

컴퓨팅 리소스 관리를 위한 표준 메타데이터 스키마 설계

이미경^{1*} · 조민희¹ · 송사광^{1,2} · 임형준¹

¹한국과학기술정보연구원 연구데이터공유센터 · ²과학기술연합대학원대학교 응용AI학과

Design of Standard Metadata Schema for Computing Resource Management

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

¹Research Data Sharing Center, KISTI · ²Dept. of Applied AI, UST

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

요 약

본 논문에서는 국가연구데이터커먼즈에서 연구데이터 분석·활용에 사용되는 컴퓨팅 리소스를 등록, 검색, 관리하기 위한 컴퓨팅 리소스 표준 메타데이터 스키마 설계 방안에 대해 소개한다. 국가연구데이터커먼즈는 연구데이터 공유·활용 극대화를 위한 연구데이터와 컴퓨팅 리소스 연합 활용 체계이다. 컴퓨팅 리소스는 연구 전 과정에서 사용하는 연구데이터를 분석·활용하는데 필요한 분석 인프라, 분석 소프트웨어 등 컴퓨팅 환경의 모든 리소스들을 말한다. KRDC 컴퓨팅 리소스 관리를 위한 표준 메타데이터 스키마는 컴퓨팅 리소스 관리를 위한 공통 필수 속성과 각 컴퓨팅 리소스 특징에 따른 속성을 고려하여 설계하였다. 컴퓨팅 리소스 관리를 위한 표준 메타데이터 스키마는 컴퓨팅 리소스 메타데이터 스키마와 컴퓨팅 리소스 제공자 메타데이터 스키마로 구성된다. 또한, 컴퓨팅 리소스와 제공자의 메타데이터 스키마는 성격에 따라 서비스 스키마와 시스템 스키마 그룹으로 구분하여 설계하였다. 표준 메타데이터 스키마는 KRDC 프레임워크를 통해 컴퓨팅 리소스 제공자와 컴퓨팅 리소스 사용자를 위한 컴퓨팅 리소스 등록, 카탈로그 검색, 컴퓨팅 리소스 관리, 워크플로우 서비스에 사용되며, 다양한 컴퓨팅 리소스 연계를 위해 확장 가능한 형태로 설계되었다.

ABSTRACT

In this paper, we introduce a computing resource standard metadata schema design plan for registering, retrieving, and managing computing resources used for research data analysis and utilization in the Korea Research Data Commons(KRDC). KRDC is a joint utilization system of research data and computing resources to maximize the sharing and utilization of research data. Computing resources refer to all resources in the computing environment, such as analysis infrastructure and analysis software, necessary to analyze and utilize research data used in the entire research process. The standard metadata schema for KRDC computing resource management is designed by considering common attributes for computing resource management and other attributes according to each computing resource feature. The standard metadata schema for computing resource management consists of a computing resource metadata schema and a computing resource provider metadata schema. In addition, the metadata schema of computing resources and providers was designed as a service schema and a system schema group according to their characteristics. The standard metadata schema designed in this paper is used for computing resource registration, retrieval, management, and workflow services for computing resource providers and computing resource users through the KRDC web service, and is designed in a scalable form for various computing resource links.

키워드

Korea Research Data Commons, Standard Metadata Schema, Research Data, Computing Resource

1. 서 론

최근 데이터 기반 과학이 성장하고 있으며 데이터 기반 R&D 연구의 중요성이 부각되고 있다. 이는 데이터 기반 과학의 대표적인 AI 분야의 저널 논문 수가 지속적으로 증가하고 있음으로 알 수

* speaker

있다. 디지털 기술이 연구 산출물의 공유를 촉진하고 데이터 접근을 확대하는데 기여하고 있으나, 여전히 연구자들은 연구에 필요한 데이터나 분석 업무를 수행하는데 필요한 분석 인프라, 소프트웨어 등은 비공식적인 방식에 의존하여 공유하고 있는 실정이다[1]. 과학 연구 분야에서 연구 결과의 재현성은 매우 중요하며, 재현성의 위기를 해결하는데 연구데이터와 연구소프트웨어의 개방과 공유가 중요하다. 오픈사이언스(Open Science)는 디지털 기술을 통해 연구의 전 과정을 보다 개방적으로 전환하려는 일련의 움직임(OECD, 2016)을 말한다. 해외 오픈사이언스 연구 움직임으로 EOSC(European Open Science Cloud), GORC(Global Open Research Commons), ARDC(Australian Research Data Commons), CST Cloud(China Science and Technology Cloud), AOSP(African Open Science Platform) 등 여러 국가에서 관련 연구를 진행 중에 있다[2,3,4,5,6]. 글로벌 연구 환경의 변화에 따라 국내 데이터 기반 연구자들도 오픈사이언스 기반의 연구데이터 커먼즈의 필요성을 느끼고 있는 실정이다[7]. 국내에서도 연구데이터 및 소프트웨어, 인프라, 실험환경, 하드웨어 등의 공유에 대한 필요성이 증대되고 있다[8].

본 논문에서는 연구데이터의 공유·활용을 극대화하는 국가연구데이터커먼즈(KRDC, Korea Research Data Commons)에 대해 소개하고, KRDC 컴퓨팅 리소스 관리를 위한 표준 메타데이터 스키마에 대해 설명한다.

II. 오픈사이언스와 연구데이터

디지털 기술의 확산과 함께 연구의 성과와 과정을 개방화하는 추세가 확대되고 있다. 오픈사이언스 정책은 오픈액세스(Open Access), 오픈데이터(Open Data), 오픈협력(Open Collaboration)으로 발전된다. 오픈액세스 저널의 성장과 온라인 논문 열람·배포가 일상화되고, 데이터 기반 과학의 성장으로 연구데이터의 개방·공유 사례가 증가하고, 개방형 연구 협력을 지원하는 디지털 인프라 및 서비스로 확대되고 있다[9]. 오픈데이터는 오픈 연구데이터를 말하며, 국내에서는 연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구 결과의 검증에 필수적인 데이터를 연구데이터로 정의한다(국가연구개발사업의 관리 등에 관한 규정 제1장 제2조).

III. 국가연구데이터커먼즈

국가연구데이터커먼즈는 연구데이터 공유·활용 극대화를 위한 연구데이터와 컴퓨팅 리소스(Computing Resource) 연합 활용 체계를 말한다. KRDC는 분산되어있는 다양한 연구데이터와 컴퓨팅 리소스

를 논리적으로 연결하여 실시간 분석·활용을 지원하는 것으로 GOSC, EOSC, ARDC 등 해외 연구데이터 협의체 및 커먼즈 등과의 상호운용성을 확보하여 국가적 오픈사이언스 역할을 강화하여 글로벌 협력 체계를 구축하는 목적을 가진다. 컴퓨팅 리소스는 데이터 기반 연구개발에 필요한 컴퓨팅 기반의 환경 및 기능적 요소를 말하며, OpenAPI, 웹서비스, 가상 머신, 데이터 저장소, 클라우드, 슈퍼컴퓨터 등을 말한다. 컴퓨팅 환경에서 연구데이터 분석·활용을 지원하는 모든 것이다. 다시 말해 KRDC는 데이터 기반 연구개발 환경을 제공하기 위한 컴퓨팅 리소스를 공유·활용하여 거대 과학, 학제간 연구, 융합 연구 등의 과학기술계의 디지털 혁신과 데이터 중심 연구개발 혁신 체제를 확립하기 위한 체계이다.

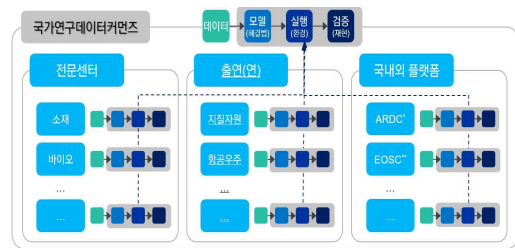


그림 1. KRDC 체계 개념도

IV. 표준 메타데이터 스키마 설계

컴퓨팅 리소스 메타데이터 스키마는 표 1과 같이 서비스 스키마와 시스템 스키마의 그룹으로 구분할 수 있다. 서비스 스키마는 컴퓨팅 리소스를 등록, 검색, 관리하기 위해서 컴퓨팅 리소스 제공자와 컴퓨팅 리소스에 대한 메타데이터이다. 서비스 스키마는 ID, 이름 등의 기본적인 정보를 가지는 기본 정보, 마케팅 정보, 카테고리 정보, 위치 정보, 관리 정보, 접근 정보의 그룹으로 나뉜다. 시스템 스키마는 KRDC 프레임워크에서 실제 서비스를 배포, 실행, 운영, 모니터링, 관리하는데 필요한 스키마들을 설계하여 그룹화하였다. 시스템 스키마는 리소스 관리, 사용자 관리, 공통 리소스 관리, 리소스 제공자 관리, 사용자 파이프라인 관리 그룹으로 구분하여 설계하였다.

표 1. 컴퓨팅 리소스 메타데이터 스키마 그룹

서비스 스키마	Basic Information
	Marketing Information
	Classification Information
	Location Information
	Contact Information
	Management Information
	Access Information

시스템 스키마	Resource management
	User Management
	Common resource management
	Resource providing management
	User pipeline management

서비스 스키마는 그림 2와 같이 컴퓨팅 리소스와 컴퓨팅 리소스 제공자를 위한 필수 메타데이터 스키마로 구성된다. 필수 메타데이터 스키마 이외에 컴퓨팅 리소스의 특징에 따라 추가되는 스키마들이 존재한다. KRDC 프레임워크를 통해 제공되는 컴퓨팅 리소스를 등록하기 위해서는 서비스 메타데이터 스키마를 기준으로 항목들이 구성되며, 서비스의 배포, 실행 및 운영을 위해서는 시스템 스키마를 기준으로 실행된다. 컴퓨팅 리소스 메타데이터 스키마는 국가연구데이터플랫폼(DataON), EOSC 등과의 연계를 위하여 필수 속성 값들은 서로 매핑되도록 설계하였다. 또한 다양한 컴퓨팅 리소스와 확장가능한 형태로 설계되었다.

Computing Resource	Service Provider
ID	ID
Name	Name
Resource Organization	Email
Resource Providers	Affiliation
Description	Location
Domain	Description
Category	
Help desk	
Contact Email	
Terms of Use	
Privacy Policy	
Access Type	

그림 2. 컴퓨팅 리소스와 컴퓨팅 리소스 제공자의 필수 메타데이터 스키마

V. 결 론

본 논문에서는 데이터 기반 연구·개발 환경 제공을 위한 국가연구데이터커먼즈 체계를 구축하기 위해서 KRDC 프레임워크에서 서비스되는 컴퓨팅 리소스를 관리하기 위한 표준 메타데이터 스키마에 대해 설명하였다. 표준 메타데이터 스키마는 서비스 스키마와 시스템 스키마로 구분하고, 서비스 스키마는 컴퓨팅 리소스와 컴퓨팅 리소스 제공자의 관점에서 7개의 스키마 그룹으로 속성 값을 그룹화하였다. 시스템 스키마는 서비스의 실행과 운영, 모니터링 관점에서 5개의 관리 그룹으로 속성 값을 그룹화하였다. 이렇게 설계된 표준 메타데이터 스키마는 KRDC 프레임워크를 통해 컴퓨팅 리소스 등록, 검색, 카탈로그 서비스에 활용된다.

Acknowledgement

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

References

- [1] B. K. Park, Y. B. Lee, “Review Key results of OECD ISSA,” *KISTEP Statistical Brief*, Vol. 5, pp. 2-13, 2021.
- [2] European Open Science Cloud (EOSC) [Internet]. Available <https://eosc-portal.eu/>
- [3] RDA Global Open Research Commons (GORC) Interest Group [Internet]. Available : <https://www.rd-alliance.org/groups/global-open-research-commons-ig>
- [4] Australian Research Data Commons (ARDC) [Internet]. Available : <https://ardc.edu.au/>
- [5] China Science & Technology Cloud [Internet]. Available : <https://www.cstcloud.net/>
- [6] African Open Science Platform (AOSP) [Internet]. Available: <https://aosp.org.za/>
- [7] M. K. Lee, “A Study on the Need for Data Commons to Promote the Sharing and Dissemination Research Data,” *KICS Fall Conference*, Yeosu, pp. 773-774, 2021.
- [8] S. K. Song, S. M. Seo, “Revitalization Plan for R&D Research Outcomes: focusing on research data,” *Korea International Conference on Convergence Content*, Jeju, Korea, 2021.
- [9] E. J. Shin, W. K. Jung, “The Spread and Implications of Open Science Policy,” *STEPI Insight*, Vol. 216, pp. 5-32, Aug. 2017.