

STEM 분야의 연구 소프트웨어 분석*

- 소프트웨어 공유, 인용, 유형의 관점에서 -

Analyzing STEM Research Software from the Perspectives of Software Sharing, Citation and Types

박 형 주 (Hyoungjoo Park)** 김 서 원 (Seowon Kim)***

김 효 은 (Hyoeyun Kim)**** 최 우 성 (Wooseong Choi)*****

목 차

- | | |
|-----------|------------|
| 1. 서론 | 4. 결과 |
| 2. 선행연구 | 5. 논의 및 결론 |
| 3. 연구 방법론 | |

초록

본 연구의 목적은 STEM(Science, Technology, Engineering, Mathematics) 분야의 리포지토리에서 확인되는 연구 소프트웨어의 공유, 인용, 공유별 인용, 저자, 고유 식별자 및 소프트웨어 유형의 현황을 분석하는 것이다. STEM의 개별 학문 분야와 STEM 분야 전체를 종합하여 살펴보았다. 분석된 STEM 분야는 천체물리학, 생명과학, 컴퓨팅, 지구과학, 수학, 기술이었다. 본 연구는 전세계 453개의 리포지토리에서 색인 및 추적되는 Data Citation Index(DCI)에 저장된 565,000개 이상의 연구 소프트웨어를 분석의 대상으로 하였다. 대부분의 STEM 분야는 소프트웨어 유형을 단순히 'software'로만 명시하고 있었으며, 이 경우 인용 역시 높았다. 예외적으로 컴퓨팅 분야와 기술 분야는 소프트웨어 유형을 상세히 명시하고 있었다. 컴퓨팅 분야는 다양한 소프트웨어 유형이 비교적 균등하게 인용을 받고 있었으나, 기술 분야는 단순히 'software'로 명시한 경우가 대부분의 인용을 받고 있었다. STEM 분야는 소프트웨어 유형을 명시하지 않고 비워 두는 경우도 있었다. STEM 분야에서 소프트웨어의 공유가 주로 이루어지는 상위 2위 리포지토리는 CRAN, Figshare였으나, 소프트웨어 인용이 높은 상위 2위 리포지토리는 CRAN, Bioconductor였다. 연구 소프트웨어는 필수 메타데이터를 대부분 제공하고 있었으나, 저자명과 출판년도 등의 매우 중요한 메타데이터가 누락된 경우도 있었다. 대부분의 소프트웨어는 기관 저자가 아닌 개별 연구자가 공동저자로 명시되어 있었다. 본 연구의 공헌은 전세계 453개의 리포지토리에서 색인되는 STEM 분야의 연구 소프트웨어의 공유 및 공식적인 인용 현황을 개별 STEM 학문 분야와 이를 종합하여 상세히 살펴보았다는 점이다.

ABSTRACT

The purpose of this study is to examine the current status of research software identified in STEM (Science, Technology, Engineering, Mathematics) repositories, specifically focusing on sharing, citations, citations based on sharing, authors, unique identifiers, and software types. This study examines both individual disciplines and the STEM field as a whole. The specific disciplines analyzed include astrophysics, biological sciences, computing, earth sciences, mathematics, and technology. The analysis was conducted on over 565,000 research software records indexed in the Data Citation Index (DCI), which indexes and tracks data from 453 repositories worldwide. The findings reveal that most STEM disciplines designate software types simply as "software," and these records tended to have high citation rates. Exceptionally, the fields of computing and technology provided more detailed specifications for software types. In the computing field, various software types received citations relatively evenly; however, in the technology field, those simply labeled as "software" accounted for the majority of citations. In some instances across STEM fields, the software type was left blank. Regarding distribution, CRAN and Figshare were identified as the top two repositories for software sharing in STEM, whereas CRAN and Bioconductor emerged as the top two repositories for software citations. While most research software provided essential metadata, some critical information—such as author name and publication year—was occasionally missing. The majority of software listed individual researchers as co-authors rather than institutional authors. The contribution of this study lies in its comprehensive and detailed examination of the formal practices of research software across more than 453 global repositories, offering insights into both specific academic disciplines and the STEM field at large.

키워드: 연구 소프트웨어, 연구데이터, Data Citation Index, STEM 분야
Research Software, Research Data, Data Citation Index, STEM Fields

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2022-NR068754).

** 충남대학교 문헌정보학과 부교수(hyoungjoo.park@cnu.ac.kr / ISNI 0000 0004 6442 7767)
(제1저자, 교신저자)

*** 충남대학교 문헌정보학과 학부생(kimseowon02@o.cnu.ac.kr / ISNI 0000 0005 2972 6750) (공동저자)

**** 충남대학교 문헌정보학과 학부생(yhks1006@o.cnu.ac.kr / ISNI 0000 0005 2972 6769) (공동저자)

***** 충남대학교 문헌정보학과 석사과정생(chldntjd375@o.cnu.ac.kr / ISNI 0000 0005 2972 7155) (공동저자)

논문접수일자: 2026년 1월 21일 최초심사일자: 2026년 1월 28일 게재확정일자: 2026년 2월 13일

한국문헌정보학회지, 60(1): 373-397, 2026. <http://dx.doi.org/10.4275/KSLIS.2026.60.1.373>

© Copyright © 2026 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

연구 소프트웨어(research software)는 기초적인 알고리즘, 소프트웨어 자체, 각각의 연구분야에 있어 연구 목적으로 개발된 스크립트, 워크플로우(workflow)이다(Goth et al., 2025). 연구 소프트웨어는 연구데이터의 한 종류로 분류되기도 하지만(Marcus & Menzies, 2010), 연구데이터와 달리 실행 가능한 창작물이다(Katz et al., 2016). 따라서 연구에서 수행하는 역할도 연구데이터와는 차이가 있다(Katz et al., 2016). 연구 소프트웨어는 연구데이터와 마찬가지로 개별적으로 인용 가능한 객체로 인식되어야 한다(Smith et al., 2016). Force11 Software Citation Working Group은 소프트웨어 인용의 원칙으로 연구의 중요성, 공로와 기여, 고유 식별, 지속성, 접근 가능성, 구성성의 6가지를 제시하고 있다(Smith et al., 2016). 연구 소프트웨어 인용은 참고문헌에 인용되거나, 본문에 저자명 등을 괄호 안에 언급하거나 본문이나 각주에 URL(Universal Resource Locator)을 제공하는 방식이 쓰이고 있다(Howison & Bullard, 2016).

STEM(Science, technology, engineering, mathematics) 분야에 사용되는 연구 소프트웨어는 과학 소프트웨어(scientific software)라고 할 수 있으며, 대규모 계산 요소를 포함하며, 의사결정을 지원하기 위한 데이터를 제공하는 소프트웨어로 정의할 수 있다(Kreyman et al., 2001). 과학 분야의 소프트웨어의 유형에 대하여 Kelly et al.(2011)는 과학적 목표를 달성하기 위한 사용자 응용 소프트웨어, 연구자가 자신의 모형을 코드로 표현하고 소프트웨어화

하여 실행할 수 있게 한 도구들로 분류하고 있다고 하였다. Carver et al.(2011)는 연구 논문 작성·출판에 활용하기 위해 개발된 소프트웨어와, 연구자나 일반 사용자의 실제 업무·서비스 환경에서 활용되는 소프트웨어로 분류하고 있다고 하였다. 연구 소프트웨어는 전 처리 단계에서부터, 대규모 데이터 처리 등 학술 작업의 모든 단계에서 사용된다(Howison et al., 2015). 예를 들어, 화학 분야의 소프트웨어는 다양한 형태를 가지는 원자와 분자의 구조와 같은 정보를 기술하고 전달하기 위한 목적으로 사용된다(Dubbeldam et al., 2019). 생명과학 분야에서는 MRI(Magnetic Resonance Imaging)와 같은 다양한 기술로부터 얻어진 데이터나 단백질 구조의 예측과 같은 방대한 데이터 세트를 분석하기 위해 소프트웨어가 사용된다(Nature Methods, 2021). 학문별로 요구사항에 맞는 소프트웨어를 개발하여 사용하는 것도 일반적이다(Bajraktari et al., 2024).

DCI(Data Citation Index)는 Clarivate Analytics에서 운영하는 Web of Science(WoS)의 데이터베이스 중의 하나이며, 엄선된 학술 자원을 선별하여 구축되었기 때문에 신뢰도가 높다고 볼 수 있다. 현재 DCI에는 전세계 453개의 데이터 리포지토리에서 수집된 소프트웨어, 데이터 세트, 데이터 스테디, 리포지토리의 카테고리로 분류된 약 17,265,000개 이상의 레코드가 색인되어 있다(Clarivate Analytics, 2026). 2026년 1월 현재 DCI는 전세계 453개의 리포지토리에서 수집된 565,000개 이상의 연구 소프트웨어를 색인하고 있다. 따라서, DCI는 연구 소프트웨어의 공유 및 인용에 대하여 포괄적으로 살펴볼 수 있는 도구를 제공한다고 볼 수 있다. 본

연구는 STEM 분야로 연구 범위를 한정하였는데, 문맥(context)이 중요하고 사람(human)을 대상으로 한 연구가 주로 행해지는 사회과학 분야와는 특징이 다르기에 별도로 살펴볼 필요가 있기 때문이었다. STEM 분야는 인문 사회분야에 비해 데이터의 재이용 의도가 두드러지는 특징이 있다(Yang & Colavizza, 2025). 특히, 연구 소프트웨어에 대한 기존의 연구는 일반적인 특성 위주의 연구가 주로 진행되어 왔다. 하지만, 학문, 학과, 대학, 국가 간의 공동연구가 행해지는 현대 연구의 특성상, STEM 분야의 개별 학문 분야와 이를 종합적으로 살펴보는 구체적인 연구가 필요함에도 불구하고, 이에 대한 연구는 활발히 진행되지 않아왔다. STEM 분야는 자연과학(hard science) 및 응용과학(applied science) 분야로 학문별로 고유한 특징이 뚜렷하므로 개별 학문 분야를 별도로 상세히 살펴볼 필요가 있다. 본 연구의 목적은 STEM 분야의 리포지토리에서 확인되는 연구 소프트웨어의 공유, 인용, 공유별 인용, 저자, 고유식별자 및 소프트웨어 유형의 현황을 분석하는 것이다. 연구문제는 다음과 같다.

- 연구 문제 1: STEM 분야에서 소프트웨어 인용 및 이용이 높은 연구 소프트웨어의 유형은 무엇인가?
- 연구 문제 2: STEM 분야 리포지토리에서의 연구 소프트웨어의 공유, 인용, 공유별 인용, 기관 저자, 고유식별자의 현황은 어떠한가?

2. 선행연구

연구데이터는 연구 과정에서 실험, 관찰, 조사, 분석 등을 통해 생성된 사실 기반의 원자료 및 2차 자료로서 연구 결과의 검증과 재현을 가능하게 하는 핵심 자원으로 정의된다(국가연구개발혁신법, 법률 제18645호). 연구 소프트웨어는 연구팀이 수행한 연구의 결과를 만들어내는 코드와 관련 구성요소의 집합으로 논문이나 학술 산출물에 보고된 연구 성과와 직접적으로 연결되는 연구 산출물이다(Gomez-Diaz & Recio, 2022). 연구 소프트웨어는 DOI 부여, 공식 출판 및 재이용이 가능한 독립적인 연구 산출물로서 다른 연구를 뒷받침하는 학술 자원으로도 기능할 수 있다(Carlin et al., 2023). 연구 소프트웨어는 지식 생산에 기여한 지적 결과물로서 데이터와 마찬가지로 학술적으로 식별되고 인용되며 관리·평가되어야 할 연구 성과이다(Soito & Hwang, 2016). 연구 결과의 재현을 위해서는 연구데이터뿐만 아니라 해당 데이터를 생성·처리·분석한 연구 소프트웨어와 계산 과정 역시 함께 공개되어야 한다(Hasselbring et al., 2020). Stall et al.(2023)은 연구 소프트웨어가 더 이상 단순한 보조 도구가 아닌 데이터와 동일한 수준의 공개성과 투명성을 요구받는 핵심 연구성과이자 연구 자원으로 인식되고 있다고 주장하였다.

STEM 분야의 연구 소프트웨어는 이론을 수치적인 예측으로 변환하고, 실험과 데이터를 제어하고 해석함으로써 과학적인 증거를 생성하는 핵심 연구 도구이다(Taschuk & Wilson, 2017). 연구 소프트웨어는 연구 전 과정에서 데이터 생성, 분석, 모델 실행, 계산, 구성요소 통합, 연구

협업 지원을 담당하는 핵심 연구 인프라로 기능한다(Chue Hong et al., 2025). STEM 분야 연구는 컴퓨터 모델링과 시뮬레이션을 통해 생성되는 계산 데이터가 핵심 자료로 활용되며, 연구 소프트웨어의 활용이 필수적이다(김지현, 2012). STEM 분야 연구데이터는 대규모 장비와 장시간의 연구 과정을 통해 생산되는 경우가 많아 생성 비용이 크고, 장기적인 보존 가치가 높다(김주섭 외, 2022). STEM 분야에서는 해양 관측 데이터, 기후 데이터와 같은 대규모 과학 데이터를 분석하기 위한 연구 소프트웨어 기반의 연구가 활발히 이루어진다(Hasselbring et al., 2020). Kuckertz et al.(2023)은 연구 소프트웨어 재이용의 저해 요인을 소프트웨어의 발견 가능성과 호환성 부족에서 찾고, 소프트웨어의 인터페이스를 기계 관독이 가능한 메타데이터로 기술하는 DataDesc 메타데이터를 활용하여 소프트웨어의 FAIR성을 향상시키고자 하였다.

소프트웨어의 메타데이터에 대한 연구들은, 논문 등 비정형 문서 및 소프트웨어로부터의 메타데이터 자동 추출에 대한 연구가 진행되었다. Gil et al.(2015)은 OntoSoft 온톨로지를 기반으로 과학 소프트웨어의 메타데이터를 체계적으로 기술하기 위한 OntoSoft 소프트웨어 레지스트리를 설계하였다. Kelley와 Garijo(2021)는 README 파일을 중심으로 과학 소프트웨어에 대한 메타데이터를 자동 추출하며, 총 23개의 소프트웨어 메타데이터의 범주를 구조화된 형태로 수집할 수 있도록 과학 소프트웨어 메타데이터를 자동으로 추출하기 위한 프레임워크인 SOMEF(SOFTWARE Metadata Extraction Framework)를 제안하였다. 반면, Mao et al.

(2019)는 소프트웨어 문서로부터 설명, 설치, 실행, 인용과 같은 핵심 소프트웨어 메타데이터를 자동으로 추출하기 위한 언어학적 기반 접근법을 제안하였다.

연구 소프트웨어의 FAIR 원칙은 FAIR4RS (FAIR for Research Software) 원칙으로 불리우며 소프트웨어의 관리, 접근성, 상호운용성, 재사용성 등을 체계적으로 제고하기 위한 원칙이다. FAIR4RS 원칙은 기존 FAIR 원칙을 연구 소프트웨어의 특성에 맞게 확장·적용하여, 검색 가능성, 접근성, 상호운용성, 재사용성을 체계적으로 제고하기 위해 수립된 원칙이다(Barker et al., 2022). Katz et al.(2021)은 연구 소프트웨어와 연구데이터의 차이점은 실행이 가능한지 여부에 있다고 규정하고, 기존 FAIR 원칙을 연구 소프트웨어의 특성에 맞게 재해석할 필요성을 제기하며, 맥락에 부합하는 FAIR 원칙을 확장·보완하였다. Druskat et al.(2022)은 FAIR 원칙에 따른 연구 소프트웨어가 연구 성과물로 인정되는 공식 출판물이어야 한다고 지적하며, 풍부한 소프트웨어 메타데이터를 자동으로 수집·검증·출판하는 HERMES를 제안하였다. Barker et al.(2022)은 FAIR4RS 원칙을 통해 연구 소프트웨어 관리 및 재이용을 위한 공통의 원칙적 틀을 제시하였다. Martín del Pico et al.(2025a)은 여러 리포지토리에 분산된 연구 소프트웨어의 메타데이터를 통합·정규화하여 FAIR 원칙에 따른 자동 평가와 관리가 가능한 Software Observatory를 제안하였다. Martín del Pico et al.(2025b)는 대형 언어 모델을 활용한 소프트웨어 메타데이터의 정체성 해소가 FAIR에 부합하는 연구 소프트웨어의 메타데이터 저장소(OpenEBench Software

Observatory) 구축에 효과적이고 확장 가능하다고 하였다.

공식적인 연구 소프트웨어의 인용(formal software citation)은 아직 과도기적 단계에 있다. Park과 Wolfram(2019)은 과학 분야의 경우 공식적인 연구 소프트웨어 인용은 아직 학술커뮤니케이션의 관행이 아니라고 하였다. Li et al.(2016)은 현재의 소프트웨어 인용의 관행은 소프트웨어의 발견 가능성을 제한하며, 결과적으로 연구 소프트웨어의 재이용과 학문적 인용을 저해할 수 있다고 하였다. Li et al.(2017)은 PLoS(Public Library of Science) 논문을 분석하여 R 소프트웨어와 패키지, 함수의 인용 방식이 전반적으로 비일관적이며 공식 인용 지침이 충분히 준수되지 않고 있음을 밝혔다. 더 나아가, 향후의 소프트웨어 인용 표준은 서로 다른 소프트웨어 개체 수준 간의 관계 또한 고려해야 한다고 하였다. 연구자들은 공식적(formal), 비공식적(informal) 소프트웨어 인용을 병행하고 있어서 연구 소프트웨어의 인용 관행은 아직 표준화되지 않은 과도기적 단계에 놓여 있다 (박형주 & Wolfram, 2021).

선행연구를 요약하면, 연구 소프트웨어는 연구데이터와는 실행 가능해야 한다는 점에서는 다른 특징이 있지만, 연구데이터와 연구 소프트웨어에 대한 범주는 아직 논의 중이다. 연구 소프트웨어는 메타데이터 및 FAIR 원칙에 따른 연구가 진행되어 왔으며, 주로 개별 학문 단위로 연구되어 왔다. 연구 소프트웨어의 FAIR 원칙인 FAIR4RS는 평가 및 관리에 대한 연구가 행해졌다. 소프트웨어의 메타데이터에 대한 연구는 리포지토리 단위의 분석 연구가 주로 행해져 왔다. 하지만, STEM 개별 학문 분

야 및 이를 종합 정리하는 구체적인 연구는 활발히 진행되지 않아 왔음을 확인할 수 있었다. 리포지토리 별 소프트웨어의 인용 및 이용에 대한 연구와, 저자, 고유식별자, 메타데이터 요소 등에 대한 구체적인 연구는 미흡한 실정이다.

3. 연구 방법론

본 연구의 목적은 STEM 분야의 리포지토리에서 확인되는 연구 소프트웨어의 공유, 인용, 공유별 인용, 저자, 고유식별자 및 소프트웨어 유형의 현황을 분석하는 것이다. 모집단은 Clarivate Analytics사의 DCI에서 전세계 453개의 리포지토리에서 색인하는 연구 소프트웨어이다. DCI는 WoS의 데이터베이스 중의 하나로 구독에 기반한 서비스를 제공한다. DCI는 453개의 리포지토리에서 565,000개 이상의 소프트웨어를 색인한다(Clarivate Analytics, 2026). DCI는 연구 소프트웨어를 별도로 색인하는 구독에 기반한 유일한 학술 데이터베이스이다. DCI에 색인된 연구 소프트웨어를 살펴보는 이점은 공식적인 소프트웨어 인용(formal software citation)을 살펴볼 수 있다는 점이다. 본 연구는 연구 분야를 STEM 분야로 한정하였는데, 인간(human) 연구가 주로 행해지고 문맥(context)이 중요한 사회과학 분야와는 특징이 다르므로, STEM 분야를 별도로 연구할 필요가 있기 때문이었다.

〈표 1〉은 본 연구에서 식별된 STEM 학문 분야와 분류 과정을 보여준다. STEM 분야를 식별하기 위하여, 미국 국립과학재단(National

〈표 1〉 국립과학재단의 학문분야와 DCI의 연구분야 비교(박형주(2025), 재인용)

| 식별된 STEM 분야 | 국립과학재단의 학문코드 | DCI의 연구분야 |
|-------------------------------|--------------------------------|---|
| 천체물리학 (Astrophysics) | Astronomy(천문학) Physics(물리학) | Astronomy & Astrophysics(천문학 및 천체물리학), Physics(물리학), Spectroscopy(분광학) |
| 생명과학 (Biological sciences) | Biological sciences (생명과학) | Genetics and Heredity(유전학 및 유전), Biochemistry & Molecular Biology(생화학 및 분자생물학), Biotechnology & Applied Microbiology(생명공학 및 응용미생물학), Cell Biology(세포생물학), Developmental Biology(발달생물학), Evolutionary Biology(진화생물학), Marine & Freshwater Biology(해양 및 담수생물학), Mathematical & Computational Biology(수학 및 계산생물학), Microbiology(미생물학), Plant Sciences(식물과학), Reproductive Biology(생식생물학), Environmental Sciences & Ecology(환경과학 및 생태학), Biodiversity & Conservation(생물다양성 및 보존), Research & Experimental Medicine(연구 및 실험의학) |
| 화학(Chemistry) | Chemistry(화학) | Chemistry(화학), Crystallography(결정학) |
| 컴퓨팅(Computing) | Computing(컴퓨팅) | Computer Science(컴퓨터 과학) |
| 지구과학 (Earth sciences) | Earth sciences (지구과학) | Geology(지질학), Oceanography(해양학), Geochemistry & Geophysics(지구화학 및 지구물리학), Meteorology & Atmospheric Sciences(기상학 및 대기과학), Water Resources(수자원) |
| 공학(Engineering) | Engineering(공학) | Engineering(공학) |
| 수학(Mathematics) | Mathematical sciences (수학) | Mathematics(수학) |
| 기술(Technology) | - | Technology(기술) |

Science Foundation)의 학문 코드(discipline code), DCI의 연구분야(research area)를 기준으로 하였다. 천문학과 물리학을 천체물리학으로 묶은 이유는, 해외의 많은 대학들이 두 학문 분야를 동일한 단과 대학에 두는 경우가 많기 때문이었다. 기술(technology)이 국립과학재단의 학문 코드에는 존재하지 않지만 추가한 이유는 STEM이라는 단어 자체가 기술(technology)을 포함하고 있기 때문이었다. STEM 분야를 식별한 후, DCI의 연구분야(research area)를 식별하였는데, 구체적으로 박형주(2025)의 연구에서 식별한 STEM 분야와 DCI의 연구분야를 활용하였다.

본 연구는 DCI에서 문서 유형(document type)이 소프트웨어로 분류된 레코드만을 수집의 대

상으로 하였다. DCI는 하나의 검색어 당 최대 10만개의 레코드를 다운로드 받을 수 있는 기능을 제공한다. 공식적인 연구 소프트웨어의 인용은 학술 커뮤니케이션의 관행이 아니므로 (Park & Wolfram, 2019), 연구 소프트웨어의 인용의 현황을 포괄적으로 살펴보기 위하여 인용이 높은 순으로 레코드를 다운로드 받은 후 마이크로소프트 엑셀에 저장하였다. 구체적으로 DCI의 'most highly cited' 순으로 정렬하여 레코드를 다운로드 받았다. DCI의 레코드는 총 4개의 카테고리로 분류되어 있는데, 구체적으로 소프트웨어, 데이터 세트, 데이터 스터디, 리포지토리이다. DCI는 소프트웨어만을 별도로 다운로드 받는 기능을 제공하지 않는다. 따라서 소프트웨어로 분류된 레코드만을 식별하기

위하여, 최대 10만개씩 다운로드 받은 STEM 학문 분야별 레코드의 문서 유형을 ‘소프트웨어’로 한정된 후, 이를 별도의 파일로 마이크로소프트 엑셀에 저장하였다. 구체적으로 필드 태그(field tag)가 DT인 열에서 ‘software’로 분류된 레코드만을 식별하였다.

DCI에서 다운로드 받은 STEM 분야 레코드 중에서, 소프트웨어로 분류된 레코드가 0(zero)인 학문 분야인 공학(engineering) 분야와 화학(chemistry) 분야는, 레코드가 없어서 분석의 대상에서 제외하였다. 최종적으로 총 6개의 STEM 학문 분야인 천체물리학, 생명과학, 컴퓨팅, 지구과학, 수학, 기술 분야가 분석의 대상이 되었다. <표 2>는 본 연구에 활용된 STEM 학문 분야별 소프트웨어의 총 레코드 수와 총 인용 수를 보여준다.

‘리포지토리’는 DCI에서 추적 및 색인하는 연구 소프트웨어가 공유되어 있는 리포지토리이다. ‘소프트웨어 인용’은 DCI에서 색인하는 공식적인 소프트웨어 인용(formal software citation)이다. 리포지토리는 ‘소프트웨어 공유 수’가 많은 순으로 정렬하였다. 학문 분야별 리포지토리를 식별한 기준은, 수집한 레코드에서 ‘소프트웨어 공유수’가 많은 순으로 리포지토리의 순위를 매겼다. 연구 소프트웨어가 공유된 리

포지토리의 수가 많은 학문 분야가 존재하여 공간 부족으로 본 연구에는 상위 10위까지만 표기하였다. ‘기관저자’는 연구 소프트웨어를 공유한 저자가 개인저자가 아닌 기관인 경우이고, DCI의 CA 필드 태그(field tag)로 식별한 후, 기관명을 확인하였다. 여기에서 기관저자는 NASA(National Aeronautics and Space Administration) 등의 저자를 의미한다. ‘소프트웨어 공유 별 소프트웨어 인용’은, 분모는 ‘개별 소프트웨어 인용 수’, 분자는 ‘전체 소프트웨어 인용 수’이며, 리포지토리에서의 소프트웨어 1건당 평균 인용수를 의미한다. ‘소프트웨어 이용’, 즉 usage(이용)은 2013년 이후의 이용수이며, DCI의 U2 필드태그로 식별하였다. ‘소프트웨어 이용률’은 전체 소프트웨어 이용 수 대비 소프트웨어 이용의 백분율이다. 수집된 레코드는 상위 10위까지의 리포지토리만을 살펴보았으므로, 리포지토리, 소프트웨어 유형에 따른 소프트웨어 인용 및 이용의 총합은 리포지토리의 총합 및 인용 및 이용의 총합과는 다를 수 있다. 수집된 연구 소프트웨어의 레코드는 엑셀의 피벗(pivot) 테이블 기능을 활용하여 각 필드를 필터링하고, 행과 값을 구하였다.

<표 2> 분석의 대상이 된 STEM 학문 분야의 연구 소프트웨어 레코드 및 인용 수

| 분석의 대상이 된 STEM 분야 | 소프트웨어의 총 레코드 수 | 소프트웨어의 총 인용 수 |
|----------------------|----------------|---------------|
| 천체물리학(astrophysics) | 1,219 | 627 |
| 생명과학(biology) | 174 | 1,394 |
| 컴퓨팅(computing) | 4,748 | 916 |
| 지구과학(earth sciences) | 269 | 420 |
| 수학(mathematics) | 24,272 | 140,593 |
| 기술(technology) | 29,518 | 202,816 |

4. 결과

결과 섹션은 천체물리학, 생명과학, 컴퓨팅, 지구과학, 수학, 기술을 개별로 살펴본 후, STEM 분야를 종합 정리한 결과를 보여준다.

4.1 천체물리학 분야

〈표 3〉은 천체물리학 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 공유 별 인용, 기관저자, 고유식별자를 비교한 표이다. 고유식별자를 살펴본 결과, 모두 DOI였다. 이는 5년전의 박형주와 Wolfram(2021)의 연구에서 천체물리학 분야의 고유식별자의 대부분은 URL이고, DOI 사용은 없었다는 것과는 다른 결과이다. 연구 소프트웨어의 고유식별자 관계가 변화되었다고 해석될 수 있다. 천체물리학 분야에서의 연구 소프트웨어 공유는 학문 특화 리포지토리인 Astrophysics Source Code Library (ASCL)에서 전체 레코드의 99.75%가 공유되고 있었다. 대부분의 소프트웨어는 개인이 공동

으로 연구한 공동저자였고 기관저자는 드물었다.

〈표 4〉는 천체물리학 분야의 데이터 공유별 인용수가 높은 연구 소프트웨어의 인용 및 이용 현황을 보여준다. 천체물리학 분야에서 수집된 레코드 수는 총 1,219개였다. 천체물리학 분야에서의 'software'는 천문 데이터를 처리하거나 우주 현상을 모델링하는 프로그램의 원시 코드를 의미하며 이를 통해 데이터 분석, 시뮬레이션, 이미지 처리 등을 수행한다(WIKIPEDIA-a, (n.d.)). 천체물리학 분야에서는 소프트웨어 유형을 상세하게 명시하지는 않는 것으로 확인되었다.

4.2 생명과학 분야

〈표 5〉는 생명과학 분야의 리포지토리별 연구 소프트웨어의 공유, 인용, 공유 별 인용, 기관저자, 고유식별자를 비교한 표이다. 생명과학 분야 연구소프트웨어의 대부분을 차지하고 있는 Bioconductor는 생물정보학용 통계 소프트웨어와 분석 도구의 R 패키지를 개발·배포하는 오픈 소스 프로젝트로, 유전체 및 고처리량

〈표 3〉 천체물리학 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용수 | 기관 저자 | 고유 식별자 |
|----|--|------------|------------|----------------|-------|--------|
| 1 | Astrophysics Source Code Library | 1,216 | 627 | 0.52 | 20 | 0 |
| 2 | eData: the STFC Research Data Repository | 3 | 0 | 0 | 0 | 3 |

〈표 4〉 천체물리학 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|---|----------|--------|----------|--------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 599 | 95.53% | 16 | 94.12% |
| 2 | software used in astronomy or astrophysics research | 28 | 4.47% | 1 | 5.88% |
| | 총합 | 627 | 100% | 17 | 100% |

〈표 5〉 생명과학 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용수 | 기관 저자 | 고유 식별자 |
|----|-------------------------------|---------------|---------------|-------------------|----------|-----------|
| 1 | Bioconductor | 173 | 1,390 | 8.03 | 10 | 173 |
| 2 | Woods Hole Open Access Server | 1 | 4 | 4.00 | 0 | 0 |

생물학 데이터의 분석을 지원하며, 주요 데이터는 유전체·전사체·유전자 발현 분석을 위한 R 패키지와 대응 메타데이터(Annotation, Experiment 등), high-throughput 생물학 데이터를 처리할 수 있는 도구 모음 등이 있다(Bioconductor, (n.d.)). 요약하면, Bioconductor 리포지토리의 고유식별자를 분석한 결과, 연구 소프트웨어는 모두 DOI를 부여받고 있었다. Bioconductor가 소프트웨어가 공유됨과 동시에 DOI를 부여하는 기능을 제공하고 있기 때문일 수 있다. 기관 저자는 드물었다.

〈표 6〉은 생명과학 분야의 소프트웨어 유형별 소프트웨어의 인용과 이용을 비교한 표이다. 생명과학 분야의 소프트웨어 유형은 상세히 명시하고 있지 않은 것으로 나타났는데, 소프트웨어 유형을 적지 않고 〈비어있음〉으로 제공하는 경우도 2.44%가 있었다.

4.3 컴퓨팅 분야

〈표 7〉은 컴퓨팅 분야의 리포지토리별 연구

소프트웨어 공유, 인용, 공유 별 인용, 기관저자, 고유식별자를 비교한 표이다. Figshare는 연구 과정에서 생성되는 데이터, 이미지, 포스터, 프레젠테이션, 코드 등 다양한 연구 산출물을 저장하고 공유할 수 있는 범용 리포지토리로서, 연구자가 자료를 업로드하면 자료에 고유한 식별자인 DOI를 자동 발급한다(Figshare, 2026). Knowledgebase of Interatomic Models는 재료과학과 계산물리학 분야에서 사용되는 interatomic 모델을 표준화된 방식으로 제공하는데, 분자동역학 시뮬레이션 연구에 특화된 리포지토리이다(National Institute of Standards and Technology, 2020). 요약하면, 컴퓨팅 분야의 소프트웨어는 범용 리포지토리(예: Figshare, Zenodo), 학문 특화 리포지토리(예: Knowledgebase of Interatomic Models, ModelDB), 기관 리포지토리(예: Edinburgh Datashare, Rutgers University Community Repository)에 동시에 공유되고 있었다. 주요 리포지토리는 Figshare, Knowledgebase of Interatomic Models, ModelDB가 있었다. 공동저자는 많았지만, 기관저자는 드

〈표 6〉 생명과학 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|----------|----------|--------|----------|------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 1,360 | 97.56% | 7 | 100% |
| 2 | (비어 있음) | 34 | 2.44% | 0 | 0% |
| | 총합 | 1,394 | 100% | 7 | 100% |

〈표 7〉 컴퓨팅 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용 수 | 기관 저자 | 고유 식별자 |
|----|---|------------|------------|-----------------|-------|--------|
| 1 | Figshare | 2,086 | 84 | 0.04 | 14 | 2,064 |
| 2 | Knowledgebase of Interatomic Models | 1,218 | 153 | 0.13 | 0 | 1,218 |
| 3 | ModelDB | 1,052 | 610 | 0.58 | 0 | 3 |
| 4 | Zenodo | 347 | 28 | 0.08 | 37 | 347 |
| 5 | Mantid Project | 41 | 41 | 1.00 | 0 | 41 |
| 6 | Rutgers University Community Repository | 2 | 0 | 0.00 | 0 | 2 |
| 7 | Edinburgh DataShare | 1 | 0 | 0.00 | 0 | 1 |
| 8 | Scholars Bank | 1 | 0 | 0.00 | 0 | 1 |

물었다. 컴퓨팅 분야는 ModelDB를 제외한 모든 리포지토리가 고유식별자를 발급하고 있었다.

〈표 8〉은 컴퓨팅 분야의 소프트웨어 유형 별 소프트웨어의 인용과 이용을 비교한 표이다. 컴퓨팅 분야의 소프트웨어 인용의 총 수는 916 이고, 소프트웨어 이용의 총 수는 77인데, 〈표 8〉은 분석에 사용된 소프트웨어의 전체 수(total numbers)인 100%가 아니라, 상위 10위까지의 인용 및 이용의 횟수만을 표기하였으므로, 총

합의 퍼센트는 76.31%로 100%가 아니다. 1위를 차지한 ‘neuron or other electrically excitable cell: computer model: software’는 컴퓨터 모델로서 단일 뉴런 혹은 전기적으로 흥분 가능한 세포 수준의 미시적 구조를 다루며, 실제 생물학적 신경 세포의 전기적 특성을 반영한다(neuron, (n.d.)). 4위인 ‘computer model: software’는 물리적, 이론적 시스템을 컴퓨터 프로그램으로 시뮬레이션하는 모델을 의미하며, 기상예측, 구

〈표 8〉 컴퓨팅 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|---|----------|--------|----------|--------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | neuron or other electrically excitable cell: computer model: software | 180 | 19.65% | 2 | 2.60% |
| 2 | software | 152 | 16.59% | 19 | 24.68% |
| 3 | realistic network: computer model: software | 104 | 11.35% | 1 | 1.30% |
| 4 | computer model: software | 80 | 8.73% | 3 | 3.90% |
| 5 | code: software | 58 | 6.33% | 21 | 27.27% |
| 6 | software: model | 43 | 4.69% | 6 | 7.79% |
| 7 | synapse: computer model: software | 25 | 2.73% | 1 | 1.30% |
| 8 | (비어 있음) | 22 | 2.40% | 2 | 2.60% |
| 9 | Code | 20 | 2.18% | 11 | 14.29% |
| 10 | Connectionist Network: Computer Model: Software | 15 | 1.64% | 0 | 0.00% |
| | 총합 | 699 | 76.31% | 66 | 85.71% |

조물 해석, 생물학적 시스템 등 광범위한 분야에서 사용된다(WIKIPEDIA-b, (n.d.)). 요약하면, 대부분의 STEM 학문 분야가 소프트웨어 유형을 'software'로 단순히 적는 것에 비하여, 컴퓨팅 분야는 상세하게 소프트웨어 유형을 제공하고 있었는데, 이는 컴퓨팅 분야와 뒤에 기술한 기술 분야만의 관례로 해석될 수 있다. 또한, 소프트웨어 유형이 다양하고, 특정한 유형이 인용률이 월등히 높다고 보다는 다양한 유형이 골고루 인용을 받고 있다고 해석될 수 있다.

4.4 지구과학 분야

〈표 9〉는 지구과학 분야의 리포지토리별 연구 소프트웨어의 공유, 인용, 공유 별 인용, 기관저자, 고유 식별자를 보여준다. 요약하면, 지구과학 분야는 대부분 개인의 공동저자가 주를

이루며, 기관 저자는 드물었다. 지구과학 분야 리포지토리에서의 평균적인 소프트웨어 공유 별 인용 수는 1.98점이었으며, Woods Hole Open Access Server가 4점으로 가장 높았다. 지구과학 분야에서 소프트웨어 공유는 269건으로 타 학문분야에 비하여 낮았다.

〈표 10〉은 지구과학 분야의 소프트웨어 유형 별 소프트웨어의 인용과 이용을 비교한 표이다. 지구과학 분야는 소프트웨어의 유형이 'software'로 단순히 적는 경우가 인용(368회, 87.62%), 이용(8회, 88.89%)로 높았다.

4.5 수학 분야

〈표 11〉은 수학 분야의 리포지토리별 연구 소프트웨어의 공유, 인용, 공유별 인용, 기관저자, 고유식별자를 보여준다. CRAN은 통계 연

〈표 9〉 지구과학 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유별 인용수 | 기관 저자 | 고유 식별자 |
|----|--|------------|------------|---------------|-------|--------|
| 1 | United States Geological Survey | 261 | 398 | 1.52 | 1 | 261 |
| 2 | Computational and Information Systems Laboratory Research Data Archive | 5 | 17 | 3.40 | 5 | 5 |
| 3 | CSIRO Data Access Portal | 1 | 0 | 0.00 | 0 | 1 |
| 4 | Woods Hole Open Access Server | 1 | 4 | 4.00 | 0 | 0 |
| 5 | World Data Centre for Climate | 1 | 1 | 1.00 | 5 | 1 |

〈표 10〉 지구과학 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|-----|----------------------|----------|--------|----------|--------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 368 | 87.62% | 8 | 88.89% |
| 2 | software source code | 1 | 0.24% | 0 | 0.00% |
| 3 | (비어 있음) | 51 | 12.14% | 1 | 11.11% |
| 총합계 | | 420 | 100% | 9 | 100% |

〈표 11〉 수학 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용수 | 기관 저자 | 고유 식별자 |
|----|---------------------------------------|------------|------------|----------------|-------|--------|
| 1 | Comprehensive R Archive Network(CRAN) | 24,272 | 140,593 | 5.79 | 30 | 0 |

어인 R의 공식 소프트웨어 배포 리포지토리로 통계 분석 도구를 제공하며, 전산 통계, 데이터 사이언스 연구자, R 개발자가 주로 사용한다. 주로 R 기본 시스템, 약 2만개 이상의 분류 학습, 시계열, 시뮬레이션 등을 포함한 R 패키지 소스 및 각 운용체제별 바이너리를 제공한다(R Core Team, (n.d.)). 요약하면, 수학 분야는 고유식별자를 제공하는 소프트웨어가 0%였는데, 이는 CRAN이 소프트웨어에 고유식별자를 제공하는 기능이 없기 때문일 수 있다. 수학 분야의 연구 소프트웨어는 100%가 범용 레포지토리인 CRAN에 공유되고 있었다. 기관저자는 드물었다.

〈표 12〉는 수학 분야의 소프트웨어 유형 별 소프트웨어의 인용과 이용을 비교한 표이다. 수학 분야는 모든 레코드가 'software'로 유형이 100% 동일하였는데, 이는 STEM 분야 중에서 수학분야에서만 보이는 특징이었다.

4.6 기술 분야

〈표 13〉은 기술 분야의 리포지토리별 연구 소프트웨어의 공유, 인용, 공유 별 인용, 기관저

자, 고유식별자를 보여준다. 기술 분야의 소프트웨어는 총 15개의 리포지토리에 공유되고 있었는데 〈표 13〉은 상위 10위의 리포지토리만 보여준다. Zenodo는 범용 리포지토리로 DOI를 자동으로 발급하여 연구 성과의 인용과 재현성을 지원한다(CERN Data Centre, Invenio RDM, (n.d.)). 요약하면, 기술 분야는 다양한 리포지토리에 소프트웨어가 공유되고 있었다. 범용 리포지토리인 CRAN과 Zenodo에 소프트웨어가 가장 많이 공유되고 있었으며, 클라우드 기반 실행 환경을 제공하는 Code Ocean, 학문 특화 리포지토리(예: nanoHub), 기관 리포지토리(예: CaltechData, Columbia University Academic Commons) 등 다양한 유형의 리포지토리에 소프트웨어가 공유되고 있었다. 기술 분야의 소프트웨어 공유가 상위 10위인 리포지토리 중에서 CRAN을 제외한 모든 리포지토리에서 DOI를 발급하고 있었다. 기술분야에서 소프트웨어 공유 별 인용수는 평균 3.52건 이었으며, CRAN의 소프트웨어 공유 1건당 인용 수는 20.68로 월등히 높았다.

〈표 14〉는 기술 분야의 소프트웨어 유형 별

〈표 12〉 수학 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|----------|----------|------|----------|------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 140,593 | 100% | 1,815 | 100% |
| | 총합계 | 140,593 | 100% | 1,815 | 100% |

〈표 13〉 기술 분야의 리포지토리별 연구 소프트웨어 공유, 인용, 공유 별 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용수 | 기관 저자 | 고유 식별자 |
|----|--|---------------|---------------|-------------------|----------|-----------|
| 1 | Zenodo | 21,993 | 60,843 | 2.77 | 865 | 21,993 |
| 2 | Comprehensive R Archive Network(CRAN) | 6,822 | 141,070 | 20.68 | 8 | 0 |
| 3 | Figshare | 305 | 373 | 1.22 | 2 | 305 |
| 4 | Code Ocean | 239 | 268 | 1.12 | 2 | 239 |
| 5 | nanoHUB | 79 | 131 | 1.66 | 0 | 78 |
| 6 | PUB - Publications at Bielefeld University | 24 | 36 | 1.50 | 0 | 24 |
| 7 | Columbia University Academic Commons | 15 | 29 | 1.93 | 0 | 15 |
| 8 | 4TU.Centre for Research Data | 10 | 10 | 1.00 | 0 | 10 |
| 9 | ETH Data Archive | 8 | 14 | 1.75 | 1 | 8 |
| 10 | EUDAT Collaborative Data Infrastructure | 7 | 11 | 1.57 | 0 | 7 |

〈표 14〉 기술 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|-------------------------|----------|---------|----------|--------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 201,735 | 99.47% | 2,860 | 99.55% |
| 2 | (비어 있음) | 667 | 0.33% | 2 | 0.07% |
| 3 | capsule: software | 224 | 0.11% | 5 | 0.17% |
| 4 | tools | 67 | 0.03% | 1 | 0.03% |
| 5 | capsule | 44 | 0.02% | 3 | 0.10% |
| 6 | compact model: software | 38 | 0.02% | 0 | 0.00% |
| 7 | tools: software | 14 | 0.01% | 2 | 0.07% |
| 8 | compact model | 12 | 0.01% | 0 | 0.00% |
| 9 | bi_dissertation | 3 | 0.00% | 0 | 0.00% |
| 10 | code | 3 | 0.00% | 0 | 0.00% |
| | 총합 | 202,816 | 100.00% | 2,873 | 100% |

소프트웨어의 인용과 이용을 비교한 표이다. 3위를 차지한 'capsule: software'는 연구용 패키지로, 코드, 데이터, 실행 환경(environment)이 하나로 묶여 저장된 형태를 의미한다(Wei et al., 2024). 6위인 'compact model: software'는 회로 시뮬레이터에서 소자의 전류-전압 동작을 빠르게 계산할 수 있도록 고안된 축약 모델이며, 전자 공학 및 집적회로 설계분야에서 소자의 내부 물리 현상을 단순화된 방정식으로 표현하여

회로 시뮬레이션이 가능하도록 하는 모델이다(Peng et al., 2013). 7위인 'tools: software'는 소프트웨어의 개발, 분석, 지원 등의 작업을 돕는 유틸리티 프로그램이다. 요약하면, 기술 분야는 소프트웨어 유형을 다양하게 기록하고 있었다. 기술 분야의 연구자는 소프트웨어의 유형으로 단순히 'software'로 적었을 때 인용을 높게 받았으며, 그 외의 소프트웨어 유형은 인용을 받는 경우가 드물었다.

있지 않다는 점에서 심각한 문제일 수 있다.

〈표 16〉은 STEM 분야 별로 본 연구에서 수집한 레코드의 연구 소프트웨어의 공유 및 인용의 연도별 변화를 비교한 표이다. 연구 소프트웨어의 공유는 최근 들어 증가 추세이기는 하나, 꾸준히 증가하는 것은 아님을 확인하였다. 연구 소프트웨어는 상대적으로 공유 기간이 오래될 수록 인용 수가 높았다. 지구과학 분야의 경우, 2016년부터 2018년까지 소프트웨어 공유 별 인용 수가 25.33회에서 46.5회로 타 학문 분야 및 연도에 비하여 월등히 높았다.

〈표 17〉은 STEM 분야 전체의 리포지토리 별 연구 소프트웨어의 공유, 인용, 공유 별 인용, 기관저자, 고유식별자를 비교한 표이다. STEM 분야 전체의 소프트웨어 공유 수는 37,860개였다. 주요 리포지토리는 범용 리포지토리인 CRAN (82.13%)이었다. 학문 특화 리포지토리에서의 소프트웨어 공유는 드물었다. 본 연구는 STEM 분야의 고유식별자의 제공이 리포지토리 별로 차이를 확인하였는데, 소프트웨어가 리포지토리에 기탁될 때 고유식별자가 자동으로 발급되는 리포지토리와 그렇지 않은 정책을 가진 리포지토리가 있기 때문일 수 있다. STEM 분야 전체를 봤을 때 소프트웨어 공유 별 인용 수는 평균 2.38이었는데, 특히 CRAN이 9.06건, Bioconductor가 8.03으로 평균보다 월등히 높았다. 즉, 연구 소프트웨어의 공유 시에, 인용 등의 공식적인 학술 크레딧을 받고자 한다면 인용 수가 높은 CRAN, BioConductor 등의 리포지토리에 기탁하는 것이, Zenodo 등의 인용 수가 낮은 리포지토리에 기탁하는 것보다 나은 전략일 수 있다. 또한, 본 연구는 범용 리포지토리인 CRAN, 학문 특화 리포지토리인 Bioconductor

의 인용 수가 높음을 확인하였는데, 이는 인용의 관점에서는 범용 리포지토리 혹은 학문 특화 리포지토리 등 리포지토리의 카테고리는 중요하지 않다고 해석될 수 있다.

〈표 18〉은 STEM 분야 전체의 상위 10위의 소프트웨어 유형 별 소프트웨어의 인용과 이용을 비교한 표이다. 소프트웨어의 총 수는 346,766이었고, 소프트웨어 이용 수는 4,798이었다. STEM 분야 전체를 살펴볼 때, 소프트웨어 유형을 단순히 'software'로 적은 경우의 인용률(99.55%)과 이용률(99.02%)이 높았다. 그 외에 소프트웨어 유형을 상세히 적어 놓는 경우, 소프트웨어 인용 수가 적었다.

〈표 19〉는 STEM 학문 분야별로 연구 소프트웨어의 메타데이터 사용 현황을 비교한 표이다. 구체적으로 저자, 제목, 출판년도, 리포지토리, 초록, 고유식별자, 소프트웨어 버전 및 유형 등의 항목을 분석하였다. 분석 결과, 고유식별자를 제공하는 레코드 수는 학문 분야에 따라 뚜렷한 차이를 보였다. 천체물리학과 수학 분야는 고유식별자 관련 메타데이터 제공 빈도가 현저히 낮았던 반면, 생명과학, 지구과학 및 기술 분야는 상대적으로 높았다. 소프트웨어 버전 정보의 경우, 전반적으로 제공 비율이 낮게 나타났다. 이는 다수의 연구 소프트웨어가 지속적인 업데이트 정보를 반영하지 못하고 있기 때문으로 풀이된다. 특히 본 연구에서는 연구 소프트웨어에 부여된 고유식별자의 100%가 DOI라는 점을 확인하였다. 한편, 제목과 리포지토리 명칭은 모든 레코드(100%)에서 제공되고 있었으나, 저자 이름과 출판년도와 같은 필수 메타데이터 요소는 매우 중요한 메타데이터임에도 불구하고 누락된 사례가 일부 확인되었다.

〈표 16〉 수집된 레코드의 연구 소프트웨어 공유 및 인용의 연도별 비교

| 학문 분야 | 천체물리학 | | 생명과학 | | 컴퓨팅 | | 지구과학 | | 수학 | | 기술 | |
|-------|-------|-----------|------|-----------|------|-----------|------|-----------|-------|-----------|-------|-----------|
| | 공유 수 | 공유 별 인용 수 | 공유 수 | 공유 별 인용 수 | 공유 수 | 공유 별 인용 수 | 공유 수 | 공유 별 인용 수 | 공유 수 | 공유 별 인용 수 | 공유 수 | 공유 별 인용 수 |
| 연도 | | | | | | | | | | | | |
| 2013 | 19 | 115 | 0 | 0 | 104 | 99 | 0 | 0 | 282 | 290 | 49 | 324 |
| 2014 | 22 | 29 | 1 | 4 | 497 | 88 | 1 | 13 | 362 | 200 | 92 | 1,373 |
| 2015 | 28 | 85 | 0 | 0 | 510 | 61 | 3 | 13 | 574 | 747 | 289 | 1,627 |
| 2016 | 26 | 90 | 0 | 0 | 568 | 72 | 2 | 93 | 880 | 1,409 | 713 | 2,927 |
| 2017 | 14 | 25 | 138 | 1,188 | 732 | 127 | 3 | 76 | 1,056 | 1,571 | 862 | 3,623 |
| 2018 | 26 | 84 | 13 | 144 | 757 | 207 | 4 | 110 | 1,361 | 1,982 | 1,317 | 5,608 |
| 2019 | 19 | 37 | 5 | 23 | 573 | 66 | 4 | 6 | 1,529 | 4,534 | 2,049 | 9,373 |
| 2020 | 28 | 45 | 0 | 0 | 264 | 102 | 3 | 4 | 2,311 | 7,109 | 3,585 | 14,586 |
| 2021 | 323 | 50 | 10 | 25 | 395 | 53 | 8 | 13 | 2,432 | 5,174 | 4,778 | 22,263 |
| 2022 | 299 | 61 | 7 | 10 | 143 | 31 | 72 | 37 | 3,477 | 23,690 | 5,548 | 32,814 |
| 2023 | 328 | 5 | 0 | 0 | 197 | 10 | 119 | 50 | 5,908 | 48,168 | 6,591 | 59,040 |
| 2024 | 86 | 1 | 0 | 0 | 7 | 0 | 50 | 5 | 4,100 | 45,719 | 3,644 | 49,257 |

〈표 17〉 STEM 분야 전체의 리포지토리별 연구 소프트웨어 공유, 인용, 공유 별 인용, 기관저자, 고유식별자 비교

| 순위 | 리포지토리 | 소프트웨어 공유 수 | 소프트웨어 인용 수 | 소프트웨어 공유 별 인용 수 | 그룹 저자 | 고유 식별자 |
|----|---------------------------------------|------------|------------|-----------------|-------|--------|
| 1 | Comprehensive R Archive Network(CRAN) | 31,094 | 281,663 | 9.06 | 38 | 0 |
| 2 | Figshare | 2,391 | 457 | 0.19 | 16 | 2,369 |
| 3 | Knowledgebase of Interatomic Models | 1,218 | 153 | 0.13 | 0 | 1,218 |
| 4 | Astrophysics Source Code Library | 1,216 | 627 | 0.52 | 20 | 0 |
| 5 | ModelDB | 1,052 | 610 | 0.58 | 0 | 3 |
| 6 | United States Geological Survey | 261 | 398 | 1.52 | 1 | 261 |
| 7 | Code Ocean | 239 | 268 | 1.12 | 2 | 239 |
| 8 | Bioconductor | 173 | 1,390 | 8.03 | 10 | 173 |
| 9 | nanoHUB | 79 | 131 | 1.66 | 0 | 78 |
| 10 | Mantid Project | 41 | 41 | 1 | 0 | 41 |

〈표 18〉 STEM 분야의 소프트웨어 유형 별 인용과 이용 비교

| 순위 | 소프트웨어 유형 | 소프트웨어 인용 | | 소프트웨어 이용 | |
|----|---|----------|--------|----------|--------|
| | | 횟수 | 퍼센트 | 횟수 | 퍼센트 |
| 1 | software | 344,807 | 99.55% | 4,725 | 99.02% |
| 2 | (비어 있음) | 774 | 0.22% | 5 | 0.10% |
| 3 | capsule: software | 224 | 0.06% | 5 | 0.10% |
| 4 | neuron or other electrically excitable cell: computer model: software | 180 | 0.05% | 2 | 0.04% |
| 5 | realistic network: computer model: software | 104 | 0.03% | 1 | 0.02% |
| 6 | computer model: software | 80 | 0.02% | 3 | 0.06% |
| 7 | tools | 67 | 0.02% | 1 | 0.02% |
| 8 | code: software | 58 | 0.02% | 21 | 0.44% |
| 9 | capsule | 44 | 0.01% | 3 | 0.06% |
| 10 | software: model | 43 | 0.01% | 6 | 0.13% |
| | 총합 | 346,381 | 100% | 4,772 | 100% |

〈표 19〉 STEM 학문 분야 별 연구 소프트웨어의 메타데이터 사용 비교

| 학문 분야 (총 레코드 수) | 메타데이터 명칭(레코드 수, 퍼센트) | | | | | | | |
|--------------------|----------------------|-----------------|------------------|-----------------|-------------------|-------------------|----------------|-------------------|
| | 저자 | 제목 | 출판년도 | 리포지토리 | 초록 | 고유식별자 | 버전 정보 | 소프트웨어 유형 |
| 천체물리학 (1219) | 1208 (99.1%) | 1219 (100%) | 1218 (99.92%) | 1219 (100%) | 1219 (100%) | 3 (0.25%) | 0 (0%) | 1218 (99.92%) |
| 생명과학 (174) | 167 (95.98%) | 174 (100%) | 174 (100%) | 174 (100%) | 1 (0.57%) | 173 (99.43%) | 0 (0%) | 158 (90.8%) |
| 컴퓨팅 (4748) | 4718 (99.37%) | 4748 (100%) | 4747 (99.98%) | 4748 (100%) | 4705 (99.09%) | 3677 (77.44%) | 160 (3.37%) | 4725 (99.52%) |
| 지구과학 (269) | 264 (98.14%) | 269 (100%) | 269 (100%) | 269 (100%) | 237 (88.1%) | 268 (99.63%) | 1 (0.37%) | 230 (85.5%) |
| 수학 (24272) | 24242 (99.88%) | 24272 (100%) | 24272 (100%) | 24272 (100%) | 24272 (100%) | 0 (0%) | 73 (0.3%) | 24272 (100%) |
| 기술 (29518) | 29067 (98.47%) | 29518 (100%) | 29518 (100%) | 29518 (100%) | 29492 (99.91%) | 22695 (76.89%) | 233 (0.79%) | 29302 (99.27%) |

5. 논의 및 결론

본 연구의 목적은 STEM 분야의 리포지토리에서 확인되는 연구 소프트웨어의 공유, 인용, 공유별 인용, 저자, 고유식별자 및 소프트웨어 유형의 현황을 분석하는 것이다. 모집단은 DCI에서 전세계 453개의 리포지토리에서 추적 및 색인하는 565,000개 이상의 연구 소프트웨어이다. 분석은 STEM 분야를 개별로 분석하고 이를 종합 정리하였다. STEM 분야는 천체물리학, 생명과학, 컴퓨팅, 지구과학, 수학, 기술 분야였다. 리포지터리별 연구 소프트웨어의 공유, 인용, 기관저자, 고유식별자를 비교 분석하였다. 소프트웨어 유형 별 소프트웨어의 인용과 이용을 비교 분석하였다. 연구 소프트웨어가 공유될 때 사용되는 메타데이터 요소에 대하여 비교 분석하였다.

첫째, 천체물리학 분야는 다음과 같다. 천체물리학 분야의 소프트웨어 인용은 낮았다. 0.23%만 고유식별자를 보유하고 있고, 해당 소프트웨어의 경우 모두 DOI였다. 약 5년전의 연구에서는 천체물리학 분야의 연구 소프트웨어의 고유식별자가 URL이고 DOI는 없는 것으로 확인되었다(박형주, Wolfram, 2021). 이는 연구 소프트웨어의 고유식별자가 DOI인 본 연구 결과와 다른 결과인데, 최근 5년동안 천체물리학 분야의 고유식별자가 발급될 경우, 고유식별자가 DOI로 표준화되고 있다고도 해석될 수 있다. 소프트웨어 공유는 학문 특화 리포지토리인 Astrophysics Source Code Library(ASCL)에서 전체 레코드의 99.77%가 공유되고 있었다. 기관저자는 드물었다. 소프트웨어 유형을 상세하게 명시하지는 않고 있었다. 본 연구는

천체물리학 분야의 소프트웨어 인용이 낮음을 확인하였다. 천체물리학 분야는 아직 소프트웨어 인용과 관련된 메타데이터에 대한 지식이나 필요성에 대한 인식이 미비하기(Allen, 2021) 때문일 수 있다. 천체물리학 리포지터리인 ASCL은 2022년을 기준으로 등록된 코드 중 55%의 코드가 인용을 명시하지 않고 있었다(Allen, 2022). 천체물리학 분야의 데이터는 용량이 매우 큰 원시 데이터이며 이를 공유하기 위한 여건이 적다(Zuiderwijk & Spiers, 2019) 예를 들어, 별, 행성, 은하수 등에 대한 고화질의 이미지, 동영상 등의 용량이 매우 큰 자료이기에, 공유 자체가 어려워서 소프트웨어의 인용에 대한 연구자들의 관심이 적기 때문일 수 있다.

둘째, 생명과학 분야는 다음과 같다. Bioconductor의 모든 연구 소프트웨어는 DOI를 부여받고 있었다. 기관 저자는 드물었다. 생명과학 분야의 소프트웨어 유형은 상세히 명시하고 있지 않은 것으로 나타났는데, 소프트웨어 유형을 적지 않고 <비어있음>으로 제공하는 경우도 2.44%가 있었다. 생명과학 분야가 타 학문 분야에 비해 소프트웨어 인용이 높았다. 생명과학 분야는 컴퓨터 기반 데이터 분석과 생물정보학의 발전으로 인해 연구 전 과정에서 소프트웨어 활용 비중이 높은 학문 분야이다(Howison & Bullard, 2016). 데이터 전처리, 분석, 시각화 등의 단계적인 분석 과정에서 여러 소프트웨어가 연속적으로 사용되며, 연구 결과의 검증과 재현을 위해서는 사용된 소프트웨어의 명확한 식별과 접근 가능성이 필수적이다. 이러한 특성으로 인해 생명과학 연구에서는 단순한 방법 서술을 넘어, 구체적인 소프트웨어를 명시하고 이를 인용할 필요성이 상대적으로 크다(Howison & Bullard,

2016). 그 결과 생명과학 분야는 타 학문 분야에 비해 연구 소프트웨어 인용이 비교적 활발하게 이루어지는 경향을 보일 수 있다.

셋째, 컴퓨팅 분야는 다음과 같다. 주요 리포지토리는 Figshare, Knowledgebase of Interatomic Models, ModelDB가 있었다. 기관 저자는 드물었다. 컴퓨팅 분야는 ModelDB를 제외한 모든 리포지토리가 고유식별자를 발급하고 있었다. 대부분의 STEM 학문 분야가 소프트웨어 유형을 'software'로 단순히 적는 것에 비하여, 컴퓨팅 분야는 상세하게 소프트웨어 유형을 제공하고 있었다.

넷째, 지구과학 분야는 다음과 같다. 기관 저자는 드물었다. 지구과학 분야 리포지토리에서의 평균적인 소프트웨어 공유 별 인용 수는 1.98점이었으며, Woods Hole Open Access Server가 4점으로 가장 높았다. 소프트웨어 공유는 269건으로 타 학문분야에 비하여 낮았다. 소프트웨어 유형을 'software'로 단순히 적는 경우가 인용수가 높았다.

다섯째, 수학 분야는 다음과 같다. 소프트웨어는 100%가 범용 레포지토리인 CRAN에 공유되고 있었다. CRAN은 고유식별자를 제공하지 않기 때문에, 수학 분야의 연구소프트웨어의 고유식별자 비율은 0%였다. 기관저자는 드물었다. 수학 분야는 모든 소프트웨어의 유형이 'software'로 유형이 100% 동일하였는데, 이는 수학분야에서만 보이는 특징이었다.

여섯째, 기술 분야는 다음과 같다. 기술 분야는 범용 리포지토리인 CRAN과 Zenodo에 소프트웨어가 가장 많이 공유되고 있었다. 소프트웨어 공유 별 인용수는 평균 3.52건이며, CRAN의 소프트웨어 공유 1건당 인용 수는

20.68로 월등히 높았다. 소프트웨어의 유형을 단순히 'software'로 적었을 때 인용을 높게 받았으며, 그 외의 소프트웨어 유형은 인용을 받는 경우가 드물었다.

일곱째, STEM 분야를 종합하면 다음과 같다. 지구과학 분야의 경우, 2016년부터 2018년까지 소프트웨어 공유 별 인용 수가 25.33회에서 46.5회로 타 학문 분야 및 연도에 비하여 월등히 높았다. STEM 분야의 소프트웨어 공유 별 인용 수는 평균 2.38이었는데, CRAN과 Bioconductor가 평균적으로 월등히 높았다. 소프트웨어 유형을 단순히 'software'로 적은 경우의 인용률이 높았다. 기술과 수학 분야의 소프트웨어 인용이 월등히 높았다. 고유식별자를 제공하는 레코드 수는 학문 분야에 따라 뚜렷한 차이를 보였다. 대부분의 소프트웨어는 기관저자가 아닌 개별 연구자가 공동저자로 명시되어 있었다. 연구 소프트웨어는 필수 메타데이터를 대부분 제공하고 있었으나, 저자명, 출판년도, 초록 등의 중요한 메타데이터가 누락된 경우도 있었다. 본 연구는 기술 분야와 수학 분야의 소프트웨어 인용이 월등히 높음을 확인하였다. 기술과 수학 분야는 연구의 핵심 결과가 이론 서술 자체보다 계산 과정, 알고리즘 구현, 수치 실험 결과에 의해 도출되는 경우가 많다. 따라서 특정 소프트웨어가 연구 결과를 재현하는 데 필수적인 요소로 사용되며, 소프트웨어를 명시적으로 인용하지 않으면 연구의 검증과 재이용이 어렵다(Katz et al., 2014). 이러한 특성으로 인해 기술 및 수학 분야에서 소프트웨어 인용 비율이 다른 학문 분야보다 높게 나타날 수 있다.

본 연구의 한계는 분석 범위를 STEM으로

한정한 것이다. 데이터 이용은 2013년 이후의 레코드만을 살펴보았는데, 이는 수집된 데이터의 한계이다. 본 연구는 전세계 453개에서 색인되는 연구 소프트웨어의 공식적인 인용 현황을 살펴보았지만, DCI에서 색인되지 않는 연구 소프트웨어는 살펴볼 수 없었다. 향후에는 연구 범위를 사회과학 분야 또는 의료 및 보건 계열로 확장할 필요가 있다. 본 연구의 공헌은 STEM 분야의 연구 소프트웨어를 직접 수집하여 실질적인 현황을 살펴보고, 학문 특화 연구와 협업 연

구가 활발한 STEM 분야를 개별 학문 분야 및 종합 정리하여 구체적으로 살펴보았다는 점이다. 또한, 학문 분야별로 소프트웨어 공유별 인용수의 순위를 밝혀냈는데, 소프트웨어 인용 등의 공식적인 학술 크레딧을 얻고자 하는 공유자의 경우, 소프트웨어 공유 별 인용수가 높은 리포지토리에 소프트웨어를 공유하고자 할 때 본 논문의 결과를 참고할 수 있다는 점에서 학술적 의의가 있다.

참 고 문 헌

- 국가연구개발혁신법 (2022) 법률 제18645호.
- 김주섭, 김선태, 한연중, 유원재, 전바울, 양성준 (2022). 학문분야별 연구자들의 연구데이터 관리 현황에 관한 연구: 문헌정보학, 통계학, 생태학 및 한국음악학을 중심으로. 한국비블리아학회지, 33(4), 229-247. <https://doi.org/10.14699/kbiblia.2022.33.4.229>
- 김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455. <https://doi.org/10.14699/kbiblia.2022.33.4.229>
- 박형주, Wolfram, D. (2021). 연구소프트웨어의 공유, 재이용 및 인용과 관련된 현재 관행의 의미. 정보관리학회지, 38(4), 65-82. <https://doi.org/10.3743/KOSIM.2021.38.4.065>
- 박형주 (2025). STEM 분야의 연구데이터 분석: Data Citation Index를 중심으로. 한국문헌정보학회지, 59(1), 489-516. <https://doi.org/10.4275/KSLIS.2025.59.1.489>
- Allen, A. (2021). Citation method, please? A case study in astrophysics. arXiv. <https://doi.org/10.48550/arXiv.2111.12574>
- Allen, A. (2022). It's your software! Get it cited the way you want! arXiv. <https://doi.org/10.48550/arXiv.2212.12683>
- Bajraktari, A., Binder, M., & Vogelsang, A. (2024). Requirements engineering for research software: a vision. Proceedings of the 32nd IEEE International Requirements Engineering Conference, 423-431. <https://doi.org/10.1109/RE59067.2024.00050>
- Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos,

- F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). Introducing the FAIR Principles for research software. *Scientific Data*, 9, 622. <https://doi.org/10.1038/s41597-022-01710-x>
- Bioconductor (n.d.). About Bioconductor. Bioconductor. Available: <https://www.bioconductor.org/about>
- Carlin, D., Rainer, A., & Wilson, D. (2023). Where is all the research software? *PeerJ Computer Science*, 9, e1546. <https://doi.org/10.7717/peerj-cs.1546>
- Carver, J., Bartlett, R., Heaton, D., & Hochstein, L. (2011). What scientists and engineers think they know about software engineering: a survey. *Computing in Science & Engineering*. SAND2011-2196J.
- CERN Data Centre & Invenio RDM. (n.d.). About Zenodo. Zenodo. Available: <https://about.zenodo.org>
- Chue Hong, N. P., Allen, A., Gonzalez-Beltran, A., Gruenpeter, M., Katz, D. S., & Garijo, D. (2025). Recognizing the value of research software: a survey of current practices. *Scientific Data*, 12, 214.
- Clarivate Analytics (2026). Web of Science Data Citation Index. Available: <https://clarivate.com/academia-government/ko/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/data-citation-index/>
- CRAN (2025). vegan: Community Ecology Package. CRAN. Available: <https://cran.r-project.org/web/packages/vegan/index.html>
- Druskat, S., Bertuch, O., Juckeland, G., Knodel, O., & Schlauch, T. (2022). Software publications with rich metadata: state of the art, automated workflows and HERMES concept. *arXiv*. <https://doi.org/10.48550/arXiv.2201.09015>
- Dubbeldam, D., Vreede, J., Vlugt, T. J. H., & Calero, S. (2019). Highlights of (bio-)chemical tools and visualization software for computational science. *Current Opinion in Chemical Engineering*, 23, 1-13. <https://doi.org/10.1016/j.coche.2019.02.001>
- Figshare (2026). About us. Figshare. Available: <https://figshare.com/about>
- Gil, Y., Ratnakar, V., & Garijo, D. (2015). OntoSoft: capturing scientific software metadata. *Proceedings of the 8th International Conference on Knowledge Capture*, 1-4. <https://doi.org/10.1145/2815833.2816955>
- Gomez-Diaz, T. & Recio, T. (2022). Research software vs. research data I: towards a research data definition in the open science context (Version 2). *F1000Research*, 11, 118. <https://doi.org/10.12688/f1000research.78195.2>

- Goth, F., Alves, R., Braun, M., Castro, L. J., Chourdakis, G., Christ, S., Cohen, J., Druskat, S., Erxleben, F., Grad, J.-N., Hagdorn, M., Hodges, T., Juckeland, G., Kempf, D., Lamprecht, A.-L., Linxweiler, J., Löffler, F., Martone, M., Schwarzmeier, M., Seibold, H., Thiele, J. P., von Waldow, H., & Wittke, S. (2025). Foundational competencies and responsibilities of a research software engineer: current state and suggestions for future directions (Version 2). *F1000Research*, 13, Article 1429. <https://doi.org/10.12688/f1000research.157778.2>
- Hasselbring, W., Carr, L., Hettrick, S., Packer, H., Tiropanis, T., & Nasser, B. (2020). From FAIR research data toward FAIR and open research software. *IT - Information Technology*, 62(1), 39-47. <https://doi.org/10.1515/itit-2019-0040>
- Howison, J. & Bullard, J. (2016). Software in the scientific literature: problems with seeing, finding, and using software mentioned in the biology literature. *Journal of the Association for Information Science and Technology*, 67(9), 2137-2155. <https://doi.org/10.1002/asi.23538>
- Howison, J., Deelman, E., McLennan, M. J. M., da Silva, R. F., & Herbsleb, J. D. (2015). Understanding the scientific software ecosystem and its impact: current and future measures. *Research Evaluation*, 24(4), 454-470. <https://doi.org/10.1515/itit-2019-0040>
- Katz, D. S., Choi, S.-C. T., Lapp, H., Maheshwari, K., Löffler, F., Turk, M., Hanwell, M. D., Wilkins-Diehr, N., Hetherington, J., Howison, J., Swenson, S., Allen, G. D., Elster, A. C., Berriman, B., & Venters, C. (2014). Summary of the first workshop on sustainable software for science: practice and experiences (WSSSPE1). *Journal of Open Research Software*. <https://doi.org/10.5334/jors.an>
- Katz, D. S., Gruenpeter, M., & Honeyman, T. (2021). Taking a fresh look at FAIR for research software. *Patterns*, 2(3), 100222. <https://doi.org/10.1016/j.patter.2021.100222>
- Katz, D. S., Niemeyer, K. E., Smith, A. M., Anderson, W. L., Boettiger, C., Hinsien, K., Hooft, R., Hucka, M., Lee, A., Löffler, F., Pollard, T., & Rios, F. (2016). Software vs. data in the context of citation. *PeerJ Preprints*, 4, e2630v1. <https://doi.org/10.7287/peerj.preprints.2630v1>
- Kelley, A. & Garijo, D. (2021). A framework for creating knowledge graphs of scientific software metadata. *Quantitative Science Studies*, 2(4), 1423-1446. https://doi.org/10.1162/qss_a_00167
- Kelly, D., Smith, S., & Meng, N. (2011). Software engineering for scientists. *Computing in Science & Engineering*, 13(5), 40-47. <https://doi.org/10.1109/MCSE.2011.103>
- Kreyman, K., Parnas, D. L., & Qiao, S. (1999). Inspection procedures for critical programs that

- model physical phenomena. *Computer Science, Engineering, Environmental Science*.
- Kuckertz, P., Göpfert, J., Karras, O., Neuroth, D., Schönau, J., Pueblas, R., Ferenz, S., Engel, F., Pflugradt, N., Weinand, J. M., Nieße, A., Auer, S., & Stolten, D. (2023). A metadata-based ecosystem to improve the FAIRness of research software. *arXiv*.
<https://doi.org/10.48550/arXiv.2306.10620>
- Li, K., Lin, X., & Greenberg, J. (2016). Software citation, reuse and metadata considerations: an exploratory study examining LAMMPS. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-10. <https://doi.org/10.1002/pra2.2016.14505301072>
- Li, K., Yan, E., & Feng, Y. (2017). How is R cited in research outputs? Structure, impacts, and citation standard. *Journal of Informetrics*, 11(4), 989-1002.
<https://doi.org/10.1016/j.joi.2017.08.003>
- Mao, A., Garijo, D., & Fakhraei, S. (2019). SoMEF: a framework for capturing scientific software metadata from its documentation. *Proceedings of the 2019 IEEE International Conference on Big Data*, 3032-3037. <https://doi.org/10.1109/BigData47090.2019.9006447>
- Marcus, A. & Menzies, T. (2010). Software is data too. *Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research (FoSER '10)*, 229-232.
<https://doi.org/10.1145/1882362.188241>
- Martín del Pico, E., Gelpí, J. L., & Capella-Gutiérrez, S. (2025a). The Software Observatory: aggregating and analysing software metadata for trend computation and FAIR assessment. *arXiv*. <https://doi.org/10.48550/arXiv.2510.05705>
- Martín del Pico, E., Gelpí, J. L., & Capella-Gutiérrez, S. (2025b). Identity resolution of software metadata using Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2505.23500>
- National Institute of Standards and Technology (2020). Overview. National Institute of Standards and Technology. Available: <https://www.ctcms.nist.gov/potentials>
- Nature Methods (2021). Computation and biology: a partnership. *Nature Methods*, 18, 695.
- Neuron (n.d.). The NEURON simulator (version 9.0.1). NEURON. Available:
<https://nrn.readthedocs.io/en/9.0.1>
- Park, H. & Wolfram, D. (2019). Research software citation in the Data Citation Index: current practices and implications for research software sharing and reuse. *Journal of Informetrics*, 13(2), 574-582. <https://doi.org/10.1016/j.joi.2019.03.005>
- Peng, J. Chen, Q., Wong, N., Meng, L., Y., Yam C. Y., & Chen, G. (2013). A multi-scale framework for nano-electronic devices modeling with application to the junctionless transistor. *IEEE International Conference of Electron Devices and Solid-state Circuits*, 1-2.

<https://doi.org/10.1109/EDSSC.2013.6628145>

R Core Team (n.d.). The Comprehensive R Archive Network. CRAN. Available:

<https://cran.r-project.org>

Smith, A. M., Katz, D. S., & Niemeyer, K. E. (2016). Software citation principles. *PeerJ Computer Science*, 2, e86. <https://doi.org/10.7717/peerj-cs.86>

Soito, L. & Hwang, L. J. (2016). Citations for software: providing identification, access and recognition for research software. *International Journal of Digital Curation*, 11(2), 48-63. <https://doi.org/10.2218/ijdc.v11i2.390>

Stall, S., Bilder, G., Cannon, M., Chue Hong, N., Edmunds, S., Erdmann, C. C., Evans, M., Farmer, R., Feeney, P., Friedman, M., Giampoala, M., Hanson, R. B., Harrison, M., Karaiskos, D., Katz, D. S., Letizia, V., Lizzi, V., MacCallum, C., Muench, A., Perry, K., Ratner, H., Schindler, U., Sedora, B., Stockhause, M., Townsend, R., Yeston, J., & Clark, T. (2023). Journal production guidance for software and data citations. *Scientific Data*, 10, 656. <https://doi.org/10.1038/s41597-023-02491-7>

Taschuk, M. & Wilson, G. (2017). Ten simple rules for making research software more robust. *PLoS Computational Biology*, 13(4), e1005412. <https://doi.org/10.1371/journal.pcbi.1005412>

Wei, X., Xi, P., Chen, M., Wen, Y., Wu, H., Wang, L., Zhu, Y., Ren, Y., & Gu, Z. (2024). Capsule robots for the monitoring, diagnosis, and treatment of intestinal diseases. "Mater Today Bio", 29, 101294. <https://doi.org/10.1016/j.mtbio.2024.101294>

WIKIPEDIA-a (n.d.). Source code. WIKIPEDIA. Available:

https://en.wikipedia.org/wiki/Source_code

WIKIPEDIA-b (n.d.). Computer simulation. WIKIPEDIA. Available:

https://en.wikipedia.org/wiki/Computer_simulation

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in the production of knowledge. *Science*, 316(5827), 1036-1039. <https://doi.org/10.1126/science.1136099>

Yang, P. & Colavizza, G. (2025). Research data in scientific publications: a cross-field analysis.

arXiv. <https://doi.org/10.48550/arXiv.2502.01407>

Zuiderwijk, A. & Spiers, H. (2019). Sharing and re-using open data: a case study of motivations in astrophysics. *International Journal of Information Management*, 49, 228-241.
<https://doi.org/10.1016/j.ijinfomgt.2019.05.024>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

Kim, Jihyun (2012). A study on university researchers' data management practices. *Journal of Korean Library and Information Science Society*, 43(3), 433-455.

<https://doi.org/10.14699/kbiblia.2022.33.4.229>

Kim, Juseop, Kim, Suntae, Han, Yeonjung, Youe, Won-Jae, Jeon, Paul, & Yang, Seong Jun (2022). A study on the current status of research data management by researchers in each academic field: focusing on Library and Information Science, Statistics, Ecology, and Korean Musicology. *Journal of the Korean Biblia Society for Library and Information Science*, 33(4), 229-247. <https://doi.org/10.14699/kbiblia.2022.33.4.229>

National Research and Development Innovation Act (2022). Act No. 18645.

Park, Hyoungjoo & Wolfram, D. (2021). The implications of current practices relating to the sharing, reuse, and citation of research software for the future of research. *Journal of the Korean Society for Information Management*, 38(4), 65-82.

<https://doi.org/10.3743/KOSIM.2021.38.4.065>

Park, Hyoungjoo (2025). Analyzing STEM research data in the Data Citation Index. *Journal of the Korean Society for Library and Information Science*, 59(1), 489-516.

<https://doi.org/10.4275/KSLIS.2025.59.1.489>

