

연구 소프트웨어를 위한 FAIR 원칙 적용 방안

조민희^{1*}, 김다솔¹, 송사광¹, 이미경¹, 임형준¹

^{1*}한국과학기술정보연구원 연구데이터공유센터

{mini, dskim, esmallj, jerryis, hjyim}@kisti.re.kr

How to apply the FAIR Principle for Research Software

Minhee Cho^{1*}, Dasol Kim¹, Sa-kwang Song¹, Mikyoung Lee¹, Hyung-Jun Yim¹

¹Research Data Sharing Center, KISTI

요약

연구 소프트웨어는 연구개발과정에서 중요한 역할을 하지만 발견, 접근, 재사용 및 상호 운용성에 어려움이 존재한다. FAIR(Findable, Accessible, Interoperable, Reusable) 원칙은 이러한 문제를 해결하기 위한 지침으로, 연구 소프트웨어의 공유 및 활용을 촉진하는 데 기여할 수 있다. 본 논문에서는 연구 소프트웨어 개발 과정에 FAIR 원칙을 적용하기 위한 연구 소프트웨어 개발 프레임워크를 제안한다.

I. 서론

연구 소프트웨어는 데이터 수집, 분석, 시각화 등 다양한 연구 활동에 사용되는 소프트웨어로 소스코드 파일, 알고리즘, 스크립트, 전산 작업 흐름 및 실행 파일이 포함된다[1]. 2014년 영국 연구 소프트웨어 설문 조사[1]에 따르면 연구자 중 90% 이상이 소프트웨어가 자신의 연구에 중요하다고 인정했으며, 약 70%의 연구자는 소프트웨어 없이는 연구가 불가능하다고 대답했다[2]. 최근 연구 데이터의 중요성이 강조됨에 따라 연구 소프트웨어의 역할도 더욱 중요해지고 있다. 하지만 현재 연구 소프트웨어는 다음과 같은 문제점을 가지고 있다.

- 발견 어려움: 소프트웨어가 적절한 메타데이터 및 식별자를 가지고 있지 않아 검색 및 발견이 어렵다.
- 접근 어려움: 접근 권한이나 권한 부여 메커니즘이 부족하여 연구 소프트웨어를 다운로드하거나 실행하기 어렵다.
- 재사용 어려움: 다른 연구자들이 개발한 연구 소프트웨어들의 문서화 부족, 재사용을 위한 적절한 라이선스 부재 등으로 인해 쉽게 재사용하기 어렵다.
- 상호 운용성 부족: 서로 다른 플랫폼 또는 환경에서 상호 운용되지 않아 연구 결과물을 효율적으로 통합하거나 활용하기 어렵다.

이러한 문제점을 해결하기 위해 FAIR4RS(FAIR for Research Software) 실무 그룹은 연구 소프트웨어 FAIR 원칙을 제안했다[1]. FAIR 원칙은 디지털자산의 발견, 접근, 재사용 및 상호 운용성을 향상시키기 위한 지침이다. FAIR 원칙은 다음 네 가지 요소로 구성된다.

- Findable(찾을 수 있는): 연구 소프트웨어를 쉽게 찾을 수 있도록 해야 한다.
- Accessible(접근 가능한): 연구 소프트웨어를 쉽게 다운로드하고 실행할 수 있도록 해야 한다.
- Interoperable(상호 운용 가능한): 연구 소프트웨어가 다른 연구 소프트웨어와 연동될 수 있도록 해야 한다.

- Reusable(재사용 가능한): 연구 소프트웨어는 라이선스에 기반하여 쉽게 실행 및 재사용할 수 있도록 해야 한다

II. FAIR 원칙 적용

현재 연구 소프트웨어에 대한 FAIR 원칙 적용은 초기 단계에 있다. 일부 연구 기관 및 단체는 연구 소프트웨어에 대한 FAIR 원칙 적용을 위한 정책 및 지침을 개발하고 있지만, 아직까지는 일반화되지 않았다. 이는 단순히 코드 저장소 이상의 기능을 제공하며 연구SW 개발, 배포 및 활용을 효율적이고 체계적으로 관리하도록 돕는다.

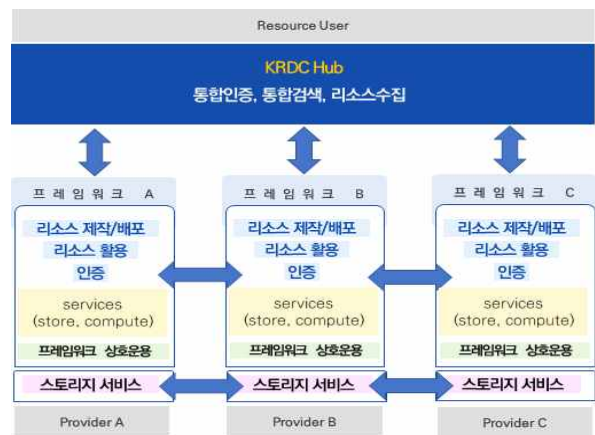


그림 1 연구SW 관리를 위한 아키텍처

본 논문에서는 연구 소프트웨어 FAIR 원칙을 적용하기 위해 KISTI가 개발하고 있는 연구 소프트웨어 개발 프레임워크를 소개한다. 그림 1의 프레임워크는 연구 소프트웨어 개발 과정에 FAIR 원칙을 통합하여 연구 결과의 공개성, 재사용성 및 재현성을 향상시키는 데 목적이 있다. 프레임워크는 다음과 같은 특징이 있다.

첫째, 연구 소프트웨어 발견 및 재사용을 향상하기 위하여 메타데이터 모

델을 정의하고, 관련정보를 표현한다[3]. 연구SW 재현에 필요한 소스 코드, 데이터파일, 메타데이터파일, 실행환경정보 등 다양한 파일들을 체계적으로 구성하고 관리하여 연구 소프트웨어 개발, 배포 및 사용 과정을 지원한다. 연구 소프트웨어를 설명하는 제목, 식별자(DOI), 버전, 제작자, 설명, 키워드, 실행방법, 라이선스 등을 포함하는 표준 메타데이터 체계를 기반으로 연구SW 메타데이터를 일관되게 등록 및 관리하므로 발견성을 높이는데 기여할 수 있다.

둘째 연구소프트웨어 접근성 및 상호운용 향상을 위해 연구SW 개발 및 공유와 관련된 전과정을 지원한다[4]. FAIR 원칙을 적용하여 코드개발을 위한 Jupyter Notebook의 적절한 사용을 보장한다. 노트북으로 개발된 코드들은 GitLab 소스코드 버전관리 저장소를 통해 검색 및 접근이 가능하고, 주피터 노트북을 통하여 메타데이터 교환하거나 API를 통한 상호작용을 통해 다른 주피터 노트북 또는 소프트웨어와 상호운용된다. 또한 코드를 실행하고 재사용이 가능하도록 돕는다. 연구 소프트웨어 제작자는 소프트웨어를 온라인에서 접근할 수 있는 APP 형태로 그림 2와 같이 변환하고, 메타데이터 모델을 기반으로 연구소프트웨어를 등록하여 공유한다.

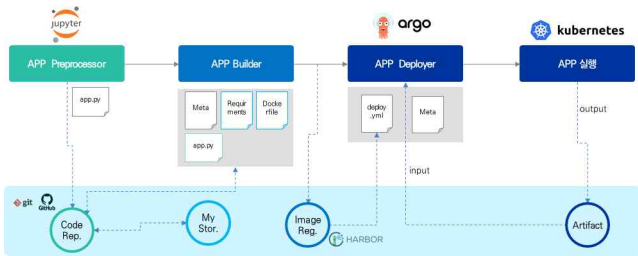


그림 2 연구SW 개발 및 배포 프로세스

셋째 연구소프트웨어를 검색하고 접근하기 위한 기능을 제공한다. 등록된 연구소프트웨어를 검색하고, 온라인에서 실행하고 재현하는 환경을 제공한다. 프레임워크에서 제공하는 검색결과를 클릭하면, ‘워크플로우 바로가기’를 통해 제작자가 등록한 입력 데이터를 참고하여 연구 소프트웨어를 재현하거나 실행할 수 있고, ‘소스코드 바로가기’ 기능을 통해 코드를 재사용할 수 있다.

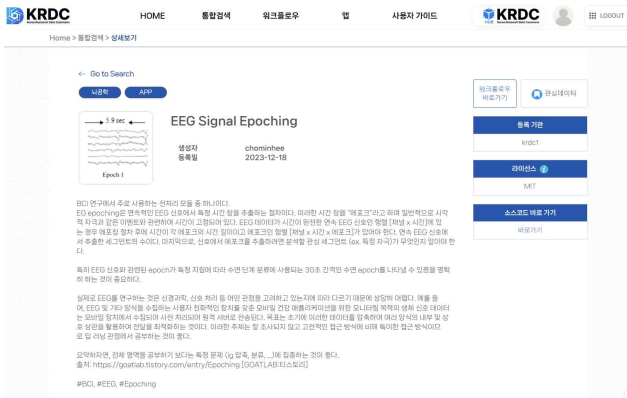


그림 3 연구SW 검색 및 접근 서비스

넷째 연구 소프트웨어 재사용 확대를 위해 온라인에서 쉽게 접근하고 실행하는 환경을 그림 4와 같이 제공한다. 이를 통해 연구자들이 소프트웨어를 쉽게 접근해 상호운용, 재사용 할 수 있다. 워크플로우 기능을 통해 공개된 소프트웨어를 자유롭게 상호 운용할 수 있고, 코드를 온라인으로 즉시 수정하여 새로운 SW로 공유할 수도 있다.

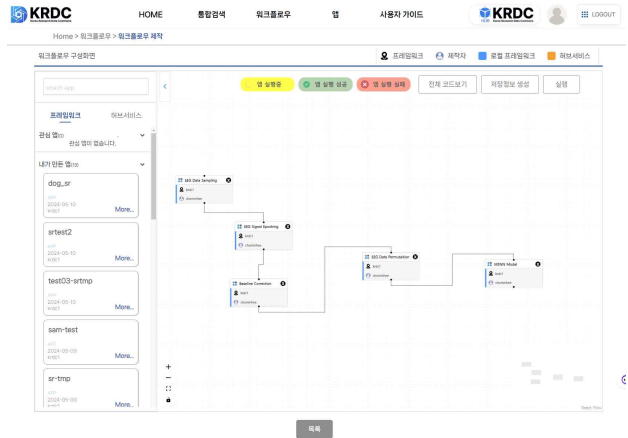


그림 4 연구SW 상호운용을 위한 워크플로우 서비스

III. 결론

본 논문에서는 FAIR 지침을 적용한 연구소프트웨어 개발 프레임워크를 제안했다. 제안된 프레임워크는 연구소프트웨어 개발자들이 FAIR 원칙을 실현하는 데 도움이 될 것으로 기대된다. FAIR 원칙을 준수하는 연구 소프트웨어는 더 쉽게 발견, 접근, 재사용 및 상호 운용될 수 있으므로, 연구 커뮤니티 전체에 혜택을 제공할 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: (KISTI)K24L1M2C3-01)

참고 문헌

- [1] NPC Hong, “FAIR Principles for Research Software (FAIR4RS Principles) 1.0”, 2022.
- [2] Hettrick, S.외, “UK Research Software Survey 2014”, Edinburgh DataShare, 2015.05.28.
- [3] 이미경, 박성은, 이상백, 임형준, “연구소프트웨어 재사용을 위한 메타데이터 스키마 설계”, 한국통신학회 학술대회논문집, 2023.
- [4] 조민희, 송사광, 임형준, 이미경, “연구소프트웨어 공유·활용을 위한 학술 인프라 설계”, 한국경영과학회 학술대회논문집, 2023.