

국가연구데이터커먼즈 포털 서비스 설계 방안

이미정*, 조민희*, 송사광**, 임형준*, 서동민**

*한국과학기술정보연구원 연구데이터공유센터

**과학기술연합대학원대학교 응용 AI학과

{jerryis, mini, imsmallj, hjyim, dmseo}@kisti.re.kr

Design of the Portal Service for Korea Research Data Commons

Mikyung Lee*, Minhee Cho*, Sa-kwang Song**, Hyung-Jun Yim*, Dongmin Seo**

*Research Data Sharing Center, KISTI

**Dept. of Applied AI, UST

요약

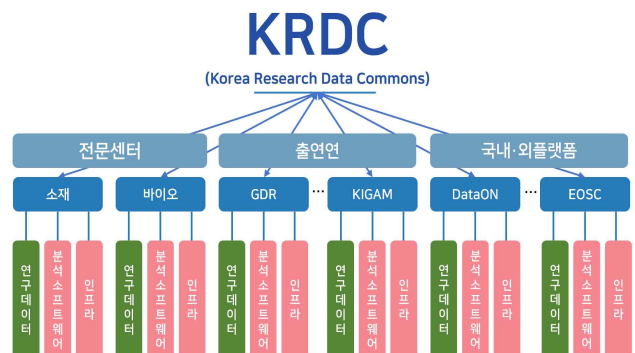
디지털 기술의 발전으로 오픈사이언스가 부상되며 데이터 기반 과학 기술이 성장하면서 연구데이터의 개방·공유 사례가 증가하고 있다. 데이터 중심 연구개발을 수행하는 연구자들은 연구데이터와 연구소프트웨어의 공유를 통해 연구 결과의 재현성을 높일 수 있게 되었다. 국가연구데이터커먼즈는 데이터 중심의 연구개발 환경을 제공하기 위해 연구데이터와 분석 소프트웨어, 인프라와 같은 컴퓨팅 리소스를 통합 및 공동 활용하는 체계를 말한다. 본 논문에서는 국가연구데이터커먼즈 체계를 실현하기 위해 개발된 국가연구데이터커먼즈 포털 서비스의 설계 방안에 대해 설명한다. 국가연구데이터커먼즈 포털 서비스는 국가연구데이터커먼즈 사용자와 KRDC 프레임워크 간을 연결해주는 역할을 한다. KRDC 포털 서비스는 사용자 대시보드를 통해 컴퓨팅 리소스 카탈로그 서비스, 컴퓨팅 리소스 마켓 플레이스, 워크플로우 서비스로 구분할 수 있다. 국가연구데이터커먼즈 사용자는 역할에 따라 KRDC 이용자, 컴퓨팅 리소스 제공자, 클러스터 관리자, KRDC 관리자의 4가지 그룹으로 나뉜다. KRDC 이용자는 카탈로그 서비스에 등록된 컴퓨팅 리소스를 사용하며, 컴퓨팅 리소스 제공자는 컴퓨팅 리소스를 마켓플레이스에 등록하여 KRDC 이용자에게 제공한다. 클러스터 관리자는 KRDC 프레임워크가 설치된 클러스터 관리 및 컴퓨팅 리소스 배포, 실행, 관리, 모니터링의 역할을 하며, KRDC 관리자는 포털 관리 및 KRDC에 등록된 컴퓨팅 리소스를 관리, 모니터링한다.

I. 서론

디지털 기술의 확산으로 최근 연구 동향은 데이터 기반 연구개발 방식으로 전환되고 있으며, 연구 과정과 성과를 개방화하는 오픈사이언스(Open Science)를 통한 연구 움직임도 활발해지고 있다. 오픈사이언스는 모든 종류의 과학적 지식을 디지털화하여 개방·공유하여 공동 연구를 활성화하는 개념으로, 디지털 기술을 활용한 연구성과와 과정을 공개하고 공유하는 것이다. 데이터 기반 연구개발 패러다임 전환으로 연구데이터를 중심으로 발간된 연구 산출물의 65%에서 데이터 집약적 연구가 활발해지고 있다[1]. 따라서 디지털 기술을 통해 연구개발 환경을 지원하는 것 뿐만 아니라 데이터 기반의 과학 성장을 위해서는 논문을 중심으로 연구데이터와 연구데이터를 분석·활용하기 위한 연구소프트웨어를 함께 제공하여 연구의 신뢰성 및 재현성을 확보해야 한다[2]. 우리는 데이터 기반 연구개발을 통해 디지털 격차를 최소화하고, 연구소프트웨어를 통한 연구생산성을 향상시키기 위하여 국가연구데이터커먼즈(Korea Research Data Commons, KRDC) 체계를 정립하여 데이터 기반 연구개발 환경을 지원하고자 한다. 본 논문에서는 연구데이터의 분석·활용을 극대화하기 위해 디지털 기반 가상 연구 환경을 지원하는 KRDC에 대해 소개하고, KRDC 체계를 실현하기 위한 KRDC 포털 서비스 설계 방안에 대해 설명하고자 한다.

II. 국가연구데이터커먼즈

연구데이터커먼즈(Research Data Commons)는 연구데이터의 활용성을 높이기 위해 상호 운용 가능한 서비스와 리소스의 연합 체계이다. KRDC는 EOSC, ARDC, GORC, GOSC 등의 해외 연구데이터커먼즈 연구 동향을 반영하여 국내 연구·개발 요구 사항에 맞춘 연구데이터 활용 체계이다. 그림 1과 같이 KRDC는 전문센터, 출연연, 연구데이터플랫폼 등에 존재하는 연구데이터와 연구소프트웨어, 분석 서비스 및 리소스들을 연합하여 연구자들에게 제공함으로써 연구데이터의 공유·활용 활성화를 목적으로 구현된다[3][4].

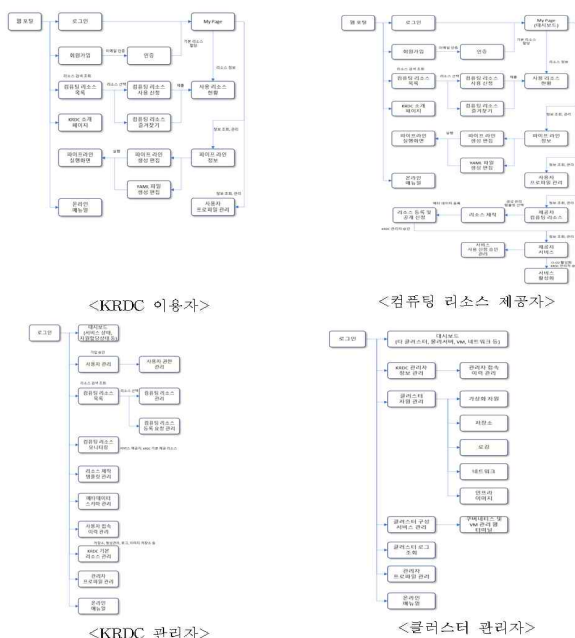


(그림 1) KRDC 개념도

KRDC에서는 연구데이터플랫폼(DataON)[5]과 같이 연구데이터와 분석 소프트웨어, 인프라를 보유하고 있는 분야별 전문센터, 출연연 및 해외 커먼즈 플랫폼들이 컴퓨팅 리소스 제공 대상이 된다. 각 기관별 연구데이터플랫폼에 연구데이터커먼즈 프레임워크를 배포하여 컴퓨팅 리소스 통합 및 공동 활용 체계를 구축한다. KRDC를 통해 연구데이터 및 리소스를 공유함으로써 연구자들이 연구에 필요한 연구데이터를 수집, 처리, 분석, 재현하는데 필요한 시간과 비용을 절감하게 되어 연구개발 주기를 단축할 수 있게 된다.

III. 국가연구데이터커먼즈 포털 서비스 설계 방안

KRDC 포털 서비스는 사용자를 위해 컴퓨팅 리소스를 등록하는 마켓플레이스, 컴퓨팅 리소스 카탈로그, 그리고 다중의 컴퓨팅 리소스를 이용하여 분석 서비스를 지원하는 워크플로우가 핵심 서비스이다. 서비스 관점에 따라 KRDC 사용자는 KRDC 이용자, 컴퓨팅 리소스 제공자, 클러스터 관리자, KRDC 관리자의 4개 그룹으로 구분하여 사용자 맞춤형 서비스를 설계하였다. KRDC 이용자는 마켓플레이스에 공유된 컴퓨팅 리소스를 사용하기 위해 포털에 접속하는 일반 이용자를 말한다. 카탈로그 서비스에서 이용자가 필요한 컴퓨팅 리소스를 검색하여 사용 승인을 받은 후, 워크플로우 서비스를 통해 분석에 필요한 다중 컴퓨팅 리소스들을 연결하여 분석 결과를 도출한다. 컴퓨팅 리소스 제공자는 직접 개발하거나 공유가 가능한 컴퓨팅 리소스들을 컨테이너 이미지로 만들어 배포하고 메타데이터와 함께 KRDC 마켓플레이스에 등록하고, KRDC 카탈로그 서비스를 통해 이용자들에게 서비스를 제공하는자를 말한다. 컴퓨팅 리소스 제공자는 대시보드를 통해 리소스 사용 승인, 리소스 관리, 운영 모니터링이 가능하다. 클러스터 운영자는 KRDC 프레임워크가 동작하는 클러스터를 관리하는 자로, 컴퓨팅 리소스에 대한 접근 권한을 부여하고 분석 서비스를 실행, 운영, 관리하고 클러스터 시스템을 운영, 관리, 모니터링하는 역할을 한다. 마지막으로 KRDC 관리자는 KRDC 포털 관리자로 KRDC 사용자(KRDC 이용자, 컴퓨팅 리소스 제공자)를 관리하고 KRDC 카탈로그에 등록된 리소스들을 운영, 관리, 모니터링하는 역할을 한다. KRDC 포털 서비스는 그림 2와 같이 사용자 그룹에 따라 서로 다른 서비스를 제공한다.



(그림 2) KRDC 사용자별 서비스 네비게이션

IV. 결론

본 논문에서는 분산되어있는 다양한 컴퓨팅 리소스를 연합하여 가상의 연구 환경 체계를 제공하는 KRDC 포털 서비스 설계 방안에 대해 설명하였다. KRDC 포털은 컴퓨팅 리소스를 등록하여 KRDC에서 공유하고 이용자들이 필요한 컴퓨팅 리소스를 검색하여 편리하게 사용할 수 있도록 지원해주는 것이 중요한 기능이다. KRDC 사용자 그룹은 성격에 따라 KRDC 이용자, 컴퓨팅 리소스 제공자, 클러스터 관리자, KRDC 관리자로 구분할 수 있다. KRDC 이용자는 카탈로그 검색하고 컴퓨팅 리소스를 사용하여 워크플로우 서비스를 수행하는 자, 컴퓨팅 리소스 제공자는 컴퓨팅 리소스를 마켓플레이스에 등록하고 모니터링 서비스를 수행하는 자, 클러스터 관리자는 등록된 컴퓨팅 리소스를 배포/실행 및 운영, KRDC 관리자는 포털 관리 및 모니터링 역할을 한다. KRDC 포털 서비스에서는 사용자 그룹에 필요한 서비스들을 맞춤형으로 제공하게 된다. 향후, 포털 서비스에 대한 사용자 요구사항을 분석하여 향상된 사용자 맞춤형 서비스를 구현할 예정이다.

ACKNOWLEDGMENT

본 연구는 2022년도 한국과학기술정보연구원(KISTI) 주요사업 “연구데이터와 인프라의 공유·활용 체제 구축(K-22-L01-C03-S01)” 과제로 수행한 것입니다.

참고 문헌

- [1] OECD, “OECD International Survey of Science,” 2020, Available: <https://oecd.org/sti/survey-of-scientific-authors.htm>
- [2] 임형준, 이미경, 송사광, 서동민, 조민희, “데이터 기반 연구개발 위한 국가연구데이터커먼즈 설계 및 적용 방안”, 특집호: 빅데이터 분석 및 시각화, 기법 및 응용, 제32권 5호, 한국지능시스템학회 논문지, 2022
- [3] 송사광, 서동민, “연구데이터 공유활용 제도 및 정책 현황”, GeoAI데이터학회 추계학술대회, pp. 118-119, 2021
- [4] 이미경, 조민희, 임형준, 박성은, 최성배, 박형선, 송사광, “연구 데이터 공유·활용 활성화를 위한 데이터 커먼즈 필요성 연구”, 한국통신학회 추계종합학술대회, 2021
- [5] 국가연구데이터플랫폼 DataON: <https://dataon.kisti.re.kr>