축의 기반이 되어야 한다. 넷째, 연구성과물과 관련된 검색서비스 및 상호운용성 확보를 위한 시스템개발의 기반구조가 되어야 한다.

3.2 메타데이터 핵심 모델의 개발

메타데이터 데이터 모델은 정보공학의 데이터 모델링

기법을 적용할 수 있다. 즉 먼저 엔터티가 될 수 있는 연구성과물을 분석하고 포함되어야 하는 속성을 파악하는 것이다. 또한 연구성과물과 관련된 다른 엔터티들도 파악한다. 다음으로 각 엔터티간의 관계를 설정하고 마지막으로 데이터 모델을 결정하는 것이다.

3.2.1 연구성과물 결정기준

우리나라 과학기술계에서 보편적으로 인정하는 유형/무형의 모든 연구성과물을 대상으로 한다. 즉 공식적인 절차 혹은 외부심사를 거쳐 성과물로 인정되는 연구 결과물 뿐 아니라, 내부 발표자료나 혹은 기술보고서 등과 같은 비공식 연구결과물(소위 회색문헌)도 연구의 성과물로 인정하고 관리대상으로 한다.

또한 연구성과물의 형태로는 일반문서는 물론 디지털 자원(전자문서, 동영상, 멀티미디어 등) 등을 포함한다. 그러나 과학기술분야의 R&D 주요 연구 결과물의 하나인 실험데이터(scientific datasets)는 향후 고려될 수 있으나 본 메타데이터 표준에서는 연구성과물에서 제외한다.

3.2.2 연구성과물 엔터티

연구성과정보는 크게 첫째, 연구성과물 자체, 둘째, 연구성과물 생산주체, 그리고 셋째, 연구성과물 유통과 관련된 요인들로 그룹화 할 수 있다. 먼저 연구성과물 그룹에는 구체적으로 연구보고서, 각종 학술논문, 각종 발표회 자료, 전시물, 단행본 및 간행, 학위논문, 산업재산권, 기술개발 등을 고려할 수 있는데, 데이터 모델링 관점에서 보면 각 연구성과물에 대하여 공통적으로 적용되는 공통요인들과 각 성과물 개개의 특성요인으로 구분할 수 있다. 공통요인을 상위(super-type) 개체로 각 성과물의 특성요인을 하위(sub-type) 개체로 구성할 수 있다. 즉 연구성과물의 데이터 요소 중 일부는 모든 개체들에 공통으로 적용되는 요소이고 나머지 데이터 요소들은 각 개체들 특성을 반영하는 독특한 데이터 요소들로 구성된다.

다음으로 연구성과물 생산주체 그룹으로 연구결과물 생산과 직접적으로 관련되는 개체로서 연구과제, 연구원, 그리고 연구기관 등이다. 이들은 해당 연구성과물의 생성배경과 관계되는 것으로 연구사업, 연구주제 및 목적, 연구자금, 연구기관 및 실제 성과물 생산에 투입된 연구원 정보 등을 포함한다.

마지막으로 연구성과물 유통과 관련되는 그룹으로 구체적으로 연구성과물의 저작권(rights)과 접근(access)과 관계된다. 연구성과물은 유통에 있어서 계약 및 지적소유권, 저작권, 개인정보 보호 등 법적인 환경과 상호연관성을 가진다. 따라서 연구성과물의 전송 및 배포를 통한 유통시 적용되어야 한 정보를 포함하고 있어야 한

다. <표 2>는 메타데이터 데이터 모델에 포함되는 엔터티와 속성들이다.

3.3 메타데이터 표준 체계 및 내용

3.3.1 데이터요소 결정기준 및 구조

메타데이터의 개발에 있어서 데이터요소는 가능한 한 기존의 표준 메타데이터 스키마로부터 데이터 요소를 상속(inheritance)받아서 구성하는 것이 바람직하다. 본 연구에서는 Dublin Core의 기본요소 및 하위요소를 근간으로 하고, 국제적으로 사실적 연구성과물 관리의 표준모델로 인정되는 CERIF에서 사용하는 요소를 수용하였다.

<표 2> 연구성과물 엔터티 및 속성들

핵심엔터티	속 성
연구성과물 공통	연구성과물 제목(연구성과물 국문 제목, 연구성과물 영문 제목), 언어, 연구성과물 분류, 날짜(연구성과물 작성일, 연구성과물 발표일), 연구성과물 작성자(연구성과물 작성참여인원수, 연구성과물 책임작성자 이름, 연구성과물 책임작성자 소속, 연구성과물 참여작성자 이름, 연구성과물 발행자(연구성과물 발행자 이름, 연구성과물 발행기관, 연구성과물 발행국), 과학기술 정보분류(계열, 대주제, 중분류), 연구성과물 주제어, 연구성과물 내용기술(연구성과물 목차, 연구성과물 요약기술), 연구성과물 저장매체 형태, 연구성과물 연결(연구성과물 접근 및 파일이름, 연구성과물 관련정보URL), 연구성과물 소장기관, 연구성과물 저작권, 연구성과물 공개여부
연구보고서	연구보고서 유형코드, 연구보고서 면수
학술논문	학술지 서명, 학술지 권 번호, 학술지 호 번호, 학술지 발행계절, 시작 페이지, 끝 페이지, 학술지 분류, 학술지 간행빈도
각종발표회 및 전시회	이벤트 명칭, 이벤트 주제, 이벤트 주최기관, 이벤트 공식언어, 세션주제, 발표자(발표방식, 발표자 이름, 발표자 소속), 행사일(이벤트 시작일, 이벤트 종료일)
간행물	ISBN, 출판사이름, 판매가격, 시작 페이지, 끝 페이지
학위논문	학위논문 식별자, 수여대학, 전공분야, 학위유형, 지도교수
산업재산권 등록	IPC code, 출원(출원자, 출원번호, 출원일자), 최종 처분, 등록(등록번호, 등록일자), 지정국
기술개발	기술활용분야, 계약당사자(기술실시 허락자, 기술실시권자), 기술이전형태, 기술실시 계약기간, 계약 금액 및 지불방식(기본료, 경상기술료)

DC로부터 15개의 데이터 요소와 qualifier를 그대로 채택하였다. 이는 국제적으로 의미적 메타데이터의 상호운용성을 확보하자는 것이다. 그리고 국내 표준(권고)으로는 ① 과학기술정보 분류 표준, ② 과학기술잡지

식별기호 표준, ③ 과학기술정보 식별자 표준, ④ 과학기술잡지 권/호 패턴 표준, ⑤ 연구보고서 발간체제 표준 등을 참조하였다.

메타데이터 데이터요소의 기술은 다음과 같다. 데이터 요소는 주요소와 관련하위요소가 있는데, 하위요소는 주요소의 구체적인 내용을 포함하고 있다. 다음으로 데이터요소의 설명(description)이 있으며 데이터타입(data type)은 비트, 텍스트, 숫자, 복합(varchar), 날짜 등이 있다. 필수여부(obligation)은 데이터요소의 값이 필수적으로 포함되어야 하는가 아닌가에 따라서 required/optional로 구분된다. 반복여부(multiplicity)는 두개 이상의 값을 가질 수 있는가 없는가를 알리는 것이다. 도메인(domain)은 해당 데이터 요소가 가질 수 있는 값은 제한하는 것이다. 마지막으로 스킴(encoding scheme)은 데이터 값의 문법적 명세(syntactic specification)으로 국제 혹은 국내의 표준에서 제안한 것을 채택을 우선하고 필요하면 개발한다.

3.3.2 범주별 데이터 요소

연구성과물 메타데이터는 다음과 같은 범주로 구분되고, 아래의 <표 3>과 같이 총 82개의 주요소, 47개의 하위요소로 구성되어 있다. 각 범주에 포함될 데이터 요소는 아래와 같은 기준으로 추출하였다. 첫째, 국제표준 기준에 맞추었다. 둘째, 기존의 코드체계 및 식별자를 활용하였다. 셋째, 정보시스템 개발을 염두에 두고 데이터 요소를 추출하였다. 마지막으로 추후 메타데이터를 XML 및 RDF로 표현을 할 것을 고려하였다.

<표 3> 메타데이터 데이터요소 구성

	주요소	하위요소
공통 메타데이터 요소		
식별기호범주	1	2
연구성과물 일반범주	14	21
개별 연구성과물 메타데이터 요소		
연구보고서범주	2	0
학술논문범주	8	0
각종발표회 및 전시회범주	7	5
간행물범주	3	0
학위논문범주	5	0
산업재산권등록범주	5	5
기술개발범주	5	4
연구주체 메타데이터 요소		
연구과제범주	11	8
연구기관범주	5	0
발주기관범주	5	0
연구원범주	11	2
총	82	47

각 범주를 구체적으로 살펴보면 다음과 같다. 연구성과물 메타데이터는 크게 3개의 그룹으로 구분되는데 “공통 메타데이터 요소”는 모든 유형의 연구성과물에 공통으로 적용되는 데이터 요소의 집합으로 연구성과물일반 범주에 포함된다.

식별기호범주에는 연구성과물을 식별할 유일한 식별자로 “연구성과물공개식별기호(resultPublicID)”와 각 조직에서 나름대로 사용하는 “연구성과물내부관리식별기호(resultInternalID)”로 구성된다. 즉 어떠한 성과물이라도 유일하게 인식할 수 있는 식별자와 기관별로 기존에 사용하던 식별자를 같이 사용할 수 있도록 하였다.

연구성과물일반범주는 연구성과물 자체에 대한 내용을 기술할 수 있게 하는 공통 데이터 요소이다. 기존의 Dublin Core의 15개 기본요소로 표현이 부족한 부분은 하위요소들로 확장하여 정의하였다. <부록 1>는 연구성과물일반범주의 메타데이터 요소들이다. 지면관계 상 본 연구에서는 하나의 범주의 메타데이터만을 보인다. <표 4>는 그 중에서 연구성과물일반범주에 포함된 연구성과물 작성자 데이터 요소에 대한 데이터 요소 정의이다.

<표 4> 데이터 요소 정의의 예

구분	내 용
주요 소명	연구성과물 작성자(dc: creator)
하위 요소	연구성과물 작성 참여인원수(numbersOfResearchers) : 연구성과물 작성에 참여한 인원수 [필수] / [스킴] : 없음
	연구성과물 책임작성자 이름(resultMainCreatorName) : 연구성과물 작성의 책임자 및 공동 책임자 이름으로 연구보고서, 학술논문, 학술대회 등의 경우 1st 저자가 이에 해당되며, 특허의 경우 발명자/고안자가 해당됨 [필수] / [스킴] : 없음
	연구성과물 책임작성자 소속(resultCreatorUnit) : 연구성과물 책임자의 소속기관 및 부서 [선택] / [스킴] : 없음
	연구성과물 참여작성자 이름(resultCoWorkersName) : 연구성과물 작성의 공동 참여자 이름 [선택] / [스킴] : 없음
	연구성과물 참여 작성자 소속(resultCoWorkersUnit) : 연구성과물 작성의 공동 참여자 소속기관 및 부서 [선택] / [스킴] : 없음

4. 메타데이터 표준의 활용

본 연구를 통하여 개발된 메타데이터 표준은 다음과 같이 다양하게 활용되어 질 수 있다.

4.1 메타데이터 레지스트리에 등록

메타데이터 레지스트는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지 관리하며, 공유되는 개념의 정형적 명세화를 통하여 메타데이터 세트 또는 메타데이터 요소간의 호환성을 유지시킨다[1]. 본 연구를 통하여 개발된 메타데이터 표준은 국가수준에서 다른 메타데이터 표준들과의 상호운용성이 확보되기 위해서는 메타데이터 레지스트리에 등록될 필요가 있다. 현재 한국과학기술정보연구원(KISTI)의 산업기술정보 메타데이터 레지스트리(IMR)와 한국정보문화진흥원(KADO)에서 추진하는 국가지식정보 메타데이터 레지스트리(NMR)에 등록 할 수 있다.

4.2 관련된 시스템 표준의 개발

본 표준은 개념적인 모델링에 기반한 메타데이터 모델이다. 따라서 이 모델이 정보시스템에 적용되기 위해서는 정보시스템의 유형에 따라 적절한 시스템 표준으로 전환되어야 한다. 일반적으로 메타데이터는 XML 기반의 언어로 저장된다. 이때는 본 표준은 XML Schema로 전환될 필요가 있다. 그리고 이것을 많은 기관/학문 분야에서 활용하려면 웹기반의 이름공간(namespace)으로 제공되어야 한다. 또한 메타데이터는 지역(local)시스템에서 관계형 데이터베이스에 1차적으로 저장된다. 따라서 본 모델은 RDB 기반의 논리적 데이터 모델로 전환하여 각 시스템의 DB설계의 기반을 제공할 수도 있다.

또한 본 메타데이터 표준의 데이터 요소 어휘를 기반으로 온톨로지를 구축할 수도 있다. 온톨로지 개발은 첫째, 연구성과물 용어의 분석 및 개념화하고, 둘째, 온톨로지 모델링을 실시하고, 셋째, 개념 추론을 위한 규칙을 개발하고, 넷째, 온톨로지를 위한 이름공간을 결정하고, 다섯째, 온톨로지의 형식론적 표현(formal description) 방법을 결정하고, 마지막으로 웹기반의 온톨로지를 구현한다. 차세대 웹 기술로 인정되는 시맨틱 웹 기술은 메타데이터를 기반으로 한 온톨로지와 지식표현(knowledge representation)이 핵심이다.

4.3 기관 및 분야별 활용

지금까지는 각 기관의 데이터베이스에 저장된 연구보고서, 저널에 발표된 논문, 프로젝트 정보 등의 연구성과 등의 정보들이 제한적으로 제공되던 정보서비스가 표준화된 메타데이터를 확보함으로써 여러 기관에 분산되고(distributed) 이질적(heterogeneous)인 시스템에 저장된 다양한 매체의 연구성과정보를 검색하고 원래의 리

소스에 접속을 가능하게 한다. 특히 국가연구정보시스템(NTIS)의 연구성과정보시스템에 적용이 가능할 것이며, 각 기관에서 체계적인 연구성과 정보관리시스템 개발에 활용될 수 있다.

만약 시맨틱 웹 기술을 기반으로 한 포털시스템이 구축된다면 기존의 분산형 데이터베이스 기술이 제공할 수 없었던 ‘지능적인(intelligent)’ 정보 서비스를 제공할 수 있을 것이다. 각 기관에서는 다음과 같은 순서로 시스템을 개발할 수 있다.

4.3.1 메타데이터 개념모델 개발

메타데이터 개념모델 개발 모듈은 정보공학의 데이터 모델링 기법을 적용할 수 있다. 즉 먼저 엔터티가 될 수 있는 연구성과물을 분석하고 포함되어야 하는 속성을 파악하는 것이다. 또한 연구성과물과 관련된 다른 엔터티들도 파악한다. 다음으로 각 엔터티간의 관계를 설정하고 마지막으로 메타데이터의 구조를 결정하는 것이다.

4.3.2 메타데이터 구현 모듈 개발

메타데이터 구현 모듈은 연구성과물의 운용성제고와 정보서비스를 고려하여 RDF와 OWL을 이용하여 메타데이터를 기계가 인식할 수 있는 언어의 형식으로 표현한다. OWL은 연구성과정보 클래스의 계층적 구조, 클래스와 클래스, 클래스와 요소, 요소와 요소간의 관계를 표현한다.

4.3.3 정보서비스 모듈

온톨로지를 이용하여 메타데이터를 생성하고 관리하는 모듈이다. 이 모듈은 각 연구기관의 연구성과정보시스템과 웹기반의 검색서비스의 기반 인프라가 된다. 각 연구기관 관점에서는 연구성과물에 관한 정보를 일관성 있고 체계적으로 관리할 수 있는 시스템 개발의 도구로 활용되어 질 수 있다. 구체적으로 연구기관별 자체 연구성과물 정보관리를 위한 표준구조로서의 역할을 하며, 연구성과물 측정 및 평가시스템용 DB 설계 시 참조될 수 있다. 국가적으로는 현재 추진 중에 있는 R&D 종합관리체제 개발에 있어서의 기반 인프라로서 활용되어 질 수 있다.

5. 결 론

현재 연구성과물에 대한 정보는 연구기관별로 서로 상이한 체제로 관리하고 있어서 연구성과물의 상호연계와 정보유통을 위한 상호 운용성이 매우 떨어지고 있다. 이와 같은 문제를 해결하기 위해서는 먼저 각 기관

별로 서로 다른 시스템하에서 관리되는 연구성과물의 메타데이터에 대한 표준화가 선행되어야 한다.

과학기술분야의 연구성과물 메타데이터 표준은 국가적으로는 현재 추진 중에 있는 R&D 종합관리체제 개발에 있어서의 기반 인프라로서 활용되어 질 것이며, 연구기관 차원에서는 연구성과물에 관한 정보를 일관성 있고 체계적으로 관리할 수 있는 시스템 개발의 도구로 활용되어 질 수 있다.

본 연구는 관리대상이 되는 연구성과물의 범위를 확장하고 이들을 통합한 표준 메타데이터를 처음으로 제시함으로써 국내 R&D 연구성과정보 관리의 기틀을 마련했다는 데 의의가 있다. 또한 Dublin Core와 CERIF 등과 같은 사실상의 국제 표준(de facto international standard)과의 호환성을 가짐으로 해서 국제적 교류도 가능하게 될 것이다.

향후 이 표준이 국가 R&D 종합관리체제 개발에 실제 사용될 수 있도록 하기위해서 국가에서 지정한 공동연구관리 규정을 준수하는 과제관리 정보에 대한 메타데이터 표준을 개발해야 한다. 그리고 개발된 과제정보와 이 연구에서 제시한 연구성과정보 메타데이터 표준을 모두 활용한 전체 연구관리 정보에 대한 메타데이터 표준을 제시하여 널리 쓰일수 있도록 노력을 기울여야 한다.

본 논문에서는 지면관계상 개발된 메타데이터의 일부분만을 소개하였으며, 도메인 모델링도 제한된 범위만을 소개하였다. 따라서 전체 메타데이터에 대하여는 한국과학기술정보연구원에서 편찬한 STISC(과학기술정보 표준위원회)의 '2005 과학기술정보표준'에서 파악이 가능하다.

참고문헌

- [1] 고영만; "메타데이터 표준화와 메타데이터 레지스트리", 국회도서관, 국회도서관보, 42(11) : 18-26, 2005.
- [2] 과학기술부; 국가 R&D 성과활용 관련 설문조사, 2005.
- [3] 과학기술통합검색; www.yeskisti.net, 2006.
- [4] 김미영; "회색문헌의 생산현황과 관리-부산대학교 내 37 연구소를 중심으로", 석사학위논문, 부산대학교, 2002.
- [5] 남영준; "디지털 시대의 회색문헌 이용 활성화에 관한 연구", 한국정보관리학회지, 19(4) : 234-255, 2002.
- [6] Almeida, Mario do G.G.; "Control Access for Grey Literature in Brazil: A Proposal," *In fourth International Conference on Grey Literature*, 1999.
- [7] Auger, C. P.; *Information Sources in Grey Literature*, 4th ed. London, 1996.
- [8] Chen, H. and Chen, C.; "Metadata Development for digital Libraries and Museums-Taiwan's Experience," DC-2001, NII, Tokyo, 2001.
- [9] CRIS; CERIF(Common European Research Information Format) <http://www.eurocris.org/en/taskgroups/cerif>, 2004.
- [10] Dublin Core Metadata Initiative; <http://dublincore.org>, 2005.
- [11] Jeffery; K.; "An Architecture for Grey Literature in a R&D," *In fourth International Conference on Grey Literature*, 1999.
- [12] Knapp, M., Dexter, S., and McLaughlin, R.; "Metadata Co-Development: A Process Resulting in Metadata about Technical Assistance to Educators," *Proceedings of the 13th International WWW Conference*, 2004.
- [13] Lightle, K.; "Using Metadata Standards to Support Interoperability," in CM Gynn & SR Acker (eds.), *Learning Objects : Contexts and Connections*, pp. 43-48, 2003.
- [14] Matthews, B., Wilson, M., and Kleese-van Dam, K., "Accessing the Outputs of Science Project," *Proceedings, 6th International Conference on CRIS*, 2002.
- [15] Miller, E.; "An Introduction to the Resource Description Framework," <http://www.dlib.org/dlib/may98/miller/05miller.html>, 1998.
- [16] Mooney S.; "Interoperability," *D-Lib Magazine*, Jan. 2001.
- [17] Steinacker A., Ghavam A., and Steinmetzm R.; "Metadata Standards for Web-Based Resource," *IEEE Multimedia*, 8(1) : 70-76, 2001.
- [18] Xie, R. and Shibaski R.; "Standardization Framework for CEOP Metadata Development and Application," *CEOP/IGWCO Joint Meeting*, University of Tokyo, 2005.

<부록 1> 연구성과물 일반범주 메타데이터

	주요소	하위요소	설명	데이터 타입(길이)	필수 여부	반복 여부	스킴
식별 기호 범주	연구성과물 식별기호 (dc: identifier)	연구성과물공개 식별기호 (resultPublicID)	연구성과물을 식별할 수 있는 유일한 식별자	char(50)	필수	불가	
		연구성과물 내부관리 식별기호 (resultInternalID)	해당조직의 내부에서 사용하는 연구성과물 식별자	char(50)	필수	불가	
연구 성과물 일반 범주	연구성과물 제목 (dc: title)	연구성과물국문제목 (resultKoreanTitle)	연구성과물의 한글 제목	char(100)	필수	불가	ISO639-2
		연구성과물 영문제목(resultEnglishTitle)	연구성과물의 영어 제목	char(100)	필수	불가	
	언어 (dc: language)		연구성과물 작성에 주로 사용된 언어	char(50)	선택	불가	
	연구성과물 분류 (dc: Type)		연구성과물의 종류	char(50)	필수	불가	
	날짜 (dc: date)	연구성과물작성일 (resultCreationDate)	연구성과물이 작성된 일자	date	선택	불가	DCMI-Period W3CDTF
		연구성과물 발표일 (resultPublicationDate)	연구성과물이 발표된 일자	date	선택	불가	
	연구성과물 작성자 (dc: creator)	연구성과물작성참여인원수 (numbersOfResearchers)	연구성과물 작성에 참여한 인원수	num	필수	불가	
		연구성과물 책임작성자 이름 (resultMainCreatorName)	연구성과물 작성의 책임자 및 공동책임자 이름	char(50)	필수	반복	
		연구성과물 책임작성자 소속 (resultCreatorUnit)	연구성과물 책임자의 소속기관(부서)	char(50)	필수	반복	
		연구성과물 참여작성자이름 (resultCoWorkersName)	연구성과물 작성의 공동 참여자 이름	char(50)	선택	반복	
		연구성과물 참여작성자 소속 (resultCoWorkersUnit)	연구성과물 작성의 공동 참여자 소속기관	char(50)	선택	반복	
	연구성과물 발행자 (dc: publisher)	연구성과물발행자이름 (resultPublisherName)	연구성과물의 발행자	char(50)	선택	반복	
		연구성과물 발행기관(resultPublishingUnit)	연구성과물의 발행기관	char(50)	선택	반복	
		연구성과물 발행국 (resultPublishingCountry)	연구성과물이 발행된 국가	char(50)	선택	불가	
	과학기술정보분류 (s&tAreaClassification)	계열(s&tSubjectArea)	주제 분류구조의 제일 첫 단계인 계열	char(50)	필수	불가	
		대주제(s&tsubjectMaster)	두 번째 단계인 대분류	char(50)	필수	불가	
		중분류(s&tsubjectDetail)	세 번째 단계인 중분류	char(70)	선택	불가	
	연구성과물 주제어 (dc: subject)		연구성과물의 주제와 색인어를 기술	char(200)	필수	반복	
	연구성과물 내용기술 (dc: description)	연구성과물 목차 (tableOfContents)	연구성과물의 주요 목차	char(200)	선택	불가	
		연구성과물요약기술 (abstracts)	연구성과물의 내용을 요약하여 기술	char(500)	필수	불가	
연구성과물 저장매체형태 (dc: format)		연구성과물의 저장형태		선택	반복		
연구성과물 연결 (dc: relation)	연구성과물접근및파일이름 (resultAccessInformation)	연구성과물이 저장된 위치경로 및 파일이름	char(50)	선택	반복		
	연구성과물관련정보URL (resultRelatedURL)	연구성과물에 관한 정보를 조회할 수 있는 홈 페이지 주소	char(100)	선택	반복	URI	
연구성과물 소장기관 (dc: source)		연구성과물의 소장기관	char(50)	선택	반복		
연구성과물 저작권 (dc: right)		연구성과물 저작권 정보 및 이용제한 정보를 기술	char(50)	필수	불가		
연구성과물 공개여부 (resultConfidentiality)		연구성과물의 공개여부	char(50)	선택	반복		