

연구데이터 활용성 극대화 위한 컴퓨팅 리소스 공유활용 체계

송사광*^{1,2} · 조민희¹ · 이미경¹ · 임형준¹

¹한국과학기술정보연구원 연구데이터공유센터 · ²과학기술연합대학원대학교

Computing Resource Sharing and Utilization System for Efficient Research Data Utilization

Sa-kwang Song*^{1,2} · Minhee Cho¹ · Mikyoung Lee¹ · Hyung-Jun Yim¹

¹Korea Institute of Science and Technology Information, Research Data Sharing Center ·

²University of Science and Technology, Dept. of Applied AI

E-mail : esmallj@kisti.re.kr / mini@kisti.re.kr / jerryis@kisti.re.kr / hjyim@kisti.re.kr

요 약

오픈엑세스, 오픈데이터, 오픈소스 등 과학기술 분야의 오픈사이언스 운동에 대한 최근의 관심 증가에 따라, 공적 자금을 투입한 연구 산출물을 공유하고 활용하려는 움직임이 구체화 및 활성화되고 있다. 이런 흐름에 따라, 최근 국내에서 연구수행의 핵심적인 자원인 연구데이터를 공유하고 활용하는 체제를 구축하고 활성화하려는 많은 노력이 진행되고 있다. 이러한 노력은 주로 연구데이터를 모아서 분야 및 기관별로 모으고 이를 국가 연구데이터 플랫폼인 DataON과 연계하여 검색 및 활용하는 것에 초점을 맞추고 있다. 그러나 선진국에서는 이러한 연구데이터 뿐만 아니라, IaaS, PaaS, SaaS, MLaaS 등 다양한 형태의 연구개발 관련 컴퓨팅 리소스를 공유하고 활용할 수 있는 체제를 구축하고 있다. EOSC (European Open Science Cloud), ARDC (Australian Research Data Commons), CSTCloud (China S&T Cloud) 등이 대표적인 사례라고 할 수 있고 국내에서는 이러한 컴퓨팅 리소스의 공유를 수월하게 수행할 수 있도록, KRDC (Korea Research Data Commons)를 설계하고, 핵심 프레임워크를 개발하고 있다. 본 연구에서는 KRDC의 필요성, 개념, 구성, 향후 계획을 소개한다.

ABSTRACT

With the recent increase in interest in the open science movement in science and technology fields such as open access, open data, and open source, the movement to share and utilize publicly funded research products is materializing and revitalizing. In line with this trend, many efforts are being made to establish and revitalize a system for sharing and utilizing research data, which is a key resource for research in Korea. These efforts are mainly focused on collecting research data by field and institution, and linking it with DataON, a national research data platform, to search and utilize it. However, developed countries are building a system that can share and utilize not only such research data but also various types of R&D-related computing resources such as IaaS, PaaS, SaaS, and MLaaS. EOSC (European Open Science Cloud), ARDC (Australian Research Data Commons), and CSTCloud (China S&T Cloud) are representative examples. In Korea, the Korea Research Data Commons (KRDC) is designed and a core framework is being developed to facilitate the sharing of these computing resources. In this study, the necessity, concept, composition, and future plans of KRDC are introduced.

키워드

Open Data, Open Science, Research Data Commons, Open Science Commons, KRDC

1. 서 론

공적자금을 투입한 R&D 산출물의 손쉬운 접근과 활용을 제고하자는 오픈사이언스 운동의 한 축인 연구데이터 및 관련 자원의 공유 및 활용을 통

* speaker

해 R&D 혁신 창출 기회를 제고하려는 국내외 요구가 커지고 있다. 연구데이터는 국가 과학기술 경쟁력의 근간이 되는 중요 자산으로 정부 주도의 체계적 관리 필요하기에, 일찍부터 미국, 영국, EU, 일본 등 주요 선진국은 국가 차원에서 이러한 자원의 관리·활용 체제를 구축하며 다양한 정책을 추진해 오고 있다. 사회적으로도, 정부의 R&D 투자 규모의 확대에 따른 연구 성과 창출 및 연구 효율성 제고에 대한 요구가 증대되고 있고, 최근 디지털 플랫폼 인프라에 대한 요구 증대와 데이터 중심의 연구개발 패러다임 변화로 인해, 그 필요성이 더욱 커지고 있다. 개인 연구자들도 연구데이터 및 소프트웨어, 연산자원, 하드웨어 등의 공유에 대한 필요성 인식이 증대되고 있다.[1,2]

II. 연구데이터 공유

국가연구개발과제의 결과로 생산되는 연구데이터에 대해 공유·활용하기 위한 움직임은 2018년 국가과학기술심의회('18.1)에서 승인된 “혁신성장 촉진을 위한 연구데이터 공유·활용 전략”을 통해 이미 다양한 제도정책 및 과제가 수행되어 오고 있다. 특히, 국가연구데이터플랫폼인 DataON[3]이 공개('20.1)된 이후 많은 국가출연연구소 및 공공기관에서 연구데이터의 수집공유를 위한 시스템 개발 및 규정마련 등이 진행되고 있다. 특히, 최근 디지털 플랫폼에 대한 많은 관심과 정책적 지원으로 향후 다양한 분야의 연구데이터를 체계적으로 수집 공유하고 활용하기 위한 국가사회적인 노력이 활발해 질 것으로 보인다.

III. 컴퓨팅 리소스 공유

연구데이터의 공유 노력에 비해, 연구개발 시에 필수적인 소프트웨어, 알고리즘, 컴퓨팅 장비, 실험장비 등의 다양한 컴퓨팅 리소스(자원)에 대한 공유·활용에 대한 체계적인 움직임은 국내에서 찾아보기 쉽지 않다. 2021년 수행(KISTI)한 ‘연구데이터 활용 환경 개선’ 설문(670명)에 따르면, 이러한 컴퓨팅 리소스의 효율적인 공유·활용을 위한 체계 구축이 필요하다는 것에는 대부분(약 85%)의 연구자들이 동의하였다. 또한 분석도구나 컴퓨팅 자원이 공유된다면 활용하고자 하는 연구자가 약 92% 이상으로 나타날 정도로 연구데이터 높은 선호도를 볼 수 있었다.

이러한 요구에 대한 국내의 제도나 정책은 거의 전무하지만, 글로벌 선진국에서는 이러한 연구데이터 뿐만 아니라, IaaS, PaaS, SaaS, MLaaS 등 다양한 형태의 연구개발 관련 컴퓨팅 리소스를 공유하고 활용할 수 있는 체제를 구축하고 있다. EU의 EOSC(European Open Science Cloud)[4], 호주의 ARDC(Australian Research Data Commons)[5], 중국의 CSTCloud(China S&T Cloud)[6] 등이 대표적인 사

례라고 할 수 있다. 국내는 KISTI에서 최근에 이러한 컴퓨팅 리소스의 공유를 수월하게 수행할 수 있도록, KRDC(Korea Research Data Commons)를 설계하고, 핵심 프레임워크를 개발하고 있다.

IV. 연구데이터 커먼즈

연구데이터 커먼즈(Research Data Commons)는 연구데이터 활용 극대화를 위해 상호 운용 가능한 고품질의 컴퓨팅 리소스에 원활한 사용성(usability)을 제공하는 신뢰할 수 있는 공통 활용 체계로 정의한다. 여기서 사용성이라 함은 FAIR(Findable, Accessible, Interoperable, Reusable)[7] 원칙, 연구과정에서 생산된 연구데이터 및 디지털 개체에 대한 신뢰가능한 재사용성 등을 의미한다. 이는 간략히 표현하면, 연구데이터와 연구데이터 관련 컴퓨팅 리소스의 연합·활용 체계로 이해할 수 있다. 이는 광의의 개념에는 연구데이터를 분석·활용하기 위한 클라우드, 가상머신, 웹서비스, 소프트웨어, 알고리즘 등 거의 모든 연구개발 관련 자원을 포함한다. 그러나, 최근에 모든 것을 서비스(XaaS: Anything as a Service)로 인식하는 개념을 적용하면 소프트웨어 서비스로 표현 가능한 컴퓨팅 리소스를 대상으로 볼 수 있다.

그림 1과 같이, 소재분야 및 바이오 분야 연구데이터 전문센터나 출연연 및 해외 연구데이터플랫폼 등에서 제공하는 연구데이터와 HW/SW 인프라를 국가 연구데이터커먼즈(Korea Research Data Commons)라는 체계를 통해 연구데이터 뿐만 아니라, 다양한 컴퓨팅 HW/SW를 찾고 활용할 수 있는 체계를 제공하고자 하는 것이다.

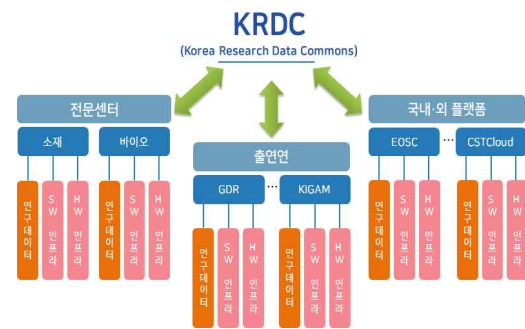


그림 1. 연구데이터커먼즈 개념도

V. 디지털 연구환경

KRDC 체계의 구축은 단기적으로는 연구데이터와 컴퓨팅 리소스의 연계·활용 활성화를 목표로 한다. 이는 궁극적으로 과학기술 연구자, 나아가 기업 및 시민 연구자까지 포함하는 사용자들에게 가

상의 디지털 연구환경을 제공해주어 그들이 손쉽게 연구개발에 집중할 수 있는 환경 구축을 목적으로 한다.

KRDC 시스템 구축은 구체적으로는 연구자들에게 분야별 사일로(장벽)을 넘는 연구데이터 및 컴퓨팅 리소스의 손쉬운 공유 환경 제공, 자유로운 가상 협업 연구 지원, 연구 재현성의 용이한 확보, 소프트웨어/알고리즘 등의 연구자산 보존 등을 실현해 낼 수 있다.

그림 2는 현재 DataON에 구축되어 있는 워크플로우 분석환경인 CANVAS(베타)로 DataON에서 사용 가능하도록 개발된 가상 디지털 연구환경이라고 할 수 있다. 비록 CANVAS가 DataON 내부에서만 사용 가능하고, 다른 플랫폼과의 호환성이 없지만, 이러한 CANVAS 환경이 기관과 국경의 장벽을 넘어 다양한 분야/국적의 연구자가 함께 연구를 수행할 수 있는 환경으로 확장되는 것이 KRDC의 목표라고 할 수 있다. 사실 EU의 EOSC와 중국의 CSTCloud에서는 이러한 목표로 두 플랫폼 간의 연계를 진행하고 있다.

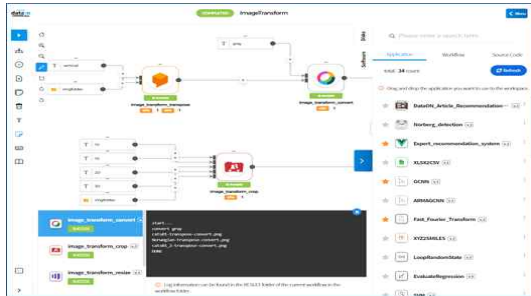


그림 2. 웹기반의 워크플로우 분석환경

V. 결론 및 향후계획

연구개발에 있어 가장 중요한 요소 중 하나인 연구데이터는 최근의 데이터 기반 연구가 주류 연구 패러다임으로 바뀌면서 더욱더 그 중요성이 커지고 있다. 이 연구데이터의 가치를 더욱더 배가하기 위해, 연구개발 컴퓨팅 리소스의 분야와 지역적 장벽을 넘는 공유활용 체계 구축이 절실하다고 할 수 있다. 이는 글로벌 선진국들의 발 빠른 준비 상황을 고려할 때 국가 과학기술 경쟁력에도 매우 중요한 비교 척도가 될 것이다. 또한, 연구자들이 가상 디지털 연구환경 위에서 다양한 분야의 연구자들과 협업 및 융합 연구 등을 효율적 수행할 수 있도록 적극적으로 지원하는 제도적 시스템적 지원이 절실하다.

이를 위해 국가 연구데이터 커먼즈 체계의 조기 구축을 통해, 연구데이터 단순 공유를 넘어 적극적인 활용 및 혁신 사례 개발이 절실하다. KISTI에서는 2021년부터 6년간 KRDC 포털, KRDC 프레임

워크, KRDC 활용 사례 개발 등 다양한 계획을 세워 국가연구데이터커먼즈 체계 구축을 위해 노력하고 있다.

2022년 KRDC 프레임워크 기반 마련 후 2023년 말까지 KRDC 프레임워크를 완성하여, 관련 기관에 배포 및 컨설팅을 수행하여, 연구데이터 및 컴퓨팅 리소스를 포함한 연구데이터 전반의 체계를 공고히 다져나갈 계획이다.

Acknowledgement

본 연구는 한국과학기술정보연구원 연구데이터와 인프라의 공유·활용 체계 구축(K-22-L01-C03-S0 1) 사업의 지원을 받아 수행된 연구임.

References

- [1] Sa-kwang Song, Songmin Seo, “Revitalization Plan for R&D Research Outcomes: focusing on research data”, Korea International Conference on Convergence Content, Jeju Special Self-Governing Province, Korea, 2021
- [2] 송사광, 서동민, “연구데이터 공유활용 제도 및 정책 현황”, GeoAI데이터학회 추계학술대회, 부산, pp. 118-119, 2021
- [3] 국가연구데이터플랫폼 DataON: [Internet]. Available : <http://www.dataon.kr>
- [4] European Open Science Cloud (EOSC): [Internet]. Available <https://eosc-portal.eu/>
- [5] Australian Research Data Commons (ARDC): [Internet]. Available : <https://ardc.edu.au/>
- [6] China Science & Technology Cloud: [Internet]. Available : <https://www.cstcloud.net/>
- [7] FAIR Principles: [Internet]. Available : <https://www.go-fair.org/fair-principles/>