

글로벌 개방형 연구데이터 커먼즈 및 시사점

송사광^{1,3}, 조민희¹, 이미경¹, 임형준²

¹한국과학기술정보연구원 연구데이터커먼즈팀 책임연구원

²한국과학기술정보연구원 연구데이터공유센터 센터장

³과학기술연합대학원대학교 응용 AI 학과 교수

esmallj@kisti.re.kr, mini@kisti.re.kr, jerryis@kisti.re.kr, hjiyim@kisti.re.kr

Global Open Research Commons and Implications

Sa-kwang Song^{1,3}, Minhee Cho¹, Mikyoung Lee¹, Hyung-Jun Yim²

¹Dept. of Research Data Commons, Korea Institute of Science and Technology Information

²Research Data Sharing Center, Korea Institute of Science and Technology Information

³Dept. of Applied AI, University of Science and Technology

요약

오픈 사이언스 운동의 활성화로 인해 다양한 연구 관련 자원들 간의 상호 운용성 확보를 위한 노력이 활발해지고 있다. 특히, 글로벌 연구데이터 커먼즈(Global Open Research Commons) 모델 개발 관련 표준화 활동이 세계 최대 연구데이터 커뮤니티인 RDA(Research Data Alliance)의 주도로 진행되어 왔고, 최근에 GORC Working Group에서 버전 1.0 모델을 오픈하였다. 이에 이 모델에 대해 살펴보고 국내의 연구데이터 커먼즈인 KRDC(Korea Research Data Commons)와 비교 및 시사점을 논하고 향후 연구 방향을 소개한다.

1. 서론

오픈 사이언스 운동의 확산으로 인해 전세계 다양한 연구자들에게 연구와 관련된 리소스의 공유 및 활용에 대한 의식이 적극적으로 확대되고 있다. 이는 국내의 연구자들에게도 동일하게 오픈 사이언스에 대한 적극적인 호응과 더불어 다양한 우려의 목소리가 나오고 있다. 연구 리소스의 공유 및 활용이라는 목적 자체는 긍정적으로 받아들이지만 지적자산권 보호 문제로부터 주도적 연구 기회의 확보 필요성 등 예민한 문제에 대한 이슈제기로 오픈 사이언스 활성화에 걸림돌로 작용하고 있는 것이 현실이다. 비단 국내뿐만 아니라 다양한 국가에서 유사한 이슈가 제기되고 있다.

그러나, 이러한 오픈 사이언스의 물결을 거부하기

에는 이미 다양한 선진 국가나 글로벌 기구에서 오픈 사이언스를 활성화하기 위한 다양한 움직임을 보이고 있다. 가장 대표적인 글로벌 기구 중 하나인 글로벌 대표 연구데이터 연합체, RDA(Research Data Alliance)[1]에서는 연구데이터 관련 다양한 리소스 공통요소인 연구데이터 커먼즈를 상호운용할 수 있도록 하기 위해 GORC(Global Open Research Commons) Interest Group을 통해 연구데이터 커먼즈의 글로벌 표준 모델 개발 작업을 진행하고 있다[2]. 동시에 ISC 산하 CODATA에서는 GOSC(Global Open Science Cloud) Initiative를 만들어 RDA의 GORC와 상호 참조하며 글로벌 오픈 사이언스의 실현을 위한 작업(Working Group)을 진행중이다[3].

이에 본 연구는 최근에 공유된 RDA GORC의

International Model[4]를 살펴보고, 국가 연구데이터 커먼즈(Korea Research Data Commons) 소개 및 시사점을 논하고자 한다.

2. 글로벌 개방형 연구데이터 커먼즈

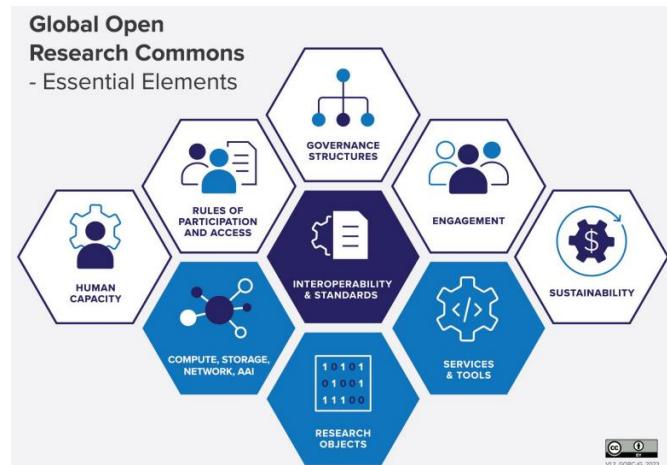
RDA GORC International Model WG(Working Group)은 GORC IG(Interest Group)의 일환으로 진행중이다. GORC IG 는 유럽 오픈 사이언스 클라우드, 아프리카 오픈 사이언스 플랫폼, 국제 가상 천문학 연합과 같은 국가, 범 국가 및 도메인별 조직들 간의 협력을 지원하기 위한 결과물을 개발하고 있으며, 이러한 조직들은 연구자들이 사회적으로 큰 도전 과제에 대응하기 위해 상호 운용 가능한 자원을 구축하고 있다. GORC의 실현을 위한 비전은 데이터, 출판물, 소프트웨어 및 컴퓨팅 자원 뿐만 아니라 모든 연구 자산, 메타데이터, 어휘 및 식별 서비스에 대한 어디서나 누구에게나 항상 자유로운 액세스를 제공하는 것이다.

2023년 9월에 공개된 GORC IM 모델의 버전 1 보고서는 이 이니셔티브의 배경 정보를 제공하며, 그 의도와 의도된 대상, 모델을 만드는 데 사용된 방법, 모델의 구조와 내용을 설명하고 있다.

GORC 국제 모델(IM) 작업 그룹(WG)은 모델을 생성하기 위해 다양한 자원과 전문가 의견을 통합하여 모델을 생성했다. 이 모델은 어떤 종류의 공통체의 개발을 수행할 때 고려해야 할 요소의 범주, 하위 범주, 속성 및 특징으로 구성되어 있으며, 이 모든 요소는 어떤 종류의 공통체이든지 어떤 단계에서든지 고려해야 한다. 범주, 하위 범주, 속성 및 특징은 중요한 것, 원하는 것 또는 선택 사항으로 표시되어 있지만, 모델은 무엇을 어떻게 구현해야 하는지를 강요하지 않으며, 어떤 것이 관련되며 어떤 자원이 투자되어야 하는지에 대한 결정은 환경 및 실행자의 우선 순위에 따라 이루어진다. 본 모델 생성을 위해 유럽 EOSC(European Open Science Cloud), 호주 ARDC(Australian Research Data Commons), 일본 NII(National Institute of Informatics), 중국 CAS(Chinese

Academy of Science) 등 다양한 기관의 커먼즈가 검토되었는데, 국내의 경우는 KISTI의 KRDC(Korea Research Data Commons)가 소개되고 검토되어 모델 개발에 반영되었다.

본 모델(IM)의 목적은 전 세계적으로 상호 운용 가능한 연구 서비스를 개발하려는 이해 관계자들에게 공통의 프레임워크와 공통 언어를 제공하는 것이다. 모델의 대상은 연구 공통체의 계획, 개발, 운영, 자금 조달 또는 사용에 관여하는 모든 이에게 해당된다. 이 모델은 지시적인 모델이 아니며, 대신 모델 사용자가 자신의 공통체가 발전하는 맥락에서 고려할 수 있는 연구 공통체의 속성 및 특징의 공통 기반을 정의하고 설정하는 데 사용된다. 본 모델의 필수 요소는 그림 1과 같이 표현된다.



(그림 1) 연구데이터 커먼즈의 필수 요소[4].

그림 1에서 파란색으로 표시된 세 가지 요소는 사람들이 상호작용하는 커먼즈의 기반 요소이다. 흰색으로 표시된 5 가지 요소는 커먼즈의 성공을 위해 필요한 사회적/인간적 요소이다. 한가운데의 진한 파란색 요소는 상호운용성과 표준이라는 커먼즈 코어에서 핵심적인 중요성을 대표하는 역할을 한다. 각 요소별 좀 더 세부적인 정보는 다음과 같다.

- 지배구조와 리더십

Governance(지배구조)는 공통의 목적을 정의하고, 그 목적을 공통의 경영 및 내부 인원이 어떻게 추

구할 것인지를 규정하는 전략, 목표, 가치 및 정책의 개발에 중점을 둔다. 이에는 미션 명세서, 가치, 조직 성과 지표, 리스크 관리 프레임워크를 비롯한 전략 계획 개발, 커뮤니티 지지 원칙 및 FAIR 및 데이터 윤리와 같은 프레임워크를 포함한다.

- 참여 및 접근 규칙(Rules of Participation and Access)

참여 및 접근 규칙은 공통체 이해 관계자들의 권리, 의무 및 책임을 정의하는 정책을 가리킨다. 참여 및 접근 규칙은 다양한 이해 관계자 그룹이 공통체와 서로 상호 작용하는 방식을 규정한다. 이러한 규칙은 공개 연구, 협력 및 투명성 등의 공통체 정의된 가치를 유지하면서, 공통체 자원 및 커뮤니티 구성원과의 책임 있는 윤리적 상호 작용을 촉진하기 위해 설계되었다.

- 지속 가능성(Sustainability)

지속 가능성은 공통체의 지속 가능성과 운영을 어떻게 보장할 것인지에 대한 모델과 협약을 포함한다. 이는 공통체가 장기적으로 유지될 수 있도록 보장하는 자금 조달 및 자원 확보 활동을 포함한다.

- 참여(Engagement)

참여는 공통체의 성공을 이루기 위해 필요한 사회/인간적 요소 중 하나이며, 공통체와 넓게 정의된 이해 관계자 커뮤니티 간의 상호 작용을 가리킨다. 참여 활동은 요구 사항 수집, 의견 수렴, 사용성 테스트, 커뮤니케이션, 이벤트 및 교육 등을 포함한 반복적인 주기로 볼 수 있어야 한다.

- 인간 능력(Human Capacity)

공통체가 모든 이해 관계자와 커뮤니티 구성원을 고려한 인간 친화적 환경을 조성하는 능력을 가리킨다. 이는 특히 사용자, 제공자 및 중개업체에게 해당하며, 공통체가 목표를 설정하고 달성하

며, 기능을 수행하고 문제를 해결하며 이 프로세스를 가능하게 하는 데 필요한 수단과 조건을 계속 발전시킬 수 있도록 하는 것을 의미한다[5].

- ICT 인프라(ICT Infrastructure)

ICT 인프라란 컴퓨터 시스템이 작동하는 데 필요한 하드웨어 및 기본 소프트웨어 구성 요소를 의미한다. 이것들은 연구를 수행하는 데 필수적이며, 프로젝트 및 기대치의 볼륨, 복잡성 및 속도가 증가함에 따라 확장될 수 있도록 설계되어야 한다.

- 상호 운용성(Interoperability)

상호 운용성은 통합되지 않는 리소스에서 나온 데이터나 도구가 최소한의 노력으로 통합되거나 함께 작동할 수 있는 능력을 의미하며, FAIR 원칙을 구현하는 데 가장 어려운 부분 중 하나로 여겨진다. 또한 상호 운용성은 다양한 공통체 사용 사례 범위를 활성화할 수 있다. 상호 운용성은 데이터의 교차 공통체 재사용을 가능하게 하며, 공통체에 중요한 역할을 한다.

- 표준 및 규칙(Standards & Conventions)

표준(Standards)은 법적인 규정에 따라 또는 사실상으로 존재할 수 있다. 법적으로 정해진 표준, 즉 법에 따른 표준(De jure standards)은 ISO 와 같은 공식적인 표준 기구에 의해 승인된다. 사실상 표준 또는 규칙(De facto standards)은 커뮤니티에서 널리 채택된다. 이러한 표준은 해당 커뮤니티에서 일반적으로 수행되는 방식의 일부가 되었을 때 생기게 된다. 사실상 표준은 형식적인 표준 기관에 제출되고 승인되면 법적으로 정해진 표준이 될 수 있다.

- 서비스 및 도구(Services & Tools)

서비스는 사용자가 자신을 대신하여 어떤 작업을 수행하기 위해 호출할 수 있는 공통체 요소다. 서비스는 일반적으로 기계가 사용하도록 고안되었으며, 주로 소프트웨어에 의해 호출된다. 도구는

연구자가 데이터를 기반으로 하여 하나 이상의 작업을 수행하도록 하는 것으로, 일반적으로 데이터를 출력물로 사용한다. 도구는 주로 인간이 사용하도록 고안되었습니다.

● 연구 개체(Research Objects)

연구 대상은 연구 과정의 결과물이지만 나중 프로세스의 입력으로도 사용될 수 있습니다. ICT 인프라와 서비스 & 도구와 마찬가지로, 이들은 공통체에서 사람들이 상호 작용하는 기본 디지털 요소이다. 여기에서의 범위는 디지털 연구 대상으로 한정되며, 연구 하드웨어 자체는 포함되지 않는다. 이 모델은 다음과 같은 다섯 가지 주요 연구 대상 범주를 식별한다.

- **출판물 및 연구 문서**: 연구 프로젝트 및 활동의 어떤 측면을 설명하거나 논의하는 디지털, 텍스트, 시각, 오디오 또는 촉각적 형태를 포함하는 인간 또는 기계가 읽을 수 있어야 한다.

- **연구 데이터**: 식별 가능하며 단일 주체에 의해 집합체로 관리되거나 게시 가능한 데이터 모음으로, 연구 활동의 결과물이다

- **연구 소프트웨어**: 연구 과정 중에 생성되거나 연구 목적을 위해 만들어진 모든 소프트웨어 구성 요소로, 컴퓨터 또는 기계에서 실행 가능한 것이다(프로토콜, 워크플로, 알고리즘 등을 설명하는 실행 가능한 연구 문서는 위의 '연구 문서' 범주에 포함된다).

- **의미론적 대상**: 고유한 식별자를 가진 설명적(descriptive) 요소들의 명명된 그룹이다. 의미론적 대상은 문서, 연구 소프트웨어 또는 연구 데이터 형태로 존재할 수 있다.

- **컬렉션**: 동일한 종류 또는 다른 종류의 연구 대상을 조합하거나 번들로 묶은 것으로, 관련성을 공유하고 디지털 개체로 취급될 가능성이 있는 것이다. 컬렉션은 참조된 각 식별자와 관련된

메타데이터로 구성된다.

● KPI 및 메트릭스(KPIs & Metrics)

KPI(KPIs)는 공통체, 공통체 이해 관계자 및 공통체 커뮤니티가 공통체 특성 및 기능의 수용, 참여 또는 사용을 측정하는데 사용할 수 있는 정성적 또는 정량적 측정 항목이다. 반면에 메트릭스(metrics)는 특정 프로세스의 진화 또는 성능을 평가하는 데 사용되는 정량적인 측정 항목으로, KPI 모니터링에 기여하거나 피드백을 제공한다.

3. 국가 연구데이터 커먼즈(KRDC)

국내의 경우는 2020년 공개된 연구데이터 플랫폼인 DataON[6]이 연구 개체의 공유활용을 목적으로 한 대표적인 커먼즈 사례이다. 또한 출판물 및 연구 문서 등의 공유활용을 위한 시스템은 AccessON[7]이 대표적이다. 그러나, 이들은 상호운용성을 목적으로 체계적으로 구축되었다가 보단 단편적인 계획에 의해 진행된 산출물이 더 적합하다고 볼 수 있다. 그러나 글로벌 동향에 맞춰 지역적, 국가적, 또는 주제별 경계를 넘어 상호 운용 가능한 커먼즈 구축을 목적으로 진행하는 프로젝트가 KRDC 라고 할 수 있다. KRDC는 GORC IM 모델 개발시 한국의 연구데이터 커먼즈 사례로 소개되고 검토된 유일한 연구데이터 커먼즈로 특히 연구 소프트웨어의 상호운용에 좀 더 강점을 두고 있는 연구라 할 수 있다. 특히, 연구활동 결과로 확보된 인공지능 알고리즘, SW, 서비스 등을 오픈 사이언스 개념에 맞춰 공유활용 할 수 있도록 지원하는 쿠버네티스 기반의 연구소프트웨어 공유활용 체계이다. 단순한 소스의 공유활용을 넘어 연구 소프트웨어의 개발 및 공유/재활용과 복수의 연구 소프트웨어를 파이프라인으로 엮어 연구 절차를 워크플로우 형태로 표현하고 수행할 수 있는 환경까지도 제공하고 있다. 이는 다른 글로벌 커먼즈들이 아직 실현하지 못하고 있는 기능이라고 할 수 있다.

4. 시사점 및 향후 연구

CODATA 의 GOSC 와 더불어 RDA 의 GORC 는 오픈 사이언스를 리딩하는 글로벌 기구로써, 연구데이터 커먼즈에 관한 De facto standard 모델을 만들기 위해 부단히 노력해 왔고, 최근에 버전 1.0 을 오픈한 것은 매우 뜻깊은 의미를 갖는다. 특히, 이 모델을 만들기까지 한국의 KRDC 를 포함하고, EU 의 EOSC, 호주의 ARDC, 중국 CAS 의 CSTCloud, 일본 NII 의 RDC 등의 발표를 포함하는 등 거의 대부분의 국가 단위 커먼즈에 대해 리뷰하였다. 각 커먼즈의 역할/미션, 로드맵, 기술적 구성/특징, 참여 규칙, 전략 등 매우 상세한 정보까지 설문/체크리스트를 통해 파악하였고 대표자들의 세부적인 발표 및 질의응답 시간을 통해 상세한 검토가 진행되었다. 이러한 과정을 거쳐 정리된 모델은 RDA 의 최종 승인을 위해 2023 년 11 월에 제출될 예정이다.

이러한 모델 개발은 글로벌 차원의 오픈 사이언스 실현하는데 매우 의의를 갖는다고 할 수 있다. 오픈 사이언스가 국가나 지역, 개별 과학 분야 등에 고립되지 않고 장벽을 넘는 개방의 효과를 구체적으로 실현하기 위한 매우 중요한 마일스톤으로써 역할을 한다고 할 수 있다. 특히, 국내의 오픈 사이언스 활동이 제한적으로 진행되고 많은 제약과 저항으로 활성화에 어려움을 겪고 있는 현실을 고려할 때, 글로벌 차원의 모델 개발이 갖는 의미는 더욱 크다고 할 수 있다. 비록 KISTI 가 연구데이터 커먼즈 구축 관련 다양한 연구를 수행해 오고 있지만, 향후 지속적이고 적극적으로 글로벌 표준화에 대응하여 오픈 사이언스 분야의 글로벌 선진국으로써 발돋움 해야할 것이다.

KRDC 를 GORC 의 국제 모델과 비교할 때, 글로벌 커먼즈들과의 상호운용을 위해서는 추가적인 연구와 지속적인 노력이 필요하다. 하지만, 연구 소프트웨어의 용이한 공유 활용이라는 관점에서는 글로벌 커먼즈들과 비교해 KRDC 가 가진 강점이 있어, 이를 강화함과 동시에 국내의 다양한 연구데이터 플랫폼/서

비스들과의 연계를 위한 국내에 특화된 표준화된 체계 구축이 필요하다.

사사

본 연구는 한국과학기술정보연구원 연구데이터와 인프라의 공유·활용 체제 구축(K-23-L01-C03-S01) 사업과 자율도전 연구과제(K-23-L05-C02-S16)의 지원을 받아 수행된 연구임

참고문헌

- [1] Research Data Alliance, <https://www.rd-alliance.org/>
- [2] GORC International Model Working Group, <https://www.rd-alliance.org/groups/gorc-international-model-wg>
- [3] CODATA Global Open Science Cloud Initiative, <https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/>
- [4] The Global Open Research Commons International Model, Version 1.0, <https://www.rd-alliance.org/group/gorc-international-model-wg/outcomes/global-open-research-commons-international-model-version>
- [5] STRATEGIC FRAMEWORK FOR HUMAN CAPACITY DEVELOPMENT IN FISHERIES, <https://www.fao.org/3/y5613e/y5613e08.htm>
- [6] 국가 연구데이터 플랫폼 DataON, <http://dataon.kr>
- [7] 국가 오픈액세스 플랫폼 AccessON, <http://accesson.kisti.re.kr>