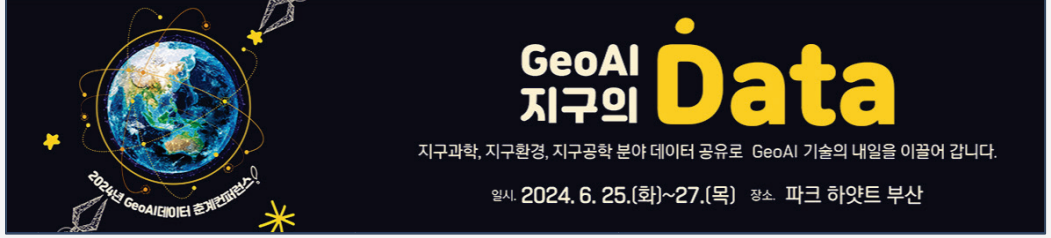


데이터학술지와 데이터저장소의 탐색적 협업 모델

안성수[1], 정영임[2]

1경상국립대학교 (GNU)
2한국과학기술정보연구원 (KISTI)

경상국립대학교
Gyeongsang National University

목차

1. 데이터학술지 서비스
    1. 데이터학술지, 데이터논문 소개
    2. 데이터논문 제출 및 **심사 서비스**
2. 데이터저장소 서비스
    1. 데이터저장소 소개
    2. 데이터저장소 (예: Zenodo) 서비스
3. 데이터학술지와 데이터저장소의 협업 모델
    1. 두 개념간 서비스 비교
    2. 두 개념간 협업 시나리오

1. 데이터학술지 서비스

연구데이터의 증가, …

- 연구활동을 통해 생산된 데이터는 조금씩 증가 추세
  - 오픈사이언스, 오픈액세스, 오픈데이터 운동 확산
    - 데이터의 중요성 인식 향상, 정보기술의 발달
  - 연구비 지원기관의 요구사항

- ***하지만,*** 연구데이터 공유와 재사용의 보이지 않는 **장벽** 존재
  - 연구데이터를 생산하는 연구자에 대한 보상과 인정 부족
    - “What‘s in it for me?”
  - 연구데이터의 낮은 품질
    - 데이터를 정확하게 서술하고 생성하는데 상당한 전문성, 시간, 노력 소요
  - 연구데이터 공유 문화 부족 등

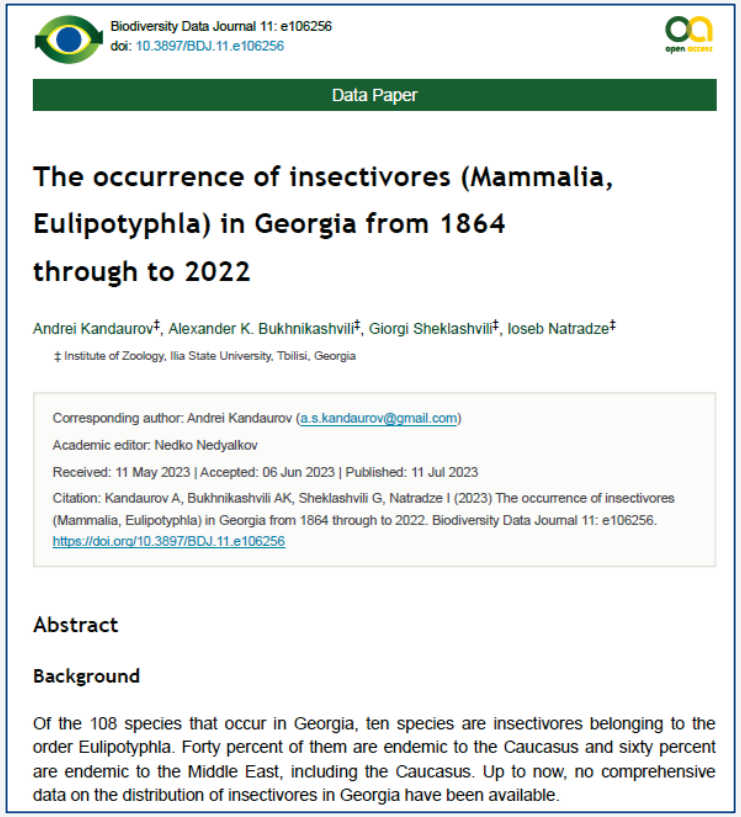
데이터학술지, 데이터논문의 등장

- 연구데이터 공유와 재사용 장벽을 깨는 **첫번째 해결책**
  - **데이터논문을 데이터학술지에 출판하여 연구성과로 인정하기**
    - Int'l Journal of Robotics Research (2009)
    - Earth System Science Data, Neuroinformatics (2011)
    - Biodiversity Data Journal (2013)
    - Scientific Data (2013) – Nature Publishing Group
    - **GEO DATA (2019)**
  - 데이터논문 정의:
    - "온라인의 특정 데이터셋을 서술하는 메타데이터 문서로 표준화된 학술정책 또는 규범에 따라 출판됨"[1]
    - "Information on the what, where, why, how, and who of the data" [2]

1 Source: Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC bioinformatics, 12, 1-12.
2 Source: Callaghan, Sarah, et al. "Making data a first class scientific output: Data citation and publication by NERC's environmental data centres." (2012)

데이터논문 사례 1

- **특정 학문분야** 또는 다학제 학술지에서 데이터



Biodiversity Data Journal 11: e106256
doi: 10.3897/BDJ.11.e106256

Data Paper

The occurrence of insectivores (Mammalia, Eulipotyphla) in Georgia from 1864 through to 2022

Andrei Kandaurov‡, Alexander K. Bukhnikashvili‡, Giorgi Sheklashvili‡, Ioseb Natradze‡

‡ Institute of Zoology, Ilia State University, Tbilisi, Georgia

Corresponding author: Andrei Kandaurov (a.s.kandaurov@gmail.com)

Academic editor: Nedko Nedyalkov

Received: 11 May 2023 | Accepted: 06 Jun 2023 | Published: 11 Jul 2023

Citation: Kandaurov A, Bukhnikashvili AK, Sheklashvili G, Natradze I (2023) The occurrence of insectivores (Mammalia, Eulipotyphla) in Georgia from 1864 through to 2022. Biodiversity Data Journal 11: e106256. https://doi.org/10.3897/BDJ.11.e106256

Abstract

Background

Of the 108 species that occur in Georgia, ten species are insectivores belonging to the order Eulipotyphla. Forty percent of them are endemic to the Caucasus and sixty percent are endemic to the Middle East, including the Caucasus. Up to now, no comprehensive data on the distribution of insectivores in Georgia have been available.

Biodiversity Data Journal (BDJ)
by Pensoft Publisher. (2013-)

GEO DATA

ORIGINAL PAPER

eISSN 2713-5004 • https://geodata.kr

June 2023 • Vol. 5. No. 2, pp. 55-59
https://doi.org/10.22761/GD.2023.0010

완도 붉가시나무림 5년간(2017–2021년)의 나비 다양성에 대한 정량적 연구

이상훈[1,†,*], 안나현[2,†]

[1]팀장, 국립생태원, 충청남도 서천군 마서면 금강로 1210, 33657, 대한민국
[2]연구원, 국립생태원, 충청남도 서천군 마서면 금강로 1210, 33657, 대한민국

Quantitative Study of Butterfly Diversity in Wando *Quercus acuta* Forest Over 5 Years (2017-2021)

Sanghun Lee[1,†,*] and Na-Hyun Ahn[2,†]

[1]Team leader, National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon-gun, 33657 Chungcheongnam-do, South Korea
[2]Researcher, National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon-gun, 33657 Chungcheongnam-do, South Korea

Received March 27, 2023
Revised April 20, 2023
Accepted May 2, 2023

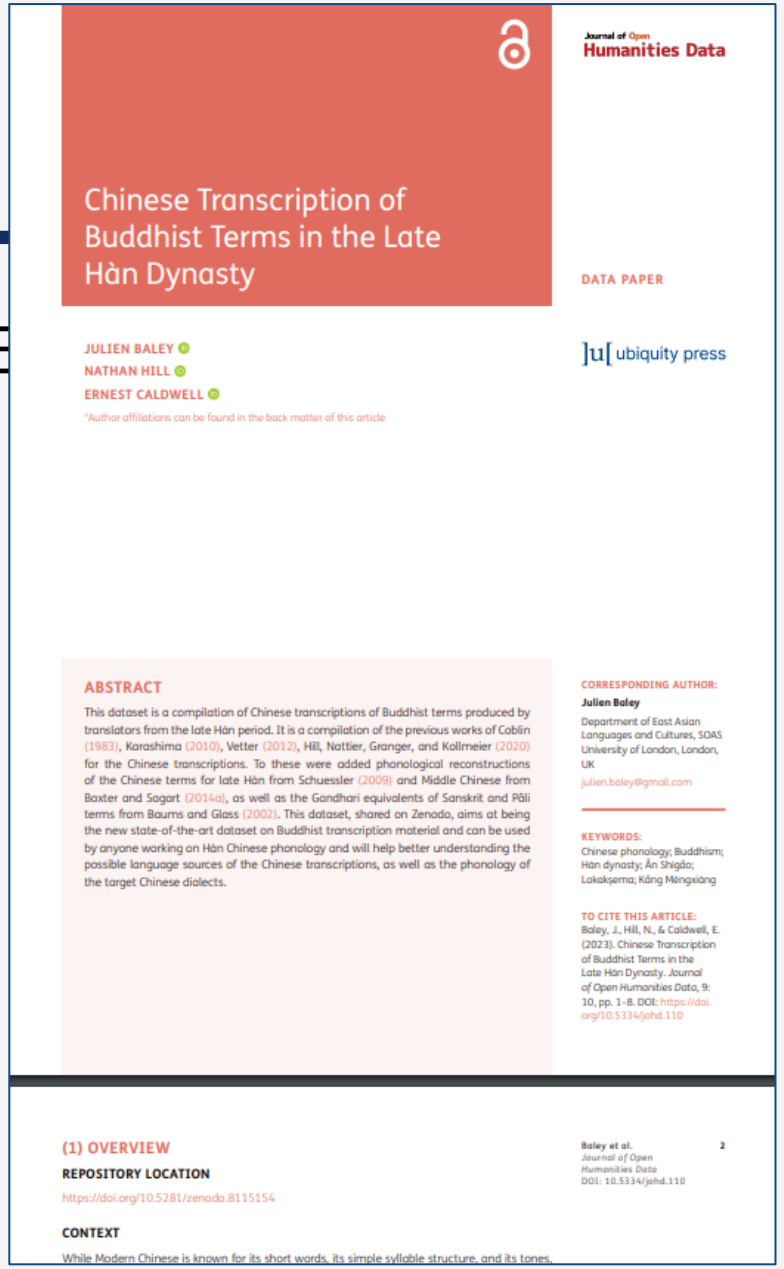
Corresponding Author
Sanghun Lee
Tel: +82-41-950-5621
E-mail: sanghunlee@nie.re.kr

†These authors contributed equally to this work.

This study presents the long-term quantitative data on butterflies in Wando Arboretum, which represents the only warm-temperate forest located in the southernmost part of South Korea. This arboretum has significant academic value as approximately 770 species of rare woody plants or herbs, such as the Japanese evergreen oak (*Quercus acuta*), found in warm temperate zones grow under natural conditions here. In this project, the butterflies in this region were studied due to their sensitivity to temperature changes. The study was conducted from March-April to October-November over 5 years (2017-2021) in the region dominated by Japanese evergreen oak. We found 1,743 individuals of 47 butterfly species belonging to five families. The acquired butterfly data could serve as a reference for the further development of a network-oriented database for assessing temporal climate changes.

Keywords: Butterfly; *Quercus acuta*; Wando island; Warm-temperate forest

Geo Data
by GeoAIData Society (2019-)



Journal of Open Humanities Data

Chinese Transcription of Buddhist Terms in the Late Hàn Dynasty

DATA PAPER

JULIEN BALEY
NATHAN HILL
ERNEST CALDWELL

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

This dataset is a compilation of Chinese transcriptions of Buddhist terms produced by translators from the late Hàn period. It is a compilation of the previous works of Coblin (1983), Karashima (2010), Vetter (2012), Hill, Nattier, Granger, and Kollmeier (2020) for the Chinese transcriptions. To these were added phonological reconstructions of the Chinese terms for late Hàn from Schuessler (2009) and Middle Chinese from Baxter and Sagart (2014a), as well as the Gandhari equivalents of Sanskrit and Pāli terms from Baums and Glass (2002). This dataset, shared on Zenodo, aims at being the new state-of-the-art dataset on Buddhist transcription material and can be used by anyone working on Hàn Chinese phonology and will help better understanding the possible language sources of the Chinese transcriptions, as well as the phonology of the target Chinese dialects.

CORRESPONDING AUTHOR:
Julien Baley
Department of East Asian Languages and Cultures, SOAS University of London, London, UK
julien.baley@gmail.com

KEYWORDS:
Chinese phonology; Buddhism; Hàn dynasty; Ân Shìgāo; Lokakṣema; Kāng Mèngxiàng

TO CITE THIS ARTICLE:
Baley, J., Hill, N., & Caldwell, E. (2023). Chinese Transcription of Buddhist Terms in the Late Hàn Dynasty. *Journal of Open Humanities Data*, 9: 10, pp. 1–8. DOI: https://doi.org/10.5334/johd.110

(1) OVERVIEW

REPOSITORY LOCATION

https://doi.org/10.5281/zenodo.8115154

CONTEXT

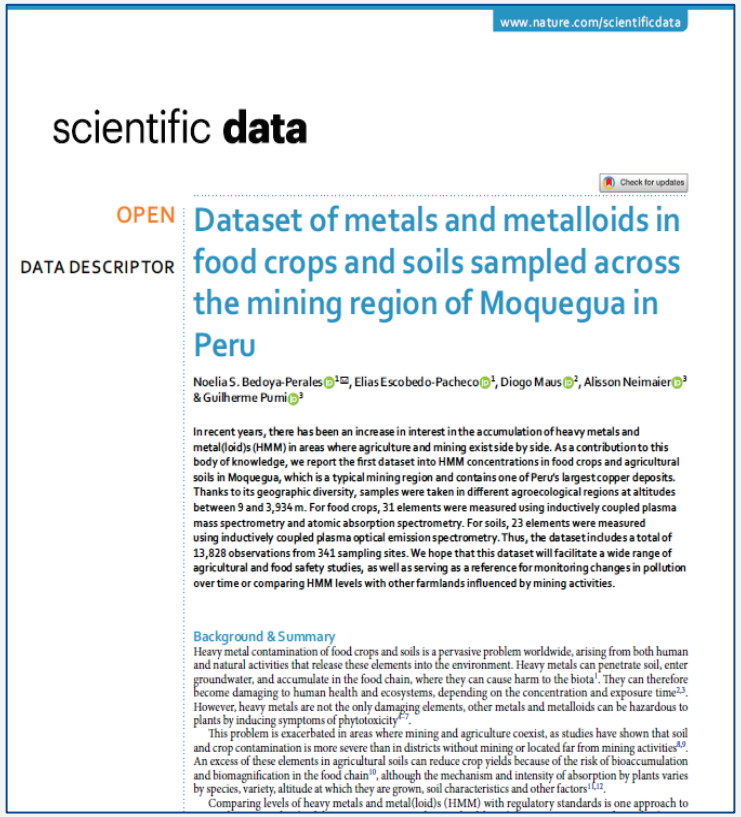
While Modern Chinese is known for its short words, its simple syllable structure, and its tones,

Baley et al. 2
Journal of Open Humanities Data
DOI: 10.5334/johd.110

JOHD (J. of Open Humanities Data)
by Ubiquity Press (2015-)

데이터논문 사례 2

- 특정 학문분야 또는 **다학제 학술지**에서 데이터논문 발간



www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

Dataset of metals and metalloids in food crops and soils sampled across the mining region of Moquegua in Peru

Noelia S. Bedoya-Perales, Elias Escobedo-Pacheco, Diogo Maus, Alisson Neimaier & Guilherme Pumi

In recent years, there has been an increase in interest in the accumulation of heavy metals and metal(loid)s (HMM) in areas where agriculture and mining exist side by side. As a contribution to this body of knowledge, we report the first dataset into HMM concentrations in food crops and agricultural soils in Moquegua, which is a typical mining region and contains one of Peru's largest copper deposits. Thanks to its geographic diversity, samples were taken in different agroecological regions at altitudes between 9 and 3,934 m. For food crops, 31 elements were measured using inductively coupled plasma mass spectrometry and atomic absorption spectrometry. For soils, 23 elements were measured using inductively coupled plasma optical emission spectrometry. Thus, the dataset includes a total of 13,828 observations from 341 sampling sites. We hope that this dataset will facilitate a wide range of agricultural and food safety studies, as well as serving as a reference for monitoring changes in pollution over time or comparing HMM levels with other farmlands influenced by mining activities.

Background & Summary

Heavy metal contamination of food crops and soils is a pervasive problem worldwide, arising from both human and natural activities that release these elements into the environment. Heavy metals can penetrate soil, enter groundwater, and accumulate in the food chain, where they can cause harm to the biota[1]. They can therefore become damaging to human health and ecosystems, depending on the concentration and exposure time[2,3]. However, heavy metals are not the only damaging elements, other metals and metalloids can be hazardous to plants by inducing symptoms of phytotoxicity[4–7].

This problem is exacerbated in areas where mining and agriculture coexist, as studies have shown that soil and crop contamination is more severe than in districts without mining or located far from mining activities[8,9]. An excess of these elements in agricultural soils can reduce crop yields because of the risk of bioaccumulation and biomagnification in the food chain[10], although the mechanism and intensity of absorption by plants varies by species, variety, altitude at which they are grown, soil characteristics and other factors[11,12].

Comparing levels of heavy metals and metal(loid)s (HMM) with regulatory standards is one approach to

Scientific Data
by Springer-Nature (2013-)

Data Article

A dataset for successful recognition of cucumber diseases

Nusrat Sultana*, Sumaita Binte Shorif, Morium Akter, Mohammad Shorif Uddi

Department of Computer Science and En

ARTICLE INFO

Article history:
Received 24 January 2023
Revised 7 April 2023
Accepted 12 June 2023
Available online 16 June 2023

Dataset link: Cucumber Disease Recognition Dataset (Original data)

Keywords:
Image recognition
Agriculture
Cucumber dataset
Deep learning
Computer vision

Specification table

Subject:	Computer Science
Specific subject area:	Identification of Plant Diseases, Deep Learning, and Computer Vision: Image
Type of Data:	
How the data were acquired:	From 6th to 30th May 2022, we collected disease-free and various disease-affected cucumber images from real fields using Nikon D5300 a Single-lens reflex digital camera to obtain raw images.
Data format:	Raw jpg
Description of data collection:	Previously, no such pre-treatment of the samples was done. With the cooperation of an expert from an agricultural institution, we gathered all the cucumber disease images along with disease free from real fields.
Data source location:	**Institution:** Cultivated cucumber field of size 4356 square feet by a local farmer, **Zone:** Basundia, Jashore **Country:** Bangladesh
Data accessibility:	**Repository name:** Mendeley Data **Data identification number (DOI number):** 10.17632/y6d3z6f8z9.1 **Link of the dataset:** https://data.mendeley.com/datasets/y6d3z6f8z9/1

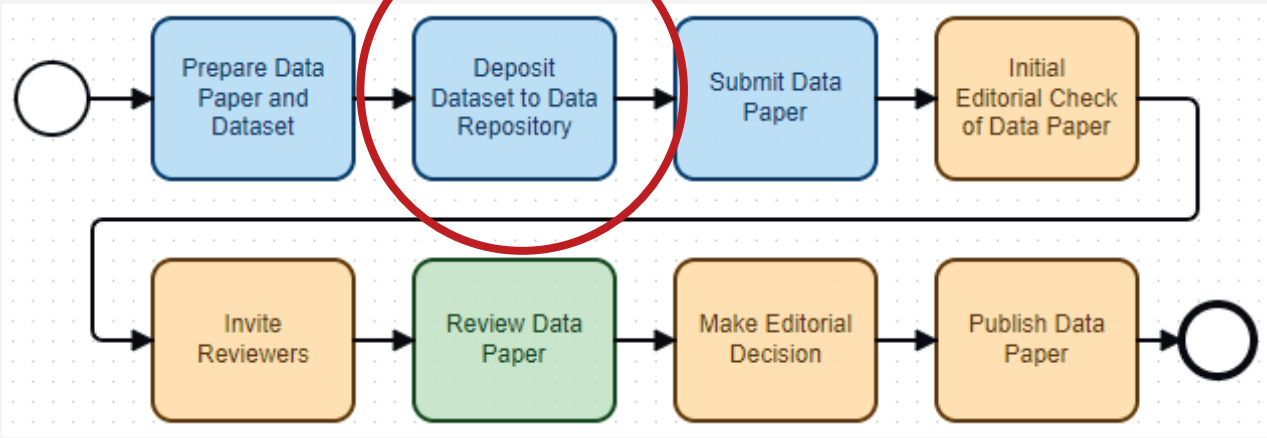
Value of the Data

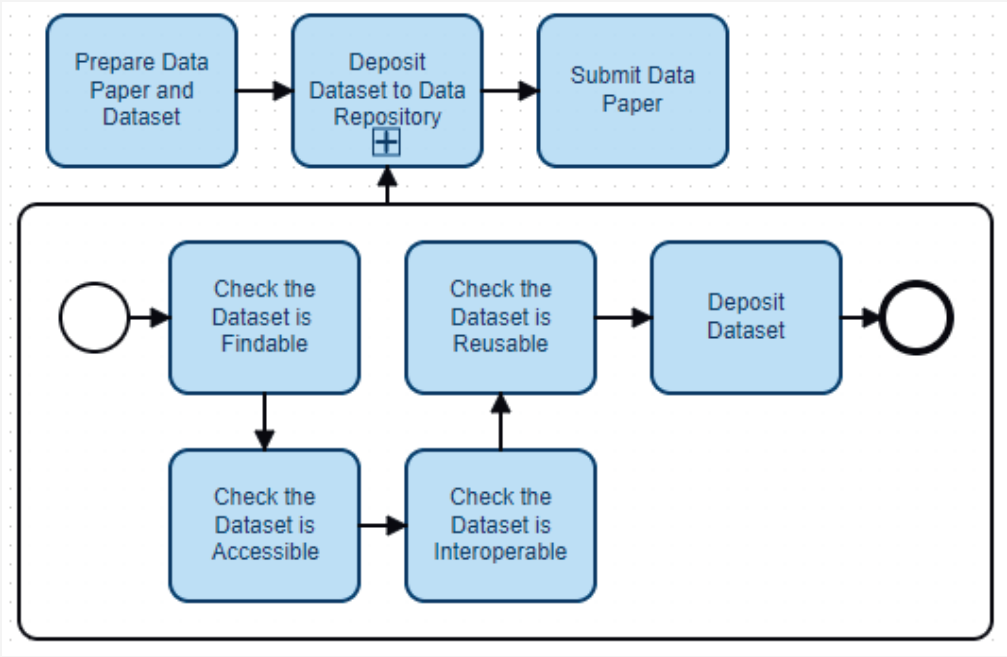
- Cucumber disease detection is typically done manually by individuals, which is time-consuming and ineffective for both farmers and retailers. For this reason, the development of a machine vision-based model is essential that will minimize human effort, expense, and production time in the agricultural industry by identifying different cucumber diseases. An effective dataset is a prerequisite for developing a better classification model.
- This dataset presents the visual appearance of cucumber diseases in a hyperspectral or-

Data in Brief
by Elsevier (2014-)

데이터논문 제출 과정

1. 데이터학술지 선정, 학술지의 **데이터논문 템플릿 다운로드**
2. 데이터논문 및 데이터셋 준비
   - 선행조건: 연구활동을 통해 데이터셋 생성 (관찰, 수집, 구축)
3. **데이터셋을 데이터저장소에 기탁 (deposit)**
4. **데이터논문 작성**, (데이터 FAIR)검토, 제

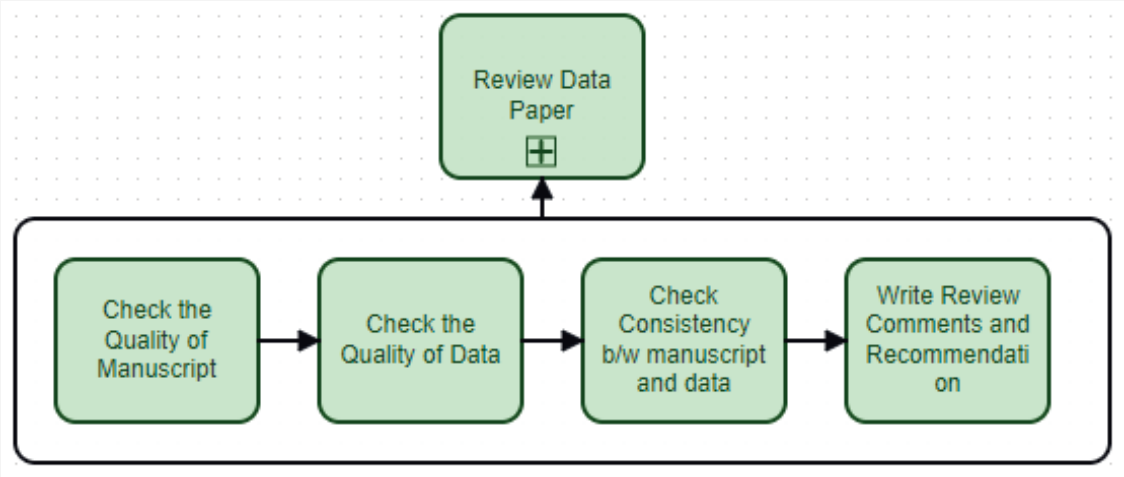




데이터논문 동료심사 프로세스

✔ 필수 적용, △암묵적 적용, 그 외 미적용

동료심사 단계	동료심사 세부단계	BDJ	Sci-Data	DIB	JOHD	Geo Data
데이터논문 및 데이터셋 준비	- 데이터논문 양식을 활용한 데이터논문 준비	✔	✔	✔	✔	△
	- 데이터셋 안내서를 참고하여 데이터셋 생성/수집	✔	✔	✔	✔	✔
데이터셋을 데이터저장소에 기탁	- 데이터셋을 데이터저장소에 기탁	✔	✔	✔	✔	✔
	- 데이터논문 내용에 FAIR와 관련한 항목 체크	✔	△	△	△	△
데이터논문 제출	- 데이터논문의 메타데이터 입력 및 논문 파일 첨부	✔	✔	✔	✔	✔
편집위원회 데이터논문 초기 체크	- 편집위원장(또는 편집팀)은 심사할 데이터논문 선별	✔	✔	✔	✔	✔
심사위원초청	- 도메인 전문가 초청	✔	✔	✔	✔	✔
	- 패널 또는 일반 전문가 초청	✔	△	△	△	△
데이터논문 동료심사	- 데이터논문 품질 심사	✔	✔	✔	✔	✔
	- 데이터셋 품질 심사	✔	△	✔	✔	✔
	- 데이터논문과 데이터셋간 관계에 대한 품질 심사	✔	✔	△	△	△
데이터논문 게재 판정	- 논문 게재 판정	✔	✔	✔	✔	✔
데이터논문 출판	- 논문 출판	✔	✔	✔	✔	✔

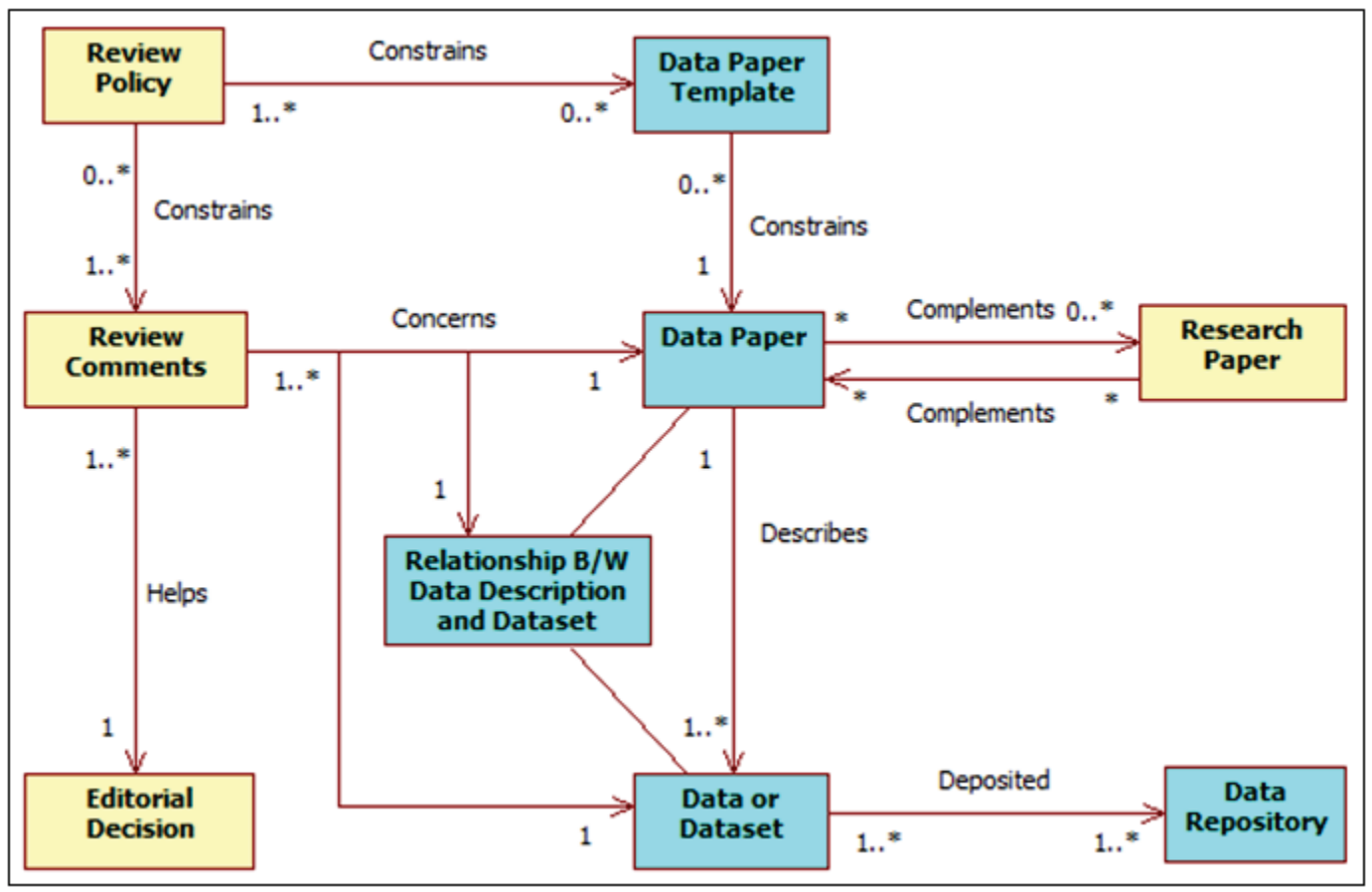


심사위원의 데이터논문 심사 단계

- 그 외 주요 동료심사 기준
  - 데이터논문의 목적과 범위
  - 데이터의 학문분야 기여 (novelty)
  - 데이터 재사용성

출처: 안성수, 조성남, 정영임 (2023). 데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링. 한국도서관·정보학회지, 54(3), 321 - 346

데이터논문 동료심사의 핵심 개념



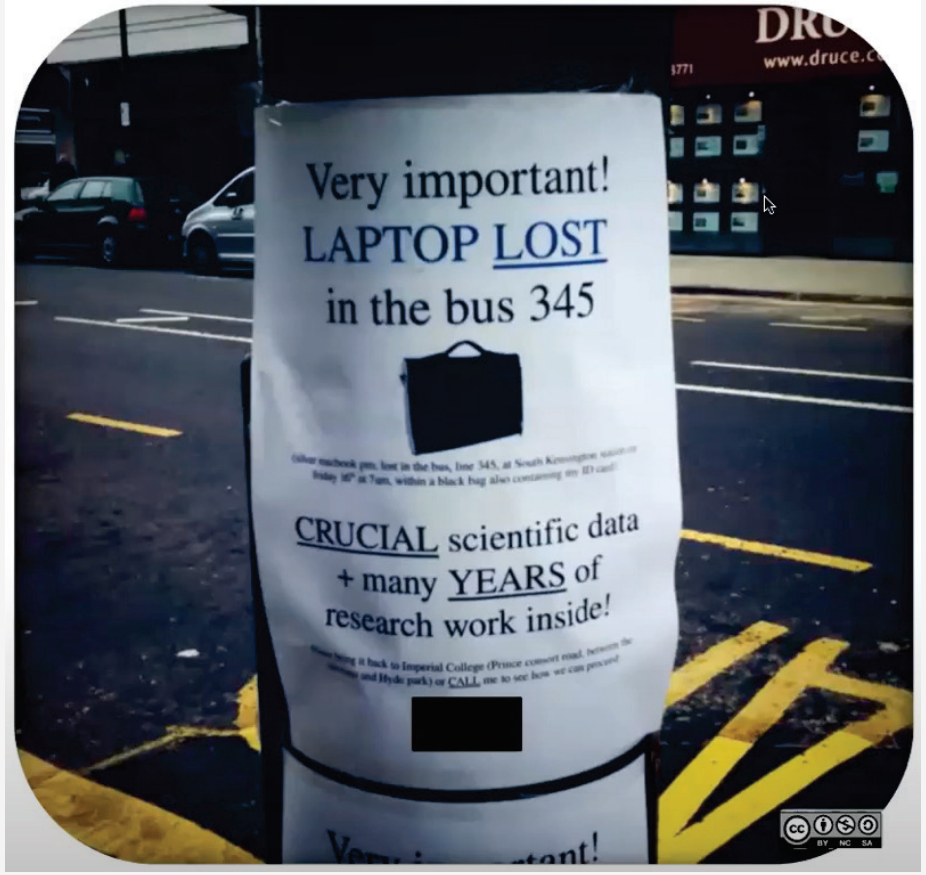
데이터논문 동료심사 개념 모델링 (클래스 다이어그램)

개념	정의
데이터논문 (Data Paper)	연구활동을 통해 생산된 데이터셋을 서술한 논문.
데이터논문 양식 (Data Paper Template)	데이터논문의 세부 형식과 구성요소에 대한 세부 설명을 제공하는 문서
연구논문 (Research Paper)	자연현상의 이해, 문제에 대한 해결책 등을 제안한 논문
데이터셋 (Dataset)	연구활동을 통해 생산된 데이터 또는 데이터의 집합
데이터논문과 데이터셋의 관계성 (Relationship between Data Paper and Dataset)	데이터논문과 데이터셋의 관계성 (예: 일관성, 무결성)
데이터저장소 (Data Repository)	연구자가 생산한 데이터셋을 저장하여 타 연구자가 접근하고 활용할 수 있도록 서비스를 제공한 인터넷 공간
심사의견서 (Review Comment)	심사위원이 논문을 심사하여 작성한 의견서
심사정책 (Review Policy)	학술지의 데이터논문 심사를 정책, 방향 또는 기준
논문판정 (Editorial Decision)	편집위원(장)의 데이터논문에 대한 최종 출판 판정

출처: 안성수, 조성남, 정영임 (2023). 데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링. 한국도서관·정보학회지, 54(3), 321 - 346

2. 데이터저장소 소개

연구활동으로 생산한 데이터가 분실된다면 …



지하철/버스에서 노트북컴퓨터를 분실하면 ….

학교/연구소에서 화재 사고가 발생하면 …

Source: Webinar | Zenodo: open digital repository | 2022.12.07. https://www.youtube.com/watch?v=yj2r8RayIX8

연구활동으로 생산한 데이터가 분실된다면 …

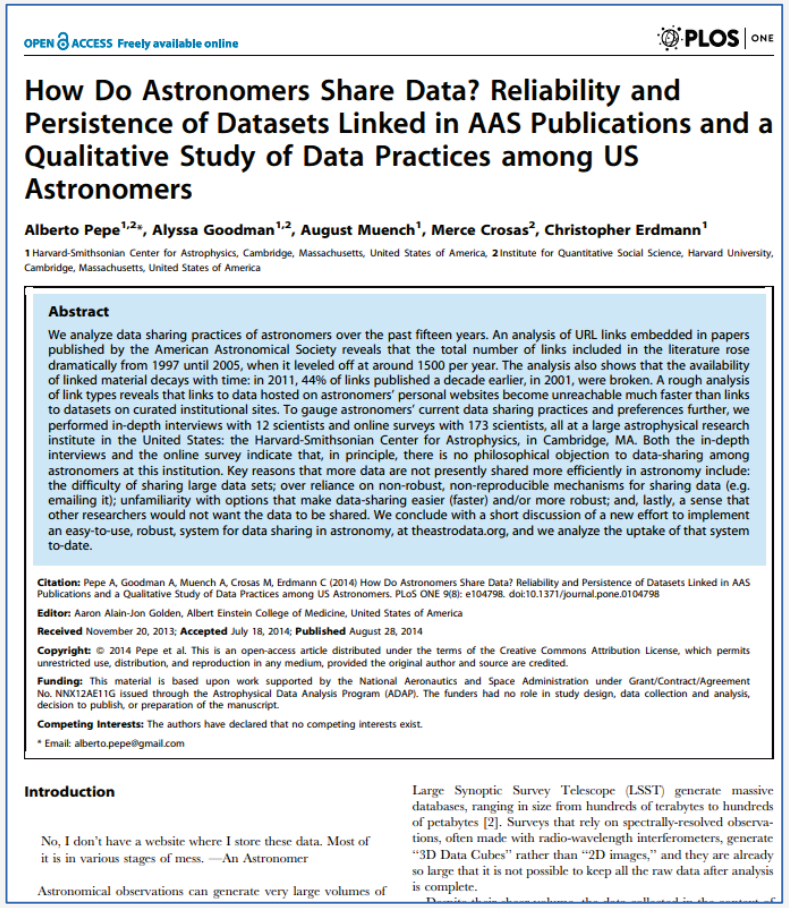
휴대용 저장장치를 분실하면 ….

프로그램 소스코드가 실수로 삭제되면 …

Source: Webinar | Zenodo: open digital repository | 2022.12.07. https://www.youtube.com/watch?v=yj2r8RayIX8

논문 공유데이터의 비접근성과 연구 비재현성

50% of the links in papers are **inaccessible** after 10 years



OPEN ACCESS Freely available online

PLOS ONE

How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers

Alberto Pepe[1,2]*, Alyssa Goodman[1,2], August Muench[1], Merce Crosas[2], Christopher Erdmann[1]

1 Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, United States of America, 2 Institute for Quantitative Social Science, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

We analyze data sharing practices of astronomers over the past fifteen years. An analysis of URL links embedded in papers published by the American Astronomical Society reveals that the total number of links included in the literature rose dramatically from 1997 until 2005, when it leveled off at around 1500 per year. The analysis also shows that the availability of linked material decays with time: in 2011, 44% of links published a decade earlier, in 2001, were broken. A rough analysis of link types reveals that links to data hosted on astronomers' personal websites become unreachable much faster than links to datasets on curated institutional sites. To gauge astronomers' current data sharing practices and preferences further, we performed in-depth interviews with 12 scientists and online surveys with 173 scientists, all at a large astrophysical research institute in the United States: the Harvard-Smithsonian Center for Astrophysics, in Cambridge, MA. Both the in-depth interviews and the online survey indicate that, in principle, there is no philosophical objection to data-sharing among astronomers at this institution. Key reasons that more data are not presently shared more efficiently in astronomy include: the difficulty of sharing large data sets; over reliance on non-robust, non-reproducible mechanisms for sharing data (e.g. emailing it); unfamiliarity with options that make data-sharing easier (faster) and/or more robust; and, lastly, a sense that other researchers would not want the data to be shared. We conclude with a short discussion of a new effort to implement an easy-to-use, robust, system for data sharing in astronomy, at theastrodata.org, and we analyze the uptake of that system to-date.

Citation: Pepe A, Goodman A, Muench A, Crosas M, Erdmann C (2014) How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. PLoS ONE 9(8): e104798. doi:10.1371/journal.pone.0104798

Editor: Aaron Alain-Jon Golden, Albert Einstein College of Medicine, United States of America

Received November 20, 2013; **Accepted** July 18, 2014; **Published** August 28, 2014

Funding: This material is based upon work supported by the National Aeronautics and Space Administration under Grant/Contract/Agreement No. NNX12AE11G issued through the Astrophysical Data Analysis Program (ADAP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: alberto.pepe@gmail.com

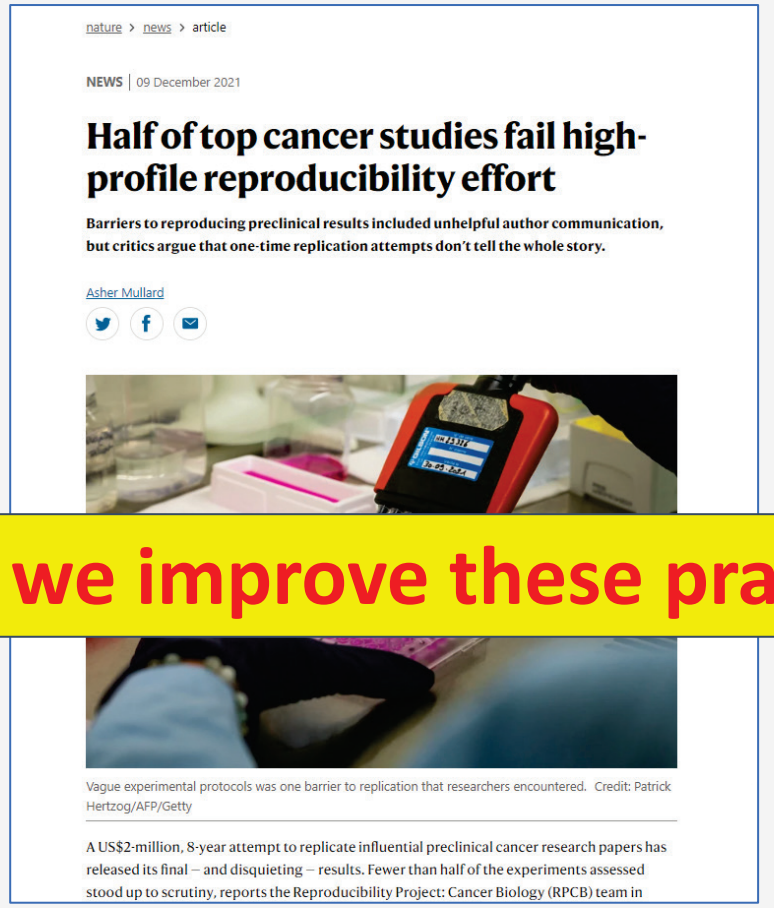
Introduction

No, I don't have a website where I store these data. Most of it is in various stages of mess. —An Astronomer

Astronomical observations can generate very large volumes of

Large Synoptic Survey Telescope (LSST) generate massive databases, ranging in size from hundreds of terabytes to hundreds of petabytes [2]. Surveys that rely on spectrally-resolved observations, often made with radio-wavelength interferometers, generate "3D Data Cubes" rather than "2D images," and they are already so large that it is not possible to keep all the raw data after analysis is complete.

89% of 53 landmark cancer research papers are **irreproducible**



nature › news › article

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.

Asher Mullard

Vague experimental protocols was one barrier to replication that researchers encountered. Credit: Patrick Hertzog/AFP/Getty

A US$2-million, 8-year attempt to replicate influential preclinical cancer research papers has released its final – and disquieting – results. Fewer than half of the experiments assessed stood up to scrutiny, reports the Reproducibility Project: Cancer Biology (RPCB) team in

How can we improve these practices?

Source: Webinar | Zenodo: open digital repository | 2022.12.07. https://www.youtube.com/watch?v=yj2r8RayIX8

데이터저장소: 데이터의 영구적인 보존, 공유와 재사용

- 데이터논문에서 서술하는 **데이터셋의 기탁 및 관리 지원**
  - 데이터셋에 대한 식별자로 영구 ID(예: DOI, Handle) 발급
  - 데이터셋에 대한 검색, 접근, 상호운용성, 재사용 즉, FAIR 지원
    - 데이터셋의 갱신 및 버전 관리 지원
    - 데이터(셋) 삭제는 관리자에게 특별히 요청할 경우만 가능

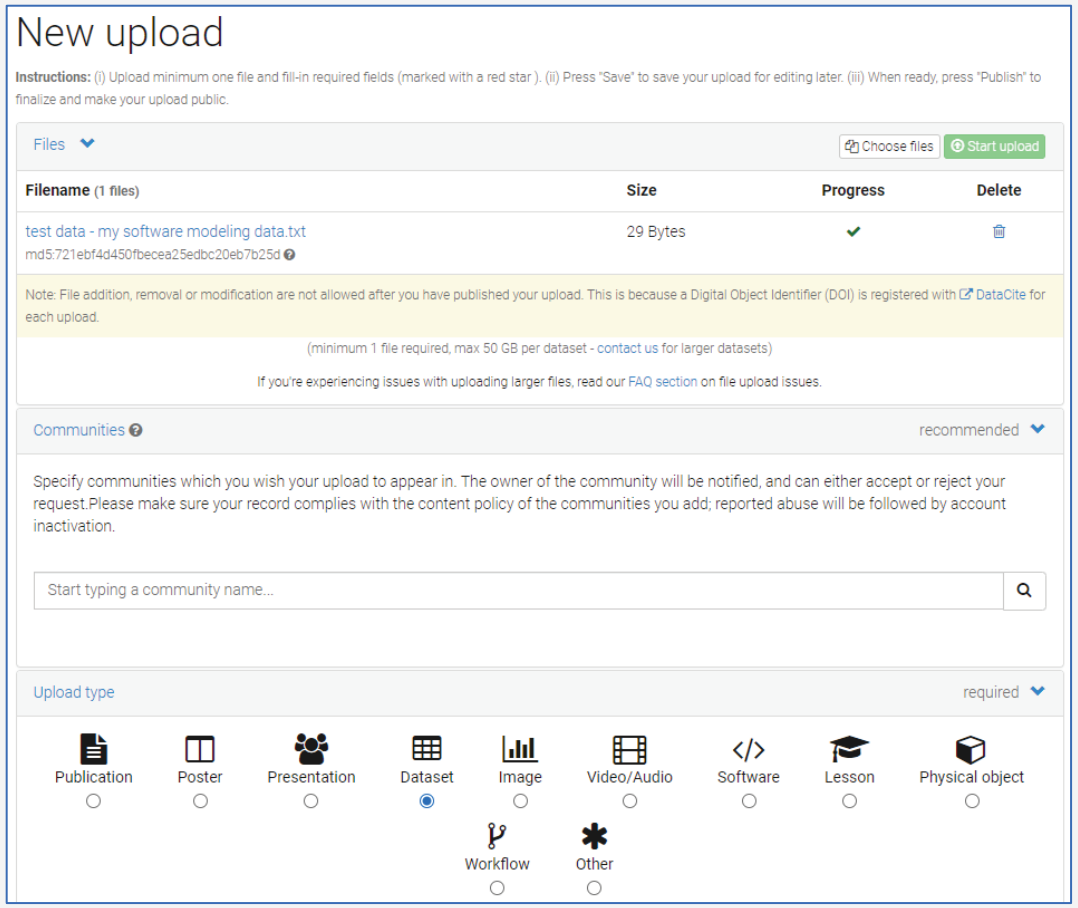
- 국내외 다양한 데이터저장소
  - **Zenodo, Figshare**
  - **Dryad, Harvard Dataverse**
  - **CODA 국립보건연구원**
  - **DataOn 한국과학기술정보연구원**

Zenodo (제노도)

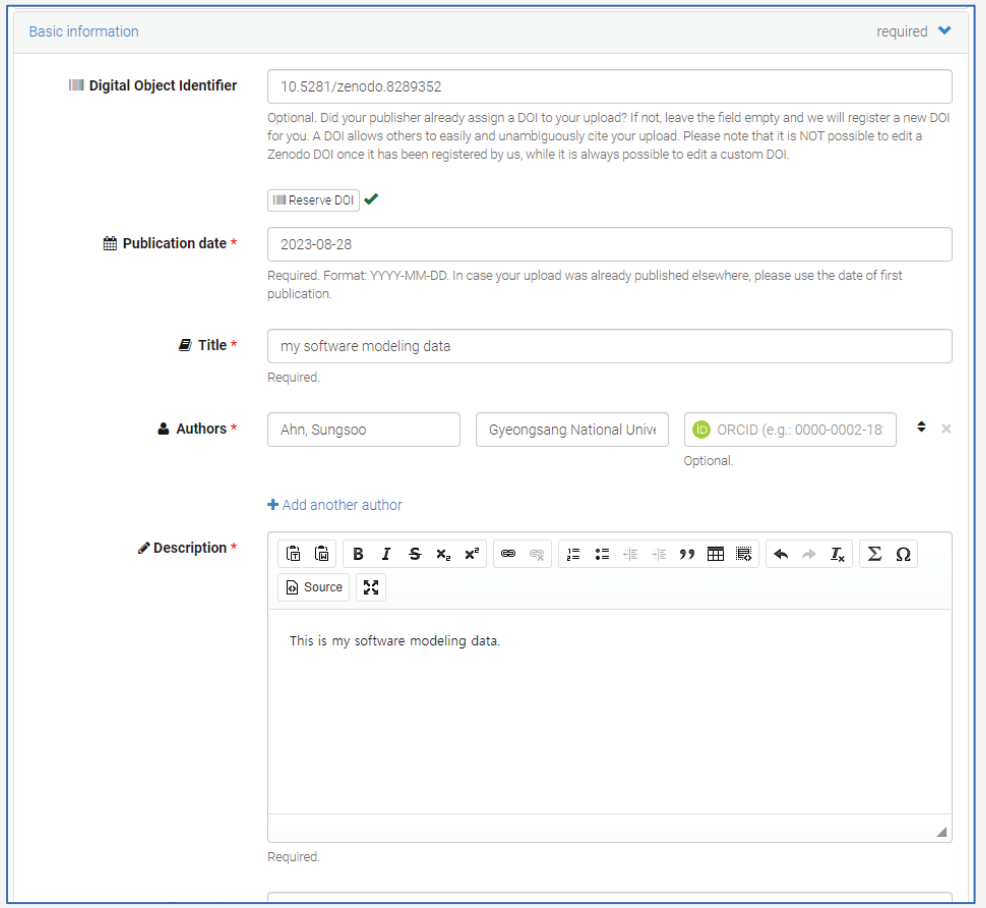
- 데이터저장소
  - 국제 정부기구인 CERN에서 운영
    - CERN이 유지되는 동안 Zenodo 운영
    - 우리나라는 비투표 회원국
- **연구데이터, 소프트웨어 등 기탁, 관리**
  - 2013. 5. 8 서비스 시작
  - **커뮤니티** 서비스 제공
- 무료로 누구나 이용 가능
  - ORCID 또는 GibHub ID로 로그인 가능
- 사용자: 여러 개의 데이터셋을 기탁 가능
  - 한 개 데이터셋은 50GB까지 기탁 가능
- 제노도 이름:
  - 알렉산드리아(도시) 고대 도서관의 첫번째 사서(librarian)인 Zenodotus의 이름에서 유래

Zenodo 누리집: https://zenodo.org/

Zenodo에 데이터셋 기탁하기 (1)

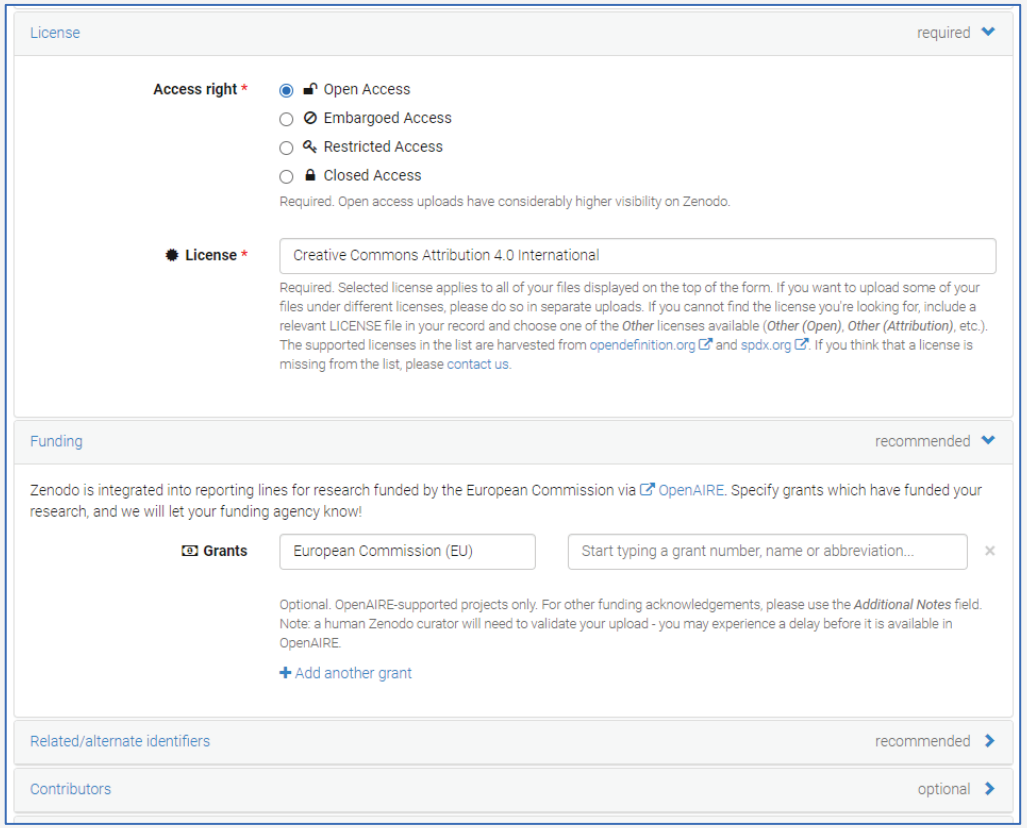


1: 데이터셋(파일) 업로드, 업로드 타입 선택

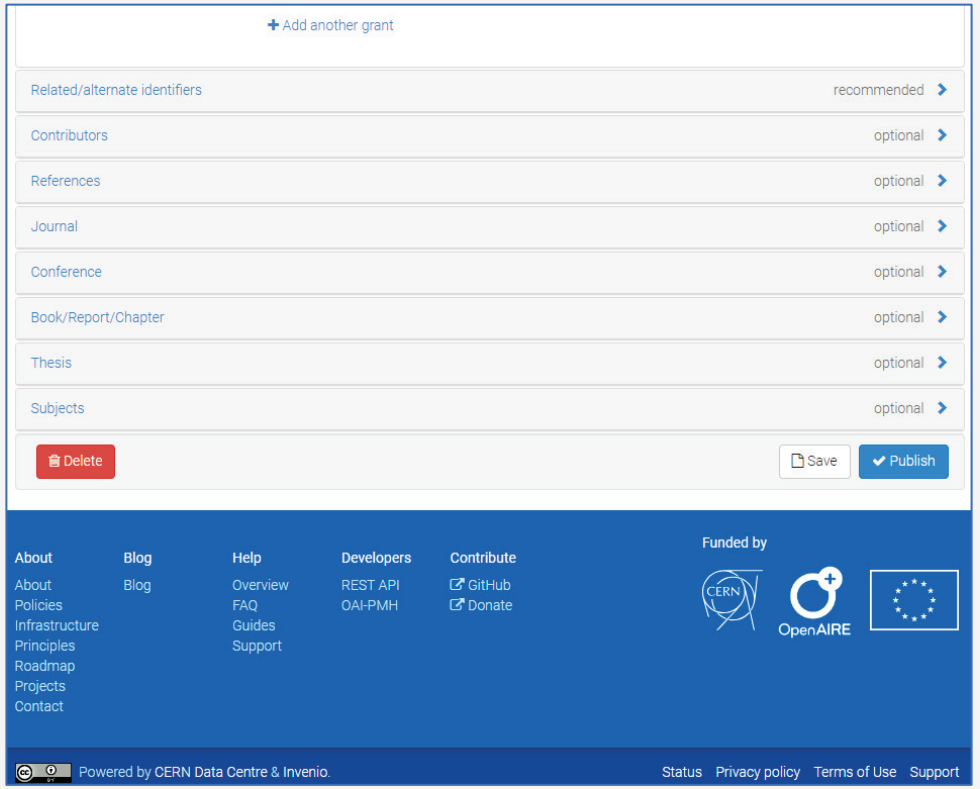


2: DOI 예약발급, 출판일자, 제목, 저자, 설명문 입력

Zenodo에 데이터셋 기탁하기 (2)

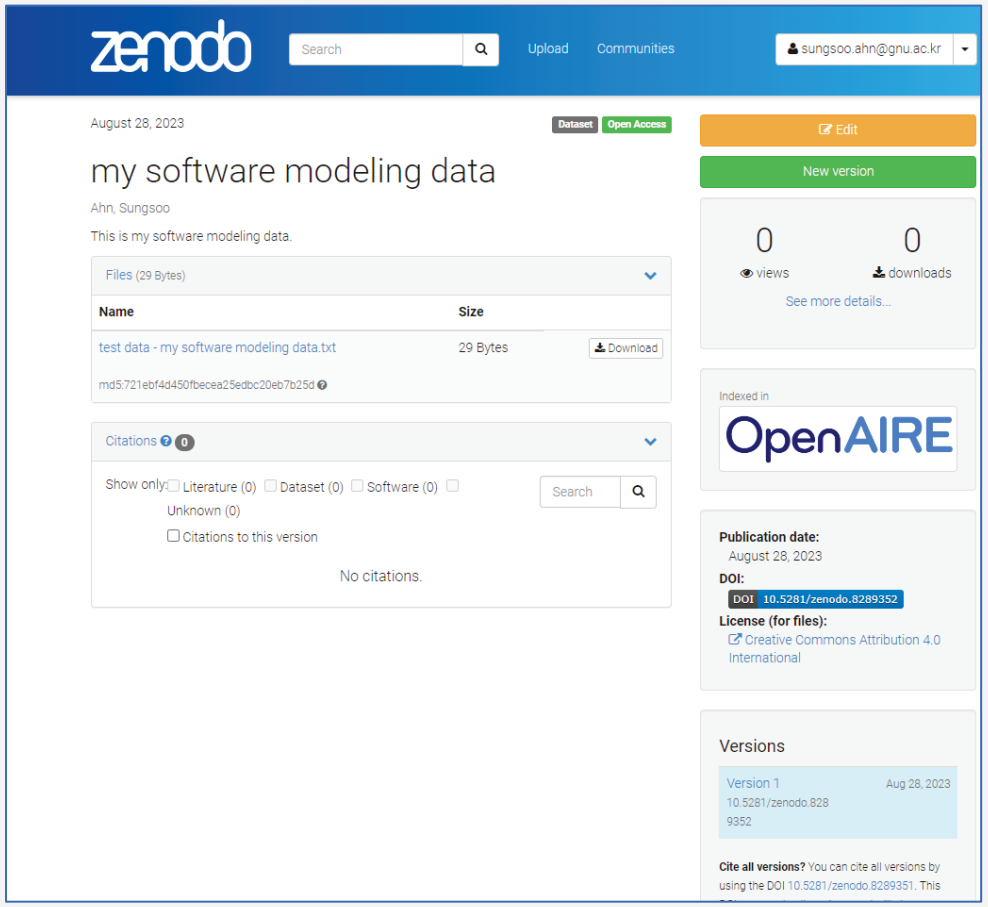


3: 데이터 접근 권한 선택, 라이선스 등



4: 선택사항 입력, 저장, 데이터 출판

Zenodo에 데이터셋 기탁하기 (3)

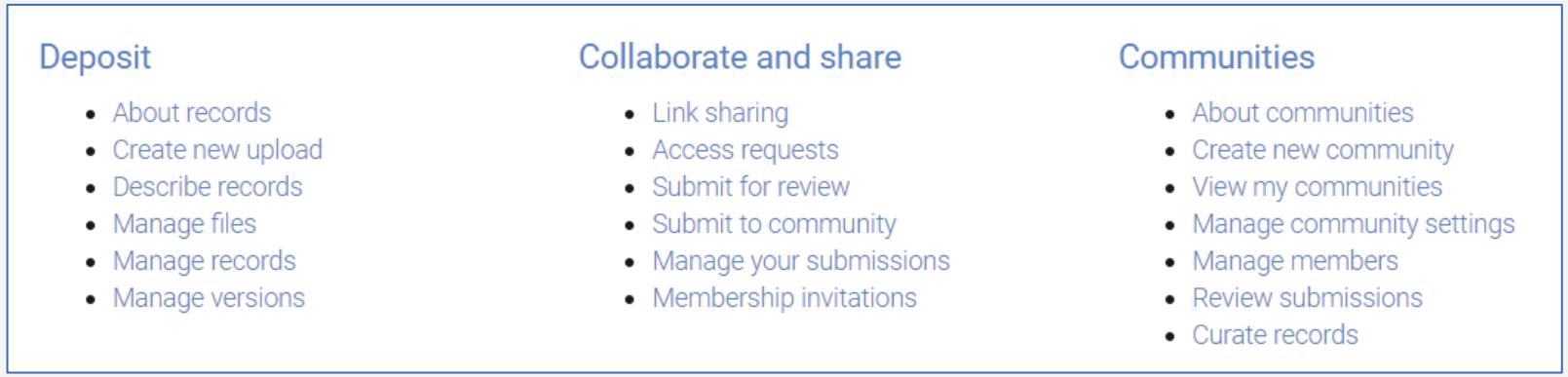


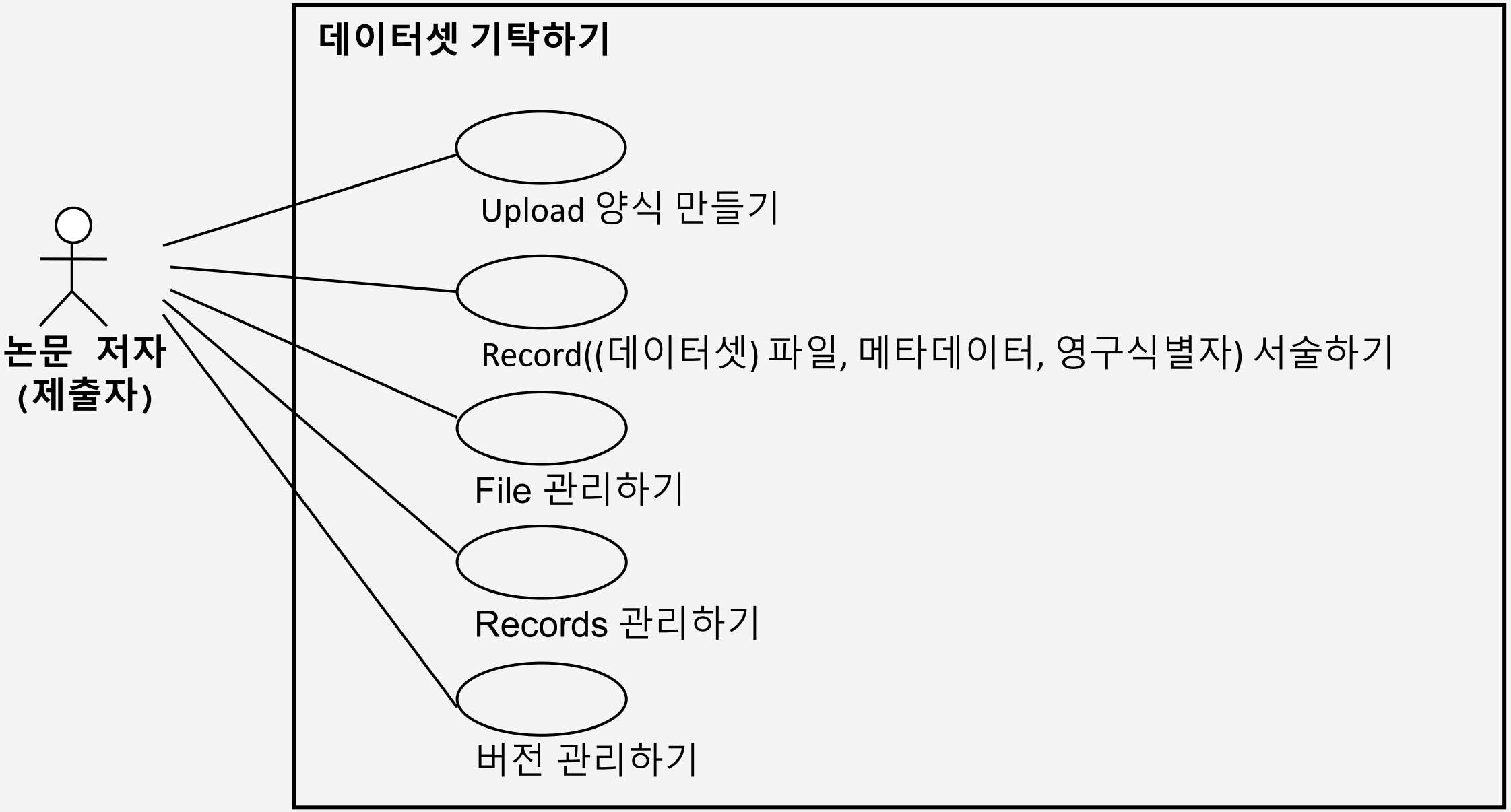
5: 데이터셋 업로드 후 화면

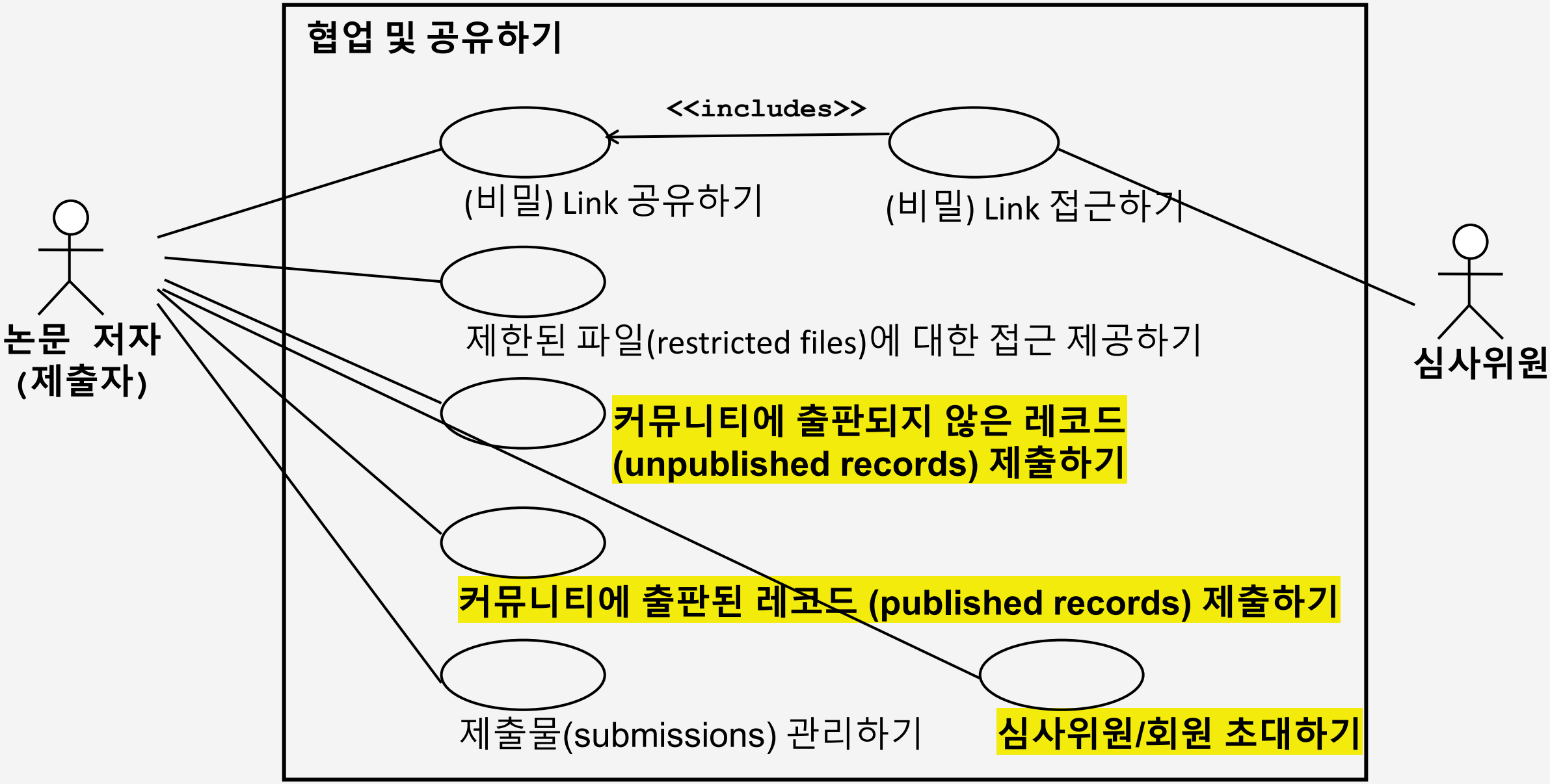
- 데이터셋 정보를 데이터논문의 데이터자원 항목에 명시
  - 출판일자: August 28, 2023
  - DOI: 10.5281/zenodo.8289352
- 관심 연구자는 이 데이터셋을 DOI로 접근하고 활용
  - https://doi.org/10.5281/zenodo.8289352
    - 데이터셋 심사
    - 데이터 다운로드, 재사용
    - 데이터분석을 통한 연구 아이디어 이해

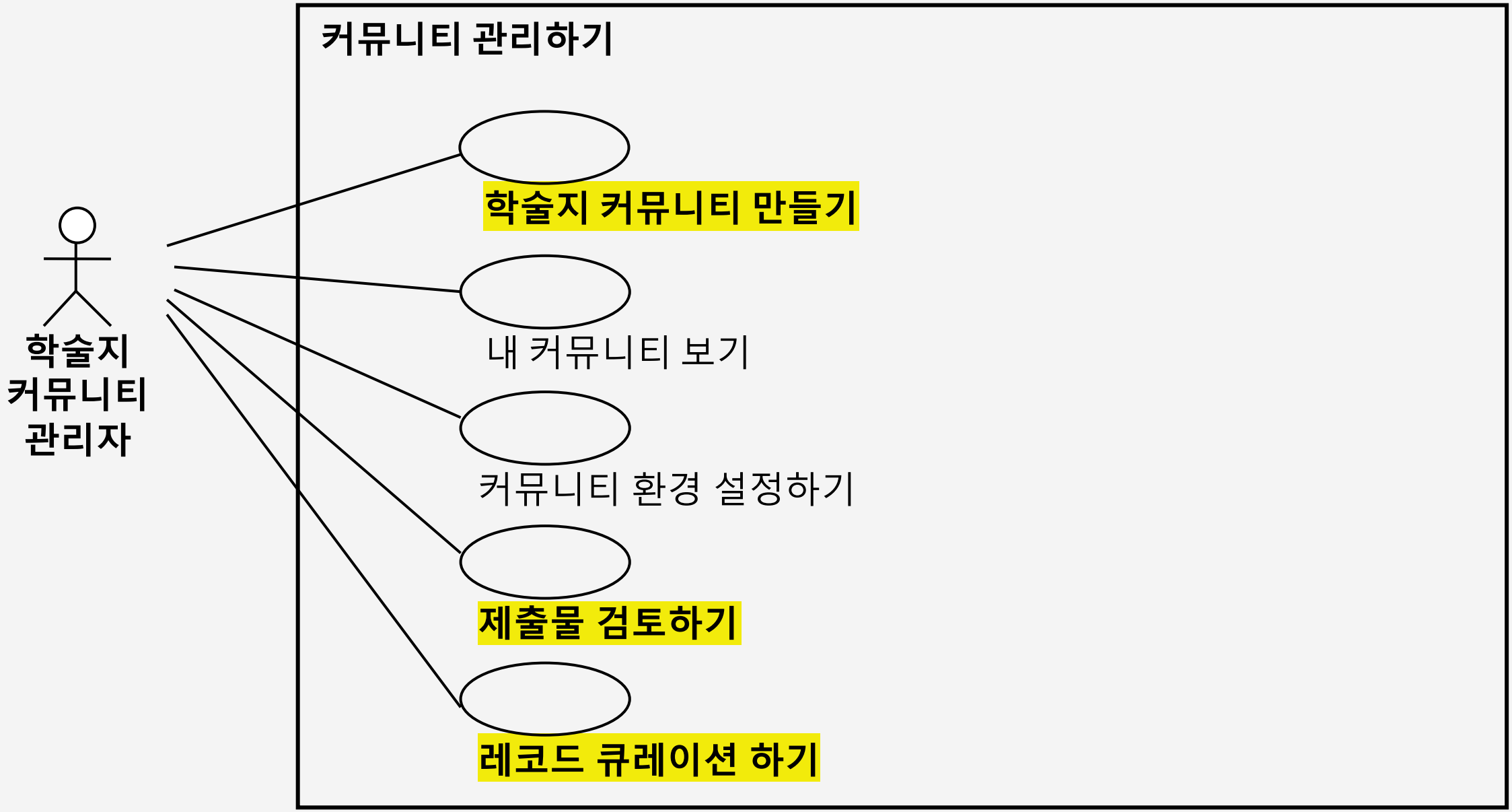
Zenodo 주요 서비스

- 기탁하기
  - 레코드(메타데이터, (데이터셋) 파일, 영구식별자) 탑재
  - 레코드, 파일, 버전 관리하기
- 협업 및 공유하기
  - 링크 공유하기
  - 검토 요청하기
  - 커뮤니티에 제출하기
- 커뮤니티 활동하기
  - 커뮤니티 만들기
  - 회원 관리하기
  - 레코드 큐레이션하기

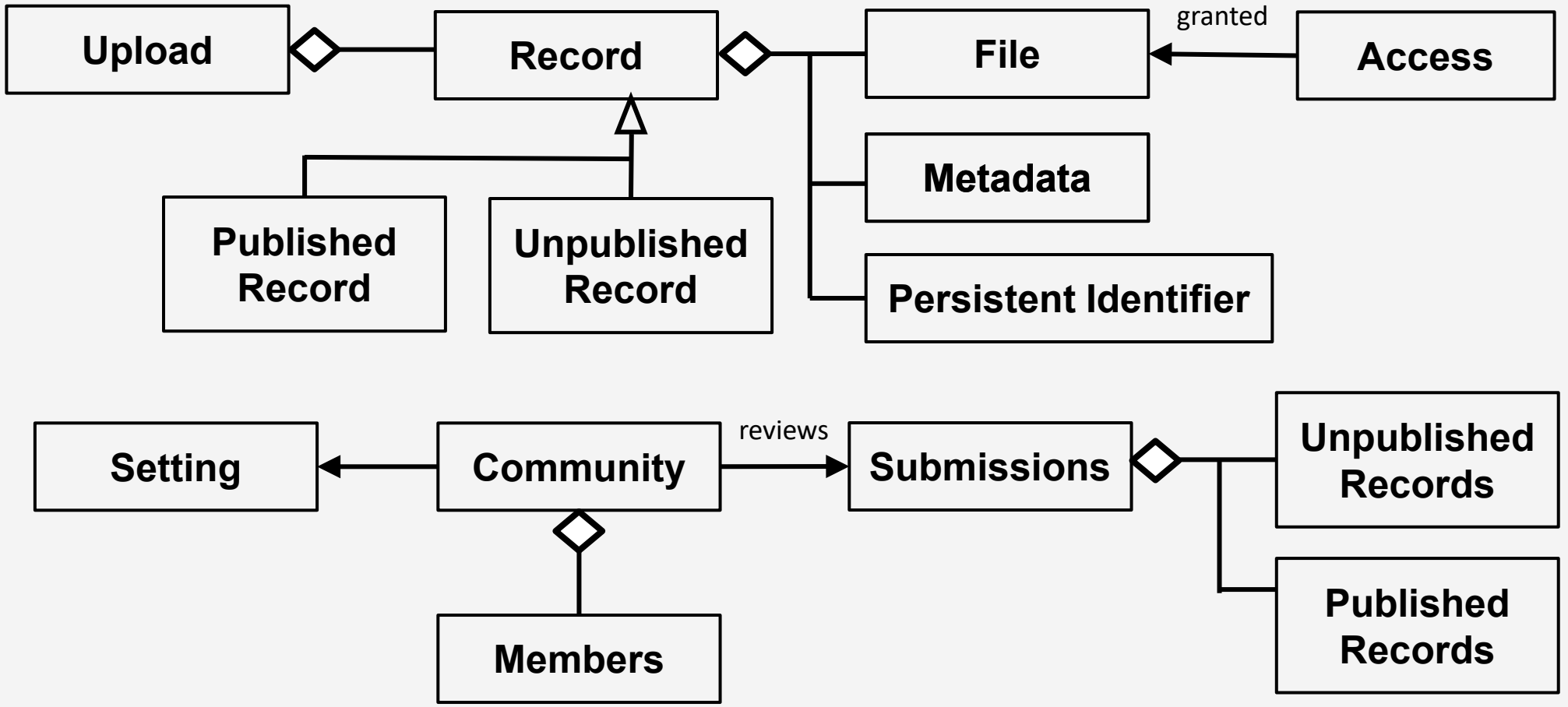



데이터셋 기탁하기
논문 저자
(제출자)
Upload 양식 만들기
Record((데이터셋) 파일, 메타데이터, 영구식별자) 서술하기
File 관리하기
Records 관리하기
버전 관리하기

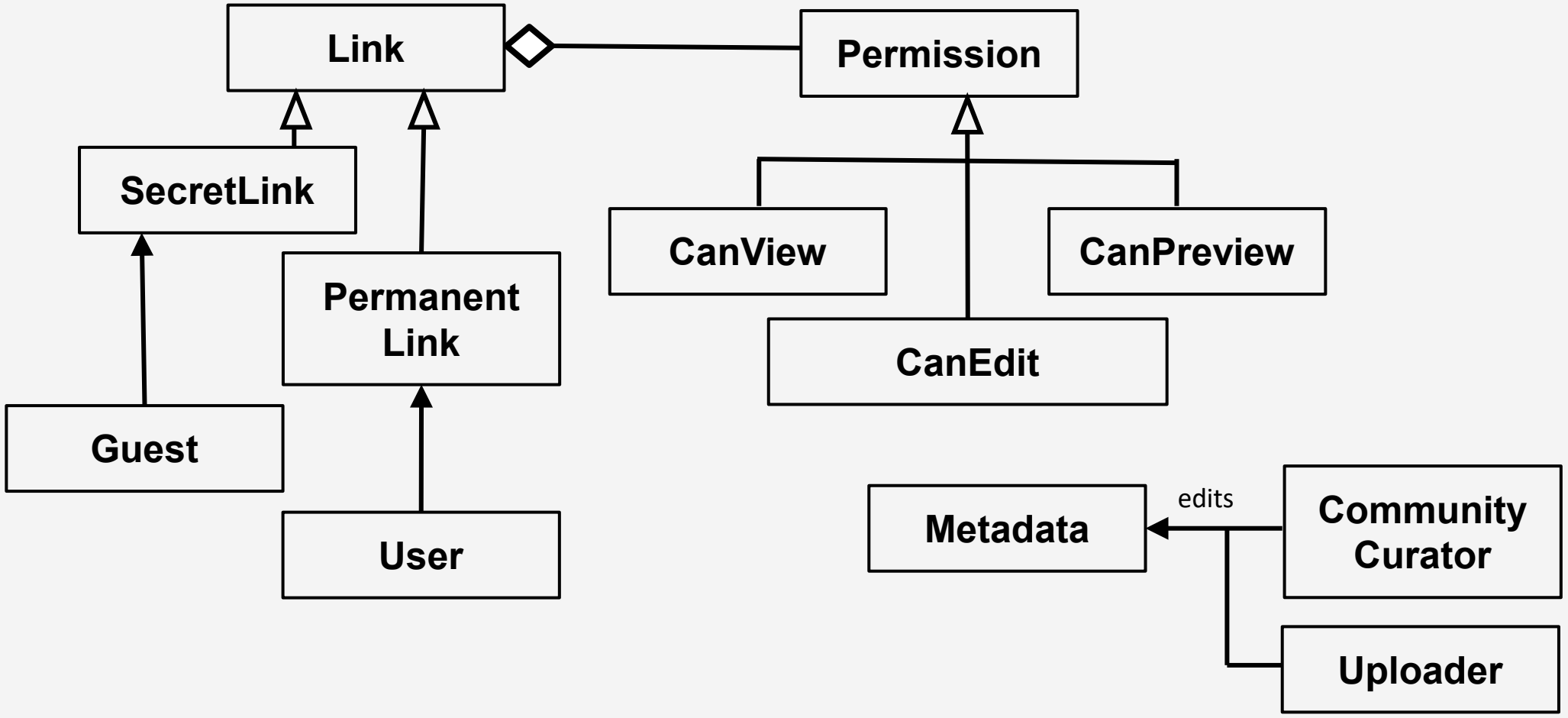

협업 및 공유하기
<<includes>>
(비밀) Link 공유하기
(비밀) Link 접근하기
논문 저자
(제출자)
제한된 파일(restricted files)에 대한 접근 제공하기
심사위원
커뮤니티에 출판되지 않은 레코드
(unpublished records) 제출하기
커뮤니티에 출판된 레코드 (published records) 제출하기
제출물(submissions) 관리하기
심사위원/회원 초대하기


커뮤니티 관리하기
학술지 커뮤니티 만들기
내 커뮤니티 보기
커뮤니티 환경 설정하기
제출물 검토하기
레코드 큐레이션 하기
학술지 커뮤니티 관리자

Zenodo 핵심 개념과 관계 모델 (진행 중)

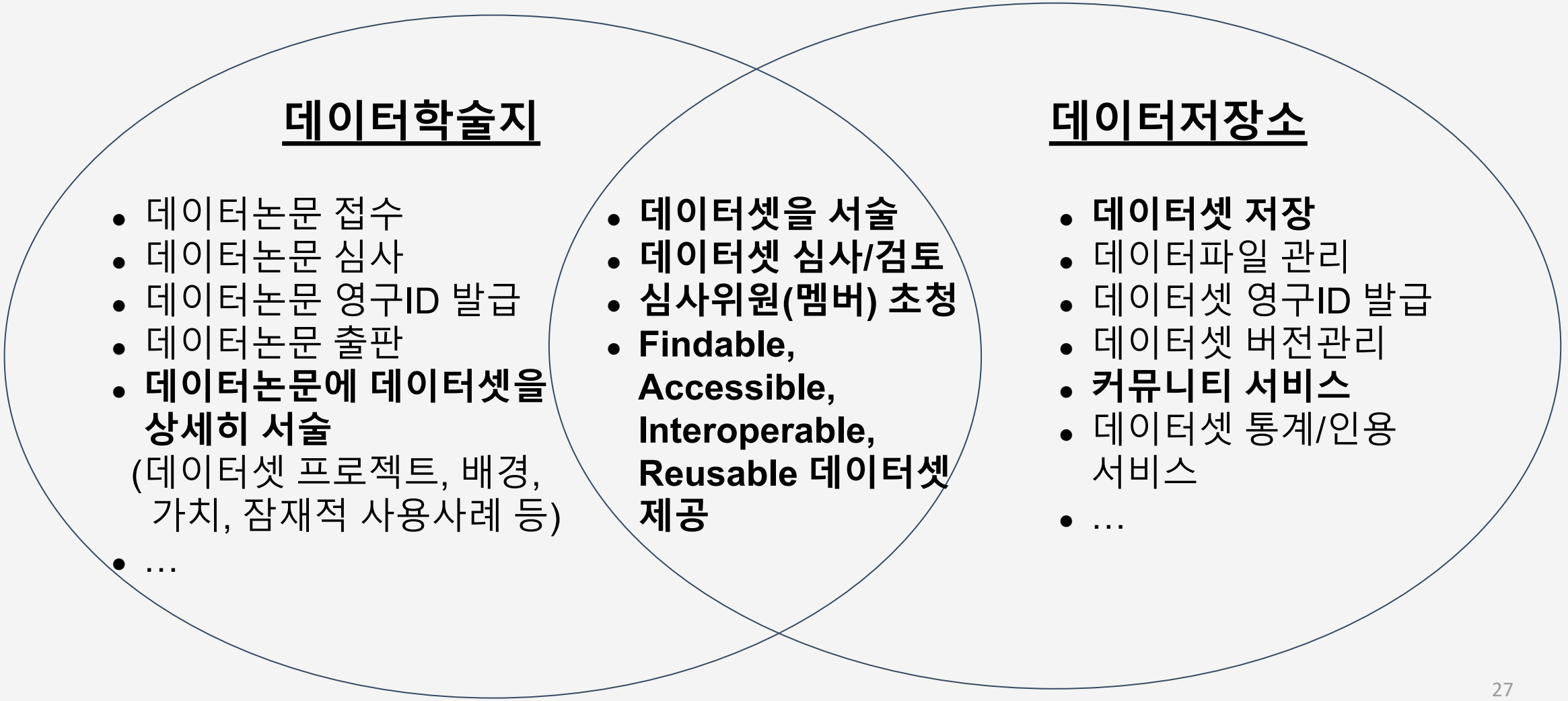


Zenodo 핵심 개념과 관계 모델 (진행 중)



3. 데이터학술지와 데이터저장소 협업 모델

데이터학술지,저장소 서비스의 공통점과 차이점



데이터저장소와 데이터학술지 비교

- 데이터저장소는 데이터셋의 보존, 공유, 재사용에 유용
  - 데이터셋의 메타데이터를 관리, 제공
  - 데이터 인용, 사용통계 서비스 등 제공
    - 일부, 커뮤니티 서비스 제공
  - 하지만, 데이터셋에 대한 심도있는 **동료심사**를 하지 않음
    - 메타데이터에 대한 필수/선택항목에 대한 체크는 진행

- 데이터학술지의 논문은 데이터셋의 문맥 정보 이해에 도움
  - 데이터셋의 메타데이터를 **더 체계적으로 관리, 제공**
    - 데이터 소개, 관련 프로젝트, 배경, 재사용 등 상세 정보
    - **데이터 생성 방법, 데이터 품질 관리 등** 문맥 정보를 이해할 수 있음
  - **동료심사를 진행**하며 데이터논문과 데이터셋의 품질 향상

협업 시나리오 (1/2)

1. 데이터학술지에 출판되는 데이터셋을 FAIR하게 관리하기 위한 **데이터저장소의 특정 커뮤니티 개설**하기
   - 데이터학술지 큐레이터 지정하기
   - 저자들에게 데이터셋 기탁(저장)/관리에 관한 안내 필요
   - 데이터논문 출판 후 데이터셋 비공개/삭제 방지

2. 데이터학술지에 출판될/된 **데이터셋을 큐레이터가 승인**하기
   - 데이터학술지 목표/범위에 맞는 데이터셋 승인
   - 데이터셋에 관한 메타데이터 기초적인 확인/검증
     - 데이터논문 심사시간 단축

협업모델 (2/2)

3. **데이터학술지 서비스**에서 **데이터저장소의 메타데이터, 주요 서비스**를 사용하기
   - 데이터논문 심사에서 활용
   - 데이터학술지 데이터논문 데이터와 데이터셋의 메타데이터 정보 제공
   - 출판된 데이터셋을 관리 (인용/통계 서비스 활용)
     - REST API를 메타데이터 접근하고 활용하기

4. 데이터학술지에서 데이터저장소로 데이터셋 파일 업로드하기
   - 데이터셋의 크기가 크지 않은 파일

데이터논문과 데이터저장소 연계 서비스 사례

- Scientific Data 학술지
  - Dryad 지원
  - Figshare에 데이터 업로드 지원

Repository Name	Information on fees/costs	Size limits	Integrated with *Scientific Data*'s manuscript submission system	Re3data / FAIRSharing entry
Dryad Digital Repository	$120 USD for first 20 GB, and $50 USD for each additional 10 GB	None stated	**Yes ✓**	view FAIRsharing entry
figshare	100 GB free per *Scientific Data* manuscript.	1 TB per dataset	**Yes ✓** - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. Download instructions.	view FAIRsharing entry
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	view re3data entry
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded	No	view FAIRsharing entry
Zenodo	Donations towards sustainability encouraged	50 GB per dataset	No	view re3data entry
Science Data Bank	Free of charge	8 GB per file, no limit to dataset size	No	view FAIRsharing entry

https://www.nature.com/sdata/policies/repositories#general

- J. of Open Humanities Data
  - 데이터 업로드 안내
    - Harvard Dataverse, 다른 데이터저장소 기탁 안내

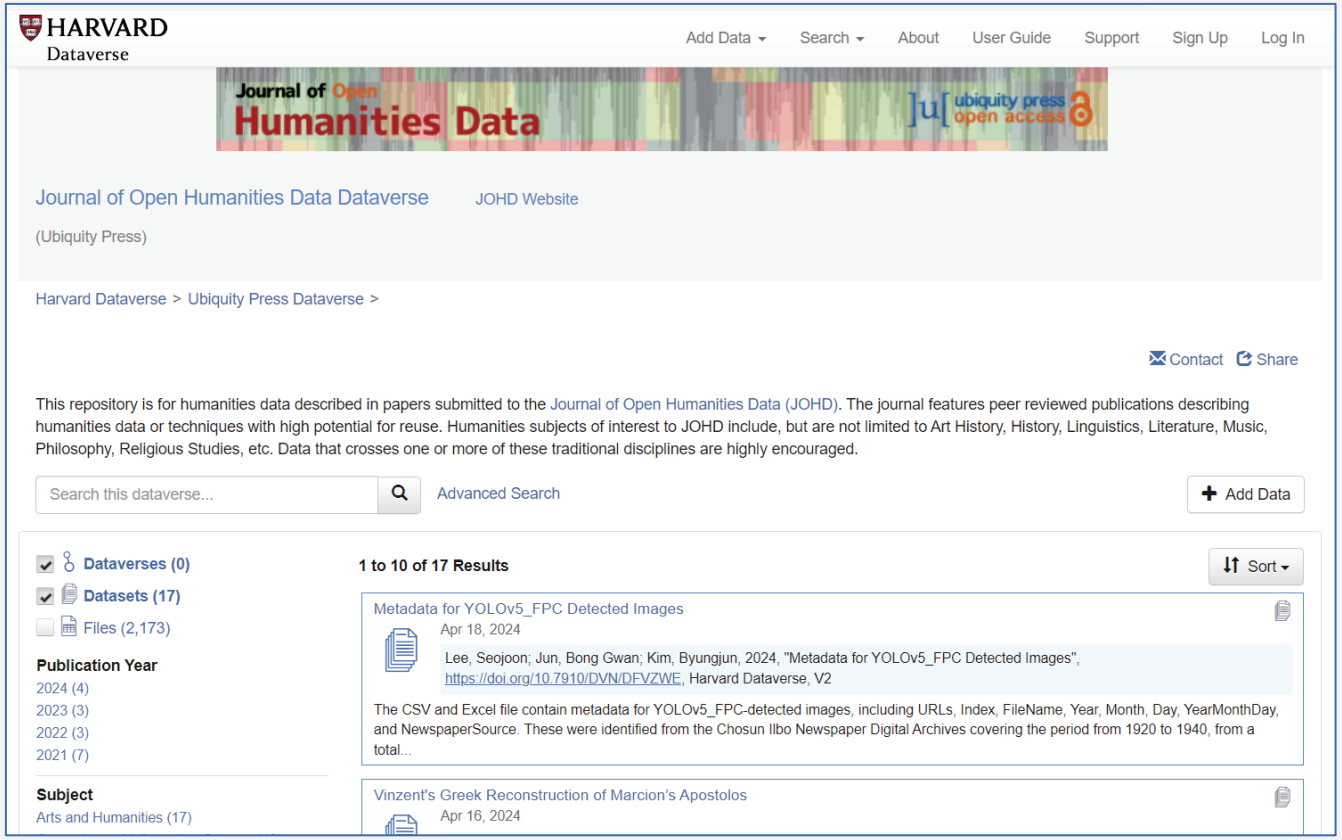
The Dataverse Network™ Project

Location	https://dataverse.harvard.edu/dataverse/johd
Focus and suitability	Data can be uploaded to the JOHD Dataverse Repository designed specifically for papers in the Journal of Open Humanities Data. We recommended this repository to authors because it is managed entirely by the JOHD editorial staff and ensures maximum interoperability between datasets and data papers.
Cost	Free for all JOHD authors.
Licenses	CC0
Identifiers used	DOI

https://openhumanitiesdata.metajnl.com/about

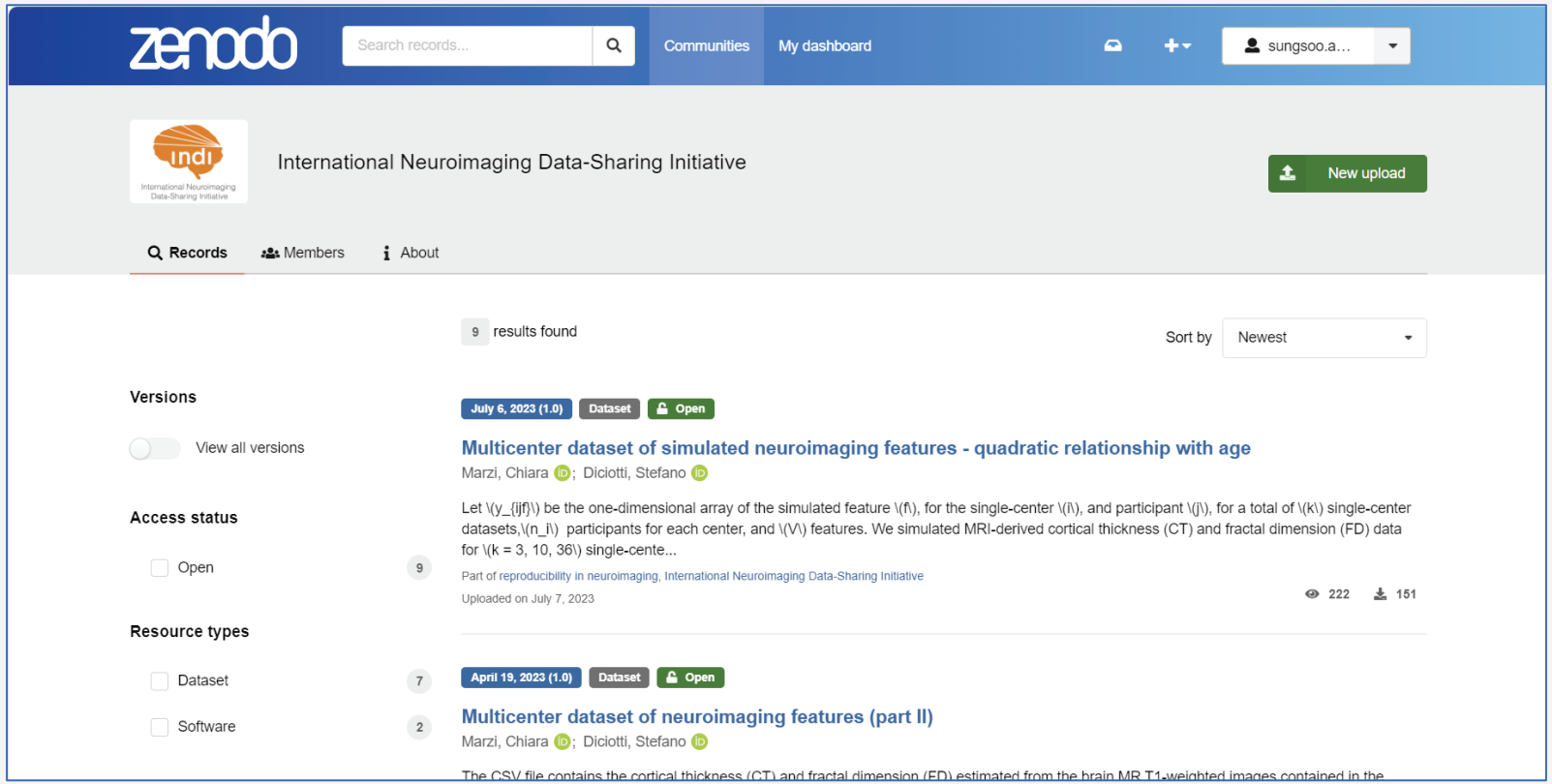
데이터논문과 데이터저장소 연계 서비스 사례

- J. of Open Humanities Data
  - Harvard Dataverse 데이터저장소에 데이터셋을 FAIR하게 관리



데이터논문과 데이터저장소 연계 서비스 사례

- International Neuroimaging Data-Sharing Initiative
  - 제노도에 연구데이터셋을 관리



결론

- 데이터학술지,저장소 서비스의 공통점
  - 데이터셋에 관한 데이터(메타데이터) 제공, 데이터셋 심사/검토
  - Findable, Accessible, Interoperable, Reusable 데이터셋 제공

- **데이터학술지와 데이터저장소간 협업모델 탐색 (시작 단계)**
  - **데이터저장소에 데이터학술지 커뮤니티 개설**하여 데이터셋 관리
    - 데이터학술지에 출판될/된 **데이터셋을 큐레이터가 승인**하기
  - 데이터학술지 서비스 확장
    - 데이터저장소로 데이터논문에서 서술하는 **데이터셋 파일 업로드하기**
  - **데이터학술지**에서 **데이터저장소의 메타데이터 기반 서비스 제공**

Any Questions?

E-mail: sungsoo.ahn@gnu.ac.kr

이 연구는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과 (2021RIS-003)입니다. 이 연구는 국가과학기술연구회 글로벌 공동연구 촉진사업 '연구데이터 동료심사 방안 및 데이터 영향력 분석 연구(글로벌-2022-012)' 과제의 지원을 받았습니다.