

연구데이터 국가 리포지터리 구축에 관한 연구

- 국내외 사례를 중심으로 -

A Study on Developing National Research Data Repository

박규리, 성균관대학교, kyurip0831@gmail.com

안병균, 성균관대학교, rns1305@gmail.com

Kyu-Ri, Park, SungKyunKwan University

Byoung-Goon, An, SungKyunKwan University

본 연구는 연구데이터가 공공재로써 관리와 공유의 주체가 국가이어야 한다는 주장으로부터 시작하였다. 국내외의 대표적인 연구데이터 리포지터리 구축 사례를 통해 연구데이터의 관리 및 공유 현황에 대해 살펴보고, 이를 통해 국내 인문사회분야 대표 리포지터리인 한국연구재단 기초학문자료센터(KRM)의 개선방향을 확인하는 것을 목적으로 하였다. 미국, 영국, 호주의 경우와 KRM의 연구데이터 관리 현황을 비교한 결과 가장 큰 차이점은 KRM의 경우 연구데이터 관리의 중요성에 대한 인식개선을 위한 교육과 실제 구축을 위한 DMP 및 메타데이터 작성 교육 등이 부재한 상황으로 연구자의 적극적 참여 유도에 긍정적 환경을 제시하지 못하고 있다는 것이다. 이외에도 장기보존을 위한 국제적 표준의 채택이나 전담 인력의 부족 또한 주요한 개선점으로 확인되었다.

1. 서론

2007년 OECD에서 발표한 Principles and Guidelines for Access to Research Data from Public Funding 보고서(Pilat·Fukasaku, 2007)가 연구데이터의 공유와 관리에 대한 중요성을 강조한 이후 10년이라는 시간이 흘렀다. 연구데이터는 인류 지식의 산물로 학제간 연구의 토대로써 지속적으로 그 중요성이 제기되어 왔다. Knowledge based Economy 보고서가 지식기반 경제로 나아가기 위해서는 정부의 적극적인 역할이 필요하다고 지적한 바와 같이(OECD, 1996), 정부의 연구비 지원을 받은 과제로부터 생산된 연구데이터는 공공재로 누구나 접근할 수 있도록 공유되어야 한다는 기본원칙과 더불어 다양한 학문분야에서 생성되는 방대한 연구데이터를 통합적으로 관

리하고 효율적으로 공유하기 위해서는 개인이나 특정 단체가 아닌 국가 차원의 주도적 관리가 이루어져야 한다는 주장이 제기되었다.(심원식, 2015)

이미 국내외의 여러 기관에서 다양한 분야에 대한 연구데이터 리포지터리를 구축하여 운영하고 있으며, 우리나라의 경우 한국연구재단의 기초학문자료센터(Korea Research Memory, 이하 KRM)가 인문사회분야의 대표적인 연구데이터 리포지터리로 재단의 지원을 받은 연구 과제들에 대해 연구데이터를 구축하고 성과물의 원문제공 이외에 관련된 자료들의 맥락 정보 또한 제공하고 있다. 그러나 기 구축된 데이터 리포지터리들과 마찬가지로 연구자의 연구데이터 공유 의식 부재, 관련 정책의 미비, 서비스 노후, 과거로부터의 소급 등과 같은 문제들을 안고 있다. 따라서 본 연구는 미

국과 영국, 호주의 연구데이터 관리 사례를 살펴봄으로써 KRM의 실질적 개선방향을 모색해 보고자 하였다.

2. 사례연구

2.1 미국의 ICPSR Data Archive

Inter-university Consortium for Political and Social Research(이하 ICPSR)의 데이터 아카이브는 미국의 대표적인 사회/행동과학분야 연구데이터 리포지터리이다. 1962년 설립된 ICPSR은 1998년부터 웹사이트를 통한 ICPSR 소장 데이터 다운로드 서비스를 제공하였으며, 2001년부터는 회원 기관을 통해 누구나 ICPSR 데이터를 이용할 수 있도록 하였다.

ICPSR은 미시간대학교 소속이나, 다양한 정부기관과 민간 재단으로부터 연구비를 지원받고 있으며, 12명의 선출직 위원으로 구성된 ICPSR Council이 ICPSR 회원들을 대표하여 행정·예산·조직상의 문제를 의결하여 관리한다.

ICPSR Data Archive는 현재 250,000개 이상의 데이터 파일을 제공하고 있는데, 이를 연구과제로 산정하면 10,000건 이상에 이르는 수이다. ICPSR가 제공하는 자료는 데이터 파일, 도큐멘테이션, 연구개요가 하나의 데이터 컬렉션을 이루고 있으며, 구축에 있어서는 연구자가 데이터와 도큐멘테이션을 ICPSR에 제출하면 이를 데이터 아키비스트가 검토 후 아카이빙하는 방식을 취하고 있다. 이 때, 원본 데이터 컬렉션의 경우 원문을 구축하고 이외의 경우에는 링크를 구축한다.

데이터 구축은 OAIS(Open Archival Information System) 참조모형 표준에 기반하며, 자료의 메타데이터 작성 시DCMI(Dublin Core Metadata Initiative) 요소 세트나 DDI(Data Documentation Initiative)와 같은 국제 표준을 따름으로써 데이터의 장기적 보존과 관리가 가능하도록 하고 있다.

이와 더불어 미국의 연구비 지원 기관들 대부분이 DMP(Data Management Plan) 제출을 의무화 혹은 권장하고 있기 때문에 ICPSR 또한 DMP 작성 가이드라인 및 컨설팅 서비스를 제공하고 있으며, 이외에도 데이터의 공유와 아카이빙에 관한 가이드라인을 출판하고, 데이터 관리와 큐레이션에 관한 섹션을 별도 구축하는 등 연구자 교육에 적극적인 입장을 보이고 있다.

2.2 영국의 UK Data Service

영국은 2012년 10월 Economic and Social Research Council(이하 ESRC)를 통해 기존의 다양한 데이터 서비스들을 국가적 차원에서 하나로 통합한 UK Data Service를 설립하였다. UK Data Service의 운영은 영국 전역의 연구데이터 전문가들에 의해 이루어진다. 그들은 연구자들이 제공하는 데이터를 품질과 보안, 큐레이션 측면에서 관리하며, 데이터 접근 뿐 아니라 데이터 분석 도구 및 이용자 지원 관련 업무도 수행한다.

ESRC에서 운영하는 것인 만큼, 사회/경제분야 연구데이터 및 인구 데이터가 구축 대상이며, 현재 7,171개의 데이터 컬렉션이 구축되어져 있다. 데이터 컬렉션은 데이터 파일과 함께 다른 연구자가 해당 데이터를 이해하고 재이용할 수 있도록 하는 충분한 정보가 담긴 도큐멘테이션으로 구성되며, 메타데이터 작성에 DDI를 이용하고 OAIS 모형에 기반하여 데이터를 처리하는 등 데이터의 장기보존을 위해 국제 표준을 이용하고 있다.

데이터 구축에 있어 UK Data Service가 보이는 특징적인 면은 구축을 신청한 데이터의 검토 후 다음의 세 가지 방식 중 하나를 권고한다는 점이다. 첫째, 해당 데이터를 바로 코어 컬렉션에 구축하는 방식, 둘째, 기관 리포지터리나 다른 장소를 이용하는 방식, 셋째,

ReSahre라는 UK Data Service의 온라인 데이터 리포지터리를 통해 연구자가 직접 구축하는 방식이다. 이는 데이터 컬렉션의 수준에 따라 다양한 구축 방식을 취함으로써 경제적·인적 자원의 제한에도 불구하고 높은 수준의 데이터 큐레이션이 가능하게 하여 재이용성을 높이고자 함이다. DMP와 관련하여서는 ESRC DMP 작성에 대한 안내는 물론 나머지 RC (Research Council)들을 비롯한 다른 DMP 작성에도 적용 가능한 내용들을 제공하고 있다.

뿐만 아니라 연구데이터와 관련된 다양한 가이드라인 및 비디오 튜토리얼을 제공하고 있으며 Data Impact Blog라는 이름의 데이터 활용과 관련된 커뮤니티를 운영하는 등 데이터 생산자와 이용자의 교육에도 활발하게 참여하고 있다.

2.3 호주의 ANDS

호주의 경우 Australian National Data Service(이하 ANDS)에서 전반적인 연구데이터 관리 체계를 규정하고 있으며, 연구데이터의 공유와 검색을 위한 웹사이트 Research Data Australia(이하 RDA)를 별도로 운영하고 있다. ANDS는 NCRIS의 지원을 받아 2008년 9월에 설립되었으며, Monash대학을 중심으로 호주국립대학과 호주연방과학산업연구기관이 협력하여 운영 중이다. 2017년 8월 기준 104개의 연구데이터 제공 기관과 협력을 맺어 132,497개의 데이터 셋을 보유하고 있으며, 수집된 데이터에 대한 장기적인 보존 방법으로는 OAIS모형을 채택하였다.

호주의 ANDS는 자체적으로 학술 데이터를 수집하기 보다는 여러 기관들에서 자체적으로 학술 연구를 지원하고 연구데이터를 구축한 내용을 링크 연결을 통해 종합적으로 제공하는 역할을 수행하고 있다. 그렇기 때문에 ANDS에서는 원문 서비스를 제공하지는 않으며, 링

크 연결을 통해서 원문 서비스가 구축된 페이지로 연결시켜주는 방식을 택하고 있다. 이때 링크 연결 될 학술 데이터에 대한 맥락정보를 제공하기 때문에 단순히 범 학문적인 검색 도구로서의 역할뿐만 아니라 연구 내용을 파악할 수 있도록 돕는 역할도 한다.

다양한 기관들로부터 학술 데이터를 제공받는 형식을 취하고 있기 때문에 ANDS에서는 여러 기관들의 학술 데이터를 통합적으로 관리하기 위한 방법이 필요하다. 이를 위한 방법으로 ANDS에서 택하고 있는 것은 두 가지로 연구데이터의 관리 단계에서는 DMP 가이드라인을, 연구데이터 공유 단계에서는 CPG(Content Providers Guide)에 대한 설명을 제공하는 것이다. ANDS의 DMP 가이드라인에서는 기관의 서로 다른 학문분야에 적합한 연구데이터 관리 내용을 제시한다. CPG에서는 협력 기관에 소속된 연구자가 제공해야 하는 학술 데이터의 기술 방식, 필수적으로 제공해야 하는 메타데이터, 학술 데이터의 포맷, 그리고 이것을 RIF-CS(Registry Interchange Format: Collection and Services) 즉, ANDS에서 사용하는 형식으로 자동 변환하는 XML 언어를 통해 기술하는 방법을 자세하게 명시한다. 이러한 방법을 통해 각 기관이 ANDS에서 요구하는 방식으로 연구데이터를 관리, 공유할 수 있도록 돕고 있다.

더 나아가 ANDS에서는 협력 기관과 연구자들이 연구데이터 관리의 중요성을 인식하고 공유를 위한 정보를 적극적으로 제공할 수 있도록 다른 연구기관의 DMP를 공개하고 기관에서 요구하는 메타데이터에 대한 자세한 설명을 기술하여 연구자가 DMP를 작성하는 데 기초가 되는 정보를 제공한다.

2.4 한국의 KRM

KRM은 한국연구재단에서 연구데이터의 관리와 공유를 위해 2006년 설립한 국가 주도 연구

데이터 리포지터리이다. 재단의 연구비 지원을 받은 연구 과제들을 대상으로 관련된 학술자료를 원자료, 중간산출물, 연구결과물로 구분하고 학문분야별 연구성과물 DB를 구축하여 서비스를 제공하고 있다. 2017년 1월 기준 122,975개 자료, 653,563개 파일이 구축되어 있다.

KRM은 인문학, 사회과학, 예술체육, 생활과학, 복합학 분야를 대상으로 하고 있으며, 연구 성과물을 단행본, 보고서, 논문, 조사자료, 고문서, 고도서, 이미지, 동영상, 녹음자료, 웹사이트, 기사자료, 낱장자료, 원문컨텐츠로 분류하고 각 유형에 맞는 메타데이터와 각 자료가 연구에 어떻게 활용되었는지에 대한 맥락정보를 제공하고 있다. 한국연구재단 지원 연구 과제와 연구 성과물에 대해 FRBR(Functional Requirements for Bibliographic Records)모델

을 기반으로 저작, 표현형, 실현형, 개별 자료의 유형을 적용하고 있으며, 이를 바탕으로 다양한 관계를 통한 학술자료 검색을 제공하고자 한다.

‘한국연구재단 지원 연구과제의 연구성과물 제출의무 관련 규정 및 조항’에서는 연구자가 연구 과정에서 수집하여 활용한 자료와 직접 생산한 자료를 제출하도록 명시하고 있다. 그러나 한국연구재단은 연구비 지원 과정에서 연구자에게 DMP를 요구를 의무화 하지 않고 있어, KRM은 연구의 수행 과정에서 발생한 모든 자료를 수집하고자 연구자들이 제출한 결과보고서와 학술 자료를 기반으로 분석을 진행하여 관련된 학술 데이터를 요청하는 방식을 취하고 있다. 연구자로부터 수집된 데이터는 기관에서 구축하며, 단행본이나 학술논문 같이 저작권의 문제가 있는 경우를 제외한 대부분의 연구데이

표 1. 4개 국가 리포지터리 비교

구분	ICPSR Data Archive	UK Data Service	ANDS	KRM
관리기구	ICPSR	ESRC 지원 영국 연구데이터 전문가 집단 운영	NCRIS 지원 Monash 대학 외 2개 기관 협력 운영	한국연구재단
창립	1998년 홈페이지 서비스 개시	2012년 10월 설립	2008년 9월	2006년 센터 설립
구축분야	사회/행동과학	사회/경제/인구	학문분야 전반	인문학/사회과학/ 생활과학/예술체육/ 복합학
1)자료규모	250,000개 이상의 데이터 파일	7,171개의 데이터 컬렉션	104개 기관의 132,497데이터셋	122,975자료, 653,563파일
DMP	●	●	●	○
원문제공	●	●	○	●
링크구축	◐	◐	●	◐
맥락정보	●	●	●	●
국제표준 및 모델	OAIS / DDI / DCMI	OAIS / DDI	OAIS /ISO2146 / ISO19115	FRBR
구축주체	기관 승인	기관 승인	협력기관 구축	기관 구축
오프라인교육	●	●	○	○
온라인교육	●	●	●	○

1) ICPSR Data Archive와 UK Data Service, ANDS는 2017.08.10. 기준이고, KRM은 2017.1.1. 기준임.

터는 연구자로부터 저작권재산권 이용 허락서를 받아 원문으로 제공한다.

KRM은 원문 및 맥락정보 서비스 등 이용자 지향적 연구데이터 관리 시스템 구축의 기반을 갖추고자 하였다. 그러나 연구자에게 연구데이터의 보존과 공유의 중요성에 대한 인식을 심어줄 수 있는 정보와 교육의 제공이 이루어지지 않고 있으며, 또한 연구와 관련된 자료에 대해 제출을 요구하기는 하지만 DMP에 대한 요구나 안내, 메타데이터 작성에 대한 교육도 미비하다. 이로 인해 연구자는 자신의 연구 데이터가 어떠한 방식으로 KRM에 구축되고 서비스되고 있는지에 대해 파악하기 어려울 뿐만 아니라 어떻게 연구 데이터를 관리해야 되는지에 대해서도 충분히 인지할 수 없다. 이로 인해 연구자가 자신의 연구임에도 불구하고 연구데이터 제공과 구축에 소극적일 수밖에 없는 상황이다.(신영란, 2012)

다양한 연구데이터가 보존되어 활용되기 위해서는 장기적이고 지속가능한 데이터 관리 실행방안과 데이터 수집, 보존을 위한 인프라와 표준이 중요하며, 이는 다양한 해외 정부 연구비지원기관의 정책을 통해서도 확인할 수 있다.(김지현, 2013) 그러나 현재 KRM은 수집된 연구데이터의 보존을 위한 국제적 표준의 적용을 명시하고 있지 않으며, 장기적 보존을 위한 인프라 또한 구축되어있지 않은 상황으로 향후 더 많은 연구데이터가 축적 될 경우 보존 측면의 문제가 발생할 수 있을 것으로 우려된다.

3. 결론 및 제언

본 연구는 해외의 연구데이터 국가 리포지터리 구축 사례와 국내 대표사례 중 하나인 KRM의 구축 사례를 비교해 봄으로써 국내 연구데이터 관리 시스템의 개선 방안을 모색해보고자 하였다.

분석결과에 기반한 개선 방안은 다음과 같다.

첫째, 연구자의 적극적 참여 유도를 위한 연구데이터의 관리 및 공유, 재이용에 대한 교육의 실시이다. 해외의 경우 오프라인 교육, 가이드, 비디오 튜토리얼, 웨비나(webinar), 온라인 페이지 등 다양한 형태로 연구데이터에 관한 교육을 실시하고 있다. 이용자 교육을 통해 데이터 생산자들이 연구데이터 관리 및 공유의 당위성을 인식하게 된다면 연구데이터 공유에 자발적으로 참여하게 될 것으로 기대된다.

둘째, 실질적 연구데이터 구축 방법에 대한 교육의 실시이다. 해외 리포지터리의 경우 워크샵, 도서, 가이드라인 등의 형태로 연구데이터 구축 절차 및 방법에 대한 교육을 실시하고 있다. 해당 교육을 통해 연구자들이 생산한 데이터 제공 이외에도 메타데이터를 직접 작성 또는 관련 정보를 제공함으로써 구축된 연구데이터의 정확성이 높아지고 이는 곧 활용성을 높이게 될 것이다.

셋째, 국제 표준을 활용한 데이터 보존 정책 수립이다. 해외 사례를 보면 공통적으로 OAIS 참조모형을 이용한 데이터의 보존 및 접근 정책을 가지고 있다. 이처럼 KRM 역시 국제 표준을 활용한 보존정책을 수립하여 명시한다면 누적된 연구데이터에 대한 영구적인 접근 및 상호운용이 가능하게 될 것이다.

참고문헌

- 기초학문자료센터 (2007. August). Retrieved from <https://www.krm.or.kr>
- 김지현 (2013). 국외 정부연구비지원기관의 연구데이터 관리정책 분석. 한국문헌정보학회지, 47(3): 251-274.
- 신영란 (2012). 인문사회 연구데이터 아카이브의 발전 방안에 관한 연구. 석사학위논문, 이화여자대학교 정책과학대학원, 기록

관리학 전공.

심원식 (2015). 국가 차원의 연구데이터 관리 체계 구축을 위한 로드맵 제안. 한국문헌정보학회지, 49(4): 355-378.

심원식, 안혜연, 변제연 (2015). 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구. 한국문헌정보학회지, 49(3): 155-183.

장금연 (2004). 학술정보유통을 위한 레포지터리 적용 방안 연구. 한국문헌정보학회지, 38(4): 291-310.

Australian National Data Service. (n.d.). Retrieved from <http://www.ands.org.au/>
ICPSR(Inter-university Consortium for Political and Social Research) (n.d.).

Retrieved from

<https://www.icpsr.umich.edu/icpsrweb/>

OECD. (1996) The knowledge-based economy. (General Distribution, OCDE/GD(96)102)

Pilat, D. and Fukasaku, Y. (2007). "OECD principles and guidelines for access to research data from public funding." Data Science Journal, 6: OD4-OD11.

Research Data Australia. (n.d.). Find data for research. Retrived from <https://researchdata.ands.org.au>

UK Data Service (n.d.). Retrieved from <https://www.ukdataservice.ac.uk/>

UK Data Service (2016, January). Collections Development Policy.