한국기록관리학회지, 18(1), 2018, pp. 101-127

연구데이터 관리를 위한 온톨로지 설계에 대한 연구*

A Study on Ontology Design for Research Data Management

박 옥 남(Ok Nam Park)**

목 차

〈초 록〉

연구데이터의 연구의 정확성이나 신뢰성 확보를 위한 정보적 가치, 연구의 재현 또는 검증, 재사용 가능성을 연구데이터에 대한 체계적 관리가 강조되고 있다. 표준 메타데이터는 연구데이터 생산, 관리, 구조화, 기탁된 데이터 추출에 핵심 역할을 수행할 것이다. 연구데이터는 연구, 연구데이터, 데이터셋, 파일 등 다양한 계층적 관계를 가지고 있으며, 인용 및 연구성과 등의 엔터티와 연계되어 있다. 이에 본 연구에서는 연구데이터 관리를 위한 온톨로지 모델을 제시하고자 한다. NTIS 사례를 제시하여 연구의 적용가능성을 제시하였다. 이를 위해 기존 연구데이터 관련 선행연구, 메타데이터 표준의 분석, 연구데이터 리포지터리 사례조사를 실시하였다.

주제어: 연구데이터, 온톨로지, 메타데이터, DDI, DataCite, DCAT, 데이터관리

〈ABSTRACT〉

The systematic management of research data is vital because it increases research data's value for research reproduction, verification, and reusability. Standard metadata will play a key role in research data registration, management, and data extraction. Research data has various structural relationships, such as research, research data, data sets, and files, and associated with entities such as citations and research results. The study proposes an ontology model for research data management. It also suggests the application of ontology to NTIS. Previous studies, metadata standard analyses, and research data repository case studies were conducted.

Keywords: research data, ontology, metadata, DDI, DataCite, DCAT, data management

* 본 연구는 2016년도 상명대학교 교내연구비를 지원받아 수행하였음
** 상명대학교 인문사회과학대학 문헌정보학과 부교수(ponda@smu.ac.kr)
■ 접수일: 2018년 1월 29일 ■ 최초심사일: 2018년 1월 31일 ■ 게재확정일: 2018년 2월 12일
■ 한국기록관리학회지 18(1), 101-127, 2018. 〈http://dx.doi.org/10.14404/JKSARM.2018.18.1.101〉

1. 서 론

1.1 연구의 필요성 및 목적

연구를 수행하는 과정에서 생산되는 데이터에 대해 최근 많은 관심이 주목되고 있다. 기존의 연구정보관리와 관련한 초점이 논문, 보고서, 특허와 같은 최종결과물에 초점을 두었다면 최근에는 연구수행 중 생성되거나 작성된 모든 데이터에 주목하고 있다. 많은 양의 데이터가 연구 수행 중에 생산되고 이용되는 과정에서 잘못 가공되고 조작되어 연구의 정확성이나 신뢰성을 저해하는 요인이 되기도 하며, 연구 수행과정에서 수집 또는 생산된 데이터는 연구의 재현 또는 검증을 위하여, 또는 향후 다른 연구의 데이터로 재활용되는 가치를 가지기 때문이다.

연구데이터는 연구자들이 연구 또는 연구 수행과정에서 연구의 부산물이자 결과물로서 생산되는 데이터로서 원 데이터와 분석된 2차 자료를 포함하며, 관찰, 조사, 실험, 경험 데이터를 의미한다(김은정, 남태우, 2012, 29). 연구데이터의 종류는 수치, 텍스트, 오디오, 컴퓨터 파일 등 다양하다. 이러한 이유로 최근 미국국립보건원(National Institute of Health, NIH), 국립과학재단(National Science Foundation, NSF), 호주 국가 연구비 지원기관 ARC(Australian Research Council) 등 연구관리기관은 최근 연구비 지원 단계에서 연구데이터 관리 계획 작성을 의무화하고 있다(심원식, 2015).

연구데이터의 가치 및 관심증가에 따라 최근 많은 연구기관에서 논문, 보고서, 특허와 같은 연구 성과물을 저장하고 관리하는 리포지터리가 아닌 연구데이터 등록 및 관리를 위한 리포지터리를 구축하고 있다. 2018년 1월 15일 기준, 연구데이터 레지스트리인 re3data.org에 등록된 리포지터리는 2,026개에 달한다. 또한 Dataverse, Zenodo, Dryad와 같은 연구데이터를 등록, 보존, 이용할 수 있는 서비스 플랫폼이 개발되고 있으며, 대학이나 연구소와 같은 기관은 물론 국가차원에서 연구데이터 서비스를 구축하여 관리되지 않는 데이터의 관리, 연결되지 않는 데이터의 연결, 재사용이 불가능한 데이터의 재사용, 검색되지 않는 데이터의 검색가능성을 지원하고 있다. 영국의 데이터 아카이브(UK Data Archive), 호주 연구데이터 서비스(Research Data Australia), 영국 데이터 큐레이션 센터(Data Curation Centre), 독일 연구데이터 리포지터리(Research Data Repository) 등이 대표적인 예이다. 이에 따라 각 분야별로 연구데이터를 효율적으로 관리하기 위한 메타데이터 표준을 개발하여 연구데이터의 체계적 관리 및 검색시스템을 지원하고 있다.

국내의 경우 역시 연구자들이 연구데이터 관리나 보존의 중요성은 인식하고 있으나 체계적으로 연구데이터를 관리할 수 있는 환경이 부족한 것으로 나타났다(김지현, 2012). 오픈 사이언스 시대를 위한 과학기술 연구지원을 위하여 오픈액세스와 연구성과 공개 플랫폼이 구축되어야 하나, 국내의 경우, 기관 리포지터리 정도만 준비되어 있고, 연구데이터 수집・관리・활용을 위한 준비는 미비한 상태이다(김순, 이보람, 김환민, 김혜선, 2017). 이를 위하여 다수의 연구에서 연구수행의 전 과정에서 생산되는 연구데이터의 등록・수집・관리・보존을 지원할 수 있

는 정보인프라의 개발이 필요함을 지적하고 있으며, 그 구성요소 중 하나가 연구데이터를 기술할 수 있는 메타데이터 표준의 개발이다(김은정, 남태우, 2012; 심원식, 안혜원, 변제연, 2015).

과학기술정통부는 최근에 들어서 연구데이터 등록 및 활용에 대한 관심을 기울이고 있다. 2018년 1월 18일 국가 연구개발 사업의 연구데이터를 관리하는 '국가연구데이터센터' 구축을 목표로 하는 '연구데이터 공유·활용 전략'을 발표하였으며, 연구데이터 관리 체계 구축을 위한 계층적 연구데이터 연계체계 마련 및 통합플랫폼 구축을 제시하였다(과학기술정보통신부, 2018.1.18). 이를 위해서 계층적 연구데이터의 지적구조를 분석하고, 계층별 연구데이터를 등록·관리할 수 있는 메타데이터의 표준화가 선결되어야 한다.

이에 본 연구는 과학기술정보통신부에서 운영하는 국가과학기술정보서비스에 적용할 수 있는 연구데이터 관리를 위한 온톨로지를 제시하는 것을 목적으로 한다. 연구데이터는 연구-데이터셋-파일 등 다양한 엔터티 사이의 지적구조에 대한 표현 및 데이터 유형, 데이터 포맷, 연구방법 등 다양한 엔터티와의 연결이 중요하다. 온톨로지 모델링은 디지털 자원의 다면성 및 풍부한 관계표현을 통해 지적구조를 체계적으로 제시하고 정보의 접근을 유용하게 한다(박희진, 박옥남, 2016). 이에 본 연구는 연구데이터 지적구조의 설계를 위하여 온톨로지를 적용하였다. 본 연구는 연구데이터와 관련한 선행연구를 조사하였으며, 연구데이터와 관련한 메타데이터 표준, 연구데이터 리포지터리 사례를 조사하였다. 본 연구에서 온톨로지 예시를 제시하여 적용가능성을 높였다.

1.2 선행연구

연구데이터의 가치 및 체계적 관리의 필요성에 대한 연구는 이미 많은 연구자에 의해 실시되었다. 연구데이트 관리 현황 및 체계의 필요성에 대해서는 김지현(2012)과 심원식(2015)의 연구를 살펴보았다. 김지현(2012)은 대학 내 연구자들의 연구데이터 관리 및 보존현황을 파악하기 위하여 설문을 실시하였다. 대부분의 응답자들은 개인 컴퓨터나 이동식 매체를 활용하여 연구데이터를 관리하고 있었으며, 연구데이터의 공유는 대부분 내부 및 연구데이터를 요청하는 연구자들에 제한되어 있는 경우가 대부분인 것으로 나타났으며, 체계적인 연구데이터 관리체계를 보유하지 못하는 대학이 대부분으로 나타났다. 연구자들의 연구데이터 관리 및 보존서비스에 대한 요구는 높게 나타나, 체계적인 연구데이터 모델이 필요함을 확인하였다.

심원식(2015)은 국내의 연구데이터 관리를 위해서 국가주도의 연구데이터 관리체계 구축이 필요함을 주장하였다. 국외의 경우, 연구데이터에 대한 체계적인 관리와 활용을 위한 인프라가 구축되어 있음에도 불구하고 국내의 경우는 이러한 인프라가 주요 대형 대학을 제외하고는 미비하여 연구데이터가 유실될 위험이 있다. 실제 개별 대학이 연구데이터의 관리 및 보존을 위한 자체 인프라를 구축하는 것이 용이하지 않으므로 국가주도의 연구데이터 관리 및 아카이빙 시스템을 구축하고, 연구데이터 생산기관과 연계할 것을 제안하였다.

연구데이터의 체계적 관리를 위해서 메타데이터와 같은 표준의 개발을 제시한 연구가 있는데, 강희종(2012)은 국가과학데이터 활용을

위하여 정책기반, 공유문화확산, 전문인력양성, 전문기업 육성 등이 필요하며, 이와 함께 과학데이터 활용기반을 구축해야 함을 강조하였다. 또한 국내의 경우, 국가차원의 과학데이터에 대한 이해부족, 국가차원의 과학데이터 활용기반 구축 및 활용을 위한 전략이 부재함을 지적하였으며, 이를 위하여 표준 및 관련 기술개발이 필요함을 제시하였다.

김은정과 남태우(2012)는 연구데이터 수집에 영향을 미치는 요인 및 활성화 방안을 제시하기 위한 설문조사를 실시하였다. 연구데이터의 동기부여 요인 강화, 장애요인 제거, 정보인프라 요인을 강화시키는 것이 필요한데, 이 중 정보인프라요인으로 연구데이터 저장 및 이용을 지원하는 관리시스템, 메타데이터 등이 영향을 미치는 것으로 나타났다. 연구의 시사점으로 정보인프라 요인으로 표준 메타데이터 개발을 제시하였는데, 표준 메타데이터의 개발은 연구데이터 구조를 명시화하고 효율적으로 관리함으로써 연구데이터의 향후 활용가능성을 높이고 지식교류 활성화, 상호운용성 고려에 있어서 중요한 요소임을 주장하였다. 이와 함께 표준 메타데이터에 출처, 목적, 시간, 지역적 장소, 생산자, 이용조건, 용어 등의 요소가 포함되어야 함을 제시하였다.

김지현(2016)은 37개 연구데이터 리포지터리에서 데이터의 접근 및 이용을 통제하는 정책요소들을 조사하였으며, 공통적으로 정책에 포함되어 있는 조항으로 저작권 및 라이선스 규정, 접근통제영역, 데이터 인용, 면책조항 및 엠바고 적용 등이 포함되어 있음을 파악하였다.

이러한 선행연구를 고려할 때, 연구데이터 수집 및 체계적 관리의 중요성에도 불구하고, 국내의 경우 연구데이터 수집・관리・활용을 위한 준비는 부족한 것을 알 수 있다. 국내의 경우 연구성과물을 등록하기 위한 기관 리포지터리 정도만 준비되어 있고, 실제 연구데이터 수집 및 관리를 위한 리포지터리는 미비함을 알 수 있으며, 다양한 요소 중에 연구데이터 관리 및 활용인프라를 갖추는 것과 이에 대한 출발점으로 메타데이터 표준의 제시가 필요함을 알 수 있다.

2. 국가과학기술정보서비스

과학기술정보통신부는 국가과학기술정보서비스(National Science & Technoogy Information Service, NTIS)를 통해 국가 R&D 성과물등록포털을 운영하고 연구자가 국가 R&D 연구과제 수행 후 발생한 9종의 연구성과물을 등록・기탁하기 위한 등록시스템을 제공하고 있다. 연구성과물 등록시스템의 근거는 「국가연구개발사업의 관리 등에 관한 규정 제25조」로, 본 규정은 연구자로 하여금 연구 성과를 지정한 기관에 등록하거나 기탁할 것과, 연구기관은 국가과학기술정보시스템과 연계하여 연구성과와 관련된 정보의 관리・유통체계를 구축 운영하도록 명시하고 있다. NTIS에서 수집・관리하는 연구성과물은 논문, 보고서 원문, 소프트웨어, 특허, 생명정보, 연구시설장비, 생물자원, 기술요약정보, 화합물, 신품종(정보), 신품종(실물)이다. 연구자가 기탁한 연구성과물은 연구성과물 관리 유통전담기관이 등록한 정보와 연계하여 과제별 성과물을 조회할 수 있도록 하고 있다. 정보연계 기관은 한국과학기술정보연구

원, 한국지식재산전략원, 한국생명공학연구원 등이다.

NTIS 시스템은 과제, 수행 기관, 사업 수행 연구자, 연구성과물의 범주에 따라 브라우징 및 탐색이 가능하도록 하고 있는데, 〈그림 1〉과 같이 각 과제는 사업명, 과제명, 과제고유번호, 과제수행기관명, - 연구책임자명, 키워드, 연구요약, 기간, 과학기술표준분류, 6T에 대한 정보 기술은 물론 관련 연구성과물에 대한 정보를 제공하고 있다.

이와 같이 NTIS를 통해 과제에 대한 상세정보, 동일 연구 책임자 수행과제 내역, 성과물 정보는 조회할 수 있으나, 연구데이터를 등록할 수 있는 시스템은 제공되고 있지 않으므로 연구데이터의 인용, 재사용성, 검증의 도구로는 활용되기 어렵다.


사업
2012 / 교육과학기술부 / 일반사업
한국원자력연구원연구운영비지원 (조사분석사업명 : 한국원자력연구원연구운영비지원)
* 본 과제에 참여한 연구자
연구책임자 이주운
과제
RT 핵심 연구시설 운영
1345194031 / 한국원자력연구원 / 주관과제 / 총 연구비 3,431.00 백만원
과학기술표준 분류 1 : 원자력 / 방사선기술 / 80%
기본정보 과제연구요원 요약문 참여연구원
과제고유번호 1345194031 당해연도 연구기간 2012-12-01 ~ 2012-12-31
(기관)세부과제번호 523200-12 총연구기간 2012-01-01 ~ 2014-12-31
대과제명 방사선 융복합 신산업 클러스터 장비
과제명 국문 RT 핵심 연구시설 운영
과제명 영문
과제수행기관 한국원자력연구원
연구관리전문기관 기초기술연구회 과제관리(전문)기관 한국원자력연구원
과제진행상태 종료 실용화대상여부 실용화비대상
연구개발단계 기타 연구수행주체 출연연구소
세부과제성격 연구관리 연구개발성격 기타개발
기술수명주기 기타 지역 전라북도
성과
* 본 과제의 성과정보
성과(물)정보 조사분석확정 성과현황
논문(6) | 특허(1) | 기술료(1)

No	논문명 / 성과년도	저자명	SCI(E) 구분
1	저준위 감마선 조사장치 조사현황 고찰('09~'11) / 2012	강태진	비SCI
2	고준위 감마선조사장치 조사현황 고찰 ('09~'11) / 2012	강태진	비SCI

〈그림 1〉 NTIS 과제상세정보

과학기술정보통신부는 최근에 연구데이터 등록 및 활용에 대한 관심을 기울이고 있다. 2018년 1월 18일 국가 연구개발 사업의 연구데이터를 관리하는 '국가연구데이터센터' 구축을 목표로 하는 '연구데이터 공유・활용 전략'을 발표하였다(과학기술정보통신부, 2018.1.18.). 이에 대한 전략의 일환으로 연구데이터 관리 체계 구축을 제시하고, 계층적 연구데이터 연계체계마련 및 원스탑으로 검색하고 활용할 수 있는 통합 플랫폼 구축을 목표로 제시하였다. 이를 위해서는 계층적 연구데이터의 지적구조를 분석하고, 계층별 연구데이터를 등록・관리할 수 있는 메타데이터의 표준화가 선결되어야 한다.

NTIS가 다양한 연구기관과 연계하여 연구성과물을 등록하는 포털로 역할을 수행하고 있다는 점과, 연구데이터 통합 플랫폼 구축을 목적으로 한다는 점을 고려할 때, NTIS의 기존 모델을 연구데이터로 수용할 수 있는 모델로의 확장이 요구되며, 이를 위한 메타데이터의 개선 역시 요구된다.

3. 연구데이터 메타데이터 표준

연구데이터 메타데이터 표준을 비교하기 위하여 Dublin Core, DDI, DataCite, DCAT를 비교하였다. re3data.org는 연구데이터 리포지터리를 등록하는 레지스트리로 2018년 1월 15일 기준 2,026개의 리포지터리가 등록되어 있다. re3data.org에 등록된 연구데이터 리포지터리에 사용된 메타데이터 표준을 살펴보면 더블린 코어가 180개, DDI가 117개, DataCite 메타데이터 스키마가 79개로 가장 많은 비중을 차지하고 있으므로 이에 대한 분석이 필요하다. 또한 DCAT는 더블린코어를 기반으로 확장을 한 메타데이터 스키마이며, DDI나 DataCite과는 달리 RDF적용을 제시하였다는 점에서 살펴볼 가치가 있다.

3.1 Dublin Core

더블린코어는 도메인에 상관없이 일반적으로 적용될 수 있는 요소의 집합으로 이루어진 메타데이터 표준이다. Dublin Core Metadata Initiative에 의해 연구되었으며, 2009년 ISO 15836 표준으로 제정되었으며, re3data.org에 등록된 리포지터리에서 가장 많이 사용된 메타데이터 표준이다. 더블린코어는 인도네시아 USU Institutional Repository(USU-IR)는 물론, 독일의 국가 고고학 및 고대연구 연구데이터센터인 IANUS Datenportal, 에딘버러 대학 데이터센터, 미국 텍사스 데이터 리포지터리 등 180개 연구데이터 리포지터리에 의해 메타데이터 표준으로 사용되고 있으며, 호주 정부데이터 기술을 위하여 고안된 AGLS 메타데이터 응용프로파일, 과학 및 의학 분야 데이터를 기술하기 위하여 고안된 Dyrad 메타데이터 응용프로파일 등 다양한 리포지터리에서 더불린 코어를 기반으로 응용 프로파일을 개발하여 적용하고 있다.

더블린코어는 제목, 출판사, 날짜, 주제어 등의 메타데이터 요소는 물론, 한정어를 사용하여 요소세목과 인코딩 스킴을 기반으로 다양한 매체 자원의 기술을 가능하도록 하고 있으며, XML(eXtensible Markup Language)와 RDF(Resource Description Framework)을 사용한

레코드 구현이 가능하다. 데이터 형식이나 구조가 단순하여 다양한 도메인의 메타데이터 개발에 적용되고 있다.

3.2 DDI

DDI(Data Documentation Initiative)는 데이터 기록화 관련 국제기관으로 2003년 설립되었다. 현재 미국, 영국, 캐나다 등 여러 국가의 통계기관 관리자 및 실무자가 워킹 그룹을 결성, 다양한 활동을 수행하고 있다. DDI는 통계 자료의 처리, 관리, 보존, 이용에 초점을 맞춘 메타데이터 표준인 'DDI 메타데이터'를 개발하였다.

DDI 메타데이터는 DDC Codebook(DDI-C)과 DDI Lifecycle(DDI-L) 두 가지 버전으로 발표되었는데, DDC Codebook은 초장기에 사회과학 연구에서 주로 생성되는 통계자료의 기술 및 관리를 위해 개발되었으므로 사회과학분야에서 많이 사용되었으나, 현재는 사회과학은 물론, 경제학, 행동학, 보건학 등에서 다양한 분야의 데이터 기술에 광범위하게 사용된다. 두 버전 모두 XML 기반이며 DDI Alliance에 의해 개발되고 유지 관리된다.

DDI-C(DDI Version 2)는 DDI의 간단 버전으로 간단한 조사데이터의 교환 및 보존 문서화를 위한 표준이다. 조사 및 통계 데이터의 해석 도구로 특정 조사에 대한 설명과 변수명 및 변수값, 변수 코드화 방법 등을 수록하는 코드북의 내용 기술을 포함하고 있다. 가장 최신버전은 2.5이다.

DDI Lifecycle(DDI 버전 3)은 DDI 3.0 이후 버전으로 데이터 개념화, 발행, 재사용에 이르는 처리, 접근 및 이용, 보존, 재구성 과정을 포괄하는 데이터 데이터의 생애주기를 반영하여 데이터 세트를 문서화하기 위하여 설계되었다. 모듈화나 확장이 가능하며 가장 최신의 버전은 DDI Lifecycle 3.2버전으로 2014년 3월에 게시되었다.

DDI 사용 예는 영국 데이터 아카이브(UK Data Archives), 사회 및 정치학 연구 데이터 아카이브(The Institution for Social and Policy Studies Data Archive), CESSDA 등이 있다. UK 데이터 아카이브는 사회과학 및 행동과학 분야 디지털 데이터 콜렉션의 리포지터리이며, ISPS는 DDI와 더블린코어를 기반으로 기관에서 생산한 학술정보 및 관련 데이터의 아카이브이다.

DDI Lifecycle의 복잡한 모듈화로 인해 DDI-C가 더 많이 사용되는데, DDI-C의 기본구조는 다음과 같다. DDI는 문서단계(docDscr), 연구단계(stdyDscrType), 파일단계(fileDscrType), 변수를 기술하기 위한 데이터 단계(dataDscr), 기타자료(otherMat)를 기본 엔터티로 범주화하고 있다.

문서단계(docDscr요소)는 서지정보를 기술하는 섹션으로 DDI 레코드 작성을 책임지고 있는 사람, 기관이 작성하여야 하며, 서지정보 작성일, 책임자 등의 내용을 기술한다. 6개의 하위요소 및 8개 속성으로 기술되며, 하위요소는 인용정보, 사용된 통제어휘, 가이드, 주기, docSrc, docStatus 등이 사용된다. 보통 문서의 헤더부분에 포함된다.

연구단계(stdyDscrType요소)는 데이터 세트 및 연구에 관한 정보를 기술하기 위한 섹션으로 8개 하위요소 및 9개 속성으로 구성된다. 연구 단계에서 기술해야 하는 정보는 연구가

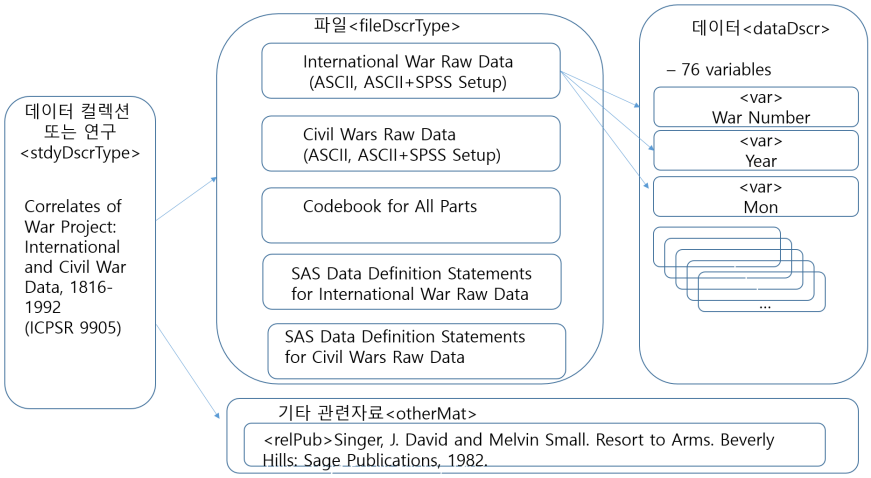
어떻게 인용되어야 하는지, 수집 및 배포 담당자, 데이터 내용에 대한 키워드, 요약, 데이터 수집 방법 및 처리 방법 등을 포함한다. 하위요소는 인용(citation), 데이터접근(DataAccs), 방법론(method), 주기(notes), 다른 연구관련 자료(otherStdyMat), 연구정보(StdyInfo), 연구권한(studyAuthorization), 연구개발(study Development) 등의 요소를 포함하며 연구단계의 세부요소 중 상당부분은 문서단계 세부요소와 중복된다.

파일 단계(fileDscrType요소)는 데이터 세트를 구성하는 데이터 파일에 대한 정보를 기술하며 다수의 파일에 대한 정보가 기술될 수 있다. 총 3개 요소 및 13개 속성으로 기술된다. 파일의 요약정보, 방법론, 다른 발행이나 인용정보, 접근정보를 포함한다.

데이터 기술 섹션(dataDscr)은 총 5개 하위요소, 8개 속성으로 기술되는데, 변수(variables)에 대한 정보를 기술하기 위하여 사용된다.

기타 관련자료(otherMat)는 연구와 관련된 다른 자료에 대한 정보를 기술하며 6개 하위요소 및 11개 속성을 포함한다. 기타 자료에 해당되는 예는 연구데이터를 활용한 다른 연구, 연구에 활용된 질문지, 코딩노트, SPSS/SAS/Stata 파일, 이용자 매뉴얼, 가이드, 샘플 컴퓨터 소프트웨어 프로그램, 용어집, 면담순서, 데이터베이스 스키마, 데이터 사전, 코딩정보, 면담스케줄 등이 대표적이다.

DDI는 통제어휘를 사용하여 요소에 허용되는 값을 지정하고 있는데, 포맷, 데이터유형, 분석, 이벤트 유형, 도구, 주제, 범위 등을 지정하기 위하여 통제어휘사용을 제시하고 있다. 자료유형의 예는 오디오, 비디오, 텍스트, 소프트웨어 등이며, 데이터 유형은 시간, 날짜, 정수, 불린 등이며, 분석단위는 개인, 조직, 패밀리 등이다. 라이프사이클 이벤트 유형은 연구제안, 파일럿 스터디, 데이터 수집, 파이널 리포트, 보존, 데이터 분석 보고서 등이다. 또한 연구 및 파일에 각각 DOI기반 식별자를 부여하고 있는 것이 특징이다.



〈그림 2〉 DDI 엔터티 구조

예시와 같이 “Correlates of War Project” 연구는 다수의 데이터 파일로 구성된 데이터셋을 가지고 있으며, 각 파일은 변수 및 변수에 대응되는 값을 가지며 이는 〈var〉 요소를 사용하여 기술된다. 또한 이 데이터세트를 사용하여 산출된 자료는 관련 출판정보 〈relPub〉 요소를 사용하여 기술된다.

3.3 DataCite

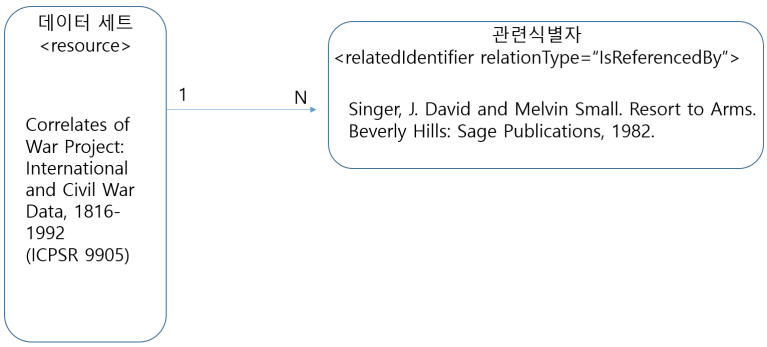
DataCite은 데이터셋의 인용 및 검색을 위한 메타데이터 요소를 정의하는 것을 목적으로 하며, DataCite 메타데이터 스키마에서 데이터셋은 가장 넓은 의미로, 숫자 데이터뿐만 아니라, 다양한 연구데이터 객체를 포함하는 개념으로 사용하고 있다.

학술연구는 많은 디지털 연구데이터를 생산하고 있으며, 연구결과 검증 및 공유, 새로운 연구의 발견을 검증하기 위해서 무엇보다 데이터가 가장 중요하다. 이를 위하여 데이터세트에 대한 접근, 식별, 공유 및 재사용을 위한 영속적인 접근이 중요한데, 이에 대한 부분은 아직 많이 부족하다. 이러한 요구에 부응하기 위하여 DataCite 국제 컨소시엄이 2009년 말에 설립되었는데 3가지 근본목표 - 1) 온라인을 통한 과학연구데이터에 대한 보다 쉬운 접근, 2) 과학에 대한 합법적이며 인용 가능한 연구 자료로서 연구데이터 수용 증진, 3) 향후 결과 검증 및 재사용을 지원하는 데이터 보존 지원 - 를 지향하고 있다.

DataCite의 이점은 도메인에 상관없이 적용 가능하므로 과학, 사회과학, 인문학 등 다양한 분야에 적용할 수 있다는 것이다. 현재 OpenAIRE, 디지털 음악 연구데이터 리포지터리(The Centre for Digital Music Research Data Repository) 등에서 사용되고 있다. DataCite 서비스의 특징은 영구 식별자의 개념인데, 영구식별자는 문자열과 파일, 파일의 일부, 사람, 조직 등을 포함하는 객체 사이의 연계, 데이터셋의 인용 촉진 등을 위하여 영구식별자로 DOI를 사용할 것을 제시하고 있다.

현재 DataCite 메타데이터 스키마의 최근 버전은 4.1이며 2017년 10월에 발행되었다. DataCite은 필수요소로 식별자(Identifier), 생산자(Creator), 제목(Title), 발행자(Publisher), 발행년(PublicationYear)을 제시하고 있으며, 주제(Subject), 기여자(Contributor), 날짜(Date), 언어(Language), 자원유형(ResourceType), 대체식별자(AlternateIdentifier), 관련식별자(RelatedIdentifier), 크기(Size), 포맷(Format), 버전(Version), 권한(Rights), 기술(Description)을 선택요소로 제시하고 있다. 또한 이와 함께 2개의 관리 메타데이터 요소를 포함하고 있는데, 최신 메타데이터 업데이트(LastMetadataUpdate), 메타데이터버전(MetadataVersionNumber)이 예시이다.

〈그림 3〉과 같이 DataCite 메타데이터 스키마는 데이터 세트 자체를 인용하기 위한 것을 목적으로 하므로, DDI에서 제시하는 파일(File)이나 변수(Variables) 단계의 엔터티에 대한 기술요소는 제공하지 않는다. 특징은 식별자 종류가 많아서 기본으로 DataCite에 등록할 때 부여되는 기본 식별자인 DOI 외에 대체식별자(alternative Identifier) 등을 허용하여 식별 가능성을 높였다. 관계를 정의하기 위한 참조(is referenced by/references), 인용(is cited by/cites)



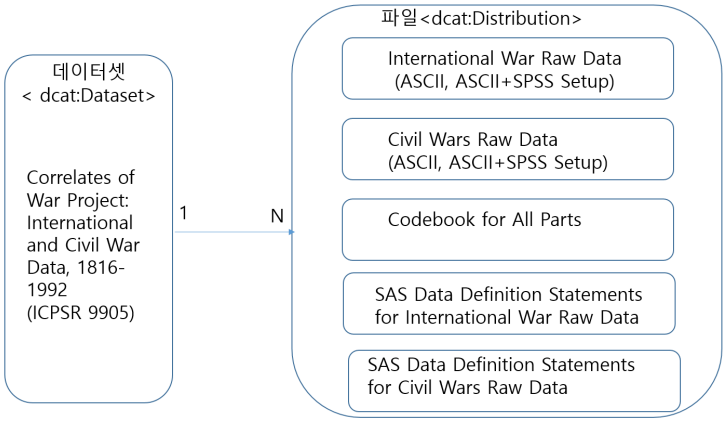
〈그림 3〉 DataCite 엔터티 구조

등의 요소를 통해 데이터 셋과 관련 있는 자료, 데이터 셋을 인용한 자료 등을 명시할 수 있도록 하였다.

3.4 DCAT

Data Catalog Vocabulary(이하 DCAT)는 DCAT를 사용하여 데이터 목록에서 데이터셋을 기술함으로써, 데이터셋의 발행자가 다수의 목록시스템에서 메타데이터를 통해 쉽게 데이터셋을 검색하고 응용할 수 있도록 하는 것을 목적으로 한다. 이 표준은 DERI에 의해 처음 개발되고 출판되었으며 많은 기관에 의해 채택되고 있다. 이 표준은 W3C eGov Interest Group에 의해 추가로 개편되었으며 W3C 정부 연계 데이터(GLD) 워킹 그룹에 의해 추가 작업이 이루어져, 현재 최종 표준 이전 상태인 후보 권고(candidate recommended)상태인 표준이다. DCAT를 적용하고 있는 리포지터리는 유럽연합 오픈데이터 포털(European Union Open Data Portal), 캐나다 공공정부(Open Governemtn Canada-Open Data), 에신 연구데이터 파인더(Etsin Research Data Finder)등이 있다. DDI나 DataCite이 XML기반의 메타데이터 스키마 및 메타데이터 레코드를 작성하고 있는 것과 대조적으로 DCAT는 RDF를 기반으로 하는 것이 특징이며, 기본적으로 기존 표준인 더블린코어, SKOS, FOAF를 활용하고 나머지 요소만 확장하였다.

DCAT의 기본 엔터티는 데이터셋(dcat:Datasets), 데이터셋의 접근 가능한 형태를 기술하기 위한 배포(dcat:Distribution), 목록에서 데이터셋 엔터티를 기술하기 위한 목록레코드(dcat:CatalogRecord)로 구성된다. 배포 클래스는 접근 가능한 형태, 즉 다운로드 가능한 파일이나 RSS Feed를 지칭하는 것으로 데이터셋의 파일에 대한 기술이 배포 클래스에서 이루어진다. 관리메타데이터에 대한 부분은 주로 목록 레코드 클래스(dcat:CatalogRecord)에서 이루어지며, 기술에 대한 부분은 데이터셋 클래스(dcat:Dataset), 구조 메타데이터에 대한 부분은 배포 클래스(dcat:Distribution)와의 관계에서 기술된다. 데이터셋 클래스에서 사용되는 속성은 제목(dct:title), 기술(dct:description), 발행일(dct:issued), 수정일(dct:modified), 식별자(dct:identifier), 주제어(dcat:keyword), 연



〈그림 4〉 DCAT 엔터티 구조

〈표 1〉 연구데이터 메타데이터 비교

표준	Dublin Core	DataCite	DDI	DCAT
모델링	자원	DataSet단계 기술	문서단계(docDscr), 연구 및 데이터셋 (stdyDscrType), 파일단계(fileDscrType), 변수(dataDscr), 기타자료(otherMat)	데이터셋(Datasets), 배포(Distribution), 목록레코드(CatalogRecord)
요소	2단계 요소 및 요소세목	3단계 요소 및 요소세목	5단계 요소 및 요소세목	2단계 요소 및 요소세목
통제어휘	자원유형(Type), 포맷(Format), 주제(Subject), 언어(Laugnage)에 통제어휘적용	기여자유형(contributorType), 날짜유형(dateType), 자원유형(resoureTypeGeneral), 관계유형(relatedIdentifierType), 분석(analysis), 이벤트유형(EventType), 데이터유형(DataType) 등	포맷(General Format), 데이터 유형(Data Type), 라이프사이클이벤트 (Lifecycle Event Type), 도구(Type of Instruments), 주제(Subject) 등	미디어유형 및 포맷유형 (MediaTypeOrExtent)
사례	- British Oceanographic Data Centre Published Data Library - Edinburgh DataShare - UK Archeology Data Service - USA Institutional Repository - IANUS Datenportal 등	- OpenAIRE - The Centre for Digital Music Research Data Repository - TCCON Data Archive - GFZ Data Services	- UK Data Archive - The Institution for Social and Policy Studies Data Archive - CESSDA - GFZ Data Services 등	- European Union Open Data Portal - Open Governemtn Canada-Open Data - Etsin Research Data Finder
포맷	XML, RDF/XML 등 사용가능	XML기반	XML기반	RDF/XML
식별자	자원에 속성을 사용하여 식별자 스킴 지정	데이터셋에 DOI부여 및 대체식별자 허용	연구(데이터셋)과 파일에 DOI부여	
엔터티 관계	relation	참조(is referenced by/references), 인용(is cited by/cites) 등 relationType 등 활용	데이터셋-파일-변수-관련자료	목록-데이터셋-파일 등을 객체속성이용 관계명시
관리메타 데이터	기여자(Contributor), 생산자(Creator), 권리(Right), 발행자(Publisher)	최신 메타데이터 업데이트(LastMetadataUpdate), 메타데이터버전(MetadataVersion Number)	문서단계(docDscr)	목록레코드(CatalogRecord)

락처(dcat:contactPoint), 언어(dct:language), 시간적 범위(dct:temporal), 공간적 범위(dct:spatial) 등이며, 배포 클래스에서 사용되는 속성은 제목(dct:title), 기술(dct:description), 발행일(dct:issued), 수정일(dct:modified), 라이선스(dct:license), 권한(dct:rights), 접근가능한 URL(dcat:accessURL), 다운로드 가능한 URL(dcat:downloadURL), 자원유형(dcat:mediaType), 포맷(dct:format), 바이트 사이즈(dcat:byteSize)이다.

DDI, DataCite과의 차이점은 관련 자원, 즉 보통 연구성과물이나 데이터셋의 인용정보와의 연결을 기술할 수 있는 요소가 포함되어 있지 않는 것이며, 배포(dcat:Distribution)에 대한 속성이 11개 요소로 각 파일에 대한 상세한 기술이 가능하도록 하고 있는 것이 특징이다.

4. 연구데이터 리포지터리 사례

연구데이터 메타데이터 표준의 적용 및 등록, 검색서비스를 살펴보기 위하여 연구데이터 리포지터리 사례를 조사하였다. 이를 위하여 ICPSR 데이터 아카이브, Harvard Dataverse, Dryad 리포지터리를 살펴보았다. ICPSR은 미국 국가 연구데이터 아카이브이며 DDI표준을 적용하고 있는 사례이며, Dataverse는 DDI표준을 기반으로 하며 오픈 소스 플랫폼으로 연구데이터를 통제, 식별, 이용, 장기보존을 포함한 웹 기반 데이터 서비스로, re3data.org 등록된 리포지터리 중 47개가 Dataverse 서비스를 사용하고 있다. Dryad는 DataCite과 더블린코어 기반의 리포지터리 서비스로 NSF는 물론 총 3개의 리포지터리에서 사용되고 있다.

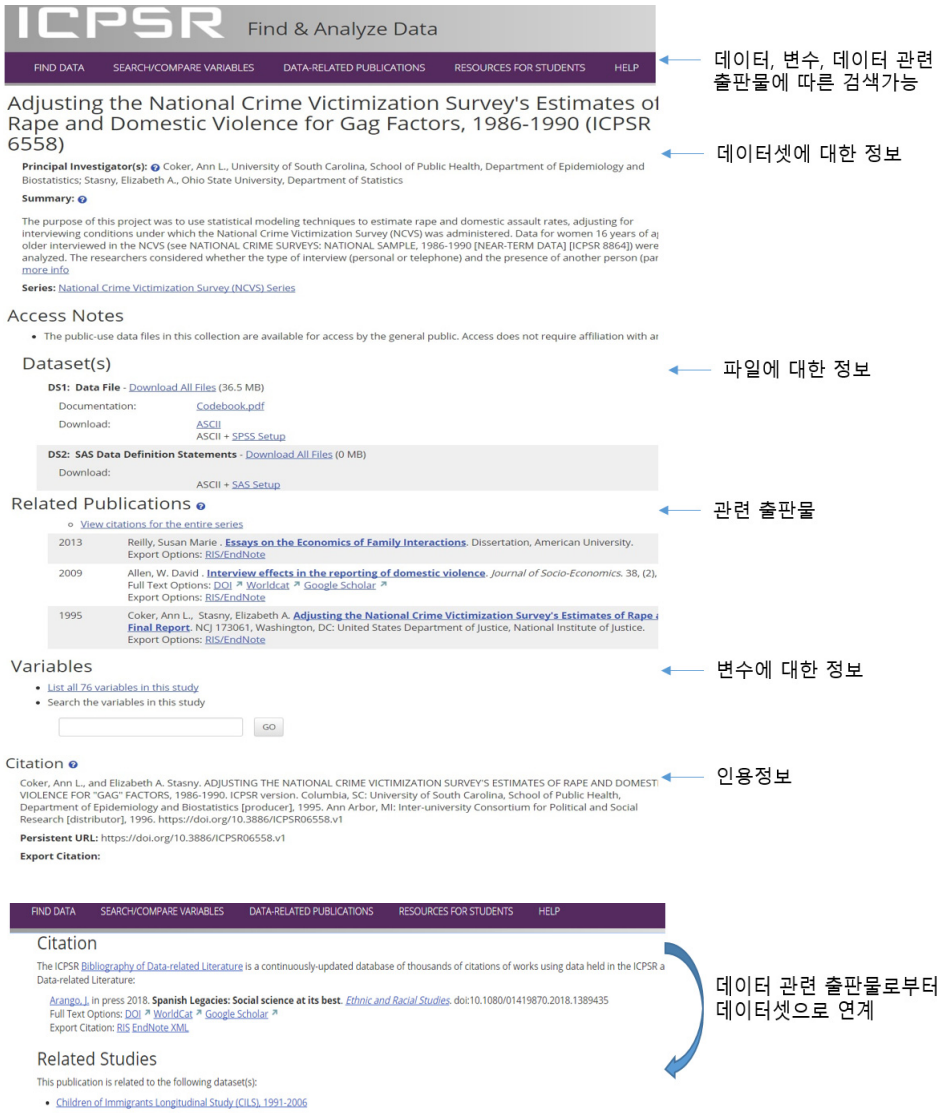
4.1 ICPSR Data Archive

ICPSR 데이터 아카이브(Inter-Consortium for Political and Social Research Data Archive)는 미국 사회과학, 인문학, 생명과학, 교육, 정치학 등 분야의 연구데이터를 수집, 관리하기 위한 리포지터리이다. 700개 이상의 학술기관 및 연구기관의 국제 컨소시엄으로 운영되고 있으며, 사회과학 분야의 10,400개 이상의 데이터셋, 500만개의 변수, 7만종 이상의 데이터 관련 출판물 등에 대한 데이터 아카이브를 유지하고 있으며, 교육, 노령화, 범죄, 약물남용, 테러리즘 등의 21개 컬렉션에 대한 데이터 수집을 제공한다.

ICPSR은 데이터 검색, 변수 검색 및 비교, 데이터 관련 출판물 등을 검색할 수 있는데, 데이터는 주제, 지역, 데이터 포맷, 분석유형, 연구시기, 제한, 데이터접근성, 업데이트 여부 등에 따라 검색이 가능하며, 데이터 포맷은 SPSS, SAS, Stata, Delimited, R, 온라인 분석으로 구분하고 있으며, 분석유형은 양적, 질적, GIS로 구분하여 접근가능하다. 데이터유형은 관리데이터, 센서스, 이벤트, 실험, 기계 가독 텍스트, 관찰, 서베이 등으로 구분하고 있다.

ICPSR은 DDI표준을 기반으로 메타데이터 레코드를 기술하는데, 〈그림 5〉와 같이 데이터셋에 대한 상세 설명 및 각 데이터 파일에 대한 기술 및 링크, 데이터셋에 사용된 변수 및 데이터(Variables), 관련 출판물에 대한 정보(Related Publications), 메타데이터 추출 및 인용정보(Citation)를 제공하고 있다. 각 Datasets은 DOI

기반의 영구식별자를 부여받고 있으며, Datasets의 활용 및 이해를 돕기 위하여 코드북(Codebook) 등의 문서(documentation)를 함께 제공하고 있다. 서지정보에서 주제어, 지역에 따른 브라우징을 활용한 정보탐색이 가능하며, 데이터 관련 출판물 검색을 통해서 출판물에서 사용한 데이터셋으로의 탐색 역시 가능한데, 이는 연구성과물을 출판 할 시에, 데이터셋의 인용정보를 기록하였기 때문에 상호참조가 가능한 것이다.



〈그림 5〉 ICPSR 연구데이터 정보

4.2 Harvard Dataverse

데이터에 특화된 리포지토리에 대한 필요가 높아짐에 따라 기존의 논문, 보고서, 발표자료가 아닌 전적으로 데이터만을 취급하는 기관 데이터 리포지토리 시스템도 개발, 배포 중이다. 대표적인 사례가 Dataverse로 하버드대학 무료로 배포하고 있는 오픈 소스 플랫폼으로 연구데이터의 통제, 식별, 이용, 장기보존을 포함한 웹 기반 데이터 서비스를 가능하게 한다(김지현, 2016). Dataverse의 특징은 데이터셋에 영구식별자를 부여하고 데이터의 서브셋에도 식별자를 부여하여 출판물이 데이터셋이나 데이터의 서브셋과 인용정보를 통해 연결될 수 있도록 하고 있는 것이다. 또한 데이터 인용 시에 저자, 발행일, 표지, 영구식별자 등의 필수요소를 포함한 인용형식을 규정하고 있다. 데이터셋의 인용예시는 다음과 같다.

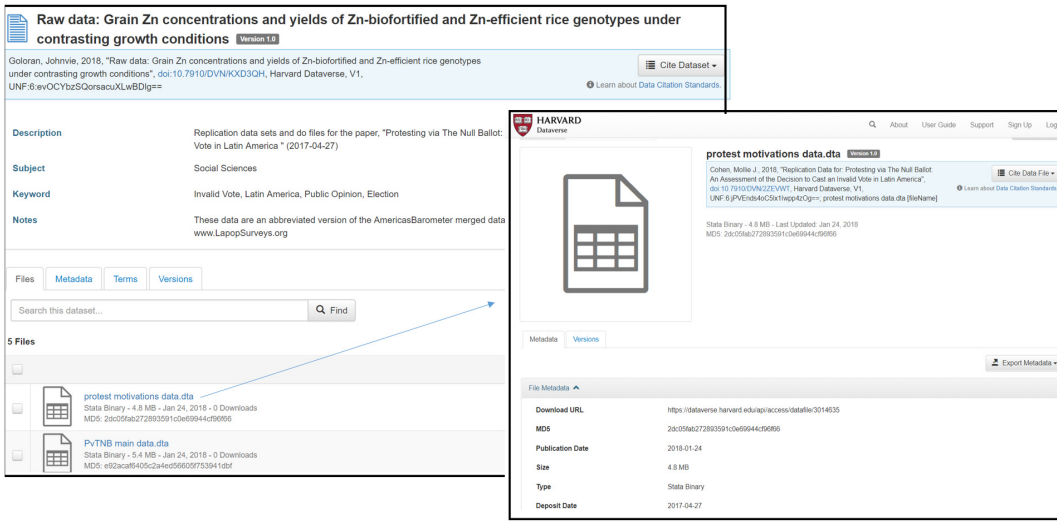
> Verba, Sidney; Lehman Schlozman, Kay; Brady, Henry E.; Nie, Norman, 1996, "American Citizen Participation Study, 1990", http://hdl.handle.net/1902.2/6635 UNF:3:aGYTy1ubiRXFTnPZBExcdA== Inter-university Consortium for Political and Social Research[Distributor] V1 [Version]

연구자가 데이터셋을 Dataverse에 업로드하기 위하여 메타데이터 입력, 데이터셋에 대한 설명, 데이터 파일의 보존 포맷으로의 변환을 요구한다. Dataverse는 XML기반의 DDI 표준을 기반으로 하며, 더블린코어, MARC, FGDC와 같은 다양한 메타데이터를 Open Archive Initiative(OAI-PMH)를 통해 수집하고 상호운용성 및 데이터 공유 프레임워크를 지원하기 위하여 노력하고 있으며, 연구데이터의 장기보존을 위하여 LOCKSS를 사용하여 다양한 리포지터리에 백업하고 복제하여 연구데이터를 분산 보존하고 있다.

Dataverse는 하버드 대학, 베이징 대학, 푸단 대학 등의 기관에서 연구데이터의 수집 및 관리를 위하여 사용하고 있으며, Data-Pass alliance와의 협력을 기반으로 하버드 대학 내의 사회과학 연구소에서 개발되어 2007년 발표된 이후 현재 Dataverse 4.0버전이 출시되어 사회과학은 물론 물리학, 생물학, 자연과학 등 다양한 학문분야의 연구데이터 수집을 위한 리포지터리로 광범위하게 사용되고 있다.

Harvard Dataverse는 〈그림 6〉과 같이 발행일, 주제, 저자, 저자소속기관 등에 따라 검색을 허용하고 있으며, 데이터셋과 파일의 메타데이터에 따라 고급검색을 활용할 수 있다. Harvard Dataverse의 데이터 유형은 설문데이터, 필드데이터, 소프트웨어, 지리데이터, 면담, 의학데이터, 엑셀파일, 실험데이터, 통계데이터, 시각화자료, 녹음자료 등을 포함한다.

Harvard Dataverse의 특징은 〈그림 6〉과 같이 데이터셋과 데이터파일에 DOI 기반 식별자를 부여하고 있으므로 데이터셋에 대한 인용을 물론, 데이터파일에 대한 인용이 가능하도록 하고 있는 것이 특징이다. 데이터셋에 대한 상세 메타데이터(식별자, 발행일, 제목, 저자, 기술, 주제, 기탁자, 기탁일 등)는 물론, 데이터파일에 대한 메타데이터(제목, 인용, 다운로드 URL, 유형, 사이즈, 변수 등)를 제공하고 있다. 데이터셋의 리포지터리로서 데이터셋을 인용한 출판물에 대한 정보는 제공하고 있지 않으며, 데이터셋과 데이터파일에 대한 메타데이터



〈그림 6〉 Harvard Dataverse 연구데이터 정보

를 DDI, DC, JSON 포맷으로 다운로드받을 수 있다

4.3 Dryad Digital Repository

Dryad 디지털 리포지터리(Dryad Digital Repository, 이하 Drayd) 는 과학 출판물의 데이터 검색, 재사용, 인용을 위한 인프라를 제공하기 위한 리포지터리 서비스이다. Drayd의 목적은 데이터의 등록, 검색, 재사용, 인용이 가능하도록 하여, 학술연구와 통합되어 지식창출을 가능하도록 하는 것이 목적이다. Drayd는 현재 노스캐롤라이나 주립대학(North Carolina State University), JISC, NSF(National Science Foundation) 등에서 활용하고 있는 플랫폼으로, Dryad는 표준문서, 소프트웨어, 구조화된 텍스트, 과학 데이터 포맷, 통계 데이터 포맷 등 광범위하고 다양한 데이터 유형을 포함한다. Dryad는 진화 생물학 및 생태학 분야의 주요 저널 및 과학 단체 그룹에서 발간한 주요 저널을 기탁하고 제공하기 위한 리포지터리로부터 시작하여 현재는 사회학, 농업, 생명과학 등 다양한 주제의 연구데이터 수집 및 관리를 위해 사용되고 있다.

고급검색을 통해 저자, 주제, 발행일, 발행처에 따른 검색이 가능하며 데이터셋은 물론 각 데이터파일에도 DOI를 부여하여 데이터파일에 대한 인용이 가능하게 한다. Dryad는 더블린코어를 기반으로 메타데이터 응용프로파일을 확장하여, 데이터 셋(Data Package)과 데이터파일(Files), 데이터셋을 인용한 출판물에 대한 정보를 제공하고 있다. 데이터 셋은 생산자, 제출일, 이용가능일, 식별자, 유형, 상태, 기술, 주제어, 시간적・공간적 배경, 외부 식별자, 관련출판물, 인용 등의 요소를 사용하여 기술하고 있으며, 파일 기술을 위해서 엠바고, 다운로드 수, 페이지 뷰, 유형, 생산자, 제목, 식별자 권한, 기술, 주제, 범위, 제출일, 포맷, 사이즈, 체크섬요소를

〈그림 7〉 Dryad 디지털 리포지터리 연구데이터 정보

사용한다. 데이터셋과 관련 발행사항과의 관계는 참조(dcterms:references)를 사용하여 기술하며, 데이터셋과 데이터의 구조적 관계는 부분-전체(decetms:hasPart/dcterms:isPartOf)를 사용하여 기술한다.

5. 연구데이터 온톨로지 설계

5.1 온톨로지 설계원칙

연구데이터 관련 표준인 더블린코어, DataCite, DDI, DCAT를 비교한 결과, 온톨로지 설계에 있어서 다음의 시사점을 도출하였다. 첫째, 연구데이터는 엔터티 구조에는 차이가 있으나 데이터셋, 데이터파일, 변수의 단계로 구조화되고 있음을 알 수 있다. DDI는 문서, 연구, 데이터로 구분하고 있으며, DCAT 역시 데이터셋과 파일로 구조화하고 있다. 둘째, 연구데이터와 관련 자료 또는 연구데이터 인용으로 파생되는 자료를 기술하기 위하여 메타데이터 요소를 사용하여 기술하고 있으며 이를 위하여 DOI와 같은 영구식별자 및 대체식별자를 부여하기 위한 요소, 인용형식을 지정하기 위한 요소가 사용되고 있다. DataCite, DCAT는 DOI를 기본 식별자로 부여하고 있으며, 대체식별자(DataCite: AlternateIdentifier)를 기술하기 위한 요소를 허용하고 있다. 또한 DDI 및 DataCite은 데이터셋 인용의 인용을 위한 요소를 제시하고 있다. 셋째, 자원유형, 데이터포맷, 관계, 이벤트, 도구 등의 값을 지정하기 위하여 통제어휘를 사용하여 기술의 활용성을 지원하고 있다. 더블린코어는 자원유형, 포맷, 주제, 언어 등에 대한 인코딩 스킴을 제시하고 있으며, DataCite은 기여자유형, 날짜유형, 관계유형, 분석, 이벤트 유형, 데이터유형을 정의하고 있다. DDI는 포맷, 데이터유형, 라이프사이클이벤트, 도구, 주제를 제시하고 있으며, DCAT는 미디어유

형, 포맷유형을 제시하고 있다. 마지막으로, 연구데이터 엔터티 간의 관계, 연구데이터와 관련 연구와의 관계를 명시하기 위한 구조메타데이터를 사용하고 있다. 더블린코어는 관계요소(relation)를 통해 구조를 표현할 수 있으며, DataCite은 참조(is referenced by/references) 및 인용(is cited by/cites) 등의 요소를 사용하고 있다.

연구데이터 리포지터리 사례를 분석한 결과 다음의 시사점을 도출하였다. 첫째, 연구데이터는 데이터셋에 대한 검색을 기본으로 하며, 데이터 관련 출판물, 데이터 파일 및 변수 검색 등으로 확대되고 있다. 데이터셋 및 데이터에 대한 인용을 위하여 인용정보를 제공하여 데이터셋 활용을 지원한다. 둘째, 메타데이터 기반 검색 시스템은 저자, 주제, 발행일, 발행처 등에 따라 검색이 가능하며 이를 바탕으로 키워드 검색 및 브라우징을 제공하고 있으므로 이에 대한 메타데이터 요소의 반영이 요구된다. 셋째, 주제, 지역, 데이터포맷, 접근성, 데이터유형 등에 대한 통제어휘를 바탕으로 검색을 제공하고 있으므로 적절한 형태의 통제어휘 수립이 요구된다. 넷째, 데이터셋은 데이터에 대한 이해 및 활용을 돕기 위한 코드북 등의 문서를 함께 제공하고 있다. 이러한 내용을 고려할 때 연구데이터의 엔터티 관계정립, 인용정보 등의 부여, 통제어휘 사용 등에 대한 사항을 메타데이터 스키마에서 고려할 필요가 있다.

이상의 시사점을 바탕으로 연구데이터 온톨로지 설계를 위하여 다음의 원칙을 적용하였다.

첫째, 연구데이터 기술을 위한 온톨로지 설계를 활용한다. 기존의 연구데이터 기술이 메타데이터로 표현되어 있어나, 연구데이터는 연구-데이터셋-파일 등 다양한 엔터티 사이의 지적구조에 대한 표현이 선결되어야 한다. 메타데이터의 온톨로지 모델링은 디지털 자원의 다면성 및 풍부한 관계표현을 통해 지적구조를 체계적으로 제시하고 정보의 접근을 유용하게 한다(박옥남, 박희진, 2016). 이에 본 연구는 연구데이터 지적구조의 설계를 위하여 온톨로지를 적용하였다.

둘째, 향후 NTIS가 다양한 기관에서 생산된 여러 분야의 데이터를 수용해야 하는 것을 감당할 때 다양한 메타데이터 표준과의 상호운용성을 지원할 필요가 있다. 이에 따라 연구데이터 기술을 위하여 가장 범용적으로 사용되고 있는 더블린코어를 기본으로 활용하되, DDI, DataCite, DCAT 및 리포지터리 사례의 시사점을 바탕으로 요소를 확장하였다. 자원에 대한 기술에 필요한 요소들 - 제목, 식별자, 주제어, 기술, 날짜, 생성자, 자원의 포맷, 자원의 유형 등 - 은 DDI, DataCite, DCAT 및 리포지터리에 상관없이 공통적으로 사용되고 있음을 파악하였다. 이에 공통적으로 적용될 수 있는 자원을 기술하는 요소는 더블린코어를 사용하였으며, 이외에 확장이 필요한 요소는 DDI, DataCite, DCAT 등 표준을 사용하여 확장하고, 기존의 메타데이터 요소로 적용될 수 있는 요소들은 추가 생성하였다.

셋째, 연구-데이터셋-파일 및 연구성과물과의 관계 및 각 단계의 연구자원 기술을 위하여 속성을 데이터유형, 객체유형으로 규정하여 각 클래스의 기술을 용이하게 하였다. 데이터 유형 속성은 속성값이 문자열, 날짜, 숫자 등의 데이터로 입력되는 경우에 사용되며, 객체속성은 속성값이 다른 클래스의 인스턴스를 지칭할 때 사용된다.

마지막으로, 명명을 위하여 클래스는 명사형-대문자로 시작하며, 속성은 (동사형)소문자를 사용하여 구분을 용이하게 하였다.

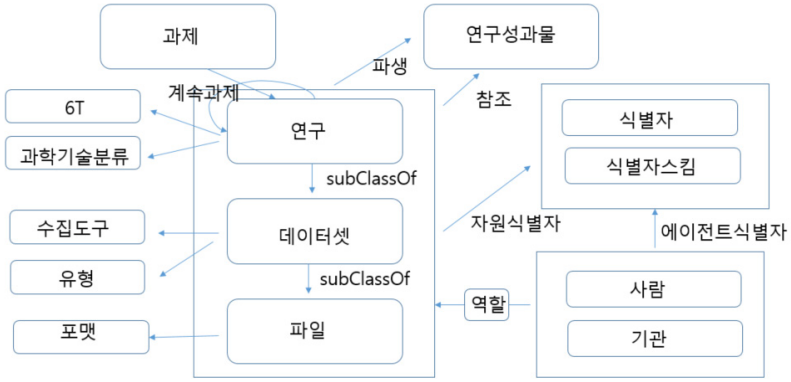
5.2 연구데이터 온톨로지

5.2.1 연구데이터 모델링

현재 NTIS의 연구성과물에 대한 등록시스템, 연구데이터 메타데이터 표준, 연구데이터 리포지터리 사례를 분석한 결과, NTIS 서비스를 연구데이터 관리 및 활용이 가능하도록 확대하기 위하여 다음이 연구데이터 개체에 대한 모델링을 도출하였다. 기존 NTIS 서비스의 기본 엔터티인 연구, 연구관련 과제 및 연구관련 성과물의 개념을 유지하고, 데이터셋과 파일을 연구의 서브클래스로 확장하였다. 계층관계로 연구-데이터셋-파일 클래스를 설정하여 연구의 기술정보가 데이터셋 및 파일에 상속되도록 설정하였다. 둘째, 연구클래스의 경우, 6T분류, 과학기술분류, 연구과제를 관련 클래스로 설정하여 범주에 따른 연구정보탐색이 가능하도록 설정하였다. 셋째, 연구, 데이터셋, 파일의 식별자를 관리하고 DOI는 물론 개별연구과제, 연구자등록번호, Orchid, ISNI 등의 연구자 식별번호 등의 스킴을 설정하기 위하여 식별자 및 식별자스킴을 클래스로 지정하여 식별자 관리 및 연구-데이터셋-파일과의 연계를 용이하게 하였다. 넷째, 사람, 기관 등의 연구자 및 연구수행기관, 데이터 수집가, 데이터 관리자 등을 관리하기 위하여 사람 및 기관 및 역할을 클래스로 지정하여 연구-데이터셋-파일과의 연계가 가능하도록 하였다. 마지막으로 수집도구, 유형, 포맷 등 데이터셋 및 파일 기술에 있어서 통제어휘가 필요한 부분은 클래스로 구성하였다.

5.2.2 연구데이터 온톨로지

연구데이터 온톨로지는 연구-데이터셋-파일의 연구데이터의 지적구조의 주요 엔터티를 중심으로 정의되었다. 연구 클래스는 연구 제목, 식별자, 연구과제가 파생된 프로젝트, 관련 계속과제, 6T분류, 과학기술표준분류, 과제관리기관, 과제수행기관, 관련 데이터셋, 과제 연구비, 키워드, 연구정보, 방법론, 연구참여자, 연구기여자, 연구수행기간, 과제연구성과 등의 속



〈그림 8〉 연구데이터 온톨로지 모델링

성을 사용하여 기술되도록 하였다. 특히 연구과제가 다년에 걸쳐 수행되는 경우가 있음을 반영하기 위하여, 계속과제(isSeriesOf) 속성을 재귀속성(recursive property)으로 정의하였으며, 과제관리기관(hasFundginAgency), 과제수행기관(hasStudyOrganization)은 에이전트 클래스의 하위클래스인 기관(Organization) 클래스와 연계하고, 연구의 하위클래스인 연구에서 생산된 데이터셋은 관련데이터셋(hasDataSet) 속성을 사용하여 정의하였다. 연구참여자(hasCreator)의 하위속성인 연구책임자(hasPrimaryResearcher) 및 공동연구자(hasResearcher)를 사용하여 에이전트 클래스의 하위클래스인 사람(Person)클래스와 연계하였으며, 연구기여자(hasContributor)는 기여자유형(ContributorType)클래스와 연계하여 연구기여자의 역할을 정의하도록 하였다.

〈표 2〉 연구클래스 속성정의

속성-하위속성	속성유형	설명	출처
제목(title)	데이터유형	연구제목	DC
식별자(hasIdentifier) - 자원식별자 (hasResourceIdentifier) - 자원식별자스킴 (hasResourceIdentifierScheme)	객체속성	연구와 관련한 정보를 식별하기 위한 요소(과제고유번호, 기관과제고유번호, DOI)	확장 (DataCite)
연구과제사업(hasProject)	객체속성	연구과제가 파생된 프로젝트 (예: 인문사회 일반공동연구자 사업)	생성
계속과제(isSeriesOf)	객체속성	연구가 계속과제일 경우 관련 계속과제 기술	생성
6T분류(has6T)	객체속성	연구가 속한 6T분류	생성
과학기술표준분류 (hasScienceCategory)	객체속성 (ScienceCategory)	연구가 속한 과학기술표준분류	생성
과제관리기관(hasFundingAgency)	객체속성 (Organization)	연구관리기관	생성
과제수행기관(hasStudyOrgnaization)	객체속성 (Organization)	연구수행기관	생성
관련데이터셋(hasDataSet)	객체속성(DataSet)	연구 관련 데이터셋	생성
과제연구비(funding)	데이터유형	연구비	생성
키워드(keyword)	데이터유형	연구관련 키워드	DC
연구정보(description)	데이터유형	연구정보	DC
방법론(methodology)	데이터유형	연구방법론, 연구수집기관 등	확장 (DDI)
연구참여자(creator) - 연구책임자 - 공동연구자	객체속성(Person)	연구에 참여한 연구자	DC
연구기여자(contributor)	객체속성(Person, ContributorType)	연구에 기여한, 기탁자, 데이터수집가, 데이터큐레이터 등	DC
연구수행기간(study_Coverage)	데이터유형	연구수행 시작일부터 종료일	DC
과제연구성과(hasStudyResults)	객체유형 (StudyPublications)	연구에서 파생된 연구성과물 기술	생성

데이터셋 클래스는 제목, 식별자, 날짜, 기술, 공간범위, 시간범위, 인용, 권한, 데이터수집도구, 데이터유형, 데이터셋 파생 연구, 인용문헌, 관련파일 등의 속성으로 정의하였다. 데이터셋도 마찬가지로 DOI 등 영구식별자를 부여하도록 하였으며, 다른 리포지터리에서 부여된 식별자를 반영하기 위하여 자원식별자 스킴 속성(hasResourceIdentifierScheme)을 사용하였다. 날짜(Date)는 리포지터리에 데이터셋 제출일(dateSubmitted), 데이터 컬렉션 수집일(StartDate), 종료일(EndDate), 데이터셋 이용가능일(dateAvailable) 등을 구분하기 위하여 하위속성을 사용하였다. 데이터 수집도구(hasInstruments)는 수집도구(Instruments)클래스와 연계하여 실험, 관찰, 설문, 면담 등의 데이터수집방법을 기술하도록 하였으며, 데이터유형(hasType)은 데이터유형(DataType)클래스와 연계하여 데이터셋의 다양한 유형, 실험데이터, 통계데이터, 관찰데이터, 추출데이터 등을 반영하도록 설정하였다. 데이터셋을 생산한 연구(isDataSet)는 연구 클래스의 데이터셋(hasDataSet) 속성과 역속성(Inverse Property)으로 정의하여 상호참조가 되도록 하였으며, 데이터셋에 포함되어 있는 개별파일은 관련파일(hasFile) 속성을 사용하여 정의하였다. 마지막으로 데이터셋을 사용하여 생산된 다른 연구성과물과의 관계는 인용문헌(isReferenceBy) 속성을 사용하여 정의하였다.

〈표 3〉 데이터셋 클래스 속성정의

속성-하위속성	속성유형	설명	출처
제목(title)	데이터유형	데이터셋 제목	DC
식별자(hasIdentifier) - 자원식별자(hasResourceIdentifier) - 자원식별자스킴 (hasResourceIdentifierScheme)	객체속성	데이터셋과 관련한 정보를 식별하기 위한 요소 식별자 스킴(DOI, URN등)	확장 (DataCite)
날짜(Date) - 제출일(dateSubmitted), 이용가능일(dateAvailable), 수집일(StartDate), 종료일(EndDate)	데이터유형	데이터셋이 리포지터리에 제출된 날짜, 이용가능날짜, 데이터 수집일과 종료일	DC
기술(Description)	데이터유형	데이터셋에 대한 설명	DC
공간범위(spatial)	데이터유형	데이터셋의 공간범위	DC
시간범위(temporal)	데이터유형	데이터셋의 시간범위	DC
인용(bibliographicCitation)	데이터유형	데이터셋 인용정보	DC
권한(rights)	데이터유형	데이터셋의 권한이나 접근제한에 대한 내용(CC0 등)	DC
데이터수집도구(hasInstruments)	객체속성(Instruments)	데이터수집방법(실험, 관찰, 설문, 면담 등)	확장 (DDI)
데이터유형(hasType)	객체속성(Type)	데이터 유형(실험데이터, 통계데이터, 관찰데이터 등)	DC
데이터셋(isDataSetOf)	객체속성(Study)	관련 연구 연계 (hasDataSets의 역속성)	생성
인용문헌(isReferencedBy)	객체속성(Publication)	데이터셋을 인용한 다른 출판물	DC
관련파일(hasFile)	객체속성(File)	데이터셋에 포함된 파일	생성

파일 클래스는 데이터셋을 구성하고 있는 개별 파일에 대한 기술이 이루어진다. 파일 클래스는 제목, 식별자, 크기, 매체, 날짜, 공간, 시간, 다운로드URL, 접근가능URL, 체크섬, 엠바고, 파일상세정보, 인용, 관련데이터셋 속성을 사용하여 정의하였다. 매체(medium)은 SPSS, Stata, TXT 등 파일의 포맷정보를 기술하기 위하여 데이터포맷(DataFormat)클래스와 연계하였으며, 특정 파일 내용에 해당하는 공간 및 시간적 범위(spatial & temporal)에 대한 요소, 파일의 다운로드 가능한 URL(downloadURL), 접근가능한 URL(accessURL), MD5 Checksum정보를 기술하기 위한 요소(checksum), 데이터 자체를 공개하고 다운로드 할 수 있는 날짜인 엠바고(embargo)정보 등을 정의하였다. 또한 파일 역시도 다른 연구자에 의해 인용될 수 있도록 인용형식(bibliographicCitation) 및 식별자(identifier)를 정의하였으며, 상위 클래스인 데이터셋과 연계를 위하여 관련데이터셋(isFileOf)을 hasFile의 역속성으로 정의하였다.

〈표 4〉 파일 클래스 속성정의

속성-하위속성	속성유형	설명	출처
제목(title)	데이터유형	파일제목	DC
식별자(hasIdentifier) - 자원식별자(hasResourceIdentifier) - 자원식별자스킴 (hasResourceIdentifierScheme)	객체속성	파일과 관련한 정보를 식별하기 위한 요소(DOI 등)	확장 (DataCite)
크기(extent)	데이터유형	파일크기(Size)	DC
매체(medium)	객체속성(Format)	파일의 포맷기술(SPSS, TXT, PDF, SAS, Stata 등)	DC
날짜(Date) - 제출일(dateSubmitted)	데이터유형	파일이 리포지터리에 제출된 날짜	DC
공간범위(spatial)	데이터유형	파일의 공간범위	DC
시간범위(temporal)	데이터유형	파일의 시간범위	DC
다운로드URL(downloadURL)	데이터유형	파일의 다운로드 가능한 URL	확장 (DCAT)
접근가능URL(accessURL)	데이터유형	파일의 접근가능한 URL	확장 (DCAT)
체크섬(checksum)	데이터유형	파일의 MD5 checksum정보 기술	DC
엠바고(embargo)	데이터유형	엠바고 일자 (데이터 자체를 공개하고 다운로드 할 수 있는 날짜)	DC
파일상세정보(description)	데이터유형	파일에 대한 상세정보(예: spss파일 이해를 돕는 codebook에 대한 설명 등)	DC
인용(bibliographicCitation)	데이터유형	파일인용정보	확장(DDI)
관련데이터셋(isFileOf)	객체속성 (DataSet)	관련 데이터셋과의 연계 (hasFile의 역속성)	생성

5.2.3 적용사례

구성된 연구데이터 온톨로지를 NTIS 예를 통해 적용함으로써 본 연구에서 도출된 온톨로지의 적용가능성을 살펴보았다. '2014년 농촌 다문화자녀들의 공동체 활용 강화방안에 대한 연구'를 기술하기 위하여 이 연구의 연구책임자('양순미'), 6T분류('생활문화기술'), 과학기술표준분류('가족복지'), 연구수행기관('국립농업과학원'), 관리기관('농촌진흥청'), 과제연구비('7500만원'), 식별자('http://doi.org/10.3866/139540')의 요소를 활용하여 값을 부여하였다. 또한 데이터셋('DS1')을 클릭하면 본 연구에서 생산된 데이터셋을 확인할 수 있는데, 데이터수집방법('설문'), 데이터유형('통계데이터'), 수집시작일('2013-10-01'), 수집종료일('2013-10-15')은 물론 제목, 상세정보, 인용정보, 데이터셋 제출일, 데이터셋에 포함된 데이터파일을 확인할 수 있다. 데이터파일('File_01')을 클릭하면 각 파일의 크기('0.2MB'), 데이터포맷('text'), 엠바고 일자('2015-10-30') 및 인용정보, 식별자, 체크섬 정보를 확인할 수 있다.

또한 계층구조로 이루어진 연구-데이터셋-파일의 구조는 상호참조(Incoming Reference)를 통해 파일의 데이터셋, 데이터셋이 생산된 연구로의 탐색이 가능하며, 통제어휘를 사용하여 클래스로 명시된 파일유형, 데이터포맷, 데이터수집방법, 데이터유형을 통해 동일한 데이터유형, 동일한 데이터포맷 등을 가진 데이터셋이나 데이터파일을 탐색할 수 있다(〈그림 9〉 참조).

〈그림 10〉은 적용 예시의 Turtle문서 중 일

〈그림 9〉 연구데이터 온톨로지 적용예시

```
NTIS:Study_1395040049
  rdf:type NTIS:Study ;
  dcterms:title "2015농촌 다문화자녀들의 공동체 활동 강화 방안 연구"^^xsd:string ;
  NTIS:Funding "75,000,000"^^xsd:string ;
  NTIS:PrimaryInvestigator NTIS:P1352 ;
  NTIS:alternativeIdentifier <http://www.ntis.go.kr/metadata#1395040049> ;
  NTIS:has6T NTIS:Society_Culture ;
  NTIS:hasDataSet NTIS:DS1 ;
  NTIS:hasFundingAgency NTIS:RDA ;
  NTIS:hasIdentifier NTIS:DOI139540 ;
  NTIS:hasProject NTIS:ArgricultuEnvironment ;
  NTIS:hasScienceCategory NTIS:FamilyWelfare ;
  NTIS:hasStudyOrganization NTIS:NationalAgricultureInstitute ;
  NTIS:hasStudyResults NTIS:Article_JECO_0401 ;
  NTIS:isSeriesOf <http://www.ntis.go.kr/metadata#1395432> ;
  NTIS:study_Coverage "연구과제  [NTIS:Study]"^^xsd:string ;
  rdfs:label "연구과제"^^xsd:string ;

NTIS:DS1
  rdf:type NTIS:Dataset ;
  dc:title NTIS:DS1 ;
  dcterms:bibliographicCitation "양순미. 농촌다문화자녀들의 공동체 활동방안연구. 국립농업과학원. 2015.
https://doi.org/10.3886/1395040049"^^xsd:string ;
  dcterms:dateSubmitted "2015-10-15"^^xsd:string ;
  dcterms:description "농촌의 다문화가구와 비다문화가구 청소년들의 인성 및 관련변인 특성 비교분석을 위한
설문문항, 설문데이터, codebook을 포함하고 있음 "^^xsd:string ;
  dcterms:rights NTIS:LicenseDocument_1 ;
  dcterms:title "DS1:다문과가구비다문화가구변인분석"^^xsd:string ;
  dcterms:type NTIS:StatistiData ;
  NTIS:endDate "2013-10-15"^^xsd:string ;
  NTIS:hasFile NTIS:File_01 ;
  NTIS:hasFile NTIS:File_02 ;
  NTIS:hasFile NTIS:File_03 ;
  NTIS:hasInstruments NTIS:Survey ;
  NTIS:startDate "2013-10-01"^^xsd:string ;


NTIS:File_01
  rdf:type NTIS:File ;
  dcterms:bibliographicCitation "양순민. 농촌다문화자녀들의 공동체활동방안 통계코드북. 2015.
http://doi.org/10.3886/DS1_F01" ;
  dcterms:extent <http://www.ntis.go.kr/metadata#0.2> ;
  dcterms:title "Codebook"^^xsd:string ;
  NTIS:checksum "1535243b"^^xsd:string ;
  NTIS:embargoedUntil "2015-10-30"^^xsd:string ;
  NTIS:hasDataFormat NTIS:TXT ;
  NTIS:hasIdentifier NTIS:DOI_DS1_F01 ;
  NTIS:hasIdentifier NTIS:DOI_DS1_File01 ;
```

〈그림 10〉 연구데이터 온톨로지 적용예시

부이다. Turtle문서는 주어, 술어(속성), 목적어(속성값)의 관계로 표현하고 있는데, 예를 들어, NTIS:Study_1395040049의 클래스 유형은 연구과제(NTIS: Study)이며, 계속과제로 〈http://www.ntis.go.kr/metadata#1395432〉, 연구성과로 NTIS:Article_JECO_0401, 데이터셋으로 NTIS:DS1을 가짐을 알 수 있다. 온톨로지에서 제시한 속성을 통해 연구과제의 연구성과물, 데이터셋, 데이터파일 등의 메타데이터 레코드와 연계되고 있다.

6. 결 론

연구데이터의 연구의 정확성이나 신뢰성 확보를 위한 정보적 가치, 연구의 재현 또는 검증, 재사용 가능성을 고려할 때 연구데이터의 관리 및 활용을 위한 체계 구축은 필요하다. 많은 국외 사례에서 데이터 리포지터리를 구축하고 연구데이터의 등록, 관리, 보존, 활용을 지원하고 있으며, 연구자의 데이터 관리 계획수립의 중요성이 강조되고 있음에도 불구하고, 국내는 아직 연구데이터 관리 및 활용을 위한 체계 구축이 미흡하다. 이를 위한 단계적 노력이 필요한데, 본 연구는 계층적 연구데이터의 지적구조를 분석하고, 계층별 연구데이터를 등록・관리할 수 있는 온톨로지 모델링을 제시하고자 하였다. 특히 국가 R&D 포털인 국가과학기술정보서비스에 적용할 수 있는 온톨로지 모델링을 제시함으로써 국가 연구성과는 물론 연구데이터의 체계적 관리를 위한 플랫폼을 구축하는 기반을 마련하고자 하였다.

이를 위하여 연구데이터와 관련한 선행연구를 조사하였으며, 연구데이터와 관련한 메타데이터 표준, 연구데이터 리포지터리의 사례를 조사하였다. 온톨로지를 예시를 제시하여 적용가능성을 높였다.

선행연구 조사결과, 연구데이터 수집 및 체계적 관리의 중요성에도 불구하고, 국내의 경우 연구데이터 수집・관리・활용을 위한 준비는 부족한 것을 알 수 있다. 국내의 경우 연구성과물을 등록하기 위한 기관 리포지터리 정도만 준비되어 있고, 실제 연구데이터 수집 및 관리를 위한 리포지터리는 미비함을 알 수 있으며, 다양한 요소 중에 연구데이터 관리 및 활용 인프라를 갖추는 것과 이에 대한 출발점으로 메타데이터 표준의 제시가 필요함을 확인하였다.

연구데이터 관련 표준인 더블린코어, DataCite, DDI, DCAT를 비교한 결과, 온톨로지 설계에 있어서 다음의 시사점을 도출하였다. 첫째, 연구데이터는 엔터티 구조에는 차이가 있으나 데이터셋, 데이터파일, 변수의 단계로 구조화되고 있음을 알 수 있다. 둘째, 연구데이터와 관련 자료 또는 연구데이터 인용으로 파생되는 자료를 기술하기 위하여 메타데이터 요소를 사용하여 기술하고 있으며 이를 위하여 DOI와 같은 영구식별자 및 대체식별자를 부여하기 위한 요소, 인용형식을 지정하기 위한 요소가 사용되고 있다. 셋째, 자원유형, 데이터포맷, 관계, 이벤트, 도구 등의 값을 지정하기 위하여 통제어휘를 사용하여 기술의 활용성을 지원하고 있다. 마지막으로, 연구데이터 엔터티 간의 관계, 연구데이터와 관련 연구와의 관계를 명시하기 위한 요소가 요구됨을 파악하였다.

연구데이터 리포지터리 사례를 분석한 결과, 연구데이터는 데이터셋에 대한 검색을 기본으로 하며, 데이터 관련 출판물, 데이터 파일 및 변수 검색 등으로 확대되고 있다. 이를 위하여 데이터셋 및 데이터에 대한 인용을 위하여 인용정보를 제공하여 데이터셋 활용을 지원한다. 둘째, 메타데이터 기반 검색 시스템은 저자, 주제, 발행일, 발행처 등에 따라 검색이 가능하며 이를 바탕으로 키워드 검색 및 브라우징을 제공하고 있으므로 이에 대한 메타데이터 요소의 반영이 요구된다. 셋째, 주제, 지역, 데이터포맷, 접근성, 데이터유형 등에 대한 통제어휘를 바탕으로 검색을 제공하고 있으므로 적절한 형태의 통제어휘 수립이 요구된다.

이상의 시사점을 바탕으로 연구데이터 온톨로지는 다음의 사항을 고려하여 도출되었다.

첫째, 향후 NTIS가 다양한 기관에서 생산된 여러 분야의 데이터를 수용해야 하는 것을 고려할 때 다양한 메타데이터 표준을 지원할 필요가 있다. 이에 따라 연구데이터 기술을 위하여 가장 범용적으로 사용되는 더블린코어를 기본으로 활용하되, DDI, DataCite, DCAT 및 리포지터리 사례의 시사점을 바탕으로 요소를 확장하였다.

둘째, 연구-데이터셋-파일 및 연구성과물과의 관계 및 각 단계의 연구자원을 기술을 위하여 속성을 데이터유형, 객체유형으로 규정하여 각 클래스의 기술을 용이하게 하였다.

셋째, 기존 NTIS 서비스의 기본 엔터티인 연구, 연구관련 과제 및 연구관련 성과물의 개념을 유지하고, 데이터셋과 파일을 연구의 서브클래스로 확장하였다. 계층관계로 연구-데이터셋-파일 클래스를 설정하여 연구데이터의 계층구조를 설정하였다.

넷째, 연구클래스의 경우, 6T분류, 과학기술분류, 연구과제를 관련 클래스로 설정하여 범주에 따른 연구정보탐색이 가능하도록 설정하였다. 연구과제 클래스는 연구 제목, 식별자, 연구과제가 파생된 프로젝트, 관련 계속과제, 6T분류, 과학기술표준분류, 과제관리기관, 과제수행기관, 관련 데이터셋, 과제 연구비, 키워드, 연구정보, 방법론, 연구참여자, 연구기여자, 연구수행기간, 과제연구성과 등의 속성을 사용하여 기술되도록 하였다. 데이터셋 클래스는 제목, 식별자, 날짜, 기술, 공간범위, 시간범위, 인용, 권한, 데이터수집도구, 데이터유형, 데이터셋 파생연구, 인용문헌, 관련파일 등의 속성으로 정의하였다. 파일 클래스는 데이터셋을 구성하고 있는 개별의 파일에 대한 기술이 이루어진다. 파일 클래스는 제목, 식별자, 크기, 매체, 날짜, 공간, 시간, 다운로드URL, 접근가능URL, 체크섬, 엠바고, 파일상세정보, 인용, 관련데이터셋 속성을 사용하여 정의하였다.

다섯째, 연구, 데이터셋, 파일의 식별자를 관리하고 DOI는 물론 개별연구과제, 연구자등록번호, Orchid, ISNI등의 연구자 식별번호 등의 스킴을 설정하기 위하여 식별자 및 식별자스킴을 클래스로 지정하여 식별자 관리 및 연구-데이터셋-파일과의 연계를 용이하게 하였다.

마지막으로 수집도구, 유형, 포맷 등 데이터셋 및 파일 기술에 있어서 통제어휘가 필요한 부분은 클래스로 구성하고 연구-데이터셋-파일 클래스와 연계하였다.

본 연구는 앞으로 연구데이터의 지적구조를 이해하고, 체계적 기술 및 관리를 위한 메타데이터 스키마를 정비하는데 기반이 될 것으로 기대된다.

참 고 문 헌

강희종 (2012). 21세기 핵심자원, 국가과학데이터 활용을 위한 정책과제. STEPI Insight, 91, 1-26.

과학기술정보통신부 (2018. 1. 18). 서랍 속 연구데이터 함께 쓰는 빅데이터로 새롭게 거듭난다. 보안뉴스. 검색일자: 2018. 1. 20. http://www.boannews.com/media/view.asp?idx=66208

김선태, 한선화, 이태영, 김용 (2010). 과학데이터 보존 및 활용모델에 관한 연구. 한국비블리아학회지, 21(4), 81-93.

김순, 이보람, 김환민, 김혜선 (2017). 오픈 사이언스 시대를 위한 과학기술 연구지원 서비스 동향 분석. 정보관리학회지, 34(3), 229-249.

김은정, 남태우 (2012). 연구데이터 수집에 영향을 미치는 요인 분석. 정보관리학회지, 29(2), 27-44.

김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455.

김지현 (2016). 연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구. 한국도서관·정보학회지, 47(3), 213-239.

박희진, 박옥남 (2016). IPAM 모형을 적용한 기록관리 메타데이터 온톨로지 설계. 한국기록관리학회지, 15(4), 99-123.

심원식 (2015). 국가 차원의 연구데이터 관리체계 구축을 위한 로드맵 제안: 영국 사례 분석을 중심으로. 한국문헌정보학회지, 49(4), 355-378.

심원식, 안혜연, 변제연 (2015). 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구: 기초학문센터를 중심으로. 한국문헌정보학회지, 48(3), 155-183.

[웹사이트]

과학기술정보통신부 국가과학기술정보서비스성과물등록포털. 검색일자: 2018. 1. 20. http://rpall.ntis.go.kr

Data Documentation Initiative. Retrieved January 10, 2018, from https://www.ddialliance.org/explore-documentation

DataCite Retrieved January 10, 2018, from https://schema.datacite.org/

DCAT. Retrieved January 10, 2018, from https://www.w3.org/TR/vocab-dcat/

Dryad Digital Repository Retrieved January 10, 2018, from http://datadryad.org

Dublin Core Metadata Initative. Retrieved January 10, 2018, from http://dublincore.org

Harvard Dataverse. Retrieved January 10, 2018, from https://dataverse.harvard.edu/

Inter-Consortium for Political and Social Research Data Archive. Retrieved January 10, 2018, from http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp

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

Kang, Hee Jong (2012). Policy Tasks for Utilization of National Science Data for Core Resources

of the 21st Century. STEPI Insight, 91, 1-26.

Kim, Eun-Jung, & Nam, Tae-woo (2012). Factor Analysis of Effects on Research Data Collection. Journal of Korean Information Science, 29(2), 27-44.

Kim, Jihyun (2012). A Study on University Researchers' Data Management Practices. Journal of Korea Library and Information Science Society, 43(3), 433-455.

Kim, Jihyun (2016). A Study on Policy Components of Data Access and Use Controls in Research Data Repositories. Journal of Korean Library and Information Science Society, 47(3), 213-239.

Kim, Soon, Lee Bo-Ram, Kim, Hwanmin, & Kim Hyesun (2017). Open Science and Technology Research Support Service Trends for Open Science Era. Journal of Korean Information Science, 34(3), 229-249.

Kim, Sun-Tae, Hahn, Sun-Hwa, Lee, Tae-young, & Kim, Yong (2010). A Study on a Model for Using and Preserving Scientific Data. The Journal of Korean Biblio Society for Library and Information Science, 21(4), 81-93.

Ministry of Science and ICT (2018. 1. 18). 21-year-old drawer newly created as big data to be used together with research data. Security News. January 20, 2018, from http://www.boannews.com/media/view.asp?idx=66208

Park, HeeJin, & Park, Oknam (2016). A Study on Metadata Ontology Design for Record Management Based on IPAM Model. Journal of Records Management & Archives Society of Korea, 15(4), 99-123.

Sim, Wonsik (2015). Developing a Roadmap for National Research Data Management Governance: Based on the Analysis of United Kingdom's Case. Journal of Korean Society for Library and Information Science, 49(4), 355-378.

Sim, Wonsik, Ahn, Hyeyeon, & Byun, Jaeyeon (2015). Developing a Roadmap for National Research Data Management Governance: Based on the Analysis of United Kingdom's Case. Journal of Korean Society for Library and Information Science, 48(3), 155-183.

[웹사이트]

Ministry of Science and ICT. National Science and Technology Information Service Result Registration Portal. Retrieved January 20, 2018, from http://rpall.ntis.go.kr