

국가연구데이터커먼즈 체계 수립을 위한 연구데이터 관리자들의 인식에 관한 연구*

A Study on the Perception of Research Data Managers to Establish a Korea Research Data Commons System

박성은 (Seong-Eun Park)** , 이미경 (Mikyong Lee)***
조민희 (Minhee Cho)**** , 송사광 (Sa-Kwang Song)*****
김다솔 (Dasol Kim)***** , 임형준 (Hyung-Jun Yim)*****

초 록

본 연구는 한국과학기술정보연구원(KISTI)에서 개발하고 있는 국가연구데이터커먼즈(KRDC)를 실제 이용할 국가과학기술연구회(NST) 산하 정부출연연구기관의 연구데이터 관리자를 대상으로 연구데이터를 분석하기 위한 인프라와 서비스의 현황을 파악하고, KRDC 체계 구축과 관련한 연구데이터 관리자들의 인식을 조사하는 것을 목적으로 하였다. 이를 위해 KISTI를 제외한 24개의 정부출연연구기관을 대상으로 설문조사 실시하였으며, 설문조사에 응답한 15개 기관 중 후속 인터뷰에 동의한 9개 기관의 연구데이터 관리자를 대상으로 인터뷰를 수행하였다. 설문 결과, 대부분의 기관들이 관련 서비스를 제공하고 있었으며, 연구데이터 활용을 위한 통합 분석 프레임워크의 도입과 외부에 공개된 분석 SW를 사용할 수 있는 체계에 대한 제공 의향 역시 높은 것으로 나타났다. 한편 후속 인터뷰를 통해 각 기관별로 제공하는 분석 서비스의 외부 공개 현황을 파악해보았을 때, 매우 소수의 기관만이 이를 외부에 공개하고 있었다. 이러한 연구 결과를 분석해보면, 프레임워크를 통해 분석 인프라와 서비스가 제공될 경우 활용하고자 하는 수요가 있으나, 각 기관에서 보유하고 있는 분석 자원을 공개 및 공유하기 어렵다는 것을 알 수 있다. KRDC 체계 구축을 위해서는 연구 현장에서의 분석 인프라와 분석 서비스의 공유가 필수적인 만큼 연구 현장에서의 인식 전환, 나아가 제도적 변화가 필요하며, 후속 인터뷰에서 제시된 시스템의 편리성, 보안, 보상체계 등을 잘 고려하는 정책을 수립하기 위해 노력할 필요가 있다.

ABSTRACT

The purpose of this study is to identify the current status of infrastructure and services for analyzing research data for research data managers at government-funded research institutions under the National Research Council for Science and Technology (NST) who will actually use the Korea Research Data Commons (KRDC), which is being developed by the Korea Institute of Science and Technology Information (KISTI) and to investigate the perceptions of research data managers related to the establishment of KRDC system. For the study, we conducted a survey targeting 24 government-funded research institutes, excluding KISTI, and interviewed research data managers from 9 of the 15 institutions surveyed who agreed to follow-up interviews. As a result of the survey, most institutions were providing related services, and their willingness to introduce an integrated analysis framework for the use of research data and provide a system for using externally released analysis software was also high. Meanwhile, when we investigated the external disclosure status of each institution's analysis services through follow-up interviews, only a minimal number of institutions were disclosing them to the outside world. The findings reveal that there is a demand to utilize analysis infrastructure and services when provided through the framework. However, it is difficult to disclose and share the analysis resources held by each organization. In order to establish the KRDC system, it is essential to share research sites' analysis infrastructure and services, and in addition, changes in the perception of research sites and institutional changes are necessary. Furthermore, there is a need to establish policies that consider the system's convenience, security, and compensation system raised in the follow-up interviews.

키워드: 국가연구데이터커먼즈, 연구데이터, 분석인프라, 분석서비스, 프레임워크

Korea Research Data Commons, research data, analysis infrastructure, analysis services, framework

* 이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구임
(과제번호: (KISTI) K24L1M2C3).

** 한국과학기술정보연구원 연구데이터공유센터 선임기술원(pse3598@kisti.re.kr) (제1저자)

*** 한국과학기술정보연구원 연구데이터공유센터 책임연구원(jerryis@kisti.re.kr) (공동저자)

**** 한국과학기술정보연구원 연구데이터공유센터 책임연구원(mini@kisti.re.kr) (공동저자)

***** 한국과학기술정보연구원 연구데이터공유센터 책임연구원, UST 응용AI학과 교수
(esmallj@kisti.re.kr, esmallj@ust.ac.kr) (공동저자)

***** 한국과학기술정보연구원 연구데이터공유센터 기술원(dskim@kisti.re.kr) (공동저자)

***** 한국과학기술정보연구원 연구데이터공유센터 센터장(hjyim@kisti.re.kr) (교신저자)

■ 논문접수일자: 2024년 2월 20일 ■ 최초심사일자: 2024년 3월 6일 ■ 게재확정일자: 2024년 3월 11일

■ 정보관리학회지, 41(1), 465-486, 2024. <http://dx.doi.org/10.3743/KOSIM.2024.41.1.465>

※ Copyright © 2024 Korean Society for Information Management

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 연구의 배경 및 필요성

2000년대 과학 연구환경의 변화에 따라 2004년 OECD(Organisation for Economic Co-operation and Development) 과학기술장관 회의에서 '공적자금에 투입된 연구데이터의 접근에 관한 선언(Declaration on Access to Research Data from Public Funding)'을 채택함으로써 연구 결과 및 과정을 개방·공유하는 오픈사이언스 개념이 대두되었다. 오픈사이언스의 개념이 등장한 것은 20년이 넘었지만, 2020년대에 들어 개방형 과학을 통한 글로벌 난제 공동해결의 필요성이 부각되고, 미국이 2023년을 '오픈 사이언스의 해(The year of Open Science)'로 선포한 데에서 볼 수 있듯이(Office of Science and Technology Policy, 2023), 오픈사이언스의 중요성이 다시금 강조되고 있다.

디지털 기술의 발전으로 연구개발 패러다임이 데이터 기반의 과학으로 변화하고 있고, 지식정보의 공유와 공정 접근을 보장하는 오픈사이언스가 중요해지는 시점에서, 연구데이터, 나아가 연구소프트웨어의 개방과 공유는 연구의 재현성 및 투명성을 향상시켜 연구의 신뢰도를 높이고, 중복 연구를 최소화하여 연구개발에 대한 투자 효율성을 증대시켜 결과적으로 국가 연구개발 경쟁력이 향상되는 결과를 가져올 수 있다.

최근 데이터 중심의 연구개발을 지향하면서 연구데이터뿐만 아니라 연구개발 전 과정에 필요한 소프트웨어, 인프라와 같은 리소스를 공유하고 활용하기 위한 정책 및 프로젝트가 선

진국을 중심으로 추진되고 있으며, 유럽 오픈사이언스 클라우드(Europe Open Science Cloud, 이하 EOSC), 호주 연구데이터커먼즈(Australia Research Data Commons, 이하 ARDC)를 대표적인 사례로 볼 수 있다.

한국과학기술정보연구원(Korea Institute of Science and Technology Information, 이하 KISTI)에서는 데이터 기반의 연구개발을 지원하기 위해 국가연구데이터플랫폼(이하 DataON)을 개발하여 2020년부터 운영하고 있으며, 현재는 DataON과 함께 분석 인프라, 분석 서비스 등을 공유하고 활용할 수 있도록 하는 국가연구데이터커먼즈(Korea Research Data Commons, 이하 KRDC)를 구축하고 있다.

연구데이터커먼즈의 구축 과정에서 심원식 외(2023)는 개방형 연구 커먼즈(Open Research Commons)의 구현을 위한 연구데이터플랫폼에 대한 연구자 요구를 분석한 바 있다. 해당 연구에서 주제로 삼은 개방형 연구 커먼즈는 전통적인 의미의 학술출판물 외에 연구데이터를 수집, 공유하는 것은 물론 데이터를 분석하기 위한 다양한 소프트웨어 도구 및 컴퓨팅 자원을 통합적으로 제공하는 것을 의미하며(Bourne et al., 2022), 본 연구에서는 연구데이터와 더불어 연구데이터를 활용할 수 있는 분석 인프라 및 분석 서비스를 제공해주는 체계의 관점으로 연구데이터커먼즈라는 표현을 사용하였다.

심원식 외(2023)는 개방형 연구 커먼즈의 잠재적 이용자들을 대상으로 실제 연구 상황에서 겪는 연구데이터, 분석도구, 컴퓨팅 자원과 관련한 주요 경험과 필요를 살펴보고, 개방형 연구 커먼즈의 필요에 대한 요구를 탐색함으로써, 개방형 연구 커먼즈의 실제 구현을 위한 기반을

모색하고자 하였다. 설문조사를 실시하여 550 명의 응답을 분석한 결과, 응답자의 약 85%가 개방형 연구 커먼즈의 필요성에 동의하였으며, 커먼즈 서비스가 제공되면 이를 통하여 본인의 연구데이터(약 77%)와 분석도구(약 84%)를 공유하고자 하는 의향을 나타냈으며, 또한 공유된 연구데이터(약 92%)나 분석도구(약 93%), 컴퓨팅 자원(약 93%)을 활용하고자 하는 의향도 나타났다. 즉 많은 연구자들이 개방형 연구 커먼즈에 대한 요구가 매우 높은 것이라 이해할 수 있다.

그러나 해당 연구에서는 잠재적 이용자를 대상으로 한 요구를 분석하였기 때문에, 실제 구현된 개방형 연구 커먼즈를 실질적으로 이용할 이용자를 비롯하여 다양한 이해관계자 집단을 대상으로 하는 후속 연구의 필요성을 제기할 바 있다.

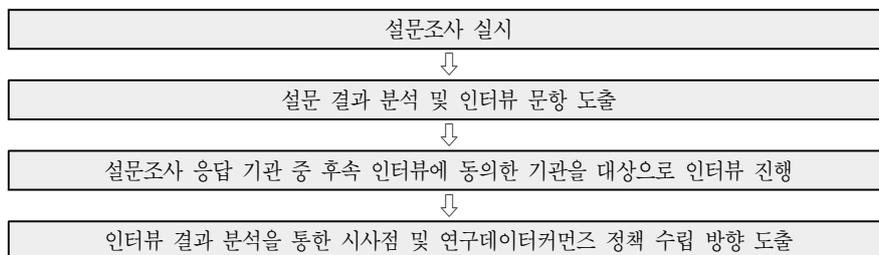
따라서 본 연구에서는 KISTI에서 개발 중인 국가연구데이터커먼즈를 실제 이용할 국가과학기술연구회(National Research Council of Science & Technology, 이하 NST) 산하 정부출연연구기관(이하 출연(연))의 연구데이터 관리자를 대상으로 설문조사 및 후속 인터뷰를 통해 연구 현장에서 활용이 되는 분석 인프라 및 서비스에 대한 구체적인 의견과 경험을 살펴

보고자 하였다. 연구데이터를 활용하는 분석 인프라와 분석 서비스를 제공하는 연구데이터커먼즈 프레임워크의 필요성을 이해하고, 이용자 요구사항이 반영된 연구데이터커먼즈 체계를 구축하는 기초를 마련하였다는 점에서 본 연구의 의의가 있다.

1.2 연구 절차

연구데이터 관리자 인식에 대한 연구데이터커먼즈 체계에 대한 인식을 조사하기 위한 절차는 다음과 같다:

첫 번째 단계에서는 NST 산하 25개 출연(연)들 중 설문 수행 주체인 KISTI를 제외한 24개 출연(연)을 대상으로 연구데이터를 활용하기 위한 분석 인프라와 서비스의 현황과 연구데이터 관리자 인식에 대해 파악하기 위한 설문조사를 실시하였다. 두 번째 단계에서는 실시한 설문에 대한 결과를 분석하고, 이를 바탕으로 인터뷰 문항을 도출하였다. 세 번째 단계에서는 설문조사 응답에서 후속 인터뷰에 동의한 9개 기관의 연구데이터 관리자들을 대상으로 인터뷰를 진행하였다. 마지막으로 인터뷰 결과를 분석하여 시사점 및 연구데이터커먼즈 정책 수립 방향을 도출하였다(〈그림 1〉 참조).



〈그림 1〉 연구 절차 도식화

2. 연구데이터커먼즈

2.1 해외 현황

주요 선진국은 새로운 지식과 가치의 창출 및 공유·확산을 위해 적극적으로 오픈사이언스 정책을 추진하고, 연구데이터 및 인프라의 공동 활용을 위한 생태계 구축을 진행하고 있다. 대표적인 예로 EOSC, ARDC를 비롯하여, 연구데이터동맹(Research Data Alliance, 이하 RDA)의 Global Open Research Commons(이하 GORC), 과학기술데이터위원회(The Committee on Data for Science and Technology, 이하 CODATA)의 Global Open Science Cloud(이하 GOSC)가 있으며, 각 사례를 하나씩 살펴보면 다음과 같다.

EOSC는 2015년 5월 유럽 위원회(European Commission, 이하 EC)가 유럽의 과학 연구 디지털 전환을 지원하기 위하여 제안한 것으로, 유럽의 기존 연구데이터 인프라를 연합하고 과학을 위한 FAIR 데이터 및 관련 서비스의 웹을 실현(Web of FAIR Data and services)하여 FAIR 원칙에 따라 연구데이터를 상호 운용하고 기계가 실행 가능하도록 만드는 것을 목표로 한다. EOSC의 역할은 연구자들에게 학문과 지리적 경계를 넘어 과학적인 발견과 협력을 지원하는 서비스를 제공하는 것이다(Budroni, Claude-Burgelman, & Schouppe, 2019).

2017년 6월 첫 번째 EOSC Summit에서 프로세스가 시작되어(약 70개국), 2018년 3월 유럽위원회에서 'EOSC 구현을 위한 로드맵'을 발

표하였으며, 2020년 EOSC의 주요 기능이 구현되어 EOSC Portal¹⁾이 오픈됨에 따라 EU 회원국에 분산된 기존 연구데이터 인프라들을 연합하고, 이에 대한 접근 기반을 마련하였다. EOSC 프로젝트는 여러 분야에 걸쳐 제공되는 연구데이터 및 서비스 시스템에 원활하고 개방적인 액세스를 가능하게 함으로써 오픈사이언스 구현에 기여한다는 특징을 가지고 있다(이미경, 조민희, 임형준, 2023).

구현 초기 단계(2018~2020)에서 EC는 약 800억 유로의 예산이 투입된 EU의 연구 및 혁신 기금 프로그램인 Horizon 2020(2014~2020)을 통해 2억 5천만 유로를 투자받았으며(European Commission, 2020), 구현 단계(2021~2027)에서는 연구 및 혁신을 위한 EU의 주요 자금 지원 프로그램으로 955억 유로의 예산을 가진 Horizon Europe 프로그램의 재정 지원으로 Strategic Research and Innovation Agenda(SRIA)에 따라 21억 8천6백만 유로를 투자받아 진행하고 있다.

다음으로 ARDC²⁾는 호주 연구 문화를 발전시키기 위한 20년 비전의 일환으로 국가 협력 연구 인프라 전략(National Collaborative Research Infrastructure Strategy, 이하 NCRIS) 이니셔티브로서 발족하여, EOSC 및 NCRIS 국가 연구 인프라 로드맵에 따라 2018년 7월 데이터 수집 역할을 하던 ANDS(Australian National Data Service), 클라우드 서비스 운영 역할을 하던 NeCTAR(National eResearch Collaboration Tools and Resources), 데이터 저장 및 관리 역할을 하던 RDS(Research Data Services)를 통

1) <https://eosc-portal.eu/>

2) <https://ardc.edu.au/about-us/who-we-are/>

합하여 2019년 5월 독립적인 법적 기관으로 공식 출범한 것이다(Barker, Wilkinson, & Treloar, 2019).

ARDC에서 정의한 연구데이터커먼즈는 데이터와 관련된 리소스(스토리지, 컴퓨팅, 모델)를 결합하여 연구자가 세계적 수준의 데이터 집약적 연구를 수행하고 협업할 수 있도록 하는 일종의 디지털 커먼즈이다. ARDC는 데이터를 통해 호주 연구자에게 경쟁 우위를 제공하고 호주 연구 품질과 영향력 극대화를 목적으로 연구자의 서비스 탐색, 접근, 효과적인 사용을 보조하는 연구 환경을 개발하는 것을 목표로 하고 있으며, ANDS, NeCTAR, RDS를 통합적으로 관리 및 운영하고 디지털 연구 인프라, 플랫폼, 기술, 데이터 컬렉션에 접근 가능하도록 지원하는 역할을 하고 있다(Finkel, 2017).

호주 정부는 NCRIS 프로그램을 통해 2016~2022년도에 ARDC 통합 전 e-Research 프로젝트에 1억 6천만 달러의 예산을 할당한 바 있고, 2023~2027년도의 ARDC 프로젝트에는 1억 2천만 달러를 편성한 바 있다.

한편 RDA는 연구데이터 분야의 가장 대표적인 글로벌 기구로써, 연구데이터와 관련된 다양한 리소스 공통요소인 연구데이터커먼즈를 상호운용할 수 있도록 하기 위해 GORC Interest Group을 통해 연구데이터커먼즈의 글로벌 표준 모델 개발 작업을 진행하고 있다(Payne et al., 2023). RDA GORC International Model Working Group³⁾은 GORC Interest Group의 일환으로 진행 중이며, EOSC, 아프리카 오픈 사이언스 플랫폼(African Open Science Platform),

국제 가상 천문대 연합(International Virtual Observatory Alliance, IVOA)과 같은 국가, 범국가 및 도메인별 조직들 간의 협력을 지원하기 위한 결과물을 개발하고 있다.

마지막으로 CODATA는 ISC(International Science Council) 산하 위원회로써, GOSC Initiative⁴⁾를 만들어 RDA의 GORC와 상호 참조하며 글로벌 오픈 사이언스의 실현을 위한 작업(Working Group)을 진행하고 있다. GOSC는 조화된 정책, 상호 운용 가능한 프로토콜, 투명한 서비스 및 지속적인 메커니즘을 통해 글로벌 협력 및 개방형 과학을 위한 대륙 간 인프라 및 가상 연구 환경을 공동 설계하고 공동 구축하는 것을 목적으로 하며, 현재 유럽, 미국, 중국, 호주, 캐나다, 한국 등이 참여하고 있다.

2.2 국내 현황

국내의 경우, 연구데이터와 함께 연구데이터 활용을 위한 분석 인프라와 서비스를 제공하는 플랫폼으로 KISTI의 DataON-CANVAS 분석 서비스, 한국화학연구원의 화학 분야 인공지능 플랫폼(ChemAI), 소재 연구데이터 지능형 플랫폼(K-MDS), 바이오소재 정보 통합플랫폼(BioOne) 등이 있다. 해당 플랫폼들은 연구데이터와 함께 데이터를 분석할 수 있는 인프라와 툴킷을 함께 제공한다.

KISTI에서는 DataON을 통해 공유되는 연구데이터의 활용을 극대화하기 위해 KRDC를 구축하고 있다(임형준 외, 2022). KRDC는 연구데이터와 리소스를 함께 연계 활용하는 체계이며,

3) <https://www.rd-alliance.org/groups/gorc-international-model-wg>

4) <https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/>

연구데이터, 연구소프트웨어, OpenAPI, 웹서비스, 가상 머신, 데이터 저장소, 클라우드 등의 공개된 리소스들을 논리적으로 연합(Federation)하여 연구데이터 활용성을 극대화하는 것을 목표로 하고 있다. 현재는 다양한 리소스 중에 연구데이터를 분석·활용하기 위한 연구소프트웨어⁵⁾의 공유·활용에 초점을 두고 연구데이터 커먼즈 체계를 개발하고 있다. 연구 결과인 논문에서 사용한 연구소프트웨어의 재현성은 과학에서 매우 중요하며, 재현성 확보를 위한 개방과 공유가 중요하기 때문이다. 또한 KRDC는 EOSC과 마찬가지로 FAIR 원칙에 따라 데이터 중심의 연구개발에 필요한 리소스들을 자유롭게 발견(Findable), 접근(Accessible), 상호운용(Interoperable), 재사용(Reusable)할 수 있도록 제공하기 위해 노력하고 있다(Wilkinson et al., 2016).

KISTI는 RDA의 GORC, CODATA의 GOSC 등의 글로벌 표준화 활동에 참여하며(송사광 외, 2023), 글로벌 연구데이터 커먼즈 기관(EOSC, ARDC, RDC 등)과 협력을 기반으로 상호운용성 확보를 통해 국가적 오픈사이언스 역할을 강화하고자 하고 있다. KRDC 상호운용성 확보를 위한 프레임워크 및 허브 서비스의 개발 역시 이러한 맥락에 있으며, 본 연구를 통해 KRDC 체계 구축을 위한 리소스 관리 및 활용을 위한 정책을 수립하고자 한다.

KRDC의 구조는 연구데이터의 분석·활용을 위한 연구소프트웨어 프레임워크와 연구소프트웨어의 공유·활용을 위한 허브 서비스로

구성된다. KRDC 프레임워크는 연구자가 다양한 유형의 연구소프트웨어를 개발하고 재사용하기 위해 상호 운용하는데 필요한 기능 및 서비스를 제공하며, 각 기관에 프레임워크 배포를 통해 연구소프트웨어 공유·활용 체계를 구축한다. KRDC 허브 서비스는 프레임워크를 통해 운영되는 리소스의 메타데이터를 등록, 수집하여 이용자들의 검색 인프라를 지원하는 카탈로그 서비스, 통합 서비스 등을 제공한다. 본 연구에서는 데이터 분석 환경 및 분석 도구를 중심으로 설문 조사 및 인터뷰를 진행하여 KRDC 프레임워크에 대한 이용자의 요구사항을 반영하고자 하였다.

3. 설문 및 인터뷰 데이터 분석

본 연구에서는 출연(연)의 연구데이터 관리자들을 대상으로 설문조사를 실시한 후 후속 인터뷰에 동의한 설문 응답자들을 대상으로 인터뷰를 수행하였다.

연구데이터를 활용하기 위한 분석 인프라와 서비스의 현황 파악 및 연구데이터 커먼즈 체계에 관한 연구데이터 관리자들의 인식을 파악하기 위해, NST 산하 25개 출연(연) 중 설문 수행 주체인 KISTI를 제외한 24개 출연(연)을 대상으로 NST의 설문조사 프로그램을 활용하여 설문지를 송부하였다. 설문조사는 2022년 7월 한 달간 실시하였으며, 24개 출연(연) 중 총 15개(62.5%) 기관이 설문조사 문항에 회신하였

5) 연구소프트웨어란 연구자들이 연구데이터를 분석·활용하기 위해 생성한 컴퓨터 프로그램, 소프트웨어, 소스 코드, 컴파일된 코드 등을 의미하며, 오픈 소스와 비교해서 볼 때 과학적 기여로 발행된 논문에서 사용한 소프트웨어로 한정할 수 있다(Gomez-Diaz & Recio, 2022).

다. 설문조사 응답 기관 중 후속 인터뷰에 동의한 기관은 9개였으며, 각 기관의 연구데이터 관리자를 대상으로 각 기관과 일정을 협의한 후, 2022년 9월 7일부터 11월 5일까지 직접 방문하여 인터뷰를 수행하였다.

3.1 설문 데이터 분석

연구데이터를 활용하기 위한 분석 인프라와 서비스의 현황 파악 및 연구데이터커먼즈 체계에 관한 연구데이터 관리자들의 인식 파악을 위한 설문조사 문항은 다음과 같이 구성되어 있다(〈표 1〉 및 〈부록 1〉 참조). 1, 2번 문항은 각 기관별 분석 인프라와 분석 서비스의 현황을 파악하기 위한 것으로, 1번 문항에서 분석 인프라를 제공하고 있지 않은 경우 1-1번 문항에 응답하도록 하였으며, 분석 인프라를 제공하고 있는 경우에는 1-2번 문항에 응답하도록 하였다. 또한 2번 문항에서 분석 서비스를 제공하고 있지 않은 경우 2-1번 문항에 응답하도록 하였으며, 2번 문항에서 분석 서비스를 제공하고 있거나, 2-1번에서 제공 의향이 있다고 응답

한 경우 2-2번 문항에 응답하도록 하였다. 3, 4번 문항의 경우 연구데이터커먼즈 체계 구축과 관련한 연구데이터 관리자들의 인식을 파악하기 위한 것이다.

설문조사에 회신한 15개 기관명은 ID로 대체하여 정리하였으며, 설문 응답 결과는 〈표 2〉와 같다.

먼저 각 기관별 분석 인프라와 분석 서비스의 현황을 파악하기 위한 1, 2번 문항에 대한 응답 결과를 살펴보면, 분석 인프라에 대한 문항에는 15개 기관 중 4개 기관(26.7%)이 분석 인프라를 제공하고 있다고 응답하였으며, 제공하는 분석 인프라는 모두 기관 보유 장비 형태라고 응답하였다. 분석 인프라를 제공하지 않는다고 응답한 11개 기관(73.3%) 중 5개 기관은 향후 분석 인프라 제공 의향이 있다고 응답하였으며, 3개 기관은 현재는 고려하지 않지만, 향후 도입할 예정이라고 응답하였다.

다음으로 분석 서비스에 대한 문항에는 15개 기관 중 5개 기관(33.3%)이 분석 서비스를 제공하고 있다고 응답하였으며, 다양한 형태의 분석 서비스를 제공하고 있었다.

〈표 1〉 설문 내용

문항 번호	설문 문항 내용
1	연구데이터 활용을 위한 분석 인프라(워크스테이션, 서버, GPU 등) 제공 여부
1-1	향후 연구데이터 활용을 위한 분석 인프라 제공 의향
1-2	제공하고 있는 분석 인프라의 형태
2	연구데이터 활용을 위한 분석 서비스 제공 여부
2-1	향후 연구데이터 활용을 위한 분석 서비스 제공 의향
2-2	연구데이터 분석 서비스를 제공하고 있거나 제공할 의향이 있는 경우, 지원하고 있거나 지원할 예정인 서비스의 수준
3	연구데이터 활용을 위한 분석 인프라 및 분석 서비스를 제공할 수 있는 프레임워크 도입 의향
4	외부에 공개된 분석 SW 사용 가능한 체계 제공 의향

<표 2> 분석 인프라 및 서비스 현황과 연구데이터 관리자 인식 파악을 위한 설문 응답 결과

설문 분항 기관 ID	1	1-1	1-2	2	2-1	2-2	3	4
	분석 인프라 제공 여부	향후 분석 인프라 제공 의향	제공하는 분석 인프라	분석 서비스 제공 여부	향후 분석 서비스 제공 의향	제공하거나 제공하고자 하는 분석 서비스*	프레임워크 도입 의향	외부공개 SW 제공 의향
I-4	X	○	-	X	X	-	○	○
I-6	X	X(향후)	-	X	○	3	○	○
I-10	○	-	기관 보유장비	○	-	1,2,3	X	○
I-11	X	X	-	X	X	-	X(향후)	○
I-12	X	○	-	X	○	-	기타(분석인 프라 구축 중)	○
I-13	○	-	기관 보유장비	○	-	4	X	X
I-14	X	X(향후)	-	X	X(향후)	-	X(향후)	X(향후)
I-15	X	X	-	X	X	-	○	○
I-17	X	X	-	X	X	-	X	X
I-18	X	○	-	X	○	-	○	○
I-19	○	-	기관 보유장비	○	-	1,2,3,4,5	○	○
I-20	X	○	-	X	○	1,2	○	○
I-21	X	○	-	X	○	1,4	X	○
I-24	○	-	기관 보유장비	○	-	-	-	-
I-25	X	X(향후)	-	○	-	2,3,4,5	X	X

* 제공하거나 제공하고자 하는 분석 서비스

- ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라 제공
- ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구 제공(R, Jupyter 등)
- ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW 제공(웹기반 or 다운로드)
- ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경 제공
- ⑤ 소속 연구자들이 직접 개발한 분석 SW를 공유하는 환경 제공

① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라를 제공한다고 응답한 기관은 I-10, I-19, I-24 3개 기관이었으며, ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구(R, Jupyter 등)를 제공한다고 응답한 기관은 I-10, I-19, I-24, I-25 4개 기관으로 가장 많은 기관이 해당 서비스 유형을 제공하고 있었다. ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW(웹기반 or 다운로드)를 제공한다고 응답한 기관은 I-10, I-19, I-25 3개 기관이었으며, ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경을 제공한다고 응답한 기관은 I-13, I-19, I-25

3개 기관이었다. 마지막으로 ⑤ 소속 연구자들이 직접 개발한 분석 SW를 공유하는 환경을 제공한다고 응답한 기관은 I-19, I-25 2개 기관으로 가장 적은 기관이 해당 서비스 유형을 제공하고 있었다. 또한 I-19 기관의 경우 설문조사에서 제시한 모든 형태의 분석 서비스를 제공하고 있음을 알 수 있었다.

I-24 기관의 경우 설문조사에서 제공하고 있는 분석 서비스에 대한 응답을 하지 않았으나, 후속 인터뷰를 통해 해당 내용에 대해 파악하였다. 인터뷰에서 I-24 기관은 주로 연구자들이 전통적으로 분석에 사용하는 SW와 Python을 활용하여 분석하고 있으며, 인프라가 부족한 측

면은 있으나 기관에서는 가상머신을 활성화하여 지원하고 있다고 하였다. 이는 설문조사에서 제시한 분석 서비스 형태를 매칭해보면 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구(R, Jupyter 등) 형태를 제공하고 있음을 알 수 있다.

한편 분석 서비스를 제공하지 않는다고 응답한 10개 기관(66.7%) 중 5개 기관이 향후 연구데이터 활용을 위한 분석 서비스를 제공할 의향이 있다고 응답하였다. I-6 기관은 ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW(웹기반 or 다운로드)를 제공할 예정이라고 응답하였으며, I-20 기관은 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구(R, Jupyter 등)를 제공할 예정이라고 응답하였다. I-21 기관은 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경을 제공할 예정이라고 응답하였으며, I-12, I-18 기관은 제공하고자 하는 분석 서비스에 대한 응답은 하지 않았다.

향후 연구데이터 활용을 위한 분석 서비스의 제공을 고려하지 않는다고 응답한 4개 기관은 I-4, I-11, I-15, I-17 기관이었으며, I-14 기관은 현재는 고려하고 있지 않으나, 향후 제공할 예정이라고 응답하였다.

다음으로 연구데이터커먼즈 체계 구축과 관련한 연구데이터 관리자들의 인식을 파악하기 위한 3, 4번 문항에 대한 응답 결과는 다음과

같다.

프레임워크 도입 의향에 대한 문항에는 15개 기관 중 6개 기관(40%)이 도입할 의향이 있다고 응답하였으며, 2개 기관(13.3%)은 현재는 고려하지 않으나 향후 도입할 예정이라고 응답하였고, 1개 기관(6.67%)은 기타 의견으로 현재 분석 인프라를 구축 중에 있다고 응답하여, 절반 이상의 기관이 프레임워크 도입에 긍정적인 응답을 하였음을 알 수 있다. 한편 15개 기관 중 5개 기관(33.3%)은 프레임워크 도입을 고려하지 않고 있다고 응답하였으며, 1개 기관(6.67%)은 해당 문항에 응답을 하지 않았다.

마지막으로 외부 공개 SW 제공 의향에 대해서는 15개 기관 중 10개 기관(66.7%)이 제공할 의향이 있다고 응답하였고, 1개 기관(6.67%)은 현재는 고려하지 않으나 향후 도입할 예정이라고 응답하여 대부분의 기관이 외부 공개 SW 제공 의향이 있음을 알 수 있었다. 한편 3개 기관(20%)은 외부 공개 SW 제공을 고려하지 않고 있다고 응답하였으며, 1개 기관(6.67%)은 해당 문항에 응답을 하지 않았다.

3.2 인터뷰 데이터 분석

24개 출연(연) 중 설문조사에 응답한 15개 기관 중 후속 인터뷰에 동의한 9개 기관의 연구데이터 관리자를 대상으로 2022년 9월 7일부터 11월 5일까지 인터뷰를 수행하였다.

기관별 인터뷰 일정 및 참여자 리스트는 <표 3>과 같으며, 9개 기관에서 1~3명의 담당자가 인터뷰에 참여하여, 최종적으로 연구데이터 관리자 15인이 참여하였다.

〈표 3〉 인터뷰 참여 현황

기관 ID	참여자 ID	인터뷰 일시
I-4	I-4-1	2022.09.26. 14:00~16:00
	I-4-2	
I-10	I-10-1	2022.11.05. 10:00~12:00
	I-10-2	
I-13	I-13-1	2022.09.21. 14:00~16:00
	I-13-2	
I-18	I-18-1	2022.09.07. 16:20~17:20
	I-18-2	
I-19	I-19-1	2022.09.27. 14:00~16:00
I-20	I-20-1	2022.09.19. 14:00~16:00
I-21	I-21-1	2022.09.16. 09:00~11:00
I-24	I-24-1	2022.09.26. 14:00~16:20
	I-24-2	
	I-24-3	
I-25	I-25-1	2022.09.13. 14:00~16:00

연구데이터커먼즈를 실제 이용할 출연(연)의 연구데이터 관리자를 대상으로 인터뷰를 진행함으로써 연구 전 과정에서 활용되는 연구 자원인 데이터, 인프라, 분석 SW 등에 대한 관리 과정과 인식을 파악하고, 연구 자원을 공동 활용하기 위하여 접근 및 재사용을 지원하는 플랫폼인 연구데이터커먼즈에 대한 요구사항을 파악하고자 하였다.

먼저 30분 정도 연구데이터커먼즈에 대한 개념을 소개한 후, 각 연구기관의 현황 및 요구사항을 파악하기 위한 인터뷰를 수행하였다. 각 기관별로 상황에 따라 인터뷰에 약 1~2시간이 소요되었으며, 준비한 인터뷰 질문지의 질문을 토대로 필요할 경우 새로운 질문을 추가하는 반구조화된 방식으로 이루어졌다.

인터뷰 질문은 설문조사를 토대로 다음과 같이 구성되었다. (1) 각 기관에서 연구데이터 활용을 위해 제공하는 분석 인프라와 (2) 각 기관

의 연구데이터 활용을 위해 제공하는 분석 서비스에 대해 향후 제공 계획 및 외부 공개 가능 여부 등을 포함하여 살펴보았다. 설문조사에서는 기관의 입장에서 연구자를 위해 제공하는 분석 인프라에 대해 확인하였다면, 인터뷰에서는 각 기관의 연구자들이 주로 사용하는 분석 인프라와 분석 서비스, 그리고 기관별 데이터 관리체계에 대해 추가로 질문하였다. 이를 통해 설문조사 문항에서는 확인할 수 없었던 기관의 개별적인 특성도 확인할 수 있었다. 마지막으로 (3) KRDC 구축 시 고려사항 및 요구사항에 대한 의견도 수집하였다.

3.2.1 기관별 분석 인프라 현황

먼저 각 기관의 연구데이터 활용을 위해 제공하는 분석 인프라에 대해 질문하였다. 분석 인프라는 워크스테이션, 서버, GPU 등을 의미한다. 〈표 4〉와 같이 인터뷰에 참여한 9개 기관

〈표 4〉 분석 인프라 현황 설문 결과

기관 ID	분석 인프라 제공 여부	향후 분석 인프라 제공 의향	제공하는 분석 인프라
I-4	×	○	-
I-10	○	-	기관 보유장비
I-13	○	-	기관 보유장비
I-18	×	○	-
I-19	○	-	기관 보유장비
I-20	×	○	-
I-21	×	○	-
I-24	○	-	기관 보유장비
I-25	×	×(향후)	-

의 설문조사 응답을 살펴보면, 4개 기관이 분석 인프라를 제공한다고 하였으며, 제공하는 분석 인프라는 모두 기관 보유 장비 형태라고 응답하였다. 분석 인프라를 제공하지 않는다고 응답한 5개 기관 중 4개 기관은 향후 연구데이터 활용을 위한 분석 인프라를 제공할 의향이 있다고 응답하였으며, 1개 기관은 현재는 고려하지 않으나, 향후 도입할 예정이라고 응답하여 인터뷰에 참여한 기관들은 분석 인프라 제공에 긍정적인 기관들임을 알 수 있었다.

설문조사에서 분석 인프라를 제공하고 있다고 응답한 기관은 4개 기관으로 이들 기관에서는 기관 보유장비 형태로 분석 인프라를 제공하고 있었다. 인터뷰를 통해 각 기관별로 제공하는 분석 인프라의 조금 더 상세한 형태를 파악하고자 하였고, 추가적으로 해당 인프라를 외부에 공개할 수 있는지에 대해서도 질문하였다.

I-10 기관의 경우 인프라는 개방형으로 제공하고 있으며, 26년까지 분석 인프라를 저장용량 46페타바이트 정도로 운영할 예정이고, 국가연구개발사업에서 해당 분야 연구데이터를 모으는 방식으로 사업이 승인되어 진행하고 있었다. 각 부처에서 수행하고 있는 해당 분야 데

이터를 모으려고 했으나 중복성 때문에 각 부처의 메타데이터만 연계하여, 해당 분석 인프라에서는 표준화된 메타데이터를 조회한 후, 원 사이트로 링크하여 다운받도록 연결하고 있다고 하였다.

I-13 기관의 경우 계산 자원, 내부 클러스터 등 시뮬레이션을 위한 원내 인프라는 있지만, 외부 사용자에게는 공개하지 않고 있으며, 정책적인 문제라 내부적으로 결정할 수 없다고 하였다.

I-19 기관의 경우 내부적으로 자원들에 대한 활용률을 높이기 위하여 기관 차원에서 활용할 수 있도록 클라우드 환경을 구축하였으며, 현재는 외부로 서비스할 계획은 없는 상태라고 하였다. I-19 기관은 대용량 데이터를 분석하는 업무를 수행하기 때문에 데이터와 분석 인프라가 상호 연계 가능한 형태로 커뮤니티를 통해 제공하고 있다고 하였다.

I-24 기관의 경우 분야 특성상 국제 커뮤니티를 통해 연구데이터와 분석 모델 공유 플랫폼이 잘 갖춰져 있다. 데이터 저장 스토리지, 클러스터 등 기관 보유 인프라가 잘 제공되고 있으며, 국제공동연구를 수행하기 때문에 커뮤니티를 통해 안정적으로 데이터, 모델, 서비스, 인

프라가 모두 제공되고 있다. 외부 공개에 대해서는 보안상 데이터의 외부 공개가 어렵고, 전문 분야 데이터이므로 데이터 공개에 따른 부담이 있다고 하였다.

분석 인프라를 제공하고 있는 4개 기관 중 3개 기관이 외부 공개에 대해 응답하였는데, 1개 기관은 해당 인프라를 외부에 공개하고 있었으며, 2개 기관은 외부 사용자에게 공개하지 않음을 알 수 있었다.

다음으로 설문조사에서 분석 인프라를 제공하고 있지 않다고 응답한 5개 기관 중 4개 기관은 향후 분석 인프라 제공 의향이 있다고 답변하였으며, I-25 기관은 현재는 분석 인프라 제공에 대해 고려하고 있지 않지만 향후 도입할 예정이라고 응답하였다.

I-4 기관의 경우 분석 작업을 주로 개인의 로컬 PC에서 수행하고, 계산 자원을 필요로 하는 컴퓨팅 환경은 필요하지 않다고 답변하였다. 다만 기관 내 데이터 관리체계로써 분야 특성을 가진 데모 프로그램을 통해 시험 결과를 정리하고 평가하며 관리하고 있다고 하였다. 데이터 특성상 보안의 문제로 인증 및 인가를 통해 원내 연구자들도 액세스가 제한되며, 참여 연구원들만 해당 프로젝트내에서 데이터를 공유할 수 있다고 하였다.

I-18 기관의 경우 분석 인프라 제공 계획이 있으나, 기관 내부에서도 분야마다 많은 다양한 소프트웨어를 사용하고 있고, 분야에 따라 사용하는 인프라와 서비스가 극명하게 다르기 때문에 인프라 환경이 필요한 내부 인원이 많지 않아서 소규모 인프라를 제공할 예정이라고 하였다. 현장에서 데이터를 취득하는 분야는 현실적으로 인프라 환경의 니즈가 없으며, 대부분 개인 PC

를 이용하여 분석을 진행한다고 하였다.

I-20 기관은 연구데이터 수집 및 공유 체계 구축에 집중하고 있으며, 향후 분석 활용 플랫폼으로 확장을 고려하고 있다고 하였다. 현재는 소규모의 클라우드 서버를 운영하고 있으며, 주로 연구자 개인별로 소규모 서버 환경을 구축하여 연구를 수행하고 있다고 하였다.

I-21 기관은 부서마다 분야 및 데이터의 특성이 달라서 공통된 데이터를 모으기가 어려운 실정이나 시범 사업을 통해 데이터 관리를 할 예정이라고 하였다. 기관 특성상 실험 데이터를 만드는 데이터 생산 업무가 주 업무이고 외부에서 사용하는 분석 업무는 없기 때문에, 분석 인프라나 사용하는 SW는 일부 연구원들이 활용하기는 하나 개인적으로 하는 것이며, 대부분 전문 분석 도구를 본인의 워크스테이션에서 돌리며, 분석은 개인의 영역이라고 하였다.

I-25 기관은 현재는 분석 인프라 제공에 대해 고려하고 있지 않지만 향후 도입할 예정이라고 응답하였는데, 속도, 보안 문제로 상용 클라우드를 분석 인프라로 제공할 예정이라고 하였다.

3.2.2 기관별 분석 서비스 현황

다음으로 각 기관이 연구데이터 활용을 위해 제공하는 분석 서비스에 대해 질문하였다. 분석 서비스는 분산 환경, 실시간 분석, 딥러닝 학습 환경, 분석 지원 도구(R, Python, Jupyter 등), 공개/상용 분석 SW 등을 의미한다. <표 5>와 같이 설문조사에서는 인터뷰에 참여한 9개 기관 중 5개 기관이 분석 서비스를 제공한다고 하였으며, 다양한 형태의 분석 서비스를 제공하고 있었다. 분석 서비스를 제공하지 않는

〈표 5〉 분석 서비스 현황 설문 결과

기관 ID	분석 서비스 제공 여부	향후 분석 서비스 제공 의향	제공하거나 제공하고자 하는 분석 서비스*
I-4	×	×	-
I-10	○	-	1,2,3
I-13	○	-	4
I-18	×	○	-
I-19	○	-	1,2,3,4,5
I-20	×	○	1,2
I-21	×	○	1,4
I-24	○	-	-
I-25	○	-	2,3,4,5

* 제공하거나 제공하고자 하는 분석 서비스

- ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라 제공
- ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구 제공(R, Jupyter 등)
- ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW 제공(웹기반 or 다운로드)
- ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경 제공
- ⑤ 소속 연구자들이 직접 개발한 분석 SW를 공유하는 환경 제공

다고 응답한 4개 기관 중 3개 기관은 향후 연구 데이터 활용을 위한 분석 서비스를 제공할 의향이 있다고 응답하여, 인터뷰에 참여한 기관들은 분석 서비스 제공에도 긍정적인 기관들임을 알 수 있었다.

설문조사에서 분석 서비스를 제공하고 있다고 응답한 5개 기관에 대해 인터뷰를 수행하여 각 기관별로 제공하고 있는 분석 서비스에 대해 조금 더 상세하게 파악하고자 하였으며, 추가적으로 분석 서비스를 외부에 공개할 수 있는지에 대해서도 질문하였다.

I-10 기관의 경우 클라우드 기반 연구 환경을 제공하여 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라를 제공하고 있으며, ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구도 제공(R, Python, Jupyter 등)하고 있다고 하였다. 마지막으로 클라우드를 통해 분석 도

구도 함께 제공함으로써 ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW도 제공하고 있었다.

I-13 기관의 경우 분석 서비스는 내부적으로 활용가능한 웹 기반 계산 플랫폼을 제공하여 ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경을 제공하고 있으며, 보안 정책상 외부 공개는 하지 않는다고 하였다.

I-19 기관의 경우 설문조사에서 제시한 모든 형태의 분석 서비스를 제공하고 있음을 알 수 있었는데, 분야 커뮤니티를 통해 데이터, 분석 모델들이 잘 공유되고 있었다. 데이터와 분석 모델들은 분야 커뮤니티를 통해 재사용하기 때문에, Jupyter 환경과 같은 서비스로 실행하면 된다고 하였다. 이처럼 연구자가 활용하는 분석 모델의 80%는 커뮤니티를 통해 이미 공유된 것을 재활용하는데, 20%의 연구를 위한 코드는 매우 다양하고 파편화가 심하기 때문에

패키징하기 어렵고, 재사용 가능성도 매우 낮은 실정이다. 바이오, 재료 등의 분야는 공통으로 활용되는 형태의 모델이 있어 공유, 활용이 가능하지만, 물리, 천문과 같은 분야는 실질적 연구를 위한 코드, 분석 모델이 파편화되어 재사용할 수 있는 체계가 구축되기 어렵다고도 하였다. 따라서 향후 재사용 가능한 코드를 기관 내에서 제공할 계획은 없으며, 데이터가 워낙 크기 때문에 데이터가 있는 곳에서 모델이 실행 되고, Jupyter와 같은 환경이 프런트엔드 서비스로 있는 경우가 일반적이라고 하였다.

I-24 기관의 경우 설문조사에서 제공 분석 서비스에 대한 응답을 하지 않았으나, 해당 내용에 대해 인터뷰를 통해 파악하였다. 주로 연구자들이 전통적으로 분석에 사용하는 SW와 Python을 활용하여 분석하고 있으며, 인프라가 부족한 측면은 있으나 기관에서는 가상머신을 활성화하여 지원하고 있다고 하였다. 이는 설문조사에서 제시한 분석 서비스 형태를 매칭해보면 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구(R, Jupyter 등) 형태를 제공하고 있음을 알 수 있다. 해외 협력 기관에서 공유하는 SW를 많이 활용하고 있으며, 분석 SW의 경우 데이터 공통 포맷으로 공유되기 때문에 커뮤니티 내에서 많은 연구자가 사용하고 있다고 하였다.

I-25 기관의 경우 ①번 항목만 제외하고 모든 형태의 분석 서비스를 제공하고 있다고 하였는데, 인터뷰를 통해 I-25 기관은 자체적으로 AI 플랫폼을 운영하고 있음을 알 수 있었다. AI 플랫폼에는 자체적 혹은 개인적으로 연구

를 진행하여 생긴 분석 모델을 논문 발표 후에 웹서비스에 퍼블리시하여 공개하고 있는데, 입력 데이터에 따라 플랫폼에서 요구하는 업로드 형식으로 통일이 어렵다는 문제점을 언급하였다. 현재 AI 플랫폼은 기관 내부 연구원 및 일반 연구자에게도 오픈되어 있다고 하였다. 그리고 기관에서 제공하고 있는 AI 플랫폼과 별개로 연구원들은 각자 Jupyter, Python 환경에서 분석 모델을 개발하여 사용하고 있다. AI 플랫폼 자체에서 모델 전후처리는 Python 기반으로 제작되고 있으며, 플랫폼 서비스 측면에서는 RESTful 기반으로 사용자가 사용할 수 있는 Python 인터페이스를 AI팀에서 개발하고 있다고 하였다.

분석 서비스를 제공하고 있는 5개 기관 중 1개 기관은 해당 서비스를 외부에 공개하고 있었으며, 3개 기관은 외부 사용자에게 공개하지 않음을 알 수 있었다.

다음으로 설문조사에서 분석 서비스를 제공하지 않는다고 응답한 기관은 4개 기관으로, I-4 기관을 제외하고 향후 연구데이터 활용을 위한 분석 서비스를 제공할 의향이 있다고 응답하였다.

I-4 기관의 경우 별도로 기관에서 제공하는 분석 서비스는 없지만, 주로 SAS 통계 프로그램을 활용하고, 통계 시각화를 하거나 실험 결과의 유의성 평가를 하는 등 계량학적 분석 및 가시화 유료 SW 등 상용소프트웨어를 주로 사용한다고 하였다. 대부분 간단한 분석이므로 계산 자원과 같은 인프라 환경은 불필요하고, 기관 전체 데이터 규모가 28GByte로 작은 편이라고 하였다.

I-18 기관은 설문조사에서 제공하고자 하는

분석 서비스에 대한 응답은 하지 않았으나, 인터뷰를 통해 AI 분석 플랫폼 개발을 준비 중에 있으며, 기관 소속 연구원들이 많이 쓰는 라이브러리를 설치한 분석 환경을 제공할 예정이라고 하였다. 이는 설문조사에서 제시한 항목에 매칭해 볼 때 ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW를 제공하고자 하는 것임을 알 수 있었다.

I-20 기관은 설문조사에서 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구(R, Jupyter 등)를 제공할 예정이라고 응답하였는데, 인터뷰를 통해 주로 Matlab, SPSS 등의 분석 SW를 구매하여 사용하지만, 기관 차원에서 분석 서비스를 제공하지는 않는다고 하였다.

마지막으로 I-21 기관은 ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라와 ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경을 제공할 예정이라고 응답하였는데, 해당 인터뷰 질문에 대해서는 답변을 하지 않았다.

3.2.3 KRDC 구축 시 고려 및 요구 사항

〈표 6〉과 같이 설문조사에서는 인터뷰에 참여한 9개 기관 중 4개 기관이 연구데이터 활용을 위한 분석 인프라 및 분석 서비스를 제공할 수 있는 프레임워크가 있다면 도입할 의향이 있다고 응답하였으며, 6개 기관이 외부에 공개된 분석 SW를 기관 연구자들이 사용할 수 있는 체계가 있다면 제공할 의향이 있다고 응답하였다.

인터뷰를 통해 KRDC의 구축에 있어서 어떤 부분을 고려해야 하는지에 대해 상세히 질문하였다. 각 기관들의 다양한 의견을 정리해보면 크게 시스템의 편리성, 메타데이터의 중요성, 데이터 품질 관리 이슈, 보안 및 보상 체계의 필요성에 대한 의견으로 구분해 볼 수 있다.

첫째, 시스템의 편리성에 대한 의견은 다음과 같다. I-10 기관은 연구자가 등록을 수월하게 할 수 있도록 실험데이터가 쌓였을 때의 로그 관리, 기록관리 도구 등의 소프트웨어 제공 필요성과 연구노트 작성처럼 데이터 등록, 메타데이터 입력 등의 자동화 필요성을 언급하였다.

I-18 기관에서는 깃허브와 같은 인터랙티브한 성격의 프레임워크의 필요성을 제시하며, 빠

〈표 6〉 프레임워크 도입 및 외부공개 SW 사용 의향 설문 결과

기관 ID	프레임워크 도입 의향	외부공개 SW 제공 의향
I-4	○	○
I-10	×	○
I-13	×	×
I-18	○	○
I-19	○	○
I-20	○	○
I-21	×	○
I-24	-	-
I-25	×	×

른 정보처리 프로세스와 다양한 앱을 통하여 사용자가 원하는 기능과 업무를 직관적으로 처리할 수 있는 인터페이스 제공을 요청하였다.

I-19 기관은 공유되는 모델들이 실행되는 VM 이미지 등의 메타데이터가 잘 구축되고, KRDC에서 분석 환경 구성 열람 및 수정이 용이한 환경이 마련되면 좋겠다는 의견을 제시하였다. 더불어 데이터 입출력이 잘 맞을 수 있도록 처리하기 위한 다양한 기능을 지원할 수 있는 기술적 지원이 있었으면 좋겠다고 하였다.

둘째, 메타데이터의 중요성에 대해서는 특히 I-10 기관에서 정확한 메타데이터를 입력할 수 있도록 시스템적인 지원이 필요하다는 의견을 제시하였다. 그 사례로 데이터를 등록할 때 정확한 정보를 넣어주면 데이터 사용성이 높아지지만, 실제로 연구자가 데이터를 등록할 때 메타데이터를 입력받는 항목이 많아서 연구자들이 입력하기 쉽지 않고 오류가 입력될 수 있기 때문에 데이터 모델링을 검증하는 경우에 대부분 실패하게 된다고 하였다.

셋째, 데이터 품질 관리 이슈에 대해서도 I-10 기관에서 언급하였다. 데이터 유형에 따라 분야 전문가가 품질 검증을 해야 한다고 하였다. 메타데이터 역시 메타데이터 전문가가 확인하여 품질을 관리해야하며, 그것이 가능할 때 연구자가 해당 데이터를 신뢰하고 활용할 수 있다고 하였다.

마지막으로 보안 및 보상 체계의 필요성에 대해서는 I-18 기관과 I-24 기관이 언급하는데, I-18 기관은 KRDC 프레임워크가 직관적으로 사용하기 쉽게 제공된다면 KRDC를 통해 분석 서비스를 제공할 의향이 있으나, 먼저 공유에 따른 보안, 보상 문제가 해결되어야 한다고 하

였다. 특히 각 기관에서 제공하는 데이터와 워크플로우 등의 보안 문제가 가장 중요하기 때문에 적절한 보상을 고민하는 것이 필요할 것이며, 보안을 위해 블록체인 기술을 활용하면 좋겠다는 의견도 제시하였다.

I-24 기관은 1차 원시데이터는 저장하지만, 국가 보안 문제 등으로 2차 분석 데이터는 저장하지 않고 개인적으로 관리하며, 공동연구기관에 한정하여 공유가 가능하기 때문에 연구데이터 보안이 중요한 문제라고 하였다.

4. 결론

본 연구에서는 KISTI에서 개발 중인 연구데이터커먼즈를 실제 이용할 NST 산하 출연(연)의 연구데이터 관리자를 대상으로 설문조사와 인터뷰를 통해 연구데이터를 활용하기 위한 분석 인프라와 분석 서비스의 현황을 파악하고, 연구데이터커먼즈 체계 구축과 관련한 연구데이터 관리자들의 인식을 조사해 보았다. 아직 국내에서는 연구데이터커먼즈에 대한 논의가 활성화되어 있지 않은 상황이지만, 선진국을 중심으로 최근 데이터 중심의 연구개발을 지향하면서 연구데이터뿐만 아니라 연구개발 전 과정에 필요한 리소스를 공유하고 활용하기 위한 정책 및 프로젝트가 추진되고 있는 실정이다. 본 연구의 결과를 통해 국내의 연구데이터 커먼즈 체계 구축을 위한 정책 수립 방안 도출을 위한 시사점을 제시하면 다음과 같다.

첫째, 설문조사에서 분석 인프라에 대한 문항에는 15개 기관 중 4개 기관이 분석 인프라를 제공하고 있으며, 분석 인프라를 제공하지 않

는다고 응답한 11개 기관 중 5개 기관은 향후 분석 인프라 제공 의향이 있다고 응답하였고, 3개 기관은 현재는 고려하지 않지만, 향후 도입할 예정이라고 응답하여, 15개 기관 중 12개 기관이 분석 인프라 제공에 긍정적인 응답을 하였음을 알 수 있다. 또한 분석 서비스에 대한 문항에도 15개 기관 중 5개 기관이 분석 서비스를 제공하고 있다고 응답하였으며, 분석 서비스를 제공하지 않는다고 응답한 10개 기관 중 5개 기관이 향후 연구데이터 활용을 위한 분석 서비스를 제공할 의향이 있다고 응답하였다. 분석 서비스의 제공을 고려하지 않는다고 응답한 4개 기관 중 1개 기관도 현재는 고려하고 있지 않으나, 향후 제공할 예정이라고 응답하여, 15개 기관 중 11개 기관이 분석 서비스 제공에 긍정적인 응답을 하였음을 알 수 있다. 그리고 프레임워크 도입 의향은 15개 기관 중 6개 기관이 도입할 의향이 있다고 응답하였으며, 2개 기관은 현재는 고려하지 않으나 향후 도입할 예정이라고 응답하였고, 1개 기관은 기타 의견으로 현재 분석 인프라를 구축 중에 있다고 응답하여, 절반 이상의 기관이 프레임워크 도입에 긍정적인 응답을 하였음을 알 수 있었다. 마지막으로 외부 공개 SW 제공 의향에 대해서도 15개 기관 중 10개 기관이 제공할 의향이 있다고 응답하고, 1개 기관은 현재는 고려하지 않으나 향후 도입할 예정이라고 응답하여 대부분의 기관이 외부 공개 SW 제공 의향이 있음을 알 수 있었다.

이러한 연구 결과를 통해 출연(연)의 분석 인프라와 분석 서비스의 Pool이 충분히 확보될 수 있으며, 프레임워크와 외부 공개 SW 사용에 대해서도 긍정적인 인식을 가지고 있음을

확인할 수 있었다.

둘째, 후속 인터뷰를 통해 각 기관별로 분석 인프라와 분석 서비스를 어떤 형태로 제공하고 있는지, 외부 공개가 가능한지 여부를 상세히 파악할 수 있었다.

설문조사에서는 대부분의 기관들이 분석 인프라와 분석 서비스를 제공하고 있었으며, 연구데이터 활용을 위한 분석 인프라 및 분석 서비스를 제공할 수 있는 프레임워크에 대한 도입과 외부에 공개된 분석 SW를 사용할 수 있는 체제에 대한 제공 의향 역시 높았으나, 분석 인프라를 제공하고 있는 4개 기관 중 1개 기관만이 해당 인프라를 외부에 공개하고 있었으며, 분석 서비스를 제공하고 있는 5개 기관 중에서도 1개 기관만이 해당 서비스를 외부에 공개하고 있었다.

이러한 연구 결과를 분석해보면, 프레임워크를 통해 분석 인프라와 분석 서비스가 제공될 경우 활용하고자 하는 수요가 있으나, 각 기관에서 보유하고 있는 분석 인프라와 분석 서비스는 공개 및 공유하기 어렵다는 것을 알 수 있다.

데이터가 잘 공유되지 않는 현실에 대한 의견을 살펴보았을 때, 데이터 보안 유지를 위해 외부 공유가 어렵기도 하고, 보편적으로 다른 정보는 공유하여 활용하고 싶지만 나의 연구 정보는 공개를 꺼린다는 것을 알 수 있었다. 또한, 출연(연) 전문가들은 기존의 분석 SW를 활용하거나 커뮤니티를 통해 분석 인프라, 분석 모델을 쉽게 확보 가능하여, 데이터 기반 연구 수행에 어려움이 없고 각 소속 기관 내의 정보 공유만으로도 충분하다고 생각하는 경향이 있었다. 이러한 상황에서 연구데이터커먼즈 체계 구축과 더불어 시급히 진행되어야 할 것은

연구 현장의 인식 전환, 나아가 제도적 변화일 것이다.

연구데이터커먼즈 체계 수립을 위해서는 글로벌 연구데이터커먼즈 기관과 협력 및 상호운용성을 확보하고 연구의 신뢰성과 활용성을 제

고하기 위해서 전반적인 연구환경에 있어서의 연구자들의 인식 전환을 위해 노력해야 할 것이며, 시스템의 편리성, 보안, 보상체계 등을 잘 고려하는 정책을 수립하기 위해 노력할 필요가 있다.

참 고 문 헌

- 송사광, 조민희, 이미경, 임형준 (2023). 글로벌 개방형 연구데이터 커먼즈 및 시사점. Annual Conference of Korea Information Processing Society 2023 학술발표대회 논문집, 30(2), 85-88.
- 심원식, 안혜연, 박규리, 송사광, 임형준 (2023). 개방형 연구 커먼즈에 대한 연구자 요구 분석에 관한 연구. 한국문헌정보학회지, 57(4), 209-232. <http://dx.doi.org/10.4275/KSLIS.2023.57.4.209>
- 이미경, 조민희, 임형준 (2023). 연구데이터 생태계 지원 통합 인프라 구축 방안. KISTI Issue Brief, 56, 1-15. <https://doi.org/10.22810/2023KIB056>
- 임형준, 이미경, 송사광, 서동민, 조민희 (2022). 데이터 기반 연구개발을 위한 국가연구데이터커먼즈 설계 및 적용 방안. 한국지능시스템학회논문지, 32(5), 392-400. <http://dx.doi.org/10.5391/JKIIS.2022.32.5.392>
- Barker, M., Wilkinson, R., & Treloar, A. (2019). The Australian research data commons. Data Science Journal, 18(1), 44. <https://doi.org/10.5334/dsj-2019-044>
- 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>
- Budroni, P., Claude-Burgelman, J., & Schouppe, M. (2019). Architectures of knowledge: The European open science cloud. ABI Technik, 39(2), 130-141. <https://doi.org/10.1515/abitech-2019-2006>
- European Commission (2020, September). Horizon 2020 Work Programme 2018-2020. 4. European research infrastructures (including e-Infrastructures). Available: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-infrastructures_en.pdf
- Finkel, A. (2017). 2016 National Research Infrastructure Roadmap.

- Gomez-Diaz, T. & Recio, T. (2022). Research software vs. research data I: Towards a research data definition in the open science context. *F1000Research*, 1-33.
<https://doi.org/10.12688/f1000research.78195.1>
- Office of Science and Technology Policy (2023, January 11). Fact Sheet: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research. Available:
https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/?utm_source=link
- Payne, K., Corrie, B., Crawley, F., Harrower, N., Macneil, R., Maxwell, L., Sansone, S.-A., Treloar, A., Woodford, C., Åkerström, W. N., & RDA GORC International Model WG (2023). The Global Open Research Commons International Model Report, Version 1 (1.0). Zenodo.
<https://doi.org/10.15497/RDA00097>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L. B., 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., 't 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., van Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van 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), 1-9. <https://doi.org/10.1038/sdata.2016.18>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Lee, Mikyoung, Cho, Minhee, & Yim, Hyung-Jun (2023). Plan to build integrated infrastructure to support research data ecosystem. *KISTI Issue Brief*, 56, 1-15.
<https://doi.org/10.22810/2023KIB056>
- Shim, Wonsik, An, Hyeyeon, Park, Kyuri, Song, Sa-Kwang, & Yim, Hyung-Jun (2023). A study on the researchers' needs for open research commons. *Journal of the Korean Society for Library and Information Science*, 57(4), 209-232.
<http://dx.doi.org/10.4275/KSLIS.2023.57.4.209>

Song, Sa-Kwang, Cho, Minhee, Lee, Mikyoung, & Yim, Hyung-Jun (2023). Global open research commons and implications. *The Annual Conference of Korea Information Processing Society*, 30(2), 85-88.

Yim, Hyung-Jun, Lee, Mikyoung, Song, Sa-Kwang, Seo, Dongmin, & Cho, Minhee (2022). Design and application of Korea research data commons for data-driven research and development. *Journal of Korean Institute of Intelligent Systems*, 32(5), 392-400.
<http://dx.doi.org/10.5391/JKIIS.2022.32.5.392>

[부록 1] 설문 문항

1. 귀 기관에서는 연구데이터 활용을 위한 분석 인프라(워크스테이션, 서버, GPU 등)를 제공하고 있습니까? ()

- ① 제공하고 있지 않음 (1-1번 항목으로 이동)
- ② 제공하고 있음 (1-2번 항목으로 이동)

1-1. 귀 기관에서는 향후 연구데이터 활용을 위한 분석 인프라를 제공할 의향이 있으십니까? ()

- ① 제공할 의향이 있음
- ② 고려하지 않고 있음
- ③ 현재는 고려하지 않으나, 향후 도입할 예정임
- ④ 기타 []

1-2. 귀 기관에서 제공하고 있는 분석 인프라는 어떤 형태입니까? ()

- ① 연구자 개인 장비
- ② 기관 보유 장비
- ③ 외부 공공 클라우드
- ④ 외부 사설 클라우드 (Microsoft, Google, AWS 등)
- ⑤ 기타 []

2. 귀 기관에서는 연구데이터 활용을 위한 분석 서비스(분산 환경, 실시간 분석, 딥러닝 학습 환경, 분석 지원 도구: R, Python, Jupyter 등)를 제공하고 있습니까? ()

- ① 제공하고 있지 않음 (2-1번 항목으로 이동)
- ② 제공하고 있음 (2-2번 항목으로 이동)

2-1. 귀 기관에서는 향후 연구데이터 활용을 위한 분석 서비스를 제공할 의향이 있으십니까? ()

- ① 제공할 의향이 있음 (2-2번 항목으로 이동)
- ② 고려하지 않고 있음
- ③ 현재는 고려하지 않으나, 향후 제공할 예정임
- ④ 기타 []

2-2. 귀 기관에서 연구데이터 분석 서비스를 제공하고 있거나 제공할 의향이 있다면, 어떤 수준의 서비스를 지원하고 있거나 지원할 예정입니까? (복수 선택 가능) ()

- ① 연구자가 직접 분석 SW를 개발하고, 실행이 가능한 대용량 분석 인프라 제공
- ② 연구자가 필요한 분석 SW를 직접 개발하여 실행할 수 있는 분석 지원 도구 제공 (R, Jupyter 등)
- ③ 기관 소속 연구자들이 주로 많이 사용하는 외부 분석 SW 제공(웹기반 or 다운로드)
- ④ 데이터 전처리, 분석 SW를 실행할 수 있는 실행 환경 제공
- ⑤ 소속 연구자들이 직접 개발한 분석 SW를 공유하는 환경 제공

3. 귀 기관에서 연구데이터 활용을 위한 분석 인프라 및 분석 서비스를 제공할 수 있는 프레임워크가 있다면 도입할 의향이 있습니까? ()

- ① 도입할 의향이 있음
- ② 고려하지 않고 있음
- ③ 현재는 고려하지 않으나, 향후 도입할 예정임
- ④ 기타 []

4. 귀 기관에서 외부에 공개된 분석 SW를 귀 기관 연구자들이 사용할 수 있는 체계가 있다면 제공할 의향이 있습니까? ()

- ① 제공할 의향이 있음
- ② 고려하지 않고 있음
- ③ 현재는 고려하지 않으나, 향후 제공할 예정임
- ④ 기타 []