

개방형 연구 커먼즈에 대한 연구자 요구 분석에 관한 연구*

A Study on the Researchers' Needs for Open Research Commons

심 원 식 (Wonsik Shim)** , 안 혜 연 (Hyeyeon An)***
박 규 리 (Kyuri Park)**** , 송 사 광 (Sa-Kwang Song)*****
임 형 준 (Hyung-Jun Yim)*****

목 차

- | | |
|----------------|---------|
| 1. 서 론 | 4. 연구결과 |
| 2. 선행연구 및 사례조사 | 5. 논 의 |
| 3. 연구방법 | 6. 결 론 |

초 록

본 연구의 목적은 개방형 연구 커먼즈(Open Research Commons)의 구현을 위하여 연구데이터 플랫폼에 대한 연구자 요구를 분석하는 것이다. 설문은 대표적인 연구데이터 커먼즈라 할 수 있는 유럽 EOSC와 호주 ARDC의 플랫폼 서비스 사례를 분석하여 도출한 주요 서비스 내용을 참고하여 설계하였다. 설문지는 크게 5개 영역(응답자 일반적 특성, 연구데이터·분석도구·컴퓨팅 자원의 활용 및 연구데이터 커먼즈의 필요)으로 구성하였다. 연구데이터 플랫폼에 대한 요구를 분석하기 위하여 관련 플랫폼을 통해 잠재적 이용자 550명이 완료한 응답을 분석하였다. 그 결과, 응답자의 약 85%가 연구데이터 커먼즈의 필요성에 동의하였으며, 커먼즈 서비스가 제공되면 이를 통하여 본인의 연구데이터(약 77%)와 분석도구(약 84%)를 공유하고자 하는 의향과 공유된 연구데이터(약 92%)나 분석도구(약 93%), 컴퓨팅 자원(약 93%)을 활용하고자 하는 의향이 나타났다. 즉 많은 연구자들이 개방형 연구 커먼즈에 대한 요구가 매우 높은 것이라 이해할 수 있다. 그러나 국내 연구데이터 서비스 분야에 있어 커먼즈에 대한 논의는 아직 생소한 것으로, 향후 주제분야, 이해관계자 특성, 커먼즈 플랫폼의 실질적 이용자 등을 다양한 측면에서의 연구가 이루어져야 할 것이다.

ABSTRACT

In order to embrace open research commons, this study looked at the requirements that researchers have for research data platforms. A survey questionnaire was developed using the results of Australian ARDC and European EOSC case studies. The questionnaire addressed five main areas: the need for research data commons, analysis tools, computing resources, consumption of research data, and general respondent characteristics. To determine the requirements for research data platforms, an analysis was conducted on the responses provided by 550 potential platform users. The findings demonstrated that more than 85% of participants concurred on the significance of open research commons. Additionally, there was a high willingness to use research data (77%), analysis tools (84%), share research data (92%), analysis tools (93%), and computing resources (93%) if commons services were provided. The results of this study show that there is a significant demand for open research commons among researchers. Discussions on research data commons are still hard to come by in Korea's research data services sector, though. Research on a range of subjects, such as subject areas, stakeholder characteristics, and real users of commons platforms, need to be conducted in the future.

키워드: 오픈사이언스, 개방형 연구 커먼즈, 오픈 리서치 커먼즈, 연구데이터 커먼즈, 연구데이터 서비스
Open science, Open research commons, Research data commons, Research data service, Data commons

* 이 논문은 2023년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구임
(과제번호: (KISTI)K-23-L01-C03-S01, (NTIS)1711198423).

** 성균관대학교 문헌정보학과 교수(wonsikshim@skku.edu / ISNI 0000 0004 6480 0867) (제1저자)

*** 성균관대학교 문헌정보학과 박사과정(gpds23@skku.edu / ISNI 0000 0004 9291 6365) (교신저자)

**** 성균관대학교 문헌정보학과 박사과정(kyurip0831@skku.edu / ISNI 0000 0005 1315 9069) (공동저자)

***** 한국과학기술정보연구원, UST 응용AI학과 교수
(esmallj@kisti.re.kr, esmallj@ust.ac.kr / ISNI 0000 0004 6098 4814) (공동저자)

***** 한국과학기술정보연구원(hjyim@kisti.re.kr / ISNI 0000 0004 7393 6934) (공동저자)

논문접수일자: 2023년 10월 22일 최초심사일자: 2023년 10월 30일 게재확정일자: 2023년 11월 10일

한국문헌정보학회지, 57(4): 209-232, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.4.209>

※ Copyright © 2023 Korean 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

디지털 환경에서 학술출판물의 접근 장벽을 제거하기 위해 2000년대 초반에 시작된 오픈액세스 운동(BOAI, 2002)은 최근 연구데이터의 공개와 공유, 오픈협력을 포괄하는 오픈사이언스라는 프레임워크로 진화하고 있다(신은정, 정원교, 2017). 기존의 인쇄 출판 환경이 급속하게 디지털 기반의 네트워크 환경으로 변화하는 과정에서 새로운 학술커뮤니케이션 시스템을 지원하기 위한 다양한 수준의 인프라가 필요하다. 예를 들어, 2002년 말에 공개된 DSpace 플랫폼은 도서관을 포함한 기관들의 오픈액세스 리포지토리를 구현하는 중요한 오픈소스 인프라 기능을 제공했다. 유사한 사례로 OAI-PMH (Open Archives Initiative protocol for Metadata Harvesting)에서 개발한 EPrints 오픈소스 소프트웨어와 European OpenAIRE 프로그램에서 개발한 오픈액세스 리포지토리인 Zenodo 시스템, 연구데이터의 공유, 보존, 인용, 분석을 지원하는 Dataverse 시스템 등이 있다.

최근에는 전통적인 의미의 학술출판물 외에 연구데이터를 수집, 공유하는 것은 물론 데이터를 분석하기 위한 다양한 소프트웨어 도구 및 컴퓨팅 자원을 통합적으로 제공하는 개방형연구커먼즈(Open Research Commons, ORC)¹⁾의 필요성이 적극 제기되고 있다(Bourne et al., 2022). 전 세계적으로 모든 사람이 사용하기 쉽고 액세스할 수 있는 공공 및 민간 부문 내에서 상호운용

가능한 개방형 연구 커먼즈의 구축은 기술적인 문제임은 물론 관련 기관 및 연구자의 목표 의식, 지속적인 참여와 이를 보장하는 거버넌스가 요구되는 정책적, 문화적인 문제라고 할 수 있다.

이미 RDA(Research Data Alliance)와 같은 기구는 Global Open Research Commons Interest Group을 구성하고 관련된 기술, 정책, 거버넌스 등을 폭넓게 논의하고 있다. 또한 Center for Open Science(COS)에서 개발한 Open Science Framework(OSF)는 전 세계 연구자들을 대상으로 연구의 전 주기(lifecycle)에 걸쳐 연구 아이디어의 개발, 연구계획의 설계, 데이터의 수집 및 분석, 결과 보고서 및 논문의 저술, 출판을 통합적으로 지원하는 워크플로우와 클라우드 기반 도구를 제공하고 있다.

UNESCO(2021)는 ‘오픈사이언스에 대한 권고안’을 통하여 디지털 기술이 발전함에 따라 연구의 성과와 과정을 더욱 광범위하고 개방적으로 공개 및 공유하고자 하는 지향과 실천까지 망라하여 오픈사이언스라고 개념화하고 있다. 국내에서도 최근 10년간 연구데이터 공유와 활용에 대한 관심이 높아졌으며, 연구데이터 관리와 리포지토리 운영 기관들은 연구자들의 연구 활동을 효율적으로 지원하기 위해서는 국내외 연구데이터를 한곳에서 모아서 공개·공유하고, 데이터를 이용한 분석 환경까지 가능하게 해주는 플랫폼이 필요하다(황미영 외, 2023)는 주장이 계속 이어져 오고 있다.

유럽연합의 European Open Science Cloud

1) 개방형연구커먼즈라는 용어는 넓게는 학술출판물, 연구데이터 등을 포함한 다양한 학술성과를 공유함으로써 오픈사이언스를 구현하는 시스템이라고 할 수 있다. 하지만, 보다 구체적으로는 연구데이터의 공유, 분석을 지원하는 시스템, 인프라를 지칭한다. 이런 면에서 개방형연구커먼즈라는 용어 대신 개방형데이터커먼즈라는 용어가 사용되기도 한다. 본 연구에서는 연구데이터의 공유, 분석에 주안을 둔 시스템으로서 개방형연구커먼즈라는 용어를 사용하고 있다.

(EOSC), 중국의 China Science & Technology Cloud(CSTCloud), 아프리카 대륙의 African Open Science Platform(AOSP), 말레이시아의 Malaysia Open Science Platform(MOSP), CODATA가 조직한 Global Open Science Cloud (GOSC) Initiative, 호주의 Australian Research Data Commons(ARDC), 캐나다의 Digital Research Alliance of Canada 등의 사례는 국가 또는 지역 내에서 오픈사이언스를 지원하는 다양한 인프라가 이미 구축되었거나 구축 중이라는 것을 보여준다.

개방형 연구 커먼즈는 구체적인 학문 분야에서 활발히 구축되고 있으며 대표적인 분야로는 의학이 있다. 미국 보건연구원(National Institutes of Health) 산하 국가암연구소(National Cancer Institute)에서는 클라우드 기반으로 연구데이터의 수집, 공유, 분석, 시각화 등을 통합적으로 제공하는 플랫폼인 Cancer Research Data Commons를 운영 중에 있고 전 세계 70개국 이상의 2,000여 개의 협력 기관이 공동으로 운영 중인 Observational Health Data Sciences Informatics(OHDSI) 플랫폼 역시 의료, 임상데이터의 공유와 분석을 위한 다양한 컴퓨팅 자원을 통합적으로 제공함으로써 새로운 형태의 개방형 의학 연구를 선도하고 있다.

이러한 다양한 오픈사이언스 특히 연구데이터의 공유 및 분석을 위한 통합적인 커먼즈 플랫폼의 출현은 향후 학술커뮤니케이션 인프라 구축이 어떤 방향으로 나아가야 하는지를 보여준다. 이미 우리나라에는 국가연구데이터플랫폼인 DataON이 구축되어 상당한 양의 데이터셋에 대한 접근을 제공하고 있다. 하지만 단순히 데이터의 저장, 접근, 공유를 넘어서 데이터

가 있는 곳에서 분석을 수행하고, 다양한 분석 도구 및 소프트웨어를 활용하고, 또한 대용량 컴퓨팅 자원을 활용할 수 있는 개방형 연구 커먼즈로 진화할 필요가 있다.

국내에서는 연구데이터 및 서비스와 관련하여 연구데이터의 수집, 공유, 활용에 관한 연구(김지현, 2015; 박미영, 안인자, 남승주, 2018)와 연구데이터에 대한 연구자 인식(유사라, 2019) 관련 연구 등이 수행되었다. 그러나 이들 연구에서 언급되는 연구데이터 서비스는 커먼즈의 개념을 포괄한 차세대 서비스로 보기에 무리가 있다. 또한 이상환(2019)이 정부 출연연구기관의 연구데이터 서비스 이해관계자 대상 설문조사를 통하여 출연연에서 산출되는 연구데이터의 빅데이터화와 관리 체계 수립 및 개선 방안 모색에 관한 논의를 이끌어내고자 인프라, 응용서비스 등과 관련한 현황 및 요구사항을 조사하였으나, 데이터를 기반으로 통합적인 연구 환경을 제공하는 개방형 연구 커먼즈 논의까지 이어지지는 못하였다.

따라서 본 연구는 개방형 연구데이터 커먼즈의 실체가 부재한 국내 상황에서 잠재적인 이용자를 대상으로 개방형 연구 커먼즈에 대한 요구사항을 수집하고 분석함으로써 향후 이러한 지식 인프라 구축에 유용한 정보를 제공하는 기반을 마련하고자 하였다.

2. 선행연구 및 사례조사

2.1 선행연구

2.1.1 연구데이터 서비스

김선태, 이정훈, 정한민(2017)은 미국 행정 관

리 예산국(OMB)이 연구데이터를 과학 커뮤니티에서 연구발견을 검증하는 데 필요하다고 인정되는 '기록된 사실 자료(recorded factual material)'로 정의하고 있으며, 연구의 재현과 검증을 위해 필요한 것이 연구데이터라고 규정하였다.

연구데이터는 연구 결과를 검증하기 위해 수집, 관찰, 생성되는 모든 정보이며, 디지털 형식과 물리적 형식 등 다양한 형식으로 산출된다. 데이터를 관리하면 보다 효율적으로 작업하고, 더 높은 품질의 데이터를 생성하고, 더 많은 연구에 노출되고, 데이터가 손실되거나 오용되지 않도록 보호할 수 있다.

따라서 연구데이터 서비스는 연구데이터를 저장, 관리 및 액세스 가능하게 만드는 문제를 해결하는 데 도움을 주기 위한 서비스라 할 수 있다.

최근 오픈사이언스에 대한 관심이 높아지면서 주요국의 연구기관, 특히 대학 내 도서관들이 선도적으로 다양한 연구데이터 서비스(Research data services: RDS)를 제공하기 위하여 노력하고 있다(Sheikh, Malik, & Adnan, 2023). 데이터의 폭발적 증가와 연구비 지원 정책의 변화, 오픈학술커뮤니케이션(open scholarship)의 확산을 계기로 대학 도서관은 다른 전문 서비스와 연구자들과 협력하여 연구데이터 관리(Research data management: RDM)를 지원하는 다양한 자문 및 기술 서비스를 개발하였다(Cox et al., 2019). 이러한 연구데이터 서비스의 사례로는 자문 및 지원 서비스, 데이터 리터러시(data literacy) 교육, 메타데이터 관리와 자원 디스커버리(metadata management & resource discovery)를 비롯한 리포지터리 관리

서비스를 들 수 있다. Bryant et al.(2017)은 연구데이터 서비스를 교육, 전문지식(expertise), 큐레이션의 세 영역으로 나누어 설명하였으며, 단, 연구데이터 서비스는 기관마다 이러한 분류가 동일하게 적용될 수 있는 일률적인 서비스가 아닌 상황에 기반한 유동적인 형태임을 강조하였다. 뿐만 아니라 연구데이터 서비스는 기관 내부는 물론 외부 환경과도 상호연결적임을 역설하였다.

오픈 사이언스 운동의 영향으로 연구데이터 관련 논의가 오픈 데이터로 전환되는 과정에서 연구데이터 서비스는 더욱 중요한 요소로 작용하게 되었다. 심원식(2021)은 오픈데이터가 연구데이터를 디지털 형태로 적절히 공개하여 이에 대한 접근성 및 활용도를 제고하는 것으로, 과학적 연구의 원천 소재로써 사용되는 사실 자료 및 연구 결과 검증을 위해 과학계에서 통상적으로 인정되는 자료를 대상으로 한다고 하였다.

2.1.2 개방형 연구 커먼즈

사회과학 분야에서 커먼즈(common)는 공동체에 귀속된 혹은 공동체가 집합적으로 소유하는 '공동의 것'을 칭하는 말이다. 공동의 것으로서 향유되어야 할 많은 것들을 분할, 교환, 상품화하는 신자유주의 자본주의 시스템에 대한 문제의식과 국가/시장 이분법을 넘어선 새로운 해법에 대한 요구 속에서 전환을 기획하는 많은 연구자와 실천가들 사이에서 커먼즈라는 개념에 대한 관심이 늘어나고 있다(안새롬, 2020).

문헌정보학에서 커먼즈는 1990년대에 등장한 용어로 정보 커먼즈(Information commons), 지식 커먼즈(Knowledge commons), 학습 커먼

즈(Learning commons), 데이터 커먼즈(Data commons), 연구데이터 커먼즈(Research Data Commons), 연구 커먼즈(Research commons) 등의 다양한 용어로 현재까지 논의되고 있다. 등장 초기 미국 대학에서 다양한 디지털 서비스를 통합하여 제공하는 장소를 일컫는 말에서 기원하여 Beagle(1999)은 정보 커먼즈를 단일한 창구인 GUI(single graphical user interface)를 통하여 다양한 디지털 서비스를 검색할 수 있으며, 검색한 서비스들에 접근할 수 있는 독립적인 온라인 환경과 통합 디지털 환경을 중심으로 작업 공간과 서비스 제공을 구성하도록 특별히 설계된 새로운 형태의 물리적 시설로 나누어 정의하였다.

미국 NSF에 의해 설립된 CyVerse에서는 데이터 커먼즈를 검색이 가능하고, 컴퓨팅 가능하며, 재사용 가능한 데이터를 위한 클라우드 기반 리소스로 정의하였다. 즉 사용자가 데이터 리포지터리에서 데이터에 접근, 생성, 관리, 선별, 게시, 검색 및 재사용할 수 있도록 지원함으로써 FAIR 원칙에 기반한 개방적인 연구 환경을 제공하고자 하는 것이라 할 수 있다.

송사광 외(2023)는 연구데이터 커먼즈(Research Data Commons)를 연구데이터의 활용을 극대화하기 위해 상호운용 가능한 고품질의 컴퓨팅 자원에 원활한 사용성(usability)을 제공하는 신뢰할 수 있는 공동 활용 체계로 정의하였다. 여기서 사용성은 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙의 개념으로, 연구 과정에서 생산된 연구데이터 및 디지털 개체에 대한 신뢰 가능한 재사용 등을 의미한다고 하였으며, 연구데이터 커먼즈는 연구데이터와 연구데이터 관련 컴퓨팅 자원의 연합·활용

체계로 이해할 수 있다고 하였다.

RDA는 글로벌 오픈 리서치 커먼즈(Global Open Research Commons: GORC)에서 “오픈 사이언스 커먼즈” 혹은 “데이터 커먼즈”를 공유 기반의 가상 공간 또는 플랫폼을 통해 데이터 및 서비스에 대한 시장을 제공하는 것이라 하였다. 그 예로는 EOSC, ARDC, AOSP, 개방형 정부 포털, 전통적인 연구 환경에서 벗어난 이니셔티브를 들 수 있다. 이들을 통해 상호운용 가능한 데이터 공유 네트워크를 활성화하고 조정하는 것이 데이터 커먼즈의 목표라 하였다.

Grossman(2019)은 개방형 연구 커먼즈를 데이터 목록(data catalog), 데이터 호수(data lake) 등 다른 데이터 플랫폼 유형과의 비교를 통해 설명하고 있다. 연구데이터의 공유를 지원하기 위한 가장 기본적인 시스템은 데이터 목록으로 데이터 자원에 대한 기초적인 메타데이터와 위치를 포함한 단순 목록이다. 데이터 목록에서 한 걸음 더 나아가 데이터 호수는 단순한 목록에 더해 데이터에 대한 디지털 식별자와 보다 풍부한 메타데이터를 제공하지만 데이터 모델이나 스키마는 갖추고 있지 않다. 이러한 배경에서 표준적인 데이터 모델에 기반하여 데이터를 수집, 접근, 공유할 수 있도록 하는 것뿐 아니라 대용량 데이터의 분석 및 처리를 위한 시스템 및 협업 도구를 통합적으로 제공하는 차세대 플랫폼인 데이터 커먼즈 개념이 등장하였다.

2.2 사례조사

대표적인 연구데이터 커먼즈는 <표 1>과 같다. 이 중 가장 역사가 오래되고 지속적으로 활발히 운영 및 관리 되고 있는 유럽 EOSC와 호

〈표 1〉 연구데이터 커먼즈 사례

국가별 커먼즈 사례	연구 분야별 커먼즈 사례	법국가적 커먼즈 사례
<ul style="list-style-type: none"> • 호주 ARDC • 유럽연합 EOSC • 중국 CST Cloud(China Science and Technology Cloud) • 말레이시아 MOSP(Malaysia Open Science Platform) • 아프리카 AOSP(African Open Science Platform) • ... 	<ul style="list-style-type: none"> • NCI Cancer Research Data Commons • OHDSI • TDC(Therapeutics Data Commons) • PSDC(Penn State Data Commons) • ... 	<ul style="list-style-type: none"> • RDA GORC(Global Open Research Commons) • CODATA GOSC(Global Open Science Cloud) • GO-FAIR • ...

* 송사광 외(2023)의 표를 수정·보완한 것임.

주 ARDC의 진화 역사와 해당 커먼즈에서 서비스하고 있는 데이터, 분석도구, 컴퓨팅 자원 및 서비스 방법 등을 조사하여 국내 오픈 리서치 커먼즈의 구성과 설립에 반영할 수 있는 점이 있는지 살펴보고자 하였다. 분야별로 다양한 연구데이터 커먼즈 사례도 있지만, 여기에서는 국가 차원 또는 지역 차원의 연구데이터 커먼즈에 국한하여 대표 사례를 제시하였다.

2.2.1 유럽 EOSC 사례

EOSC²⁾는 모든 연구 분야의 유럽 연구자들이 생산하는 연구데이터를 공유하기 위해 구상한 Horizon Europe의 오픈사이언스 영역 핵심 프로젝트이다. 2015년 5월 유럽연합 집행위원회(European Commission; EC)에서 유럽의 기존 연구데이터 인프라 통합 및 데이터를 비롯한 연구 관련 서비스에서의 FAIR 원칙 실현을 목적으로 하는 EOSC에 대한 초기 아이디어가 제안되었다.

EC는 EOSC의 필요성을 다음과 같이 5개로

정리하고 있다. 1) 연구데이터 생산량 급증, 2) 연구와 혁신의 데이터 집약화에 따른 오픈 데이터 커먼즈 수요 증대, 3) 유럽의 기존 인프라 연결 및 각 분야별 분산된 연구데이터 통합의 필요, 4) 연구데이터와 강력한 컴퓨팅 기술, 신속한 네트워크 기술 간 연결 필요, 5) 유럽의 오픈사이언스와 오픈이노베이션을 위한 정책적 필요

EOSC 프로젝트는 국가와 여러 분야에 걸쳐 제공되는 연구데이터 및 서비스 시스템에 원활하고 개방적인 접근을 가능하게 함으로써 오픈 사이언스 실천에 기여하는 것을 특징으로 한다. Horizon 2020 프로젝트를 통해 2020년까지 3.2억 유로의 재정 지원을 받아 프로토타입 개발이 진행되었으며, 이후 2021년부터 2027년까지는 Horizon European의 재정 지원을 통해 고도화가 진행될 예정이다.

EOSC는 유럽의 모든 과학정보 및 자원에 대한 단일한 접근점으로서 EOSC 포털을 개발하여 EOSC에서 제공하고자 하는 가상 연구 환경에 대한 게이트웨이 역할을 부여하였다. EOSC

2) <https://eosc-portal.eu/about/eosc>

포털에서는 정보와 자원뿐 아니라 EOSC의 거버넌스와 역할, 관련 프로젝트, 국가 정책, 연구비 신청 기회, EOSC 활용 사례를 제공한다. 이때 데이터와 서비스와 같은 연구 자원은 EOSC 포털 내 마켓플레이스를 통하여 접근할 수 있게 한 것이 특징적이다.

또한 EOSC 포털의 마켓플레이스는 데이터셋, 출판물, 소프트웨어와 같은 연구성과물의 발견, 관리, 출판을 가능하게 할 뿐 아니라 다른 연구자 혹은 기관의 프레임워크, 연구 장비, 분석도구, 컴퓨팅 자원, 저장 공간 등을 탐색하고 이에 접근할 수 있게 하고 있다.

2.2.2 호주 ARDC 사례

2018년 7월 호주 NCRIS(National Collaborative Research Infrastructure Strategy)의 ANDS(Australian National Data Service), Nectar(National eResearch Collaboration Tools and Resources), RDS(Research Data Services)를 통합하여 출범한 ARDC³⁾는 데이터를 통해 호주 연구자에게 경쟁 우위를 제공하고, 호주의 대표적인 데이터 플랫폼들을 통합적으로 관리 및 운영하고 있다. 이를 통해 호주 연구자들에게 경쟁 우위를 제공하고 호주 연구 품질과 영향력 극대화를 목적으로 연구자의 서비스 탐색, 접근, 효과적인 사용을 보조하는 연구 환경을 개발하고자 하였다.

ARDC에서 제공하는 대표적인 서비스는 Nectar Research Cloud, Research Data Australia, Identifier Services, Research Vocabularies Australia 등이다.

1) Nectar Research Cloud: 호주 최초의 연합 연구 클라우드로, 연구 커뮤니티에 컴퓨팅 인프라와 소프트웨어를 제공하고 원격으로 신속하고 자율적으로 데이터를 저장, 액세스, 실행이 가능하게 함.

2) Research Data Australia: 연구데이터 및 관련 프로젝트, 연구원 및 데이터 서비스에 대한 검색을 제공하는 온라인 포털로서 호주 연구기관, 정부 기관 및 문화 기관의 연구데이터를 검색, 접근, 재사용 서비스를 제공함.

3) Identifier Services: 연구데이터, 연구 샘플, 파일, 문서 등 디지털 개체에 대한 영구 식별자 부여 및 관리 서비스를 제공하며 DOI, DOI 관리, IGSN, RAiD 서비스 등을 제공하고, 이를 통해 객체와 컨텍스트 연결을 가능케 하여 데이터에 가치를 부여함.

4) Research Vocabularies Australia: 연구에 사용되는 통제어휘 검색 및 사용을 지원하고 호주 연구기관이 자체적으로 통제어휘를 출판하고 관리함.

유럽의 EOSC의 영향을 받아 개발된 ARDC는 FAIR 원칙을 따르며, 국제적 연구 협력 관계를 유지하고, 국제적 리더십 역할과 연구에 필요한 사회적 인프라, 기술적 인프라 구현을 위한 촉매제 역할을 수행하고 있다. 또한 글로벌 연구의 필수적인 요소인 연구 소프트웨어를 위한 Research Software Alliance를 공동 창설하여 협력을 통한 데이터 및 클라우드 인프라 프로젝트에 참여함으로써 확장 가능한 데이터 환경 구축을 위한 연구에 노력을 기울이고 있다.

3) <https://ardc.edu.au/>

3. 연구방법

본 연구에서는 개방형 연구 커먼즈의 구현을 위하여 설문을 통해 연구자들의 요구를 살펴보고자 하였다.

3.1 설문지 구성

설문 내용은 앞선 EOSC와 ARDC 사례 분석을 통해 수집된 연구데이터, 분석도구, 컴퓨팅 자원 관련 서비스 내용을 참고하여 응답자 일반적 특성, 연구데이터의 활용, 분석도구의 활용, 컴퓨팅 자원의 활용, 커먼즈의 필요 등 크게 5개 영역으로 구성하였다. 우선 연구데이터의 활용, 분석도구의 활용, 컴퓨팅 자원의 활용 영역 설문을 통해 연구자들이 실제 연구 상황에서 겪는 주요 경험(활용/어려움)과 필요에

대해 살펴보고자 하였다. 각 항목은 측정을 위하여 4점 리커트 척도를 활용하였으며, 자유 기입이 가능한 기타 응답란을 두어 다양한 의견을 수용하고자 하였다. 설문의 마지막 영역은 커먼즈의 필요에 대한 요구를 반영하기 위한 문항들로 구성하였다(〈표 2〉 참조).

설문조사를 실시하기 전, 연구데이터 플랫폼 전문가 5인으로부터 설문 내용과 진술 방식 등의 적절성에 대해 피드백을 받았으며, 그 결과를 바탕으로 일부 설문 항목을 수정하여 최종 설문 문항을 도출하였다.

3.2 자료 수집 및 분석

본 연구의 설문조사는 한국과학기술정보연구원의 협조를 받아 데이터온(DataON) 및 사이언스온(ScienceON) 홈페이지에 배너를 설치

〈표 2〉 설문 내용

영역	설문 문항 내용
응답자 일반적 특성	• 연구과제 참여 시 주 역할, 직위, 전공 분야
연구데이터의 활용	• 연구 수행 시 연구데이터 생산 여부 • 생산한 연구데이터의 저장 혹은 관리 시 겪는 어려움 • 연구데이터 분석 수행 시 겪는 어려움 • 연구데이터의 처리 및 분석을 위한 컴퓨터 프로그래밍 작업 필요 여부 • 타 연구자 생산 연구데이터 활용 경험 • 타 연구자 생산 연구데이터 활용 시 겪는 어려움 • 타 연구자 생산 연구데이터 필요성
분석도구의 활용	• 타 연구자 생산 분석도구 활용 경험 • 타 연구자 생산 분석도구 활용 시 겪는 어려움 • 타 연구자 생산 분석도구 필요성
컴퓨팅 자원의 활용	• 타 연구자 소유 컴퓨팅 자원 활용 경험 • 타 연구자 소유 컴퓨팅 자원 활용 시 겪는 어려움 • 타 연구자 소유 컴퓨팅 자원 필요성
커먼즈의 필요	• 커먼즈의 필요성 • 연구데이터/분석도구 및 컴퓨팅 자원 관련 서비스 요구 • 커먼즈를 통한 본인 생산 연구데이터/분석도구 공유 의향 • 커먼즈를 통한 타 연구자 생산 연구데이터/분석도구 활용 의향 • 커먼즈를 통한 타 연구자 소유 컴퓨팅 자원 활용 의향

하고, 데이터온 및 사이언스온 가입 회원 중 설문조사에 동의한 회원에게 설문 링크가 담긴 이메일을 발송하는 온라인 방식으로 진행되었다. 설문지는 온라인 설문조사 플랫폼인 '서베이몽키'(https://ko.surveymonkey.com/)를 활용하여 제작·배포하였다. 설문조사는 2021년 10월 6일부터 2021년 10월 17일까지 실시되었으며, 총 550명이 최종적으로 응답을 완료하였다.

수집된 자료는 R 프로그래밍 언어 기반 오픈 통계 소프트웨어인 Jamovi를 사용하여 통계적 분석을 실시하였다.

3.3 응답자의 일반적 특성

설문 응답자의 직위는 정부출연 연구소 연구

원이 209명(38.0%)으로 가장 많았으며, 민간기관 연구원이 153명(27.8%)으로 뒤를 이었다. 이후 대학원생이 78명(14.2%), 대학 연구원이 43명(7.8%), 공공기관 및 공기업 연구원이 19명(3.5%), 대학 전임 교원이 14명(2.5%), 대학생이 4명(0.7%)이며, 공공도서관 사서, 감사 등의 기타 직위도 30명(5.5%) 있었다.

다음으로 연구과제에 참여할 때 주로 맡은 역할에 대한 응답은 참여연구원이 242명(44.0%), 연구책임자가 174명(31.6%)으로 대부분을 차지하였으며, 뒤이어 공동연구원이 86명(15.6%), 연구보조원이 43명(7.8%), 자문, 감독관 등의 기타 직무 5명(0.9%) 순으로 나타났다.

응답자의 전공 분야는 공학이 286명(52.0%)으로 압도적으로 많았으며, 이후 순위도 생명

〈표 3〉 응답자의 일반적 특성

구분		빈도(명)	비율(%)
직위	대학 전임 교원	14	2.5
	대학 연구원	43	7.8
	정부 출연연 연구원	209	38.0
	(준)공공기관 연구원	19	3.5
	민간기관 연구원	153	27.8
	대학원생	78	14.2
	대학생	4	0.7
	기타	30	5.5
연구과제 참여 시 주 역할	연구책임자	174	31.6
	공동연구원	86	15.6
	참여연구원	242	44.0
	연구보조원	43	7.8
	기타	5	0.9
전공 분야	자연과학	83	15.1
	생명과학	111	20.2
	공학	286	52.0
	인문과학	6	1.1
	사회과학	52	9.5
	복합학	12	2.2

과학 111명(20.2%), 자연과학 83명(15.1%)으로 이공계열 비율이 높았다. 그 외에 사회과학 52명(9.5%), 복합학 12명(2.2%), 인문과학 6명(1.1%) 순으로 인문·사회계열 및 복합학 계열 응답자가 포함되었다.

4. 연구결과

4.1 연구데이터의 활용

4.1.1 연구데이터 생산 및 분석 경험

연구데이터를 생산해본 적 있다는 응답은 484명(88.0%)이고, 생산해본 적 없다는 응답은 66명(12.0%)으로 나타났다. 이 중 연구데이터 생산 경험이 있는 응답자(484명)를 대상으로 연구데이터 관리 및 분석과 관련된 구체적인 경험을 물어보았다. 연구데이터의 저장 및 관리에 따른 어려움을 겪었다는 응답은 매우 그러함 76명(15.7%), 그러함 218명(45.0%) 등 총 294명으로 전체의 약 61%에 해당하였다.

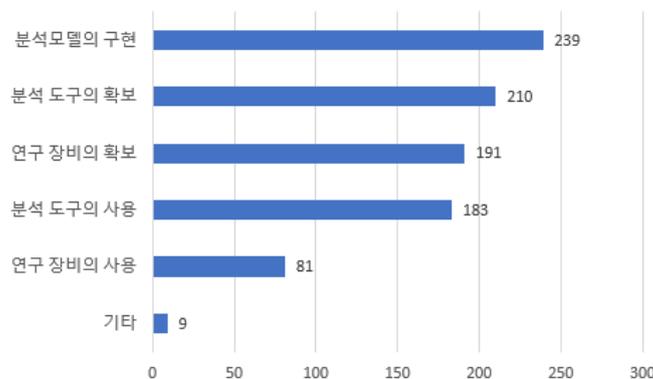
연구데이터 분석 시 경험하는 어려움과 관련

해서는 분석모델의 구현, 분석 도구의 확보 및 사용, 연구 장비의 확보 및 사용 등을 조사하였는데, 분석 도구와 연구 장비 모두 사용보다는 확보에 어려움을 겪고 있었다. '기타' 항목을 통해 제시된 항목 외에 응답자들이 토로한 어려움은 '분석을 위한 다양한 참고모델 및 관련 예시 미보유', '인력', '데이터의 확보', '분산되어 있는 데이터 명세, 수집, 관리', '분석 환경 설정의 어려움', '전처리' 등이 있었다(〈그림 1〉 참조).

연구데이터 분석 시 컴퓨터 프로그래밍 작업이 필요한지 묻는 문항에서는 매우 그러함 97명(20.0%), 그러함 215명(44.4%)으로, 필요하다고 응답한 수가 그렇지 않음(126명, 26.0%) 혹은 전혀 그렇지 않음(40명, 8.3%)이라고 응답한 수의 약 2배로 나타났다. 기타 의견도 '필요에 따라', '필요하나 사용할 줄 모름' 등 프로그래밍 작업의 필요성 자체에는 동의하는 의견이 주를 이루었다(〈표 4〉 참조).

4.1.2 연구데이터 활용 경험

다른 연구자 혹은 연구팀이 생산한 연구데이터를 활용해본 적 있다는 응답은 자주 활용함



〈그림 1〉 연구데이터 분석 시 겪는 어려움(N=484)

〈표 4〉 연구데이터 생산 경험

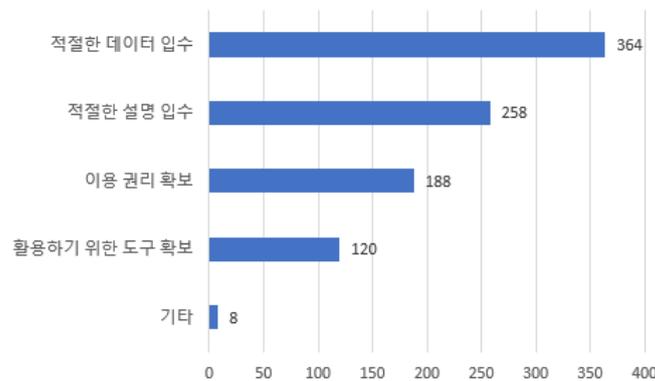
구분	연구데이터 저장/관리 어려움	프로그래밍 작업 필요 여부
매우 어려움	76 (15.7%)	97 (20.0%)
어려움	218 (45.0%)	215 (44.4%)
그렇지 않음	166 (34.3%)	126 (26.0%)
전혀 그렇지 않음	23 (4.8%)	40 (8.3%)
기타	1 (0.2%)	6 (1.2%)
합계	484	484

146명(26.7%), 가끔 활용함 269명(49.2%), 거의 활용하지 않음 116명(21.2%)으로 총 531명(97.1%)이었다. 이들을 대상으로 활용 시 겪었던 어려움을 조사해보니 적절한 연구데이터 입수(364명, 68.7%), 연구데이터에 대한 적절한 설명 입수(258명, 48.7%), 연구데이터 이용에 대한 라이선스, 권리에 대한 정보 확보(188명, 35.5%), 연구데이터를 활용하기 위한 도구를 확보(120명, 22.6%) 순으로 나타났다. 이외에 '데이터 품질', '데이터 신뢰성', '대규모 데이터에의 접근 및 전송' 등이 어려움으로 꼽혔다. 타인 혹은 타 연구팀에 의해 산출된 연구데이터가 필요하다는 물음에는 매우 필요함 146명(26.7%), 어느 정도 필요함 338명(61.8%)으로

필요하다는 응답이 상당히 높은 편이었다(〈그림 2〉 참조).

4.2 분석도구의 활용

본 조사에서 분석도구란 연구데이터의 활용을 위해 사용되는 소프트웨어, 스크립트를 포함한 컴퓨터 코드 등을 의미한다. 타 연구자 혹은 연구팀이 만든 분석 도구 활용 관련 경험을 묻는 질문에는 자주 활용한다는 응답이 90명(16.4%), 가끔 활용한다 275명(50.0%), 거의 활용하지 않는다 149명(27.1%), 전혀 활용하지 않는다 36명(6.5%)으로 활용한다는 응답이 전체의 약 64%를 차지하였다. 활용 시 어려움은 적절



〈그림 2〉 연구데이터 활용 시 겪는 어려움(N = 531)

한 설명 입수(349명, 68.0%), 존재 여부 파악(219명, 42.7%), 이용 권리 확보(190명, 37.0%), 활용하기 위한 인력 확보(186명, 36.3%) 순으로 응답자가 많았다. 기타 의견으로는 '분석가와와의 커뮤니케이션', '분석 정보의 타당성 인정 여부', '내가 가진 자원과의 적합성 여부', '자주 변화하는 버전 및 분석 환경에 대한 불일치'가 있었다. 분석 도구의 필요성은 매우 필요함과 어느 정도 필요함이 각 113명(20.5%), 347명(63.1%)으로 전체의 80%를 넘었다(〈표 5〉, 〈그림 3〉 참조).

4.3 컴퓨팅 자원의 활용

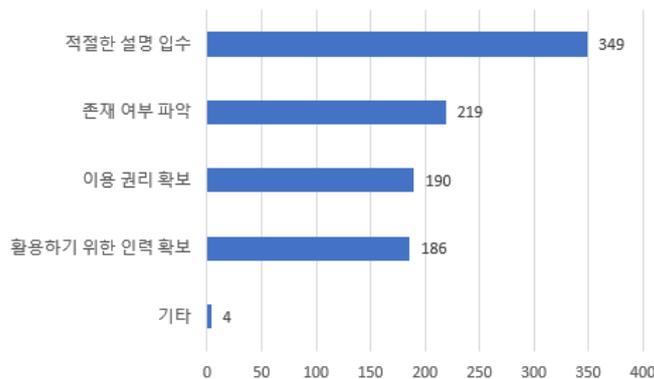
본 연구에서 컴퓨팅 자원은 연구자 또는 연구자가 소속된 실험실에서 보유하고 있는 컴퓨팅

장비 수준을 넘어서는 클라우드 컴퓨팅, 슈퍼컴퓨팅과 같은 자원을 의미한다. 연구를 수행하는 데 컴퓨팅 자원을 활용한 경험 관련 질문에 자주 활용함 96명(17.5%), 가끔 활용함 209명(38.0%), 거의 활용하지 않음 166명(30.2%), 전혀 활용하지 않음 78명(14.2%)으로 활용한 적 있다는 응답과 활용하지 않는다는 응답이 각각 55.5%와 44.4%로 나타났다. 기타 의견은 '컴퓨팅 자원이 없음'이었다.

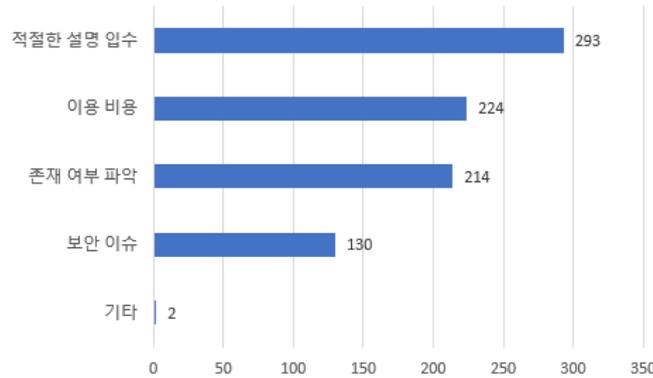
컴퓨팅 자원 활용 시 경험한 어려움은 적절한 설명 입수(293명, 62.6%), 이용 비용(224명, 47.9%), 존재 여부 파악(214명, 45.7%), 보안 이슈(130명, 27.8%) 순으로 나타났으며, 기타 의견으로 '활용을 위한 인적 자원의 부족', '자원 부족'이 있었다(〈그림 4〉 참조).

〈표 5〉 타 연구자 생산 자원 활용 경험

구분	타인 생산 데이터 활용 경험	타인 생산 분석도구 활용 경험	타인 소유 컴퓨팅 자원 활용 경험
자주 활용함	146 (26.7%)	90 (16.4%)	96 (17.5%)
가끔 활용함	269 (49.2%)	275 (50.0%)	209 (38.0%)
거의 활용하지 않음	116 (21.2%)	149 (27.1%)	166 (30.2%)
전혀 활용하지 않음	16 (2.9%)	36 (6.5%)	78 (14.2%)
기타	0 (0.0%)	0 (0.0%)	1 (0.2%)
합계	547	550	550



〈그림 3〉 분석도구 활용 시 겪는 어려움(N=513)



〈그림 4〉 컴퓨팅 자원 활용 시 겪는 어려움(N=468)

컴퓨팅 자원의 필요성 관련 문항에는 매우 필요함 138명(25.1%), 어느 정도 필요함 295명(53.6%), 별로 필요하지 않음 95명(17.3%), 전혀 필요하지 않음 22명(4.0%)으로 필요하다는 응답이 전체의 78.7%에 달했다(〈표 6〉 참조).

이 된다면 연구 개발에 있어서는 유리하지만, 경쟁사로의 데이터 유출이 우려되어 관련 플랫폼에 대한 논의 자체가 어려움', '국가 차원보다는 기존의 연구자들이 많이 공유하고 있는 각 유무상의 플랫폼들과의 연계하여 활용할 수 있는 환경이 필요함'이 있었다.

4.4 커먼즈의 필요성

연구데이터 커먼즈의 필요성에 대해서는 매우 필요함 147명(26.7%), 필요함 323명(58.7%), 필요하지 않음 63명(11.5%), 전혀 필요하지 않음 14명(2.5%)으로 필요하다는 응답이 전체의 85.4%를 차지하였다. 기타 응답으로는 '사안에 따라 필요성 있음', '관련 정보 통합/연계/활용

4.4.1 서비스 요구

연구데이터 커먼즈 서비스에 대한 요구를 연구데이터 관련 서비스와 분석 도구 및 컴퓨팅 자원 관련 서비스로 나누어 살펴보았다. 먼저, 연구데이터 관련 서비스에 대한 요구는 연계, 공유 서비스(378명, 71.2%), 저장 공간(340명, 64.0%), 검색 및 브라우징 서비스(320명, 60.3%),

〈표 6〉 타 연구자 생산 자원 및 커먼즈 필요성

구분	타인 생산 데이터	타인 생산 분석도구	타인 소유 컴퓨팅 자원	커먼즈
매우 필요함	146 (26.7%)	113 (20.5%)	138 (25.1%)	147 (26.7%)
어느 정도 필요함	338 (61.8%)	347 (63.1%)	295 (53.6%)	323 (58.7%)
별로 필요하지 않음	62 (11.3%)	75 (13.6%)	95 (17.3%)	63 (11.5%)
전혀 필요하지 않음	1 (0.2%)	15 (2.7%)	22 (4.0%)	14 (2.5%)
기타	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (0.5%)
합계	547	550	550	550

메타데이터 등록 및 관리 서비스(249명, 46.9%), 영구 식별자 생성 및 관리 서비스(189명, 35.6%) 순으로 나타났다. 이외에도 '보안프로그램', '보안, 특허 침해 가능성 여부', '보안 데이터에 대한 처리 방식이 매우 중요함, 연구 특성상 타인의 데이터는 보고 싶고 내 데이터는 감추고 싶은 경향이 매우 컸음', '데이터 보안, 권리 보호', '공개, 분석, 사용에 관한 License 이슈 해결' 등 데이터 보안 및 사용 권리에 대한 요구와 '개인별 연계 연구자료 업데이트 기능', '큐레이션', '접근성이 좋아야 함', '유용한 데이터의 수집방안 제공, 수집된 데이터의 경량화 방안', '연구 데이터에 대한 구체적인 설명, 사용을 위한 매

뉴얼 등'이 있었다(〈그림 5〉 참조).

다음으로 분석 도구 및 컴퓨팅 자원 관련 서비스에 대한 요구는 소프트웨어 검색 및 공유 서비스(364명, 68.5%), 대용량 데이터 분석, 처리 환경(330명, 62.1%), 컴퓨터 코드, 알고리즘 검색 및 공유 서비스(236명, 44.4%), 교육 및 훈련 기능(235명, 44.3%), API 검색 및 공유 서비스(230명, 43.3%), 협업 환경(191명, 36.0%), 개발 환경(147명, 27.7%), 컨테이너 서비스(101명, 19.0%) 순으로 나타났다. 기타 의견으로 '기존에 많이 사용하고 있는 GitHub, Stack Overflow 등 여러 외부 서비스와의 연동이 잘 되었으면 좋겠음'과 '가시화'가 있었다(〈그림 6〉 참조).



〈그림 5〉 연구데이터 관련 서비스 요구(N=531)



〈그림 6〉 분석 도구 및 컴퓨팅 자원 관련 서비스 요구(N=531)

4.4.2 커먼즈 활용 의향

연구데이터 커먼즈 활용 의향을 알아보기 위하여 커먼즈를 통해 연구데이터와 분석도구를 공유할 의향이 있는지와 커먼즈에서 제공하는 연구데이터와 분석도구, 컴퓨팅 자원을 사용할 의향이 있는지 살펴보았다. 먼저, 커먼즈가 생기면 이를 통해 연구데이터를 공유하겠다는 질문에 매우 동의함 94명(17.1%), 동의함 328명(59.6%), 동의하지 않음 85명(15.5%), 전혀 동의하지 않음 14명(2.5%)로 동의한다는 응답이 전체의 약 77%를 차지하였다. 기타 의견으로는 공유 의향에 있어서는 소속기관의 정책 혹은 상황에 따라 유동적일 수 있다는 전제를 바탕으로 부분 혹은 전체에 대해 동의하고 있다([부록 2] 참조).

커먼즈를 통해 본인 혹은 소속 연구팀이 생산한 분석도구를 공유하겠다는 응답 또한 매우

동의함 94명(17.1%), 동의함 370명(67.3%), 동의하지 않음 64명(11.6%), 전혀 동의하지 않음 10명(1.8%)으로, 동의한다는 응답이 전체의 약 84%로 나타났다.

다음으로 커먼즈를 통해 공유된 연구데이터, 분석도구, 컴퓨팅 자원을 활용할 의향이 있다는 응답을 살펴보면 <표 8>과 같다. 활용하는데 매우 동의 혹은 동의한다는 응답이 연구데이터는 전체의 약 92%, 분석도구와 컴퓨팅 자원은 약 93%를 차지하였다.

5. 논의

본 연구에서는 연구자들을 대상으로 실제 연구 상황에서 겪는 연구데이터, 분석도구, 컴퓨팅 자원 관련 주요 경험(활용/어려움)과 필요

<표 7> 커먼즈를 통한 자원 공유 의향

구분	연구데이터 공유 의향	분석도구 공유 의향
매우 동의함	94 (17.1%)	94 (17.1%)
동의함	328 (59.6%)	370 (67.3%)
동의하지 않음	85 (15.5%)	64 (11.6%)
전혀 동의하지 않음	14 (2.5%)	10 (1.8%)
기타	29 (5.3%)	12 (2.2%)
합계	550	550

<표 8> 커먼즈를 통해 공유된 자원 활용 의향

구분	연구데이터 활용 의향	분석도구 활용 의향	컴퓨팅 자원 활용 의향
매우 동의함	123 (22.4%)	116 (21.1%)	124 (22.5%)
동의함	382 (69.5%)	393 (71.5%)	385 (70.0%)
동의하지 않음	33 (6.0%)	33 (6.0%)	33 (6.0%)
전혀 동의하지 않음	6 (1.1%)	5 (0.9%)	7 (1.3%)
기타	6 (1.1%)	3 (0.5%)	1 (0.2%)
합계	550	550	550

를 살펴보고, 커먼즈의 필요에 대한 요구를 탐색해보았다. 즉, 개방형 연구 커먼즈의 실제 구현을 위한 기반을 모색해 보고자 한 것이다. 설문은 정량적 측정 방식을 통해 요구의 정도를 측정하였으며, 이와 더불어 자유기입 방식의 기타(의견)를 두어 가능한 다양한 의견을 수렴하고자 하였다. 따라서 본 연구의 결과는 정량적, 정성적 응답을 기반으로 다음과 같이 요약할 수 있다.

연구데이터 커먼즈의 주요 구성 요소인 타 연구자가 생산 혹은 소유한 연구데이터, 분석도구 및 컴퓨팅 자원의 활용 수준은 각각 약 76%, 약 66%, 약 55%로, 세 가지 요소 모두에 대해 반 이상의 응답자가 자주 혹은 가끔 활용한다고 답하였다. 커먼즈의 필요성에 대해서는 응답자의 약 85%가 필요하다는 답변을 하였으며, 구성 요소별 필요성은 데이터 입수(약 89%)와 분석도구(약 84%), 컴퓨팅 자원(약 79%)의 순으로 대부분의 항목이 높게 나타났다. 마지막으로 커먼즈 활용과 관련하여 데이터(약 76%), 분석도구(약 84%)의 공유 의사를 보여주었으며, 활용 의사는 데이터, 분석도구, 컴퓨팅 자원 모두 90% 이상으로 공유보다 높은 수준의 기대가 도출되었다.

본 연구에서 논의하고자 하는 커먼즈에 관한 직접적인 연구는 찾아보기 어렵다. 앞서 언급된 이상환(2019)의 연구 보고서는 국내 연구데이터 관리 및 활용에 관해 수행한 연구로, 실질적인 연구데이터 플랫폼의 개선을 위한 요구를 포함하고 있어 이들과 본 연구의 결과를 비교하여 살펴보는 것은 연구자들의 커먼즈에 대한 요구를 명확하게 하는 데 의미가 있다고 하겠다.

이상환(2019)의 연구에서는 국가과학기술연구회 소속 출연(연) 과제 책임자 301명을 대상으로 국가 R&D 과제에서 생산된 연구데이터 관련하여 설문조사를 실시하였다. 설문 결과 중 연구데이터 생산에 관한 응답은 생산 경험 있음 226명(75.1%)으로 보고하고 있다. 본 연구의 결과에서 약 88%의 응답자가 연구데이터를 생산해본 적 있다고 답하였다.

선행 연구 보고서에서 연구데이터 관리 애로사항은 전담부서, 예산, 시스템, 제도의 부재가 언급되었으며, 특히 '기타' 의견을 통해 제시된 '대량 데이터 사이즈에 대한 관리 어려움'은 본 연구의 연구데이터의 저장 및 관리의 어려움에 관한 문항에서 약 61%의 응답자가 어려움을 겪었다고 응답한 것과 맥락을 같이 한다고 볼 수 있겠다.

또한, 연구데이터 활용 경험과 관련하여 이상환(2019)의 보고서는 타 연구자 데이터 활용 애로사항으로 데이터 미공개, 존재/보유 위치 모름이 높은 순위를 차지하였다. 본 연구에서도 적절한 연구데이터를 입수하는 것을 가장 큰 어려움으로 언급하였으며, 분석도구의 존재 여부 파악 역시 활용 시 경험하는 어려움 중 높은 순위를 차지하였다. 컴퓨팅 자원 또한 약 46%의 응답자가 존재 여부 파악을 활용 시 경험한 어려움으로 꼽았다. 이와 같은 자원의 존재 여부 파악 관련 문제는 FAIR 원칙 중 발견 가능성(Findable)과 접근 가능성(Accessible)을 논의함으로써 해결방안을 모색해 보아야 할 것이다.

다음으로, 필요성에 대하여 이상환(2019)의 조사에서 타 연구자의 데이터가 필요하다는 응답이 약 82%로, 본 연구의 약 88%와 유사한

수준을 보이고 있다. 즉, 연구자들에게 선행 연구자의 데이터는 필요 가치가 있는 것으로 추정할 수 있다. 이에 더하여 이상환(2019)은 연구 보고서에서 연구데이터의 공유·활용 시스템 구축에 필요한 중요 기능으로 최적화된 검색 기능, 데이터 처리 분석 기능, 분석 환경 제공, 저장 공간의 제공 등의 요구를 수집한 바 있다. 이는 연구데이터 외에도 분석도구, 컴퓨팅 자원 등 커먼즈 서비스의 구성 요소를 의미하는 것으로, 연구자들이 원하는 차세대 플랫폼 서비스의 모형은 커먼즈 기반의 개방형 연구데이터 서비스임을 보여주는 것이라 하겠다.

본 연구에서는 커먼즈 기반 개방형 연구데이터 서비스의 구현을 위해 연구자들의 요구를 우선적으로 확인하고자 하였다. 그러나 연구데이터 커먼즈의 구성 요소에 대한 현재 활용 수준, 필요성, 활용 의사에 대한 응답자 특성별(직위, 연구과제에서의 역할 및 전공 분야) 통계적인 차이가 있는지 분산분석(ANOVA)을 실시한 결과 유의미한 차이는 발견되지 않았다. 이는 응답자의 특성과 상관없이 설문 문항에 대한 전반적으로 높은 긍정적인 응답에 기인한 것이라 할 수 있다. 이러한 결과는 연구데이터 커먼즈에 대한 높은 기대 수준을 반영한다고 할 수 있지만, 한편으로는 상당수의 연구자가 연구데이터 커먼즈를 본격적으로 경험하지 않은 상황이라는 점도 고려될 필요가 있다.

따라서 향후 연구데이터 서비스와 관련한 다양한 분야에서 커먼즈에 대한 활발한 논의가 필요하며, 실제 구현된 커먼즈 플랫폼을 실질적으로 이용한 이용자를 비롯하여 커먼즈와 관련된 다양한 이해관계자 집단을 대상으로 하는 후속 연구가 수행되어야 할 것이다.

6. 결론

학술연구의 기반이 되는 데이터 공개 및 공유의 중요성이 강조되면서 데이터 학술연구를 지원하기 위한 지식기반의 구축이 오픈사이언스 프레임워크에서 구상되고 있으며, 단계적으로 실현되고 있다. 연구데이터 지식기반은 단순히 시스템적인 요소뿐 아니라 법률, 정책적 기반과 이것을 운영하는 인적 기반 등을 포함한다. 2016년 발표된 FAIR 원칙(Wilkinson et al., 2016)은 연구데이터의 관리와 공유에 관한 전반적인 지침 역할을 하고 있다. 연구데이터의 공유와 활용에 대한 정책은 주로 미국, 영국 등의 국가에서 선도적으로 시행되었지만(심원식, 2021), 우리나라도 최근 국가연구개발사업에서 생산되는 연구데이터를 체계적으로 관리하고 공유, 활용할 수 있는 법적 기반이 마련되고 있으며(과학기술정보통신부, 2023), 입법 예고된 사항 중에는 연구데이터 통합플랫폼의 구축 및 연계에 관한 내용이 포함되어 있다.

전 세계 연구데이터 리포지토리의 목록인 re3data.org에 2023년 10월 초 기준 등록된 3,915개의 국가별 리포지토리 중에 미국 1,171개, 독일 505개, 캐나다 395개, 영국 321개, 프랑스 128개, 호주 103개, 일본 66개이며, 우리나라의 리포지토리는 13개로 파악된다. 이는 우리나라의 연구데이터가 체계적으로 관리 및 활용되려면 다양한 주제 분야별, 국가 차원의 데이터 플랫폼의 구축이 시급함을 보여준다.

본 연구는 향후 연구데이터 지식기반을 구축하는데 있어 보다 통합적인 플랫폼으로 제시되고 있는 연구 커먼즈 모델에 대한 국내 연구자의 요구를 통해 연구데이터, 분석도구, 컴퓨팅

자원 활용 경험을 살펴보고 개방형 연구 커먼즈에 대한 요구사항을 파악함으로써 향후 구축될 국내 연구데이터 지식기반의 토대를 마련하고자 하였다.

설문 결과 이미 많은 연구자가 타 연구자가 생산 혹은 소유한 연구데이터와 분석도구, 컴퓨팅 자원 등 연구 활동에 필요한 자원 활용의 필요를 인식하고 있으며, 이와 더불어 자원의 공유에 대해서도 긍정적인 요구와 기대를 가지고 있음이 밝혀졌다. 이러한 요구는 연구데이터 서비스 플랫폼의 진화 즉 연구 커먼즈에 대한 요구라 볼 수 있다.

본 연구에서 도출된 분석 결과뿐만 아니라 이와 더불어 앞선 유사 연구의 결과와 비교하여 살펴봄으로써 연구데이터 서비스가 차세대 서비스라고 할 수 있는 커먼즈 형태로 발전될

필요가 있음을 확인하였다는 점에서 연구의 의의가 있다. 또한, 관련 분야에서 커먼즈에 대한 논의가 적극적으로 이루어지지 않고 있는 상황에서 본 연구의 설문은 여전히 유의미한 가치를 가지고 있다고 볼 수 있다.

따라서 연구데이터 지식기반의 구축은 장기적인 과제로 향후 커먼즈를 포함한 연구데이터 인프라에 대한 활발한 논의가 필요하며, 실제로 커먼즈를 구현하기 위해서는 실제 이용자 집단을 대상으로 심층적인 후속 연구와 다양한 이해관계자 집단의 기대와 요구사항에 대한 파악과 진단이 필요하다. 이와 더불어 연구데이터 지식기반은 연구성과물 지식기반 구축 경험에 기반하여 연구성과물과 데이터의 복잡하고 미묘한 차이를 고려한 세심한 접근이 필요하다.

참 고 문 헌

- 과학기술정보통신부 (2023. 9. 27.). 국가연구데이터 관리 및 활용 촉진에 관한 법률 제정안 입법예고, 과학기술정보통신부공고 제2023-0890호.
- 김선태, 이정훈, 정한민 (2017). 연구데이터 이해와 관리. 대전: 한국과학기술정보연구원.
- 김지현 (2015). 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 한국문헌정보학회지, 49(3), 413-436. <http://doi.org/10.4275/KSLIS.2015.49.3.413>
- 박미영, 안인자, 남승주 (2018). 과학기술분야 출연연구기관 연구데이터 관리 및 공유 사례 분석 연구. 한국비블리아학회지, 29(4), 319-344. <https://doi.org/10.14699/kbiblia.2018.29.4.319>
- 송사광, 조민희, 이미경, 김다솔, 박성은, 이상백, 임형준 (2023). 컴퓨팅 리소스 공유활용을 위한 연구데이터 커먼즈 프레임워크 구축 방안 연구. 한국경영과학회 2023년 춘계 공동학술대회 논문집, 1305-1314.
- 신은정, 정원교 (2017). 오픈사이언스정책의 확산과 시사점(STEPI Insight 216). 과학기술정책연구원.
- 심원식 (2021). 국가 연구데이터 클라우드 구축 기획 연구(K-21-L01-C04). 한국과학기술정보연구원.

- 안새롬 (2020). 전환 담론으로서 커먼즈: 대기 커먼즈를 위한 시론. *ECO*, 24(1), 333-370.
<http://www.doi.or.kr/10.22734/ECO.24.1.202006.009>
- 유사라 (2019). 차세대 학술연구 데이터 공유 활성화를 위한 연구기록의 구조적 요건에 대한 연구. *한국문헌정보학회지*, 53(3), 101-120. <https://doi.org/10.4275/KSLIS.2019.53.3.101>
- 이상환 (2019). 출연(연) 연구데이터 관리·활용 방안 연구(2018-09). 한국과학기술정보연구원.
- 황미녕, 박상배, 신영호, 신성호, 서동민 (2023). 연구데이터 공유·관리·분석을 위한 국가연구데이터 플랫폼(DataON) 서비스 운영. *한국콘텐츠학회논문지*, 23(1), 36-47.
<http://doi.org/10.5392/JKCA.2023.23.01.036>
- Beagle, D. (1999). Conceptualizing an information commons. *Journal of Academic Librarianship*, 25(2), 82-89. [https://doi.org/10.1016/S0099-1333\(99\)80003-2](https://doi.org/10.1016/S0099-1333(99)80003-2)
- BOAI(Budapest Open Access Initiative) (2002). Read the Declaration.
 Available: <https://www.budapestopenaccessinitiative.org/read/>
- Bourne, P. E., Bonazzi, V., Brand, A., Carroll, B., Foster, I., Guha, R. V., Hanisch, R., Keller, S. A., Kennedy, M. L., Kirkpatrick, C., Mons, B., Nusser, S. M., Stebbins, M., Strawn, G., & Szaly, A. (2022). Playing catch-up in building an open research commons. *Science*, 377(6603), 256-258. <https://doi.org/10.1126/science.abo5947>
- Bryant, R., Lavoie, B., & Malpas, C. (2017). The Realities of Research Data Management: Part Two: Scoping the University RDM Service Bundle. OCLC. <http://doi.org/10.25333/C3Z039>
- Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Saffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, 75(6), 1432-1462. <https://doi.org/10.1108/JD-12-2018-0211>
- Grossman, R. L. (2019). Data lakes, clouds, and commons: a review of platforms for analyzing and sharing genomic data. *Trends in Genetics*, 35(3), 223-234.
<https://doi.org/https://doi.org/10.1016/j.tig.2018.12.006>
- Sheikh, A., Malik, A., & Adnan, R. (2023). Evolution of research data management in academic libraries: a review of the literature. *Information Development*.
<https://doi.org/10.1177/02666669231157405>
- UNESCO (2021). UNESCO Recommendation on Open Science.
<https://doi.org/10.54677/MNMMH8546>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G. A., Axton, M., Baak, A., Blomberg, N., Boiten, J., Santos, L. B. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C., Hooft, R., Kuhn,

T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Lei, J., Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1-9. <https://doi.org/10.1038/sdata.2016.18>

• 국문 참고자료의 영어 표기

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

- Ahn, Saerom (2020). Commons as transition discourse: a conceptual basis for the study on atmosphere as commons. *ECO*, 24(1), 333-370. <http://www.doi.or.kr/10.22734/ECO.24.1.202006.009>
- Hwang, Mi-Nyeong, Park, Sang Bae, Shin, Young-Ho, Shin, Sungho, & Seo, Dongmin (2023). DataON service operation for research data sharing, management and analysis. *Journal of the Korea Contents Association*, 23(1), 36-47. <http://doi.org/10.5392/JKCA.2023.23.01.036>
- Kim, Ji Hyun (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. <http://doi.org/10.4275/KSLIS.2015.49.3.413>
- Kim, Suntae, Lee, Jeonghoon, & Jung, Hanmin (2017). *Understanding and Managing Research Data*. Daejeon: Korea Institute of Science and Technology Information.
- Korean Ministry of Science and ICT (2023, September 27). Notice of legislative proposal for the Act on Promotion of Management and Use of National Research Data. Notification No. 2023-0890.
- Lee, Sanghwan (2019). *Research Data Management Status of Science and Technology Research Institutes in Korea (2018-09)*. Korea Institute of Science and Technology Information.
- Park, Miyoung, Ahn, Inja, & Nam, Seungjoo (2018). A study on the analysis of research data management and sharing of science & technology government-funded research institutes. *Journal of the Korean Biblia Society for Library and Information Science*, 29(4), 319-344. <https://doi.org/10.14699/kbiblia.2018.29.4.319>
- Shim, Wonsik (2021). *Strategic Planning for the Development of Korea Research Data Cloud (K-21-L01-C04)*. Korea Institute of Science and Technology Information.
- Shin, Eunjung & Jung, Wonkyo (2017). *Diffusion and Implications of Open Science Policies*

(STEI Insight 216). Science & Technology Policy Institute.

Song, Sa-kwang, Cho, Minhee, Lee, Mikyoung, Kim, Dasol, Park, Seongeun, Lee, Sangbaek, & Yim, Hyung-Jun (2023). A study on the research data commons framework for computing resource sharing and utilization. In Proceedings of the 2023 Spring Joint Conference of the Korean Operations Research and Management Science Society, 1305-1314.

Yoo, Sarah (2019). Reconsideration of research framework for RRM in the perspective of linked open data. Journal of the Korean Society for Library and Information Science, 53(3), 101-120. <https://doi.org/10.4275/KSLIS.2019.53.3.101>

[부록 1] 항목별 기타 응답 종합(1)

구분	질의	기타 응답
연구데이터의 활용	분석의 어려움	<ul style="list-style-type: none"> • 결과값에 대한 분석의 어려움이 있음 • 분석을 위한 다양한 참고모델 및 관련 예시 미보유 • 분산되어 있는 데이터 명세, 수집, 관리 • 데이터의 확보 • 인력 • 분석 환경 설정의 어려움 • (데이터의) 전처리
	활용의 어려움	<ul style="list-style-type: none"> • 데이터 품질 • 데이터 신뢰성 • 대규모 데이터에 대한 접근 및 전송
	컴퓨터 프로그래밍 작업의 필요 여부	<ul style="list-style-type: none"> • 필요에 따라 • 필요하나 사용할 줄 모름
분석도구의 활용	활용의 어려움	<ul style="list-style-type: none"> • 분석가와의 커뮤니케이션 • 분석 정보의 타당성 인정 여부 • 내가 가진 자원에의 적합성 여부 • 자주 변화하는 버전 및 분석 환경에 대한 불일치
컴퓨팅 자원의 활용	활용의 어려움	<ul style="list-style-type: none"> • (활용 가능한) 컴퓨팅 자원이 없음 • 자원 부족 • 활용을 위한 인적 자원의 부족

[부록 2] 항목별 기타 응답 종합(2)

구분	질의	기타 응답
커먼즈의 필요	커먼즈 서비스 (플랫폼)의 필요	<ul style="list-style-type: none"> • 사안에 따라 필요성 있음 • 관련 정보 통합/연계/활용이 된다면 연구 개발에 있어서는 유리하지만, 경쟁사로의 데이터 유출이 우려되어 관련 플랫폼에 대한 논의 자체가 어려움 • 국가 차원보다는 기존의 연구자들이 많이 공유하고 있는 각 유무상의 플랫폼들과의 연계하여 활용할 수 있는 환경이 필요함
	서비스에 대한 요구	
	연구데이터 서비스에 대한 요구	<ul style="list-style-type: none"> • 큐레이션 • 연구데이터에 대한 구체적인 설명, 사용을 위한 매뉴얼 등 (필요) • 접근성이 좋아야 함 • 개인별 연계 연구자료 업데이트 기능 • 유용한 데이터의 수집방안 제공, 수집된 데이터의 경량화 방안 • 데이터 보안, 권리 보호 • 보안, 특허 침해 가능성 여부 • 보안프로그램 • 보안 데이터에 대한 처리 방식이 매우 중요함, 연구 특성상 타인의 데이터는 보고 싶고 내 데이터는 감추고 싶은 경향이 매우 큼 • 공개, 분석, 사용에 관한 License 이슈 해결
	분석 도구 및 컴퓨팅 자원에 대한 요구	<ul style="list-style-type: none"> • 기존에 많이 사용하고 있는 GitHub, stackoverflow 등 여러 외부 서비스와의 연동이 잘 되었으면 좋겠음 • 가시화
	커먼즈를 통한 자원 공유 의향	
연구데이터 공유 의향	<ul style="list-style-type: none"> • 상황이 중요함 • 일부 데이터 공유에 동의함 • 공유 데이터의 범위 선택권 필요 • 공유 가능한 부분을 검토 해봐야 할 것 같음 • 주제와 내용의 공개범위에 따라 결정될 듯 • 분야에 따라 다름. 협업이 필수적인 분야에서는 공유하고 그렇지 않으면 안 함 • 동의 하지만, 국외유출을 대비한 보안을 강화 • 공유할 의향은 있으나 권한이 없음 • 소속기관의 검토와 승인에 따라 공유 • 회사 내규 및 연구개발 보안 규정에 따름 • 사기업이므로 대표이사 승인 하에 가능 • 기관 허락 획득 후 공유 • 회사 정책상 가능한 범위 내 공유 가능 • 상사 결제 필요 • 회사 정책에 따른다(법무팀 의견필요) • 기관 정책에 따라 달라짐 	

구분	질의	기타 응답
커먼즈의 필요	연구데이터 공유 의향	<ul style="list-style-type: none"> • 특히관련 제외 공유 가능 • 손쉬운 공유 보다는 공유 받는 사람이 누구인지와 활용 목적이 무엇인지가 중요함 • 이미 기관에서 공유 중 • 영리집단이라 영업보안에 관련된다면 어렵겠으나 학회나 협회에 발표한 연구자료는 가능한 정도 • 기관 특성상 공유가 불가능한 자료는 공유 불가, 일부 자료는 가능 • 정출연이라 공유가 현실적으로는 어려울 것으로 판단됨 • 의료기관의 개인정보가 포함된 진단데이터가 많아 IRB 등 별도 절차 없이 공개가 전혀 불가능함 • 경쟁사로의 유출 우려로 공유가 쉽지 않음 • 동의하나 회사 사정상 쉽지 않아 보임 • 정보공유 계약이 필요함 • 엠바고 기간 필요 • 데이터 암호화 수준 및 보안사항에 따라 상이 • 잘 모르겠음
	분석 도구 공유 의향	<ul style="list-style-type: none"> • 표준계약 검토 필요 • 관련 법, 법령에 근거하여 가능한 범위 내 • 회사 내규 및 연구개발 보안 규정에 따름 • 보안규정 등에 따라 상이 • 사안에 따라 공유할 의향이 있음, 결정은 연구원(소속 기관)에서 함 • 기관 허락 획득 후 공유 • 회사 정책상 가능한 범위 내 공유 가능 • 공유 가능한 부분을 검토 해봐야 할 것 같음 • 주제 대상 기능에 따라서 공유 가능한 범위 내용의 것은 공유 • 경쟁사로의 유출 우려로 쉽지 않음 • 개인 연구자가 자체적으로 만든 도구의 완성도는 타 연구자가 사용하기에 높지 않아 선뜻 공유하기 어렵습니다 • 손쉬운 공유 보다는 공유 받는 사람이 누구인지와 활용 목적이 무엇인지가 중요함
커먼즈를 통한 자원 활용 의향		
	연구데이터 활용 의향	<ul style="list-style-type: none"> • 적절한 데이터 확보 여부 및 데이터의 활용 가능 정도 등 따라 상이 • 퀄리티가 보증된다면 사용하겠음 • 신뢰도와 데이터 획득 관련 구체적 정보가 제시된 연구데이터일 경우 동의함 • 조건부 동의 • 연구 중에서 보안자료가 존재하여 일률적으로 판단하기 어려움
	분석 도구 활용 의향	<ul style="list-style-type: none"> • 분석도구의 활용 용이성 및 적용 가능성 등 따라 상이 • 연구목적에 따라 커스터마이징 가능한 경우에 한해 활용 • 보안자료에 대한 접근권한에 따라 상이
	컴퓨팅 자원 활용 의향	<ul style="list-style-type: none"> • 보안상 어려울 것으로 간주