

과학기술분야 기관 연구데이터 리포지터리 운영 활성화 방안 연구*

A Study on Strategies to Promote the Activation of Institutional Research Data Repositories in the Field of Science and Technology

김 예 현 (Ye Hyeon Kim)**

김 지 현 (Jihyun Kim)***

초 록

본 연구의 목적은 과학기술분야 연구기관에서 운영되는 기관 연구데이터 리포지터리 운영 현황을 파악하고 활성화 방안을 제시하는 것에 있다. 이를 위해 문헌 연구와 사례 분석, 국내외 기관 리포지터리 담당자와의 인터뷰를 수행하였으며, 리포지터리 규정 및 정책 수립, 연구데이터 공유 인식 개선, 연구데이터 품질 관리 강화를 골자로 하는 기관 연구데이터 리포지터리 운영 활성화 방안을 제안하였다. 첫째, 리포지터리 규정 및 정책 수립 측면에서는 현재 연구데이터와 관련한 규정인 국가연구개발정보 처리기준의 지위 향상과 리포지터리 근거 규정의 명시가 필요하다고 보았다. 둘째, 연구데이터 공유 인식 개선 측면에서 전반적인 연구데이터 교육과 우수 사례 발굴의 필요성을 제안하였다. 셋째, 연구데이터 품질 관리 강화 측면에서 연구자-담당자-위원회의 상호작용과 표준화 작업, 장기 보존을 위한 준비의 필요성을 제안하였다.

ABSTRACT

The purpose of this study is to identify the current status of institutional research data repositories in the field of science and technology and to suggest ways to activate them. The study conducted literature research, case analysis, and interviews with repository managers both domestically and internationally. The study suggested strategies with a focus on establishing repository regulations and policies, improving awareness of research data sharing, and enhancing research data quality management. First, in terms of repository regulations and policy establishment, it was considered necessary to promote the status of the National R&D Information Processing Standards, a regulation related to research data, and clarify repository basis regulations. Second, to enhance awareness of research data sharing, the need for comprehensive research data education and the identification of exemplary cases were suggested. Third, in terms of strengthening research data quality management, the need for preparation for interaction between researchers-persons in charge-committees, standardization work, and long-term preservation was suggested.

키워드: 기관 연구데이터 리포지터리, 연구기관, 연구데이터, 데이터 리포지터리, 데이터 플랫폼

Institutional Research Data Repository, Research Institution, Research Data, Data Repository, Data Platform

* 이 논문은 이화여자대학교 일반대학원 석사학위논문(2023.08)을 축약한 것임.

** 이화여자대학교 일반대학원 문헌정보학과 기록관리학전공 석사, 순천대학교 부속 천안병원 의학도서관 사서 (yehyun.k@ewhain.net) (제1저자)

*** 이화여자대학교 사회과학대학 문헌정보학과 교수(kim.jh@ewha.ac.kr) (교신저자)

논문접수일자 : 2023년 8월 14일 논문심사일자 : 2023년 8월 17일 게재확정일자 : 2023년 9월 2일
한국비블리아학회지, 34(3): 109-134, 2023. <http://dx.doi.org/10.14699/kbiblia.2023.34.3.109>

※ Copyright © 2023 Korean Biblia Society for Library and Information Science

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

1. 서론

1.1 연구 배경 및 목적

현재 학술 커뮤니케이션 생태계의 주된 흐름인 오픈 사이언스는 출판물과 연구데이터, 소프트웨어와 사회적 자원을 포함한 포괄적인 개념이다. 오픈 사이언스 확산 및 연구데이터 공유 활성화를 위한 각종 정책은 국내외를 막론하고 등장하고 있다. 2022년 유럽연합 집행위원회는 연구데이터의 공유 및 사용과 보안을 골자로 하는 데이터법(Data Act) 초안을 제시하였다(European Commission, 2022). 프랑스 오픈 사이언스 위원회는 데이터 리포지터리 선택을 데이터 공유의 핵심 요소로 보았다(Le comité pour la science ouverte, 2022). 미 과학기술정책국(Office of Science and Technology Policy, OSTP) 역시 연구 성과와 데이터를 기탁할 적절한 리포지터리에 대한 지침을 제공할 것을 언급하였다(Office of Science and Technology Policy, 2022).

국내의 경우 2016년 빅데이터 민관합동 T/F를 통해 빅데이터 공유 및 활용과 관련한 의견을 수렴하고 이를 지원하기 위한 논의를 시작하였다(신은정 외, 2019). 또한 윤석열 정부 110대 국정과제 중 과학기술정보통신부와 농림축산식품부 과제로서 연구데이터 개방·공유 플랫폼 구축이 명시되었다(제20대대통령직인수위원회, 2022). 특히, 과학기술정보통신부는 연구개발 디지털화 촉진방안에서 데이터 저장소와의 연계 강화를 연구데이터 수집 확대 방안으로 제시하였다(과학기술정보통신부, 2022). 또한 주요 연구데이터의 공유·활용을 위한 통

합 연구데이터 플랫폼 구축도 추진하고 있다. 다만 이러한 플랫폼의 근간을 이루는 기관의 연구데이터 리포지터리에 대한 구체적인 정책이나 연구는 찾아보기 어렵고 실제 운영하는 기관 역시 내부 공개의 방식으로 운영되고 있다.

정리해보면, 오픈 사이언스 확산 및 연구데이터 공유 활성화의 배경에는 특히 과학기술분야 위주의 정책이 큰 비중을 차지한다. 따라서 데이터 보존 및 공유를 의무로 규정할 수 있는 공공성을 띤 기관에서 가장 활발히 연구데이터의 보존 및 공유가 이루어질 것으로 예상할 수 있다. 특히 전 주기적 연구 활동 지원과 연구데이터 공유 활성화에 있어 연구데이터 보존을 위한 필수적 인프라인 기관 연구데이터 리포지터리를 구축·운영하는 기관의 현황을 살펴볼 필요가 있다.

한편, 기관 연구데이터 리포지터리와 관련하여 데이터 리포지터리, 연구데이터 플랫폼 등의 용어가 혼재되어 사용되고 있다. 본고에서는 출판된 논문을 위주로 하는 일반적인 기관 리포지터리와 구분하기 위해 기관 연구데이터 리포지터리(institutional research data repository)라는 표현을 사용한다. 이는 기관에서 생산되는 연구데이터에 집중한다는 측면에서 해외 대학 및 연구기관에서도 보편적으로 사용된다.

본 연구는 과학기술분야 국가 부처/연구기관의 연구데이터 리포지터리를 대상으로 국내외 현황을 파악하고 연구데이터 보존, 공유 및 재이용과 관련된 문제점을 토대로 기관 연구데이터 리포지터리 활성화 방안을 제안하고자 한다.

1.2 연구의 방법

본 연구는 과학기술분야 국가 연구기관의 연구데이터 리포지터리 운영 현황을 파악하고 기관 차원의 연구데이터 리포지터리 활성화 방안을 제시하는 것을 목적으로 한다. 이에 문헌 연구와 사례 분석, 리포지터리 운영 담당자 인터뷰의 방법으로 연구를 진행하였다.

첫째, 문헌 연구에서는 연구데이터 관리 서비스에 관한 도서관 실무자들의 인식과 연구데이터 공유에 관한 연구자들의 인식, 그리고 연구데이터 리포지터리에 관한 연구의 세 가지로 나누어 살펴보았다.

둘째, 국내외 과학기술 분야 연구데이터 리포지터리 운영 사례를 조사하였다. 사례조사 기관을 선정하기 위해 국내 과학기술분야 기관 연구데이터 리포지터리 보급 현황을 파악하여 공개된 홈페이지를 운영하는 4개 기관을 대상으로 사례 분석을 진행하였다. 해외 사례는 Nature의 Scientific Data에서 권장하는 리포지터리 목록과 국제적인 연구데이터 리포지터리 데이터베이스인 re3data를 참고하여 국내 사례와 유사한 주제 분야의 해외 기관 연구데이터 리포지터리 13개를 선정하였다.

셋째, 국내외 기관 연구데이터 리포지터리 업무 담당자를 대상으로 인터뷰를 진행하였다. 인터뷰 기준은 1) 리포지터리 담당자를 찾을 수 있거나 2) 외부에서 접근할 수 있는 별도의 주제 연구데이터 플랫폼을 운영 또는 운영이 논의되고 있는 기관으로 선정하였다. 그에 따라 별도 홈페이지를 구축하여 운영하는 국내 과학기술분야 기관 연구데이터 리포지터리 13개 중 9개를 선정하였다. 이 중 인터뷰를 수락

한 7개 기관을 대상으로 대면으로 6개 기관, 비대면으로 1개 기관 인터뷰를 진행하였다. 국외 기관 연구데이터 리포지터리 인터뷰 대상 모집단은 사례 분석을 실시한 13개 기관으로 한정하였으며, 모두 인터뷰 기준을 만족하였다. 이 중 인터뷰를 수락한 3개 기관에 대해 서면 인터뷰를 진행하였다.

1.3 선행연구

데이터 리포지터리에 대한 선행연구로는 우선 김태중, 이상환, 신진섭(2012)의 데이터 리포지터리 검토 조건 연구를 들 수 있다. 이들은 과학 데이터의 저작권에 특별한 권리(sui generis right)를 부여하고 콘텐츠의 범위와 메타데이터 관련 사항을 결정하는 등 데이터 리포지터리 구축 시 고려해야 할 조건을 제시하였다. Waddington et al.(2013)은 클라우드 기반 저장방식이 자체적인 데이터 리포지터리를 만들기 어려운 중소 연구 조직에 유용할 것으로 보았다. 박규리, 안병균(2017)은 국가 연구데이터 리포지터리 사례를 비교, 분석하여 연구데이터 관련 이용자 교육 시행, 데이터 구축 방법 교육 시행, 국제 표준을 준용한 데이터 보존 정책 수립을 개선안으로 제안하였다. Xafis와 Labude(2019)는 보건 분야 연구데이터 리포지터리를 통한 데이터 공유 과정에서 발생하는 윤리적 문제를 연구하여 개방적 데이터 공유 과정에서 선택할 수 있는 옵션들을 제안하였다. 김우중 외(2021)는 re3data에 등록된 자연과학 분야 연구데이터 리포지터리 속성을 분석하여 국내 데이터 리포지터리 활용도를 높이기 위한 방안으로 인식 개선, 인프라 지원, 분야별 특화된 기

준의 제공을 주장하였다. 김주섭, 양성준, 김신태(2022)는 CTS(CoreTrustSeal)를 중심으로 인증 체계를 분석하여 데이터 리포지터리의 국제적 신뢰성 확보를 위한 자체 평가, 필요 정책 및 가이드라인 파악과 개발, 신청서 작성의 네 가지 단계별 추진 전략을 제공하였다. 박성은, 고영만(2022)은 국가 연구데이터 플랫폼(DataON)과 바이오 연구데이터 플랫폼(K-BDS) 간 메타데이터 상호운용성 확보를 위한 연구를 수행하였다. 두 플랫폼의 메타데이터 요소가 전부 매칭됨으로써 상호운용성 확보의 가능성을 보여주었으며, 향후 협의를 통한 세밀한 매칭 조정 역시 필요할 것으로 보았다. Das et al.(2022)은 캐나다 몬트리올 신경 연구 병원의 임상 생체표본 및 유전자 리포지터리인 C-BIG 리포지터리를 분석하여 향후 데이터 쿼리 최적화, 사용자 경험을 바탕으로 한 지속적인 업데이트, 표준화 등의 개선안을 제안하였다. Thoegersen과 Borlund(2022)는 데이터 리포지터리를 통한 데이터 공유에 대한 연구자들의 태도를 분석하여 데이터 공유라는 용어의 모호성과 각기 다른 정의가 연구자의 태도 분석을 어렵게 만든다고 보았으며, 데이터 공유 유형을 구체화한다면 연구자의 태도를 이해하는 데에 더 도움이 될 것으로 보았다. Xu et al.(2022)은 텍사스 지역 고등 교육 기관의 통합 연구데이터 리포지터리인 Texas Data Repository(TDR)를 사용한 적 없는 연구자들을 대상으로 업무를 수행하도록 하였다. 그 결과 많은 연구자가 메타데이터 추가 업무 수행에 실패하였으나 전반적인 이용 경험에 대해 긍정적인 표시를 나타냈으며 이에 따라 적절한 도서관의 교육이 필요함을 시사하였다.

이상을 정리해보면 특정 리포지터리를 대상으로 한 사례 연구, 리포지터리에 등록되는 연구데이터의 유형·표준·형식 등에 관한 연구, 그리고 연구데이터 공유 과정에서 리포지터리의 역할에 관한 연구로 나누어볼 수 있다. 사례 연구는 주로 국가 차원에서 운영하는 리포지터리나 대학의 리포지터를 대상으로 삼는 경우가 많았다. 과학기술정보통신부 주관 한국과학기술정보연구원(KISTI)에서 발표한 국가연구데이터플랫폼 구축 및 운영 보고서에서 제시된 국가연구데이터플랫폼 개념도에서는 기관의 연구데이터 리포지터리와의 상호 연계를 명시함으로써 통합 연구데이터 공유 플랫폼의 근간에도 결국 기관의 데이터베이스가 자리하고 있음을 분명히 하고 있다(송사광 외, 2019). 그럼에도 국내에서 기관의 연구데이터 리포지터리, 특히 대학 외의 연구기관에 대한 연구데이터 리포지터리 연구는 부족한 실정이므로 해당 영역의 연구가 보다 활발히 이루어질 필요가 있다.

2. 과학기술분야 기관 연구데이터 리포지터리

2.1 기관 연구데이터 리포지터리

기관 연구데이터 리포지터리의 정의에 대해서는 국내와 해외에서 통용되는 정의의 범위가 다소 차이를 보이는데, 이는 기관 연구데이터 리포지터리의 발전 과정에서 그 원인을 찾을 수 있다. 먼저 해외의 경우 2000년대 ePrints를 시작으로 DSpace, Digital Commons 플랫폼 출시와 함께 기관 리포지터리가 본격적으로 널

리 받아들여지면서(Bluh & Hepfer, 2013), 기관 리포지터리에 대한 다양한 정의가 등장하였다. Lynch와 Lippincott(2005)는 기관 리포지터리에 대해 두 가지 다른 견해의 정의가 있음을 언급하였다. 하나는 주로 출판물의 e-prints와 같이 오픈액세스 출판물 제공에 초점을 맞추어 리포지터리를 바라보는 관점이다. 다른 하나는 기관 리포지터리를 출판물을 비롯한 데이터 세트, 영상, 학습 개체, 소프트웨어 등 지적 작업의 결과물을 보관하는 곳으로 바라보는 관점이다. Clobridge(2010)는 리포지터리를 디지털 자산 관리 시스템, 기관 리포지터리, 디지털 리포지터리로 나누면서 그 중 기관 리포지터리에 대해 단일 기관의 디지털 개체 및 관련 메타데이터 집합체로 정의하였다. 이러한 정의를 보면 일반적으로 해외에서는 기관 리포지터리에서 관리되는 자료로 연구데이터에 해당하는 데이터세트나 이미지, 영상 등을 포함하고 있다.

한편, 국내에서는 국립중앙도서관에서 OAK(Open Access Korea) 국가 리포지터리 보급 사업을 추진하며 기관 리포지터리 확산을 위한 노력을 펼치고 있다. OAK 리포지터리는 공공 기관, 대학, 연구소 등의 기관에서 생산한 디지털 지식정보를 보존 및 관리하여 국내외 이용자들의 자유로운 접근과 활용으로 지식정보를 확산하기 위한 오픈액세스 아카이브를 말한다(OAK 국가리포지터리, 발행년불명). 그러나 이러한 포괄적 정의와 달리 실제 보급대상으로 선정되어 기관 리포지터리를 구축한 기관들은 모두 소속 연구자들이 출판한 논문을 주요 대상으로 학위논문, 보고서 일부를 등록하고 있다. 기관의 연구데이터는 이와 별개로 기관 연구데이터 리포지터리, 데이터 플랫폼 등의 명칭 하

에 관리되고 있는데, 이는 해외와 달리 기관 리포지터리와 기관 연구데이터 리포지터리가 각기 다른 중심축을 두고 확산되었기 때문이다. 기관 리포지터리가 대학도서관을 중심으로 이미 출판된 논문과 학위논문의 수집·공유를 목적으로 발전해왔다면, 기관 연구데이터 리포지터리는 연구기관을 중심으로 데이터 혹은 성과 부서에서 연구 성과 관리의 일환으로 연구데이터를 관리하고자 구축된 것이라 볼 수 있다. 따라서 이론상으로는 포함 관계에 있는 기관 리포지터리와 기관 연구데이터 리포지터리가 실제 국내에서는 구분되어 사용되고 있는 실정이다.

따라서 본 연구에서는 국내 상황을 고려하여 기관 리포지터리와는 구분된 기관 연구데이터 리포지터리는 표현을 사용하였다. 즉 기관의 연구자가 연구 수행 과정 전반에서 생성·수집한 데이터이자 연구 결과를 검증하고 향후 재이용하는 데에 필요한 모든 연구데이터를 관리, 보존, 공유하기 위해 기관 단위에서 운영하는 디지털 저장소로 기관 연구데이터 리포지터를 정의하였다.

2.2 기관 연구데이터 리포지터리의 역할과 기능

보다 포괄적으로 기관 리포지터리의 역할과 기능을 먼저 살펴보면 미국도서관협회(American Library Association, ALA)에서는 기관 리포지터리의 역할로 학술 커뮤니케이션의 위기에 대한 대응, 학술 정보에의 접근 확대, 지적 결과물의 깊이와 폭 강조, '정보 자유화' 운동에의 가속화, 차세대 대학의 전자 저널, 저자의 비용 감소의 여섯 가지를 기술하였다(Bluh & Hepfer, 2013). Kindling et al.(2017)은 연구데이터 리

포지터리가 연구데이터의 공유 및 재이용을 실현하는 데 크게 이바지한다고 보았다.

또한 대형 출판사들은 연구자들의 연구 결과물을 출판할 때 과학 연구데이터의 제출을 요구하는 추세에 있다(김선태 외, 2010). 이러한 출판사들의 요구는 리포지터리의 필요성을 뒷받침해주며, 저널에서 제시하는 '적절한' 리포지터리의 기준은 리포지터리가 수행해야 할 역할과 기능을 안내해주는 기준으로 바라볼 수 있다. Scientific Data에서 요구하는 적절한 리포지터리의 기능은 총 다섯 가지로, 최소 5년간의 장기 지속·보존 보장, 안정적 영구 식별자 제공, 대중에의 장벽 없는 접근성 제공, CC0·CC-BY 및 이와 동등한 라이선스 지원, 필요한 경우에 한한 오픈 액세스 데이터 제공이 있다(Scientific Data, n.d.).

기관 연구데이터 리포지터리는 기관 리포지터리와 연구데이터 리포지터리의 역할과 기능을 모두 수행한다는 점에서 기관의 연구자들이 연구데이터를 보존, 공유, 재이용할 수 있는 각종 제반 서비스를 제공하고 나아가 기관 자체의 학문적 우수성을 가시화하여 홍보 도구로서의 역할을 가능케 한다. 또한 기관이 개별 연구자 수준에서 사장될 수 있는 연구데이터를 체계적으로 관리하여 장기 보존 및 공유를 가능하게 함으로써 연구데이터 자체에 대한 가치를 향상시킨다. 이러한 관리가 원활히 이루어질 때 시간이 지나도 해당 연구데이터를 찾아보고 이해할 수 있으며, 동일한 연구데이터를 재생산하지 않아도 된다는 점에서 연구자 본인에게도, 다른 연구자에게도 경제적 효율성을 가져다준다. 더불어 공유한 연구데이터를 재이용함으로써 새로운 연구 성과가 나타날 수 있다는 점에서 과

학 발전의 미래에도 긍정적인 역할을 한다.

2.3 기관 연구데이터 리포지터리 소프트웨어

현재 국내 기관 연구데이터 리포지터리 기반 시스템으로는 KISTI에서 개발한 표준형 기관 연구데이터 리포지터리인 NaRDA가 있다. NaRDA는 연구자의 연구데이터 관리 활동 주기를 고려하여 설계 및 구현되었으며, 그에 따라 연구데이터 관리 계획 단계·연구 수행 중 연구데이터 관리 단계·연구데이터 등록 및 공유 단계로 구분된다(엄정호, 2021). 연구데이터 관리 계획 단계에서는 DMP 작성 양식 및 등록을 지원하며, 연구 수행 중 단계에서는 연구데이터의 등록 및 관리, 타 연구자의 연구데이터 신청 및 다운로드를 지원한다. 또한 iRODS 시스템을 기반으로 사용자에게 연구데이터 저장 기능을 제공한다(엄정호, 2021). 연구데이터 등록 및 공유 단계에서는 연구자 종료된 후에 메타데이터를 입력하고 파일을 첨부하여 연구데이터를 등록하는 단계로 DOI 출판 기능을 함께 제공하고 있다. 또한 DataCite, 더블린 코어, DCAT 등 여러 메타데이터 표준을 참고한 자체 표준을 만들어 사용하고 있으며, 보급 기관별로 다른 메타데이터를 적용할 수 있다. NaRDA 시스템은 현재 과학기술분야 정부출연연구기관을 위주로 23개 기관에 보급되어 있으며, 향후 타 연구기관과 대학으로의 확장 및 신뢰성 확보를 위한 CTS 인증을 준비하고 있다.

이 외에도 별도의 유료 구독을 통해 연구데이터 리포지터리를 구축할 수 있는 DataWorks 소프트웨어가 있다(아르고넷, 발행년불명). 다만 아직까지는 DataWorks 단독 도입 사례 1건

을 제외하면 NaRDA를 도입한 기관에서 기관 맞춤형 커스터마이징을 진행한 사례가 대부분이다. 이러한 국내 기관 연구데이터 리포지터리 소프트웨어의 개발과 보급은 대학보다는 연구기관 위주로 이루어지고 있으며, 연구 활동 주기 전반을 포괄하는 시스템의 특성상 내부 연구 성과 관리 시스템과 연계되면서 내부 연구 데이터 관리용 리포지터리로 운영되는 특징을 보인다. 또한 도서관이 아닌 성과 관리 부서, 연구데이터 담당 부서를 중심으로 운영되고 있어 기관 연구데이터 리포지터리라는 표현 대신 데이터 플랫폼, 데이터 센터 등의 상이한 표현들이 혼용되는 상황이다.

3. 기관 연구데이터 리포지터리 사례 분석 및 현황 조사

3.1 국내외 기관 연구데이터 리포지터리 사례 분석

3.1.1 국내 사례

2022년까지 NaRDA가 보급된 기관은 총 23개이며 이 중 5개는 독자적 리포지터리 홈페이지를 운영하는 것이 아닌 KISTI의 DataON 서비스에 포함된 형태로 운영되고 있다. 이에 국내 사례는 현재 기관 연구데이터 리포지터리 사이트를 독자적으로 운영하는 18개 기관 중 대학과 비(非) 과학기술분야 6개 기관을 제외하고 현재 외부 이용자에게 리포지터리 홈페이지가 공개된 극지연구소, 한국과학기술정보연구원, 농촌진흥청, 한국지질자원연구원의 4개 기관을 사례 분석 대상으로 선정하였다.

3.1.2 국외 사례

해외의 주요 저널에서는 논문 투고 시에 적절한 리포지터리에 연구데이터를 기탁할 수 있도록 안내하고 있다. 이 중 Nature의 Scientific Data의 권장 리포지터리 목록을 중심으로 연구데이터 리포지터리 등록 사이트 re3data를 참고하였다. Scientific Data는 저널 영향력지수 8.5로 상위 25%에 해당할 만큼 권위 있는 연구데이터 학술지라는 점과 주제별 연구데이터 기탁을 위한 신뢰할 수 있는 리포지터리 사이트를 제공한다는 점에서 주요 참고원으로 선정하였다. re3data에 등록된 리포지터리는 법인에 의해 운영되는 지속 가능한 기관이어야 하고, 무엇보다 수집의 초점이 연구데이터에 있어야 한다는 점에서 기관의 연구데이터 리포지터를 대상으로 하는 본 연구에 참고하기에 적절하다고 판단하였다. 현재까지 re3data에 등록된 리포지터리는 3,000건 이상으로, Scientific Data를 비롯해 Horizon 2020의 과학 출판 및 데이터에 대한 오픈액세스 가이드라인에서도 사용이 권장된다.

현재 Scientific Data에서 권장하는 122개의 리포지터리 목록 가운데, 국내 과학기술분야 기관 연구데이터 리포지터리와 유사한 주제 분야(환경, 극지, 지질, 한의학, 화학)의 리포지터리 5개를 일차적으로 선정하였다. 이 중 한의학 분야는 임상의학으로 포괄하였으며, NERC Data Centres의 경우 리포지터리 개편으로 다섯 개의 개별 연구데이터 리포지터리로 세분되었다. 그에 따라 그 중 국립환경과학원과 가장 주제 연관성이 높은 Environmental Information Data Centre를 선택하였다.

Scientific Data에서 동일 주제를 찾을 수 없

는 경우 re3data의 브라우징 검색 기능을 이용하여 유사한 주제 분야를 검색하였다. 각 주제별로 산림과 농업 분야 각 148개, 공학 분야 674개, 전자통신 분야 3개, 운송 분야 20개, 핵에너지 분야 54개 기관이 도출되었다. 그런 뒤 필터 기능을 통해 영어권 리포지터리로 한정하였다. 국내와 마찬가지로 대학은 제외하였으며, 필터에 'institutional'로 분류되지 않아도 특정 기관에서 서비스하는 경우는 기관 연구데이터 리포지터리로 보았다. 이러한 과정을 통하여 산림과 농업, 공학 분야에서 각 2개, 운송과 핵에너지

분야에서 각 1개 등 8개의 기관 연구데이터 리포지터리를 선정하였다.

최종적으로 캐나다 산림청의 National Forestry Database, 미국 농무부의 Forest Service Research Data Archive를 비롯하여, 총 13개 기관을 국외 사례 분석 대상으로 선정하였다(〈표 1〉 참조). 〈표 1〉에서 기관 연구데이터 리포지터리의 명칭이 없거나 확인되지 않은 국내 기관은 '기관명 + 연구데이터 리포지터리' 형태로 기재하였다.

〈표 1〉 주제별 국외 기관 연구데이터 리포지터리 목록

| 국내 기관 연구데이터 리포지터리 | 주제분야 | 국외 기관 연구데이터 리포지터리* |
|-------------------------------------|---|--|
| DataFos | 산림 (Forestry) | R National Forestry Database (NFD) R Forest Service Research Data Archive |
| 국립환경과학원 연구데이터 리포지터리 | 환경 (Environment) | S Environmental Information Data Centre (EIDC) |
| Korea Polar Data Center (KPDC) | 극지 (Polar) | S Australian Antarctic Data Centre (AADC) |
| 농촌진흥청 연구데이터 리포지터리 | 농업 (Agriculture) | R Ag Data Commons R Data INRAE |
| 한국기초과학지원연구원 연구데이터 리포지터리 | 기초공학 (Engineering Sciences, Bio-materials, Computer Science) | R Materials Data Repository (MDR) |
| 한국과학기술연구원 강릉분원 연구데이터 리포지터리 | | R Rossendorf Data Repository (RODARE) |
| DataON | | |
| EDR (ETRI Data Repository) | 전자통신 (Electronic communication) | - (조건에 맞는 기관 없음) |
| GDR (Geoscience Data Repository) | 지질 (Geology) | S British Geological Survey (BGS) |
| 한국철도기술연구원 연구데이터 리포지터리 | 운송 (Transportation) | R Repository & Open Science Access Portal (ROSA P) |
| KMDR (한의학 연구데이터 리포지터리) | 한의학, 임상 (Clinical Science) | S ClinicalTrials.gov |
| 한국핵융합에너지연구원 연구데이터 리포지터리 | 핵에너지 (Nuclear Energy) | R Atomic and Molecular Data Information System (AMDIS) |
| 한국화학연구원 연구데이터 리포지터리 | 화학 (Chemistry) | S ioChem-BD Computational Chemistry Datasets |

*약어 풀이: S Scientific Data R re3data

3.1.3 분석 결과

국내 기관 연구데이터 리포지터리 홈페이지는 키워드 검색과 고급 검색, 브라우징 검색을 모두 지원하는 반면에 국외 기관은 전반적으로 고급 검색을 지원하지 않는 경향을 보인다. 연구데이터 등록 기능에서도 국내 기관은 모두 이를 포함하는 반면에 일부 국외 기관 연구데이터 리포지터리는 등록 기능을 찾아볼 수 없다. 다만 로그인 한 상태에서만 연구데이터 등록 탭이 보이는 경우가 있을 수 있어 무조건적으로 연구데이터 등록 기능을 제공하지 않는다고 단정하기에는 어려운 측면이 있다.

이용자를 위한 가이드 제공 여부는 국내와 국외 각 한 개씩을 제외하고 가이드를 제공한다. 홈페이지 내에서 HTML 형식으로 간단한 이용 가이드를 제공하는 방식부터 상세한 매뉴얼 페이지를 별도로 운영(DataON, MDR)하거나 영상 가이드를 제공(EIDC)하는 등 가이드의 상세함이나 제공 방식은 리포지터리마다 다양한 모습을 보인다.

연구데이터에 대한 DOI 제공 여부와 관련하여, 국내에서는 모두 DOI 발급을 지원하고 있으나 국외 일부 리포지터리는 DOI 발급을 지원하지 않고 있다. ClinicalTrials.gov는 예외적으로 DOI 대신 자체 식별자인 NCT를 사용하여 NCT 뒤에 번호를 부여하는 방식으로 연구데이터를 관리하고 있다.

리포지터리 홈페이지에 정책 게시 여부와 관련하여 국내 한 개, 국외 두 개 리포지터리를 제외하고 전부 홈페이지에서 정책을 확인할 수 있다. 다만 국내 리포지터리에서 게시하고 있는 정책이 일반적인 연구데이터 등록 및 개인정보 보호 정책인 반면, 국외 리포지터리는 연

구데이터 등록과 인용, 주제어, 관련 법령 등의 상세한 정보를 폭넓게 포함하고 있다는 것이 그 차이이다.

그 외 사항과 관련하여서는 국내의 기관 연구데이터 리포지터리 모두 기관 특성에 맞추어 메타데이터를 둘 이상 조합하여 사용하는 경향을 보인다. 데이터 품질 인증을 받은 기관 연구데이터 리포지터리는 세 곳뿐인데 국내에서는 지오빅데이터 오픈 플랫폼이 유일하게 한국데이터산업진흥원의 데이터 품질 인증을 받았고, 국외에서는 EIDC와 AADC가 CTS 인증을 받은 상태이다. 이상을 정리하면 <표 2>와 같다.

3.2 국내외 기관 연구데이터 리포지터리 담당자 인터뷰

국내외 과학기술분야 기관 연구데이터 리포지터리 운영 현황을 파악하기 위해 담당자 인터뷰를 실시하였다. 홈페이지 조직도 또는 리포지터리 구축 사업 보고서에서 담당자가 명시되어 있거나, 오픈 연구데이터 플랫폼을 운영하는 기관을 기준으로 삼았다. 그 근거는 주로 기관 내부를 대상으로만 리포지터리를 공개하는 상황에서 외부에 담당자를 명시하는 것은 리포지터리가 단기 사업으로 그치지 않을 가능성이 크고, 오픈 연구데이터 플랫폼의 운영은 단순히 국내 연구데이터를 수집하는 역할 이상으로 연구데이터 공유에 의지를 갖고 있는 기관이라고 판단함에 있다.

국내의 경우 독자적으로 홈페이지를 구축하여 운영하는 기관 13개 중 상기 조건을 만족하는 9개 기관을 인터뷰 대상으로 선정하였다. 최종적으로 인터뷰를 수락한 7개 기관의 담당자

〈표 2〉 국내외 기관 연구데이터 리포지터리 기능 및 제반 사항

| 구분 | 홈페이지 제공 기능 | | | | | | | 기타 제반 사항 | | |
|-----------|--------------------------------------|-------|------|--------|-----|-----|------|----------|---------------------------|-----|
| | 키워드 검색 | 고급 검색 | 브라우저 | 데이터 등록 | 가이드 | DOI | 정책 | 메타데이터 | 인증 | |
| 국내 | KPDC | ○ | ○ | ○ | ○ | X | ○ | ○ | DIF | X |
| | DataON | ○ | ○ | ○ | ○ | ○ | ○ | ○ | DCAT, DataCite, 자체 개발 | X |
| | 농촌진흥청 기관 연구데이터 리포지터리 | ○ | ○ | ○ | ○ | ○ | ○ | X | 자체 개발 | X |
| | 지오빅데이터 오픈 플랫폼 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | DC, 자체 개발 | 국내 |
| 국외 | NFD | X | X | ○ | X | ○ | X | X | FGDC/CSDGM | X |
| | Forest Service Research Data Archive | ○ | X | ○ | ○ | ○ | ○ | ○ | FGDC/CSDGM, ISO 19115 | X |
| | EIDC | ○ | X | ○ | ○ | ○ | ○ | ○ | DC, DataCite, ISO 19115 | CTS |
| | AADC | ○ | X | ○ | ○ | ○ | ○ | ○ | DIF, ISO 19115 | CTS |
| | Ag Data Commons | ○ | X | ○ | ○ | ○ | ○ | ○ | DCAT, DataCite, ISO 19115 | X |
| | Data INRAE | ○ | ○ | ○ | X | ○ | ○ | ○ | DDI, DataCite, ISA-Tab | X |
| | MDR | ○ | X | ○ | ○ | ○ | ○ | ○ | 자체 개발 | X |
| | RODARE | ○ | X | △** | ○ | X | ○ | ○ | DC, DataCite | X |
| | BGS | ○ | X | ○ | X | ○ | X | ○ | ISO 19115 | X |
| | ROSA P | ○ | ○ | ○ | ○ | ○ | ○ | ○ | DC, DCAT | X |
| | ClinicalTrials.gov | ○ | ○ | ○ | ○ | ○ | X*** | ○ | DC | X |
| | AMDIS* | △ | △ | △ | X | △ | X | X | 알 수 없음 | X |
| ioChem-BD | ○ | X | ○ | ○ | ○ | ○ | ○ | DC | X | |

* 리포지터리 내에 데이터베이스별로 검색 기능 및 가이드 제공 여부 상이

** 이용자, 담당자가 구성한 컬렉션이 일부 브라우저 기능을 수행

*** DOI 대신 자체 식별자(NCT) 사용

8인을 대상으로 인터뷰를 진행하였다. 인터뷰는 2023년 4월 중 실시되었으며, 7인은 대면, 1인은 실시간 비대면 인터뷰로 진행되었다. 인터뷰 방식은 사전에 공통된 질문지를 송부한 뒤 인터뷰 기관의 특성이나 담당자 답변에 따라 추가 질의를 수행하는 반구조화된 방식으로 수행되었다. 모든 인터뷰는 응답자들의 동의를 받아 본 연구에 사용될 목적으로만 녹취하였으며 평균 약 85분이 소요되었다.

국외 기관 연구데이터 리포지터리 인터뷰 모집단은 사례 분석을 실시한 13개 기관으로 한정하였으며, 모두 인터뷰 기준을 만족하였다. 이중 3개 기관의 담당자가 인터뷰를 수락하여 메

일을 통해 서면 인터뷰를 진행하였다. 국내외 전체 인터뷰 대상자와 관련한 정보를 종합하여 정리하면 〈표 3〉과 같으며, 전공으로 인해 기관 및 담당자가 특정될 수 있는 경우 포괄적인 전공으로 수정하였다.

인터뷰 질문지의 구조는 운영 사항, 담당자 업무 및 이용자 인식, 정책, 연구데이터 관련 사항의 네 측면으로 구분하였다(〈표 4〉 참조). 운영 측면에서는 구축 시기와 계기, 형태, 구축 현황 등의 기본적인 사항을 포함하였다. 담당자 업무 및 이용자 인식에 관한 질문은 대학 연구자들의 연구데이터 관리 및 공유 인식을 연구한 김지현(2015)과 연구자의 연구데이터

〈표 3〉 기관 연구데이터 리포지터리 인터뷰 대상자

| 번호 | 직종 | 전공 | 리포지터리 업무 담당 기간 |
|-------|------------------|-----------------------------|----------------|
| 응답자 A | 연구원 | 생명과학 | 2년 |
| 응답자 B | 연구원 | 생명과학 | 2년 |
| 응답자 C | 연구원 | 전자계산, 대기과학(석사) | 8년 |
| 응답자 D | 연구원 | 화학, 전산(석사), 생명공학(박사) | 6년 |
| 응답자 E | 연구원 | 컴퓨터공학 | 5년 |
| 응답자 F | 연구원 | 데이터베이스 | 5년 |
| 응답자 G | 연구원 | 데이터베이스 | 5년 |
| 응답자 H | 사서 | 문헌정보학 | 2년 |
| 응답자 I | Administration | 미응답 | 미응답 |
| 응답자 J | Data Coordinator | 식물학 | 25년 |
| 응답자 K | Data Curator | 문헌정보학(석사) (데이터 큐레이션 수료증) | 미응답 |

〈표 4〉 인터뷰 질문지 구조

| 주제 | 구분 | 질문 사항 | 문항 수 |
|-----------------|--------|--|------|
| 인적사항 | - | 소속 기관, 부서명, 직급/직함, 담당 업무 및 기간, 전공 | - |
| 운영 사항 | 현황 파악 | 리포지터리 구축 시기 구축 동기 및 원인 운영 형태 이용률 및 데이터양 기탁되는 데이터 유형 리포지터리의 특징 제공 서비스 역할 수행정도에 대한 자가 평가 | 15 |
| | 활성화 방안 | 소속 기관 및 국가의 지원을 바라는 사항 | 1 |
| 담당자 업무 및 이용자 인식 | 현황 파악 | 일상적으로 수행하는 업무 가장 많은 노동력을 필요로 하는 업무 운영 시 겪은 어려움 협력·협조가 필요했던 부서나 기관 이용자 요구 사항 및 인식 정도 기탁된 연구데이터 활용 사례 이용 교육 및 홍보 담당자 역량 | 9 |
| | 활성화 방안 | 이용자 교육 보완 사항 홍보 계획 리포지터리 관련 부서의 업무 분장 및 어려움 | 3 |
| 정책 사항 | 현황 파악 | 전반적인 운영 정책 연구데이터 의무 기탁 정책 연구데이터의 권리 및 책임 규정 인센티브 제도 연구데이터 공유·활용을 유도하는 제도 | 5 |

| 주제 | 구분 | 질문 사항 | 문항 수 |
|---------------------------|--------|-------------------------|------|
| 연구데이터 관련 사항 | 활성화 방안 | 연구데이터 의무 기탁 정책의 효과 | 3 |
| | | 인센티브 제도의 효과 | |
| | | 그 외 연구데이터 공유·활용 제도의 필요성 | |
| | 현황 파악 | 품질 관리 및 장기 보존 조치 | 6 |
| | | 보안 조치 | |
| | | 메타데이터 표준 | |
| | | 이용자의 메타데이터 기술 정도 | |
| | | 큐레이션 제공 여부 | |
| | 활성화 방안 | 품질 관리 및 장기 보존을 위한 개선 사항 | 5 |
| | | 보안 측면의 개선 사항 | |
| 메타데이터 기술 관련 개선 사항 | | | |
| 큐레이션의 어려움 및 보완점 | | | |
| 그 외 연구데이터 재이용 활성화를 위한 서비스 | | | |
| 종합 | - | 기관 연구데이터 리포지터리의 발전 방향 | 1 |

관리 관행과 기관 및 사서의 지원 요건을 제안한 Tenopir et al.(2020)의 연구를 참고하여 구성하였다. 정책 측면의 질문은 공공부문 연구데이터 활용 방안을 제안한 신은정 외(2019)의 연구를 참고하여 연구데이터 권리 보장 현황과, 연구데이터 공유 정책에 관한 질문을 포함하였다. 연구데이터 관련 사항은 연구데이터 관리 핵심 요소와 기능적 요구사항을 도출한 김주섭, 김선태, 최상기(2019)의 연구와 해양과학분야 연구데이터 리포지터리 JOISS를 FAIR 원칙의 관점에서 평가한 송태윤 외(2022)의 연구, 그리고 NaRDA 연구데이터 리포지터리 개발 보고서(엄정호, 2020)를 참고하여 연구데이터 품질 관리와 보존 조치에 관한 질문으로 구성하였다.

3.2.1 운영 사항

우선 국내 기관 연구데이터 리포지터리 구축은 연구데이터 보존 의지가 있던 기관에서 과학기술정보통신부의 연구데이터 활용체계 구

축 전략, 연구개발혁신법의 제정 등을 근거 삼아 시작된 것이 일반적이었다.

운영 형태의 경우 전체 외부 공개(2), 일부 외부 공개 형태(5), 내부 공개 형태(3)로 구분되었다. 전체 외부 공개로 운영한다고 답한 응답자 K는 이러한 형태로 운영할 때 개인 식별 정보나 민감 정보, 국가 안보 관련 정보가 포함되어 있지 않은지 확인해야 한다고 조언하였다. 나머지는 외부 공개라고 하더라도 내부 이용자와 외부 이용자가 이용할 수 있는 기능에는 차이가 있었다. 전체 외부 공개가 어려운 이유로는 해당 연구과제가 종료 후 논문 출판이 예정된 경우, 품질/리스크 관리의 필요성, 특허 문제 등을 그 이유로 꼽았다.

연구데이터 축적 과정에 있어서 국내 기관 대다수가 기관 내 연구 관리 시스템과 구조적·기술적으로 연계되어 DMP 작성과 연구데이터 등록 단계에서 자연스럽게 리포지터리를 이용하도록 유도하는 경향을 보였다. 리포지터리에서 승인되어 워터마크가 찍힌 DMP만을 연구 계획서

자료로 인정하거나(응답자 H), 성과 지표로 인정받기 위해서는 리포지터리를 이용할 수밖에 없도록 유도하는 기관(응답자 B)도 있었다.

이렇게 기탁되는 연구데이터는 기관을 막론하고 주로 엑셀 형태의 실험데이터가 큰 비중을 차지하고 있었으며, 그 외에 전문 소프트웨어에서 추출된 파일, 이미지, 보고서 등이 주로 기탁되고 있었다. 인터뷰 기관 중 여섯 개 기관이 동일한 시스템에 기반함에도 불구하고 기관별 주제 분야에 따라 해당 분야에서 주로 생산되는 연구데이터에 특화된 분류 체계를 갖추고 있었으며 응답자 D와 F는 신청자에 한하여 분석 서비스도 제공하고 있었다. 국내 기관 연구데이터 리포지터리의 경우 DMP 제출 대상 과제들은 연구데이터 제출이 의무화되는 만큼 DMP 입력과 최종 연구데이터 등록 단계에서 가장 활발히 연구데이터 리포지터리가 이용되고 있었다.

3.2.2 담당자 업무

리포지터리 관련 일상적인 업무와 관련하여서는 시스템 관리 및 유지 보수, 작성된 메타데이터 검토, 관련 문서와의 연계 작업 등이 언급되었다. 많은 노동력을 필요로 하는 업무로는 담당자 다수가 연구자와의 소통 부분을 꼽았으며, 연구데이터 전처리, 표준화, 교육, 오류 수정 등이 언급되었다.

리포지터리 운영 과정에서의 어려움으로는 국내외를 막론하고 인력 및 예산 부족을 토로하였으며, 국내 응답자들은 제도적 근거 부족과 기관장의 의지에 달린 정책의 비일관성, 표준화, 업무에 대한 보상 및 비전 부재 등을 추가로 언급하였다. 또한 당장 성과를 보여줄 수 있는 사업이 아닌 만큼 규정 외의 근거도 부족하다는

점이 운영상의 어려움으로 작용하였다.

“At our current technology levels all the data curation tasks are very manual and require human labor and intervention. ... Our unit lacks the financial and technical resources to solve many of these. (현재 기술 수준에서 모든 데이터 큐레이션 작업-민감 정보 검토, 메타데이터 요건 충족 여부 검토, DOI 생성 등-은 매우 수동적이며 인력과 개입이 필요합니다. ... 우리 부서는 이러한 문제를 해결할 재정 및 기술 자원이 부족합니다.)” - 응답자 K

이러한 어려움은 전반적인 운영 활성화 방안에 대한 답변에서도 반복하여 나타나 해당 문제에 대한 해결의 필요성이 두드러졌다. 추가적인 활성화 방안으로는 연구데이터 관련 기관장 평가 항목 신설(응답자 A)과 주제 리포지터리의 확장 필요성(응답자 G)이 언급되었다.

리포지터리 이용자의 주요 요구로는 기초적인 도움 요청과 이용자 편의성 관련 요구가 있었다. 응답자 I는 사용자 조사에서 검색 옵션 개선, 표준화를 통한 연구데이터의 형식·내용 개선, 제출 프로세스 간소화 등의 피드백을 받았다고 답하였다.

인식 측면에서 젊은 연구자 위주로 리포지터리에 대한 긍정적인 인식과 이용이 나타나고 있으나, 전반적으로 연구자가 능동적으로 연구데이터를 등록하고 활용하려는 움직임이 아직은 부족하다고 보았다.

리포지터리에 등록된 연구데이터의 재이용 사례와 관련해서는 경진대회를 개최했던 기관에서 단백체 데이터를 활용한 신약 개발 애플

리케이션 제작 사례(응답자 F)와 지리 정보 데이터를 바탕으로 한 초등학생용 야외 수업 교재 제작 사례(응답자 G)를 확인할 수 있었다.

리포지터리 홍보 및 이용 교육과 관련하여 다섯 기관이 교육을 시행하고 있었으며 상시 교육을 진행하는 기관의 경우 전반적인 연구데이터 관리 프로세스와 성과 지표 등을 모두 포함하여 교육하고 있었다(응답자 B). 연구데이터 등록 체계가 직관적이고, 연구과제 수행 과정에서 리포지터리를 자연스럽게 접하게 되는 만큼 매뉴얼 외에 별도의 교육이나 홍보를 수행하지 않는 기관이 많았다. 향후 필요한 교육으로는 연구데이터의 정의와 공유 필요성에 대한 교육이 중요하다고 보는 응답자들이 많았으며, 저작권/소유권 또는 품질 관리에 대한 교육의 필요성도 언급되었다.

리포지터리 운영과 관련하여 필요한 담당자 역량으로는 많은 담당자가 해당 연구 분야에 대한 전반적인 이해도와 연구데이터 분석·품질 관리·활용 능력을 꼽았다. 이 외에도 연구데이터 공유 인식 개선의 노력과 서비스 정신, 업무에 대한 비전, 커뮤니케이션 스킬 등이 필요 역량으로 제시되었다.

3.2.3 정책

한 개 기관을 제외하고는 전부 리포지터리와 관련한 내부 지침을 가지고 있었으며, 대개 연구데이터 수집, 관리, 공유에 대한 지침 안에 리포지터리 관련 조항을 포함하는 형태로 존재하고 있었다.

리포지터리에 기탁되는 연구데이터의 권리 및 책임에 대해서도 기관 대부분이 CC-BY 라이선스를 기반으로 하고, 연구데이터의 소유권은 기

관 혹은 정부에 있었다. 연구데이터 출판을 지원하는 기관의 경우 연구데이터를 등록할 때 연구자별 기여 지분과 연구데이터 생성 과정에서의 역할을 작성하도록 하여 연구데이터에 대한 책임 소재와 지분을 명확히 하고 있었다(응답자 A).

리포지터리에 연구데이터를 기탁하거나 재이용한 연구자에 대한 보상 제도는 국내 다섯 기관이 시행 이력을 갖고 있었으나 현재까지 유지하는 기관은 두 기관뿐이었다. 대체로 이미 연구데이터를 활용한 논문에 성과 점수를 주고 있는 만큼 중복 보상에 대한 우려 및 평가 기준의 모호함을 근거로 신중한 태도를 보였다. 정량 평가의 경우 연구데이터 '한 건'의 정의가 모호하고, 정성 평가 역시 특정 연구데이터가 더 유용하다거나 질적으로 우수하다는 판단을 내리기 어려운 측면이 있다. 현재 보상 제도를 운영하는 기관의 경우 연구데이터를 리포지터리에 등록하거나 서비스와 연계할 때 과제 평가 가산점을 부여하거나 연구데이터 출판 성과 지표를 개발하여 리포지터리에 연구데이터를 등록·공개하도록 유도하는 방식으로 운영하고 있었다. 한편 의무 제출 대상인 연구데이터를 미등록하는 경우 감점 처리 혹은 차후 자금 지원 시 불이익을 주는 정책을 운영하는 기관도 있었다.

연구데이터 기탁 의무화 정책이 리포지터리 활성화 방안에 미치는 영향에 대해서는 응답자 2인이 중립적 반응을 보였으며 나머지 응답자들은 전부 긍정적인 입장을 보였다. 의무화 정책이 필요하다고 답한 응답자 G는 과거 리포지터리 구축 실패 원인이 규정이나 정책의 부재에 있다고 보았다. 또한 응답자 E와 K도 현재

대다수 연구자가 리포지터리를 이용하는 원인 역시 DMP의 제출 의무화와 규정예의 리포지터리 이용 명시에 있다고 보았다.

반면 증립적 입장의 담당자들은 제도 자체의 필요성은 인정하면서도 연구데이터 공개까지 의무화하는 것은 시기상조로 오히려 중요한 연구데이터를 아예 등록하지 않는 부작용을 우려했다(응답자 H). 또한 탑-다운 방식에 대해 회의적이거나 반발을 보이는 연구자가 있다는 점(응답자 F)에서도 신중한 태도를 보였다.

연구데이터 기탁 및 재이용에 대한 보상 정책이 리포지터리 활성화에 미치는 영향에 대해서도 응답자 대부분이 긍정적으로 답변하였으나 실제 운영에 있어서는 신중한 자세를 취하였다.

“논문처럼 성과로 인정되면 쉽게 될 건데, 부작용이 있을 수밖에 없을 것 같아요. ... 예를 들어 SCI 상위 저널에 게재가 쉬운 분야가 있고 어려운 분야가 있고, 이 일을 하면 논문이 한 편만 나오는 분야가 있는데, 다른 일을 하면 작년에 썼던 논문에서 숫자만 바뀌도 새로운 논문이 되는 분야가 있고...” - 응답자 E

인센티브의 종류와 관련해서도 성과 점수 반영, 금전적 포상, 인사 고과 반영, 연구데이터의 권리화 등 다양한 의견이 제시되었다. 보상 정책에 부정적 견해를 보인 응답자 H는 중요 연구데이터는 보상 정책이 있어도 공개하지 않을 것이므로 큰 효과가 없을 것으로 보았다. 마찬가지로 응답자 K도 인용, 표창, 비현금성 상만으로는 연구자에게 충분하지 않을 수 있다고 답하였다. 다만 그럼에도 정부 기관이 비현금성 인

센티브를 마련할 필요는 있으며, 대학 수준에서부터 교육과 연구데이터 출판 지표 반영이 이루어져야 한다고 답하였다.

3.2.4 연구데이터

연구데이터의 품질 관리와 관련해서는 대체로 인력 부족으로 인해 필수 메타데이터 내용 및 파일 실행 여부 정도만 확인하고 있었다. 반면 메타데이터 입력 시 시스템에서 문제를 자동으로 감지할 수 있는 기능이 있는 자체 검토프로세스를 갖춘 기관, 메타데이터의 DOI 발급 절차를 자체적으로 강화하여 양질의 메타데이터 입력을 지원하고, 연구기능별 데이터 표준 매뉴얼을 작성·배포하여 수집 단계부터 통일된 방식을 따르도록 하는 기관과 같은 사례도 일부 있었다. 또한 자체적으로 표준 품질 관리 매뉴얼을 만들어 세부 연구 분야별 수집 및 조사 가이드라인을 제공하고 단위, 날짜 표시 등을 표준화하려는 계획을 세우고 있는 응답자도 있었다.

“저희 같은 경우에는 이 데이터에 대한 500자 기술을 하게끔 되어 있어요. ... 이 데이터는 왜 만들었고, 엑셀 파일이면 컬럼이 설명하는 내용은 뭐고 아니면 이 이미지가 의미하는 게 뭐고, 이런 거에 대한 설명을 500자 기술을 하지 않으면 DOI 발급이 안 돼요.” - 응답자 A

연구데이터의 내용 자체에 대한 질 관리 측면에서는 기관 내 인력 부족으로 등록된 연구데이터를 있는 그대로 둘 수밖에 없다는 견해(응답자 K)와 위원회를 통해 충분히 질 관리가 가능하다는 견해(응답자 G)가 모두 존재하였다.

장기적인 연구데이터 보존과 보안 측면에서 취해지는 조치는 모든 기관이 백업을 기본으로 하고 있었다. 인터뷰 대상 기관이 전부 정부 기관 또는 정부 부처 산하 연구기관인 만큼 기관의 보안 요구 정도가 높아 보안 측면의 우려는 상당히 낮았다.

리포지터리의 메타데이터 표준과 관련하여서는 대체로 기관 특성에 맞게 여러 표준을 조합하여 사용하고 있었다. 동일 기관 내에서도 연구데이터 유형에 따라 작성해야 하는 메타데이터가 다르기도 했다. 이러한 메타데이터 입력 성실도에 있어 연구자별 편차가 심했는데, 필수 요소라 할지라도 연구 성과와 관련되는 경우 일부러 틀리게 입력하는 연구자도 있다고 답하였다(응답자 D). 이 외에도 연구자의 메타데이터 작성을 허용하지 않고 담당 직원이 메타데이터를 생성하는 경우(응답자 K)도 있었다.

연구데이터 리포지터리 품질 관리 및 장기적 보존, 보안 측면의 개선 사항에 있어서는 표준화를 꼽은 담당자가 가장 많았으며, 정기적 시스템 관리와 연구데이터 품질 관리의 필요성 등이 언급되었다. 보안 측면에서는 지식재산권 확보 및 민감 정보에 대한 암호화 등에 대한 세부적인 기준과 민감 정보를 사전에 감지할 수 있는 기술적 도구, 연구데이터 관리자의 수준 향상이 필요하다고 보았다.

메타데이터 기술과 관련한 개선 사항에 있어 연구자의 연구 특색에 맞는 작성을 지원하고 연구자들이 같은 내용을 여러 번 기술하지 않아도 되도록 하는 시스템 자동화, 가이드라인 작성, machine-actionable한 시스템 개발 등이 제안되었다. 특히 연구자들이 연구자 본인만 알아볼 수 있는 내용으로 기술하거나 지나치게

간략하게 기술하는 경우가 있어 누구나 해당 연구데이터를 이해할 수 있는 방식으로 기술하도록 안내하는 것이 필요하다고 보았다.

최종적으로 기관 연구데이터 리포지터리가 가져야 할 목표에 대해 담당자들은 연구데이터를 (준)영구적, 체계적, 종합적으로 관리 및 활용할 수 있도록 해야 하며, CTS 인증 준비 등 신뢰성을 인증받기 위한 노력과, 장기적으로 해당 주제 분야의 연구데이터 센터로의 역할을 수행할 필요가 있다고 보았다.

4. 과학기술분야 기관 연구데이터 리포지터리 운영 개선방안

지금까지 국내의 과학기술분야 기관 연구데이터 리포지터리 사례조사와 인터뷰 분석을 통해 기관 연구데이터 리포지터리의 운영 현황 및 담당자 요구사항을 확인하였다. 이를 바탕으로 기관 연구데이터 리포지터리의 안정적이고 지속적인 운영과 신뢰성을 보장하기 위하여 리포지터리 규정 및 정책 수립, 연구데이터 공유 인식 개선, 연구데이터 품질 관리 측면의 개선방안을 제시하고자 한다.

4.1 리포지터리 규정 및 정책 수립

국내 기관 연구데이터 리포지터리 담당자들은 인터뷰 질문 전반적으로 근본적으로 정책의 부재를 문제 삼았다. 정책 수립은 박규리, 안병준(2017) 등 다수의 선행연구에서도 언급되었다. 다만, 이전 연구들이 데이터 보존 등 특정 측면에서의 정책 필요성을 논의한 데 반해 본

고에서는 좀 더 근본적으로 리포지터리의 구축 근거 및 인력, 조직 측면의 법제화가 필요함을 제안한다.

첫째, 국가연구개발정보처리기준의 지위 향상과 리포지터리 근거 규정의 명시가 필요하다. 연구데이터와 관련한 제도적 규정인 국가연구개발정보처리기준은 현재 과학기술정보통신부 고시 수준으로 외부적 구속력이 없다. 이러한 낮은 위상은 기관으로 하여금 연구데이터 관리의 중요성을 인식시키기에 부족하며, 이마저도 리포지터리에 관한 조항은 없다. 따라서 기관의 규모나 성격 등 특정 기준을 만족하는 경우 연구데이터 리포지터리를 운영해야 함을 명시하여 안정적인 운영 기반을 마련해야 한다. 또한 구체적으로 기관의 규모에 따른 최소한의 리포지터리 담당 인력을 규정하고, 각 담당자의 업무 범위도 명시해야 한다.

둘째, 기관 수준에서도 리포지터리 담당 부서 및 담당자 배치의 필요성을 포함한 상세한 리포지터리 운영 정책을 수립하여야 한다. 비교적 상세하게 리포지터리 정책을 제시한 국외 기관도 담당 부서나 인센티브 정책 등은 부재하였다. 또한 인터뷰 과정에서 많은 응답자가 리포지터리 업무에만 전념하지 못하고 있다고 답하였다. 상세한 관내 정책의 수립은 기관장의 의지에 따라 조직이 임의로 개편되거나 축소되는 위험을 줄일 수 있고 지속적이고 안정적인 리포지터리 운영을 가능케 한다. 연구데이터 관련 법의 지위가 향상될 때 이러한 기관의 정책은 더욱 안정성을 가진다.

또한 인센티브나 패널티 정책을 마련할 필요가 있다. 다만 인터뷰에서 확인했다시피 그 기준에 있어서는 많은 고려가 필요하다. 해당 제

도를 시행 중인 기관 사례를 참고할 때 인용을 기준으로 하는 것이 가장 적절해 보이거나 연구데이터를 기반으로 출간된 논문과의 이중 인센티브 부여에 대한 형평성을 고려해야 한다. 또한 연구 분야별로 연구데이터의 추출이나 출판이 상대적으로 용이한 분야와 그렇지 못한 분야가 있을 수 있어 학제 간 불균형 역시 고려 대상이다. 기존의 논문이나 저널 인용 지수만큼 연구데이터 출판과 관련된 지수가 충분하지 않은 상황이므로 구체적으로 인용 지수를 어떻게 계산할 것인가에 대해서는 신중을 기해야 한다.

4.2 연구데이터 공유 인식 개선

연구자들은 연구데이터 공유에 있어서 오사용(Cragin et al., 2010), 소유권과 제어권(Zuk et al., 2020), 기밀 유지(Mallasvik & Martins, 2020) 등에 대해 우려하며, 공유에 대한 일반적인 관행이 정착되지 않았다(Tenopir et al., 2020)고 느끼고 있다. 한편, 인터뷰 응답자들도 정책을 이용한 하향식 접근의 필요성을 주장하는 동시에 연구자들 간에 연구데이터의 정의와 공유의 필요성에 대한 인식 확산의 중요성을 인지하고 있었다.

그에 따라 첫째, 연구자를 대상으로 연구데이터의 정의와 공유의 필요성에 대한 인식을 심어주는 주기적인 교육이 필요하다. 이러한 교육의 필요성은 박규리, 안병균(2017), 김우중 외(2021), Xu et al.(2022) 등 여러 선행연구에서도 제안되었다. 특히 연구데이터의 공유가 연구자 본인에게도 유익할 수 있다는 사실을 알리는 것이 중요하다. 또한 DOI 발급을

통해 연구데이터 인용을 추적할 수 있다는 점, 기밀 유지가 필요할 시 유예 기간을 설정할 수 있다는 점 등을 함께 교육하여 연구데이터 공유 시의 각종 우려 사항을 해소시켜야 한다. 이러한 교육은 앞서 제시한 인센티브/패널티 정책과 함께 시행된다면 더욱 효과적일 수 있다.

한편 리포지터리 이용교육과 관련한 기존 연구에서는 연구자들이 데이터의 검색과 메타데이터 입력(Suhr, Dungi, & Stocker, 2020), 데이터 저장(Perrier & Barnes, 2018)에 어려움을 겪고 있음이 나타났다. 반면 인터뷰 응답자들은 별도의 교육이 필요하지 않으며 요청도 받지 않았다고 답하였다. 이러한 차이는 담당자와 연구자의 인식 차이 때문일 수도 있고, 현재 리포지터리의 기능이 단순하거나 이용자가 적기 때문일 수도 있다. 다만 앞으로 리포지터리의 이용이 활발해지고 다양한 기능을 제공할수록 연구자들의 교육 요구는 점점 증가할 것이다. 이러한 이용교육은 연구데이터 정의와 공유의 필요성에 관한 교육과 함께 수행될 수 있으며 내부 직원 대상 필수 교육으로 포함한다면 교육 참여도를 높일 수 있다. 또한 시스템 자동화와 프로세스로 개선을 통해 기능적 편의성도 향상하여 이용 장벽을 낮출 필요가 있다.

둘째, 연구데이터 우수 활용사례를 수집하고 소개함으로써 이용 실황을 가시적으로 보여줄 필요가 있다. 사례를 통해 연구자들은 연구데이터의 활용 가능성을 깨달을 수 있으며, 나아가 연구데이터가 실제로 어떻게 활용되는지 확인함으로써 불특정 다수를 대상으로 한 공유의 불안감을 줄이고, 효능감과 보람을 느낄 수 있

다. 현재 경진대회를 개최하는 지질자원연구원과 KISTI는 수상자에게만 상금을 지급하는 형태로 운영되고 있지만, 수상 결과물에 활용된 연구데이터를 공유한 원작자에게도 인센티브를 준다면 연구데이터 공유를 격려하는 동시에 인센티브 정책의 일환이 될 수 있다.

아울러 이러한 인식 개선을 위해 리포지터리 담당자 스스로도 역량을 키울 필요가 있다. 응답자들은 공통적으로 데이터 이해 능력과 커뮤니케이션 능력, 기관의 연구 분야에 대한 이해를 필요 역량으로 꼽았다. 기관의 연구 분야와 전산, 데이터에 대한 폭넓은 이해는 담당자들이 해당 연구 분야에 적절한 리포지터리를 구성하고 기능을 개선하는 데에 도움이 된다. 뿐만 아니라 커뮤니케이션 능력도 담당자가 갖추어야 할 중요한 역량 중 하나이다. 리포지터리라는 개념이 문헌정보학 외의 전공자들에게는 낯설 수 있는 만큼 기본적으로 리포지터리의 정의와 데이터 공유의 필요성을 연구자들에게 이해시키기 위해서 커뮤니케이션 능력은 매우 중요한 역량이라고 볼 수 있다.

4.3 연구데이터 품질 관리 강화

리포지터리에 실제 기탁되는 연구데이터와 관련하여서는 품질 관리가 제대로 수행되어야 한다. 연구데이터를 열람 및 활용 가능성을 높이기 위해서는 첫째로 연구자, 담당자, 위원회의 긴밀한 상호작용을 통해 연구데이터의 품질을 높여야 한다. 우선 연구자 스스로 연구데이터를 등록하고 메타데이터를 입력하는 과정에서 성실하게 임해야 한다. 다만 연구자는 자신의 연구에 익숙하므로 본인이 알고 있는 것

을 다른 이들도 알고 있을 것으로 단정할 수 있다.

이를 리포지터리 담당자가 보완해야 한다. 담당자들은 대개 연구자의 연구 분야 전공자가 아니므로 기술 내용이 충분한지 비교적 쉽게 알 수 있다. 담당자는 그 외에도 연구자가 놓칠 수 있는 세세한 사항들을 살펴보고 수정해야 한다. 연구자가 상세하게 메타데이터를 기술하고 책임감을 가질 수 있도록 교육을 진행하며 관련 성과 지표를 연구하는 것 역시 담당자의 업무이다.

연구데이터의 내용 자체를 검증하는 역할은 위원회에서 수행해야 한다. 위원회는 해당 분야에 대한 이해가 깊은 연구자들로 구성되어야 하며 형평성을 위해 내외부 위원을 고루 위촉할 필요가 있다. 이렇게 구성된 위원회는 등록된 연구데이터의 내용을 검증하여 결과를 재현하는 데 오류가 없고 신뢰할 만한 연구데이터 인지를 평가해야 하며, 이러한 평가를 바탕으로 연구자가 더 나은 연구데이터를 제공하고 메타데이터를 입력할 수 있도록 피드백을 제공해야 한다.

연구데이터 품질 관리를 위한 두 번째 요소로는 연구데이터 표준화가 있다. 이는 Das et al.(2022)의 연구와 담당자 인터뷰에서 여러 차례 언급되었던 것으로 기관의 연구데이터의 통일성 유지뿐 아니라 다른 리포지터리와의 연계를 위해서도 필수적이다. 박성은, 고영만(2022)의 연구는 연구데이터 플랫폼 간 상호 연계 가능성을 보여주었으나 기관별 다양한 주제별 메타데이터 요소를 고려할 때 더욱 상세한 표준화 작업이 요구된다. 이를 위해 매뉴얼과 교육을 통해 기본적인 형식을 통일시키고 접근성이

좋은 곳에 가이드를 게시하여 연구데이터의 수집부터 등록 단계까지 연구자를 지원할 필요가 있다.

그와 별개로 운영 부서 차원에서도 표준화 작업을 진행할 필요가 있다. 날짜나 성별 표기는 캘린더나 콤보박스 등 기술적 개선으로 표기를 통일할 수 있으며, 국외 사례와 같이 해당 주제 분야의 대표어와 유사어를 정리한 용어집을 홈페이지에 게시하여 통일된 용어 사용을 장려할 수 있다. 또한 메타데이터 요소 간 원활한 연계를 위해 공통 필수 요소를 두고, 선택 요소는 서로 유사한 요소와 대응될 수 있도록 하는 작업 역시 표준화 작업에서 수행되어야 한다.

세 번째로, 연구데이터의 품질 관리를 위해서는 장기 보존에 대비해야 한다. 장기 보존은 먼 미래에도 해당 연구데이터를 열람하고 활용할 수 있어야 함을 의미한다. 현재 모든 인터뷰 기관에서 수행되고 있던 백업만으로는 장기적 활용이 보장되지 않는다. 특히나 연구기관의 연구데이터는 해당 연구 분야의 전문 소프트웨어를 통해 추출되는 파일의 확장자들이 존재한다. 이러한 파일의 장기적 보존을 위해 Ag Data Commons에서는 오픈 소스가 아닌 소프트웨어를 사용하여 산출된 연구데이터의 경우 그 코드를 등록하도록 하고 있다. 국내 인터뷰 기관 중에서도 리포지터리에 소프트웨어와 코드가 등록 가능한 기관이 있다. 이러한 방식을 참고하여 오픈 소스 소프트웨어 또는 장기 보존용 포맷 사용을 권장하고 연구데이터에 따라 애플리케이션 및 마이그레이션 전략을 수립·적용하여 연구데이터에 대한 지속적인 접근을 보장해야 한다.

이 외에도 CTS 인증 항목을 확인하여 연구데이터 품질 관리를 위해 살펴야 할 추가적인 사항을 참고할 수 있다. 2023년 6월 기준 CTS 인증을 받은 리포지터리는 전 세계적으로 81개 기관에 불과하며, 아직 국내에는 CTS 인증을 받은 기관이 없다. 따라서 국내에서 CTS 인증을 받아 유지해 나간다면 적절한 연구데이터 품질 관리가 수행되고 있음을 입증받을 수 있을 뿐만 아니라 국제적으로 신뢰성 있는 리포지터리로 발돋움하는 계기가 될 수 있을 것이다.

5. 결론

오픈 사이언스 확산 및 연구데이터 공유의 흐름의 일환으로 연구데이터를 관리하고 공유할 수 있는 신뢰성 있는 리포지터리에 대한 논의가 이루어지고 있다. 최근 국내에서는 국가 차원에서 주요 주제에 관한 연구데이터 플랫폼을 구축하고자 노력하고 있다. 이러한 플랫폼 구축은 개별 기관이 소속 연구자들의 연구데이터를 수집하여 상위 플랫폼에 연계하는 방식으로 구성된다. 그럼에도 국내에서는 기관 연구데이터 리포지터리에 관한 연구가 활성화되지 못하고 있고, 명확한 근거 규정도 부재하다.

국내에서는 출판 문서 위주로 운영되는 국립중앙도서관의 OAK 기관 리포지터리 보급 사업과 별개로 KISTI에서 개발한 NARDA가 기관 연구데이터 리포지터리로 보급되고 있다. 본 연구에서는 후자에 집중하여 국내 기관 연구데이터 리포지터리의 운영 현황을 파악하고 국내외 기관 연구데이터 리포지터리 홈페이지

분석을 통한 사례조사와 담당자 인터뷰를 통해 리포지터리 규정 및 정책 수립, 연구데이터 공유 인식 개선, 연구데이터 품질 관리 강화의 세 가지 운영 활성화 방안을 제안하였다.

첫째로, 리포지터리 규정 및 정책 수립의 측면에서 고시 수준의 법령 지위 향상과 리포지터리 근거 규정 명시가 필요하다고 보았다. 현재 국가연구개발정보처리기준은 외부적 구속력이 없고, 연구데이터 리포지터리에 대한 규정도 부재하다. 따라서 그 지위를 승격시킬 필요가 있으며, 구체적으로 연구데이터의 등록과 관리를 담당할 플랫폼으로써 리포지터리의 필요성을 명시하여 명확한 기반을 마련해야 한다. 한편 기관 수준에서도 기관장의 의지 여하와 상관없는 안정적인 운영을 위하여 리포지터리에 대한 자체 정책을 수립하고 보완해야 한다.

두 번째로 연구데이터 공유와 관련한 인식 개선 측면에서 이용 교육을 포함한 전반적인 연구데이터 교육의 필요성을 제안하였다. 연구데이터의 정의와 필요성에 대한 근본적인 교육을 통해 연구자들 간에 연구데이터 공유에 대한 긍정적인 인식을 심어줄 필요가 있으며, 경진대회 등을 통해 실제 활용사례를 발굴하여 연구데이터 공유와 관련한 우려를 잠재울 필요가 있다. 이러한 교육과 사례는 연구자들에게 동기부여가 될 수 있으며, 각종 인센티브 정책과 병행할 때 시너지 효과를 기대해볼 수 있다.

세 번째로 연구데이터 품질 관리 강화 측면에 있어서는 연구자-담당자-위원회의 상호작용과 표준화 작업, 장기 보존을 위한 준비의 필요성을 제안하였다. 구체적으로 연구자는 스스

로 책임감을 가지고 상세하게 연구데이터를 기술해야 하며, 담당자는 연구데이터 외적인 면을 보완해야 하고, 위원회에서는 연구데이터의 내용적인 면을 검증하고 평가해야 한다. 또한 내부적인 통일성과 외부와의 연계성을 고려하여 표준화 작업에 힘써야 하며, 시간이 흐른 뒤에도 지속적으로 연구데이터에 대한 접근 및 활용 가능성을 유지하기 위해 장기 보존을 위한 준비를 시행해야 한다.

본 연구는 그간 연구가 드물었던 기관 연구

데이터 리포지터리에 집중하여 실무자 관점에서 국내 운영 현황을 살펴보고 기관의 연구데이터 수집 및 공유에 대한 활성화 방안을 제시하였다는 의의를 가진다. 이에 대한 지속적인 연구와 실무적 개선은 기관의 연구자들이 생산한 가치 있는 연구데이터를 장기적으로 이용할 수 있게 할 뿐 아니라 기관 연구데이터 리포지터리가 그 자체로 전문 주제 리포지터리 역할을 수행할 수 있게 한다.

참 고 문 헌

- 과학기술정보통신부 (2022). 연구개발 디지털화 촉진방안 발표. 대한민국 정책브리핑.
출처: <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156526649>
- 국가연구개발정보처리기준. 과학기술정보통신부고시 제2020-102호(2020).
- 김선태, 한선화, 이태영, 김용 (2010). 과학데이터 보존 및 활용모델에 관한 연구. 한국비블리아학회지, 21(4), 81-93.
- 김우중, 김형준, 박희진, 최상욱 (2021). 자연과학 분야 연구 데이터 리포지터리 속성분석: 운영수준, 주제, 유형, 활용도를 중심으로. 기술혁신학회지, 24(4), 577-595.
- 김주섭, 김선태, 최상기 (2019). 연구 데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건. 한국문헌정보학회지, 53(3), 317-344. <https://doi.org/10.4275/KSLIS.2019.53.3.317>
- 김주섭, 양성준, 김선태 (2022). 데이터 리포지터리 인증 체계 분석 및 인증 전략에 관한 연구: Coretrustseal을 중심으로. 한국문헌정보학회지, 56(2), 209-229.
<https://doi.org/10.4275/KSLIS.2022.56.2.209>
- 김지현 (2015). 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 한국문헌정보학회지, 49(3), 413-436. <https://doi.org/10.4275/KSLIS.2015.49.3.413>
- 김태중, 이상환, 신진섭 (2012). 과학데이터 리포지터리에서 검토되어야 할 권리와 조건. 한국기술혁신학회 추계학술대회, 481-488.
- 박규리, 안병군 (2017). 연구데이터 국가 리포지터리 구축에 관한 연구: 국내외 사례를 중심으로. 한국정보관리학회 학술대회 논문집, 33-38.

박성은, 고영만 (2022). 국가 연구데이터플랫폼과 바이오 연구데이터플랫폼의 메타데이터 상호운용성에 관한 연구. *정보관리학회지*, 39(2), 159-202.

<https://doi.org/10.3743/KOSIM.2022.39.2.159>

송사광, 김재수, 김민기, 김성찬, 김인, 김진영, 김호용, 문명환, 박선경, 박성원, 박수현, 서준아, 송영현, 신성호, 신영호, 윤영미, 이민호, 이상환, 이승복, 조민희, 진승교, 홍승균, Tran Quang Khai, Kien Ngoc Mai, 김철연, 고영만, 류재철, Adrian Button (2019). 국가연구데이터플랫폼 구축 및 운영. 한국과학기술정보연구원.

송태운, 이지운, 김우람, 박소예나, 노태근 (2022). 해양과학 데이터저장소 JOISS는 FAIR한가?. *지오데이터*, 4(2), 47-56. <https://doi.org/10.22761/DJ2022.4.2.005>

신은정, 손수정, 서지영, 김영린 (2019). 공공부문 연구데이터의 소유·활용제도 개선 방안 (정책연구 2019-16). 과학기술정책연구원.

아르곤넷 [발행년불명]. DataWorks (연구데이터 리포지터리).

출처: <https://argonet.co.kr/dataworks/>

엄정호 (2020). 전주기 연구활동 지원을 위한 연구데이터 리포지터리 개발 보고서. KISTI.

엄정호 (2021). 연구자의 데이터 관리를 지원하는 연구데이터 리포지터리, NaRDA. *지식정보인프라*, 81, 8-11.

제20대대통령직인수위원회 (2022). 윤석열정부 110대 국정과제. 대한민국 정책브리핑.

출처: <https://www.korea.kr/docViewer/skin/doc.html?fn=196876242&rs=/docViewer/result/2022.07/27/196876242>

OAK 국가리포지터리 [발행년불명]. OAK 소개, OAK 리포지터리.

출처: <https://oak.go.kr/about/aboutOak.do?menuSeq=79>

Bluh, P. & Hepfer, C. (2013). *The Institutional Repository: Benefits and Challenges*. [n.p.]: Association for Library Collections & Technical Services, American Library Association.

Clobridge, A. (2010). 1 - Introduction. In A. Clobridge (Ed.), *Building a Digital Repository Program with Limited Resources*. [n.p.]: Chandos Publishing.

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038.

<https://doi.org/10.1098/rsta.2010.0165>

Das, S., Abou-Haidar, R., Rabalais, H., Sun, S. D. L. W., Rosli, Z., Chatpar, K., Boivin, M.-N., Tabatabaei, M., Rogers, C., Legault, M., Lo, D., Degroot, C., Dagher, A., Dyke, S. O. M., Durcan, T. M., Seyller, A., Doyon, J., Poupon, V., Fon, E. A., Genge, A., Rouleau, G. A., Karamchandani, J., & Evans, A. C. (2022). The C-BIG repository: an institution-level

- open science platform. *Neuroinformatics*, 20(1), 139-153.
<https://doi.org/10.1007/s12021-021-09516-9>
- European Commission (2022). Data Act: Commission proposes measures for a fair and innovative data economy. Available:
https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
- Kindling, M., Pampel, H., van de Sandt, S., Rücknagel, J., Vierkant, P., Kloska, G., Witt, M., Schirmbacher, P., Bertelmann, R., & Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, 23(3/4).
<https://doi.org/10.1045/march2017-kindling>
- Le comité pour la science ouverte (2022). Partager les données liées aux publications scientifiques. Guide pour les chercheurs. Available:
https://www.ouvrirelascience.fr/wp-content/uploads/2022/04/Guide_Partager_les_donnees_web.pdf
- Lynch C. A. & Lippincott, J. K. (2005). Institutional repository deployment in the United States as of early 2005. *D-lib Magazine*, 11(9), 1-11.
<https://doi.org/10.1045/september2005-lynch>
- Mallasvik, M. L. & Martins, J. T. (2020). Research data sharing behaviour of engineering researchers in Norway and the UK: uncovering the double face of Janus. *Journal of Documentation*, 77(2), 576-593. <https://doi.org/10.1108/JD-08-2020-0135>
- Office of Science and Technology Policy (2022). Memorandum for the heads of executive departments and agencies. Available:
<https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>
- Perrier, L. & Barnes, L. (2018). Developing research data management services and support for researchers: A mixed methods study. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 13(1).
<https://doi.org/10.21083/partnership.v13i1.4115>
- Scientific Data [n.d.]. Data Repository Guidance. Available:
<https://www.nature.com/sdata/policies/repositories>
- Suhr, B., Dunjl, J., & Stocker, A. (2020). Search, reuse and sharing of research data in materials science and engineering: A qualitative interview study. *PloS one*, 15(9), e0239216.
<https://doi.org/10.1371/journal.pone.0239216>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf,

- R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PloS one*, 15(3), e0229003.
<https://doi.org/10.1371/journal.pone.0229003>
- Thoegersen, J. L. & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: a meta-evaluation of studies on researcher data sharing. *Journal of Documentation*, 78(7), 1-17. <https://doi.org/10.1108/JD-01-2021-0015>
- Waddington, S., Zhang, J., Knight, G., Jensen, J., Downing, R., & Ketley, C. (2013). Cloud repositories for research data: addressing the needs of researchers. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1), 13.
<https://doi.org/10.1186/2192-113X-2-13>
- Xafis, V. & Labude, M. K. (2019). Openness in big data and data repositories. *Asian Bioethics Review*, 11(3), 255-273. <https://doi.org/10.1007/s41649-019-00097-z>
- Xu, Z., Watts, J., Bankston, S., & Sare, L. (2022). Depositing data: A usability study of the texas data repository. *Journal of eScience Librarianship*, 11(1), e1233.
<https://doi.org/10.7191/jeslib.2022.1233>
- Zuk, P., Sanchez, C. E., Kostick, K., Torgerson, L., Muñoz, K. A., Hsu, R., Kalwani, L., Sierra-Mercado, D., Robinson, J. O., Outram, S., Koenig, B. A., Pereira, S., McGuire, A. L., & Lázaro-Muñoz, G. (2020). Researcher perspectives on data sharing in deep brain stimulation. *Frontiers in Human Neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.578687>

• 국문 참고자료의 영어 표기

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

- Argonet [n.d.]. DataWorks (Research Data Repository). Available:
<https://argonet.co.kr/dataworks/>
- Kim, J. (2015). A study on the perceptions of university researchers on data management and sharing. *Journal of the Korean Society for Library and Information Science*, 49(3), 413-436.
<https://doi.org/10.4275/KSLIS.2015.49.3.413>
- Kim, J., Kim, S., & Choi, S. (2019). The functional requirements of core elements for research data management and service. *Journal of the Korean Society for Library and Information Science*, 53(3), 317-344. <https://doi.org/10.4275/KSLIS.2019.53.3.317>
- Kim, J., Yang, S. J., & Kim, S. (2022). A study on data repository certification scheme analysis and certification strategy: focused on coretrustseal. *Journal of the Korean Society for*

- Library and Information Science, 56(2), 209-229.
<https://doi.org/10.4275/KSLIS.2022.56.2.209>
- Kim, S. T., Hahn, S. H., Lee, T., & Kim, Y. (2010). A study on a model for using and preserving scientific data. *Journal of the Korean Bibliography Society for Library and Information Science*, 21(4), 81-93.
- Kim, T. J., Lee, S. H., & Sin, J. S. (2012). Rights and conditions to be reviewed in scientific data repositories. *Proceedings of the Korea Technology Innovation Society Conference*, 481-488.
- Kim, W. J., Kim, H., Park, H., & Choi, S. (2021). An analysis on operational level, subjects, content types and utilization of research data repositories in natural science. *Journal of Korea Technology Innovation Society*, 24(4), 577-595.
- Ministry of Science and ICT (2022). Announcement of R&D digitization promotion plan. Korea Policy Briefing.
<https://www.korea.kr/briefing/pressReleaseView.do?newsId=156526649>
- National R&D Information Processing Standards. Ministry of Science and ICT Notice No. 2020-102.
- OAK National Repository [n.d.]. Introduction to OAK, OAK repository. Available:
<https://oak.go.kr/about/aboutOak.do?menuSeq=79>
- Park, K. R. & An, B. G. (2017). A study on developing national research data repository. *Proceedings of the Korean Society for Information Management Conference*, 33-38.
- Park, S. E. & Ko, Y. M. (2022). A study on metadata interoperability between the national research data platform and the bio research data platform. *Journal of the Korean Society for information Management*, 39(2), 159-202. <https://doi.org/10.3743/KOSIM.2022.39.2.159>
- Shin, E. J., Sohn, S. J., Seo, J. Y., & Kim, Y. L. (2019). Considerations for data rights in publicly funded research (Policy Research 2019-16). Science and Technology Policy Institute.
- Song, S. K., Kim, J. S., Kim, M. K., Kim, S. C., Kim, I., Kim, J. Y., Kim, H. Y., Moon, M. H., Park, S. K., Park, S. W., Park, S. H., Seo, J. A., Song, Y. H., Shin, S. H., Shin, Y. H., Yoon, Y. M., Lee, M. H., Lee, S. H., Lee, S. B., Cho, M. H. Jin, S. G., Hong, S. G., Tran Quang Khai, Kien Ngoc Mai, Kim, C. Y., Ko, Y. M., Ryu, J. C., & Button, A. (2019). Construction and operation of research data platform. Korea Institute of Science and Technology Information.
- Song, T. Y., Lee, J. Y., Kim, W. R., Park, S., & Rho, T. K. (2022). Is the ocean science data repository, JOISS able to be FAIR?. *Geo Data*, 4(2), 47-56.
<https://doi.org/10.22761/DJ2022.4.2.005>

The 20th Presidential Transition Committee (2022). The 110 national tasks of the Yoon Seok-Yeol administration. Korea Policy Briefing. Available:

<https://www.korea.kr/docViewer/skin/doc.html?fn=196876242&rs=/docViewer/result/2022.07/27/196876242>

Um, J. H. (2020). Research data repository development report for full cycle research activity support. KISTI.

Um, J. H. (2021). "NaRDA", a research data repository that supports researchers' data management. Knowledge Information Infrastructure, 81, 8-11.