데이터 리포지터리 연계 Publish-Review-Curate 모델

2024. 6. 27

정 영 임[1, 2] 안성수[3]

[1] 한국과학기술정보연구원
[2] 연합과학기술대학교대학원
[2] 경상국립대학교

KiSTi www.kisti.re.kr 한국과학기술정보연구원 Korea Institute of Science and Technology Information

KISTI SINCE 1962

KISTI SINCE 1962

데이터 저널과 데이터 논문의 정의

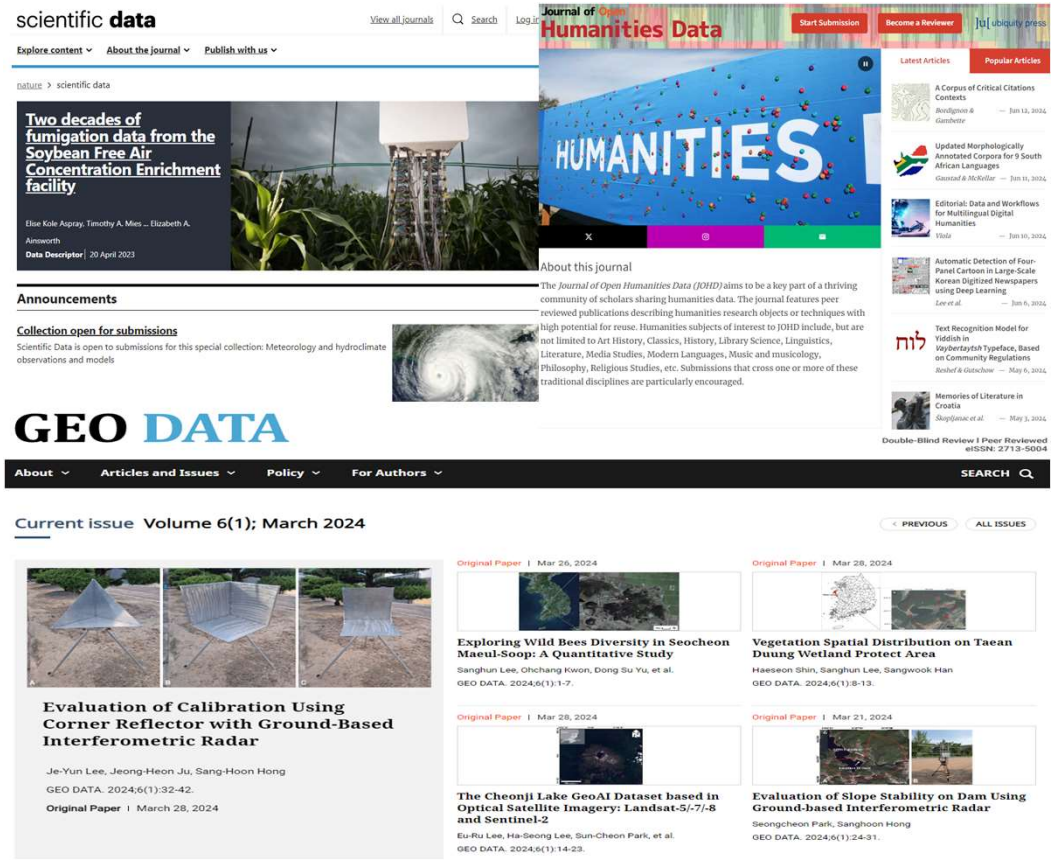
- 데이터 논문은 데이터 기술을 목적으로 동료심사 저널에 출판되는 논문으로 저널별로 다양하게 지칭되고 있음 1)
- 데이터 논문만을 포함하는 순수 데이터 저널과 기존 연구 논문을 포함하는 하이브리드 데이터 저널

저널별 데이터 논문의 명칭 및 정의

Terminology	Definition	Journals
data article	Data articles are brief, peer-reviewed publications about research data. Data articles are peer-reviewed, citable papers that describe research data.	Chemical Data Collections, Data in Brief, Genomics Data, Data Science Journal
data descriptor	Scientific Data primarily publishes Data Descriptors, a new type of publication that provides detailed descriptions of research datasets, including the methods used to collect the data and technical analyses supporting the quality of the measurements.	Scientific Data
data note	Data notes briefly and concisely describe sound research data with the aim of increasing visibility and transparency, supporting the reuse of valuable research data and helping authors to comply with funder mandates on data sharing.	BMC Research Notes
data paper	- presenting a data set, the method of its generation, intended use and, possibly, short description of the results of such use - A data paper is a peer-reviewed document describing a dataset, published in a peer-reviewed journal.	Pensoft journals, Biomedical Data Journal, Ecological Archives, Ecology, Ecological Research, Geoscience Data Journal
data-based modeling and models paper	presenting novel approaches to modelling on the basis of experimental or clinical data and resulting models	Biomedical Data Journal
database article	Database articles should describe a novel biomedical database likely to be of broad utility.	BMC journals
dataset paper	(no definition)	Hindawi journals
full-length article	(no definition but by journal description) - The journal is devoted to the publication of tables and graphs of general usefulness to researchers in both basic and applied areas. - The journal is mostly produced from Evaluated Nuclear Structure Data File (ENSDF), a computer file maintained by the US National Nuclear Data Center	Atomic Data and Nuclear Data Tables, Nuclear Data Sheets

1) 정영임 외(2019)

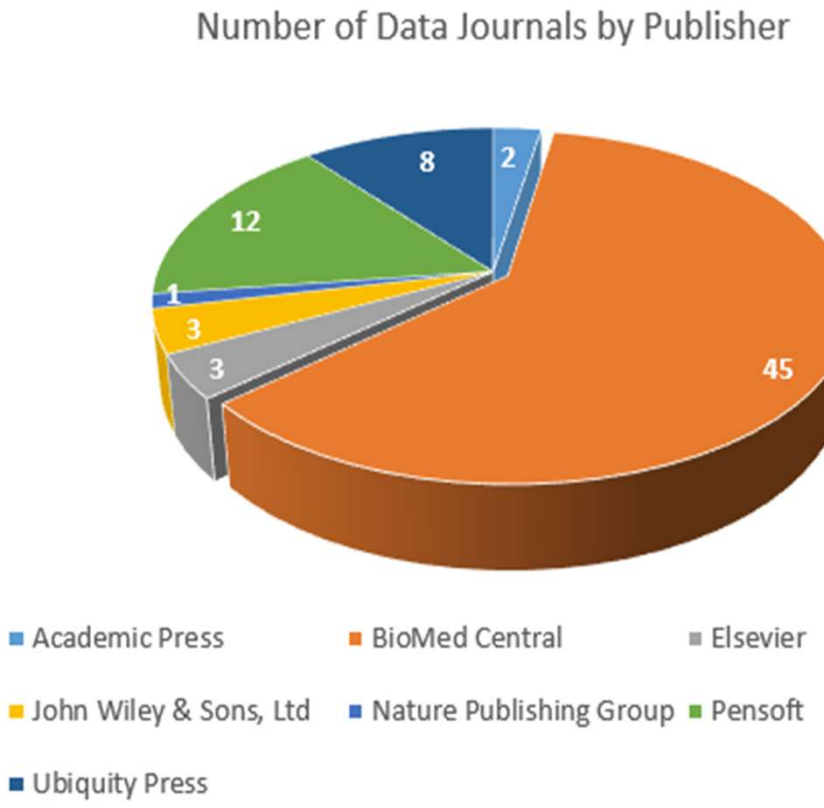
데이터 저널



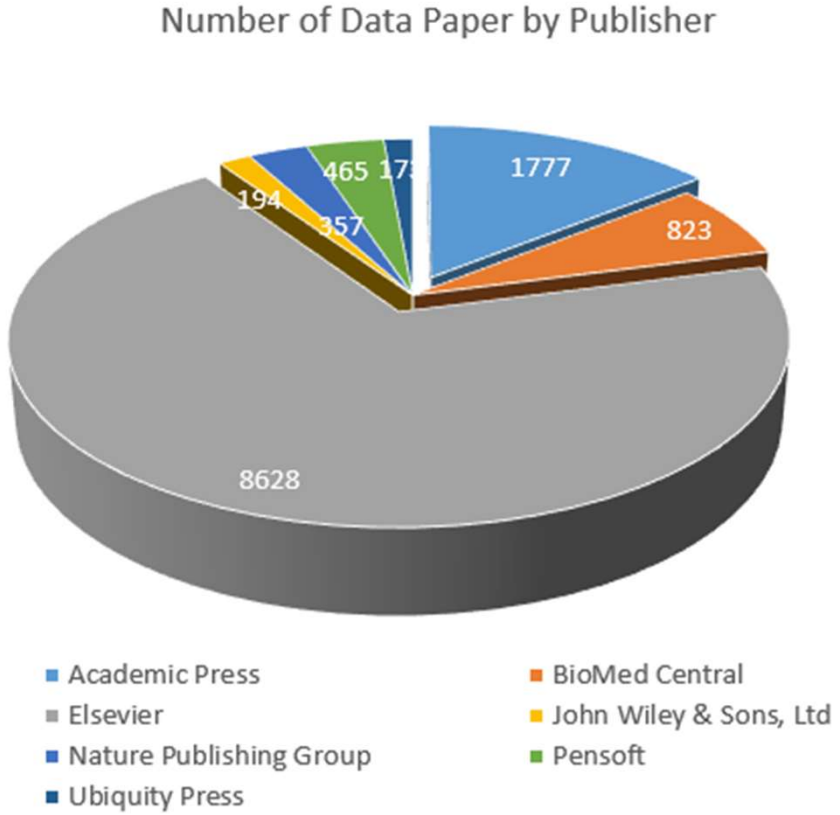
글로벌 데이터 저널 규모 및 출판 현황

- 데이터 저널은 BMC, Pensoft, Ubiquity Press 등 신생 OA 출판사에서 활발하게 발간 [1)]
- 출판된 데이터 논문 건수로 보면 Elsevier, Academic Press, NPG 등 전통적 거대 출판사 논문 수가 압도적

출판사별 데이터 저널 발간 현황



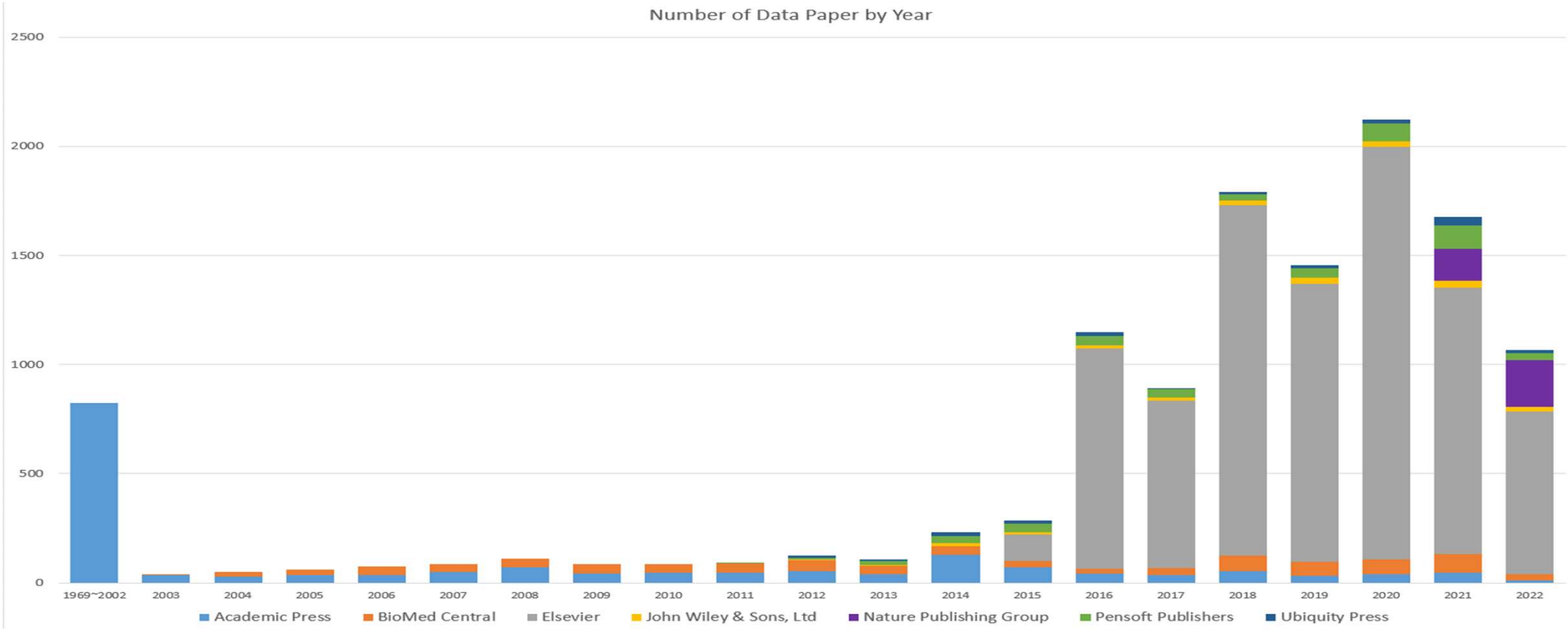
출판사별 데이터 논문 출판 현황



1) 정영임 외(2019) 생태학 분야 데이터 저널 발간 전략 수립

글로벌 데이터 논문 출판 추이 2)

- Academic Press는 Nuclear Data Sheets와 Atomic Data and Nuclear Data Tables을 통해 지속적으로 출판
- BMC는 BMC Plant Biology, BMC Microbiology을 통해 2000년 초부터 꾸준히 데이터 논문 출판
- "메가" 데이터 저널인 Data in Brief와 Scientific Data 등장으로 데이터 논문 출판의 기하급수적 성장



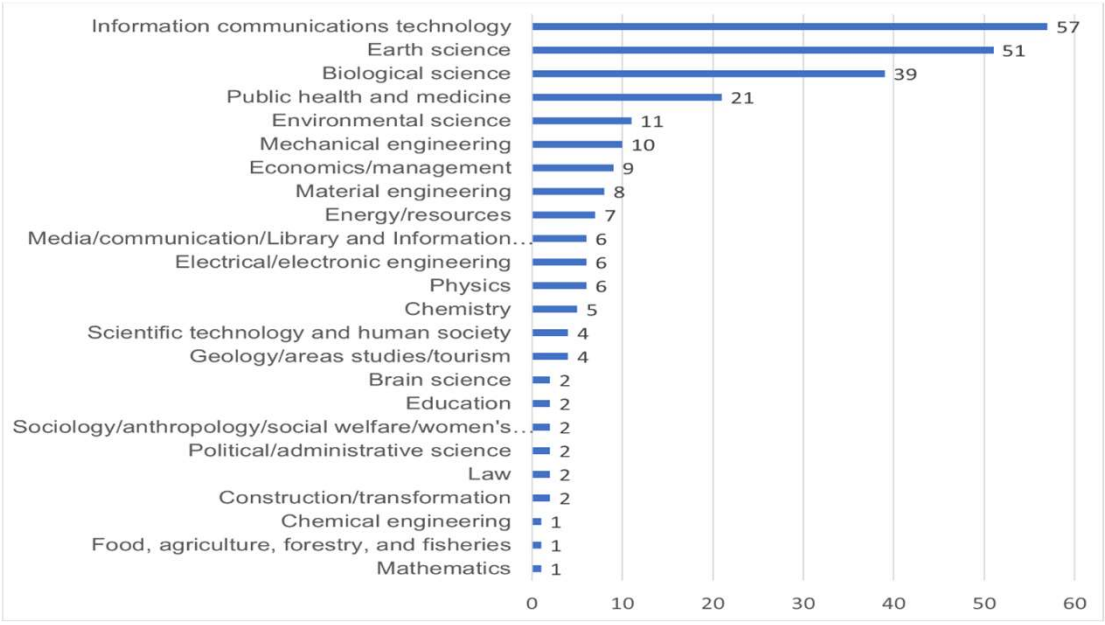
2)정영임 외(2022) 데이터 논문 출판 현황 분석

데이터 논문 출판 인프라 구축 수요 증가 3)

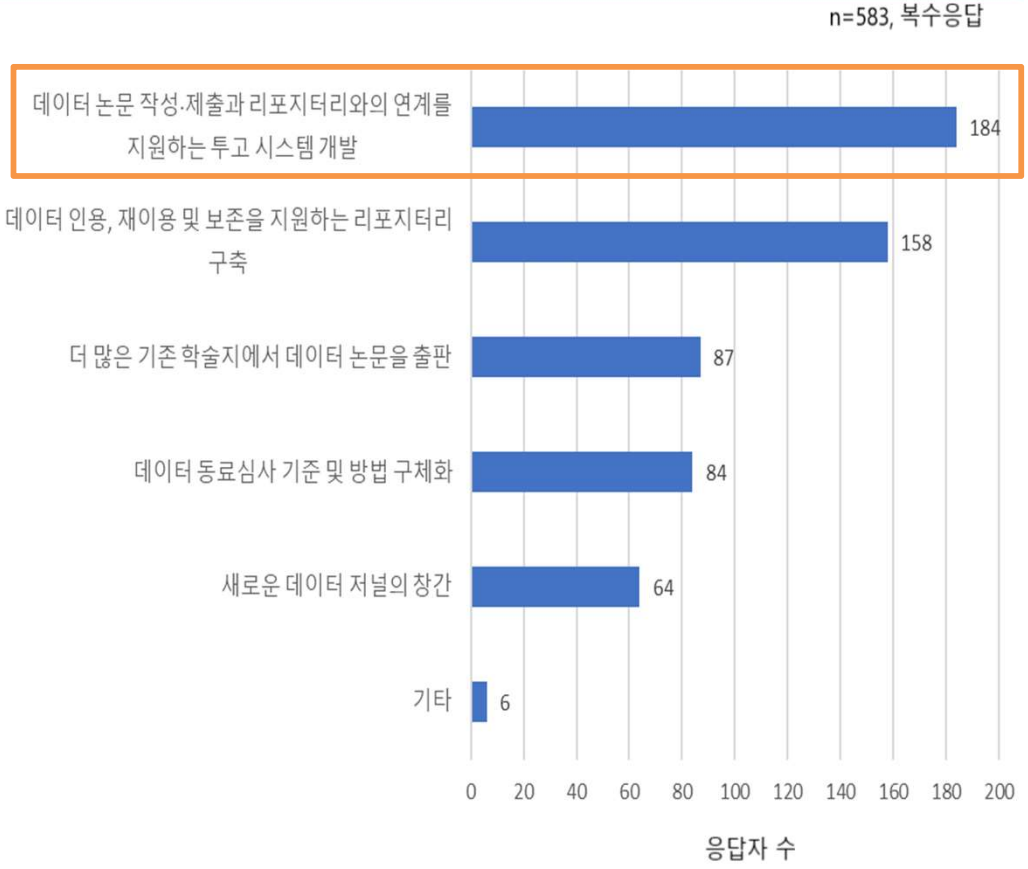
☑국내 연구자 대상 설문조사 결과 연구데이터 출판 촉진 요인으로 데이터 논문과 리포지터리와의 연계 지원 투고시스템 개발, 데이터 인용, 재이용 및 보존을 지원하는 리포지터리 구축 등 인프라 구축을 중요 요인으로 인식함.

데이터 출판 인식 설문조사

- 주제: 연구자 대상 데이터 출판에 대한 인식과 태도
- 기간: '22.06.15 ~'22.07.05
- 대상: 과기 분야 5개 출연연 및 GeoAIData 학회 소속 연구자
- 응답: 총 259명 응답



설문조사 결과



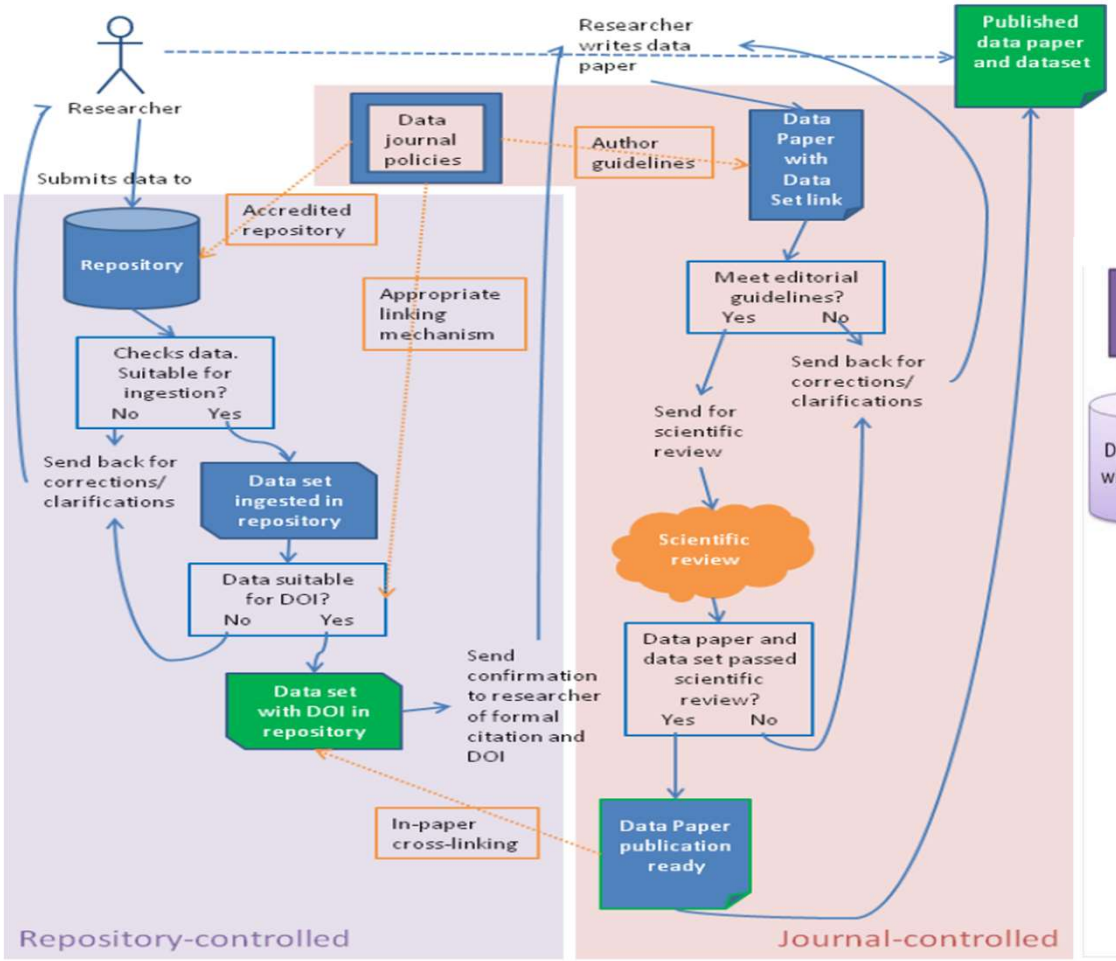
3) 황혜경 외(2023) 과학기술분야 연구자들의 데이터 출판경험 및 인식 연구

데이터 리포지터리 연계 데이터 논문 동료심사 모델 개발

- 신속한 동료 심사를 통한 데이터 논문 출판 주기 단축 및 데이터 논문과 연구데이터 가시성 증대 필요
- 데이터 논문과 데이터에 대한 전문적 기술 검토, 일관성 검토 절차 필요
- 데이터 리포지터리와 데이터 저널 시스템 간 기여자별 역할 정의 및 데이터 교환 요소 기술 필요
- 해당 요구 사항을 만족하는 데이터 논문 동료심사 모델 개발

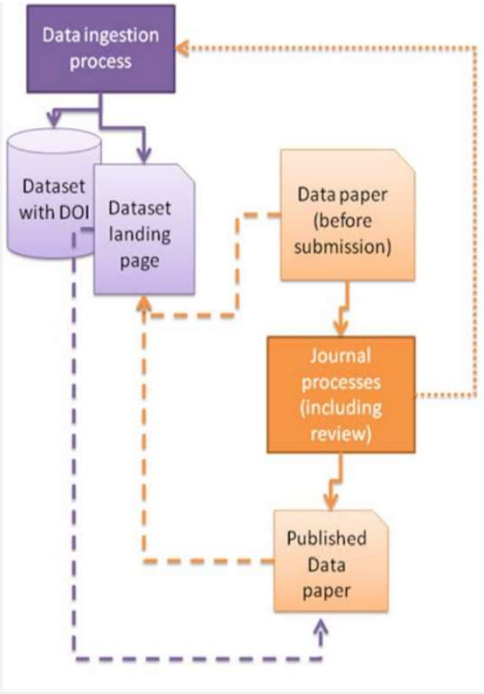
데이터 리포지터리와 데이터 저널 시스템 연계 워크플로우 4)

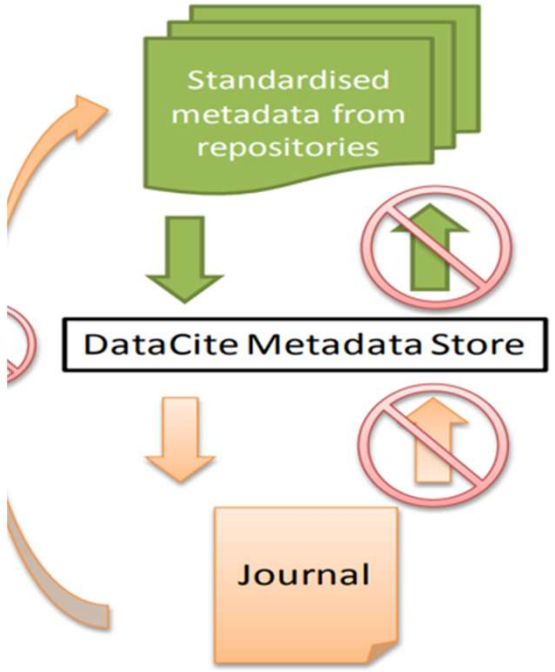
- 데이터 논문과 기술(참조)하는 데이터 간 **교차 링크 부여 방안만 제시**
- **다단계에 걸친 중복된 데이터 검토** 및 데이터 논문 심사 후 출판 승인 절차로 **데이터 논문 출판 지연**



3 main areas of interest (in orange)

1. Workflows and cross-linking between journal and repository
2. Repository accreditation
3. Scientific peer-review of data

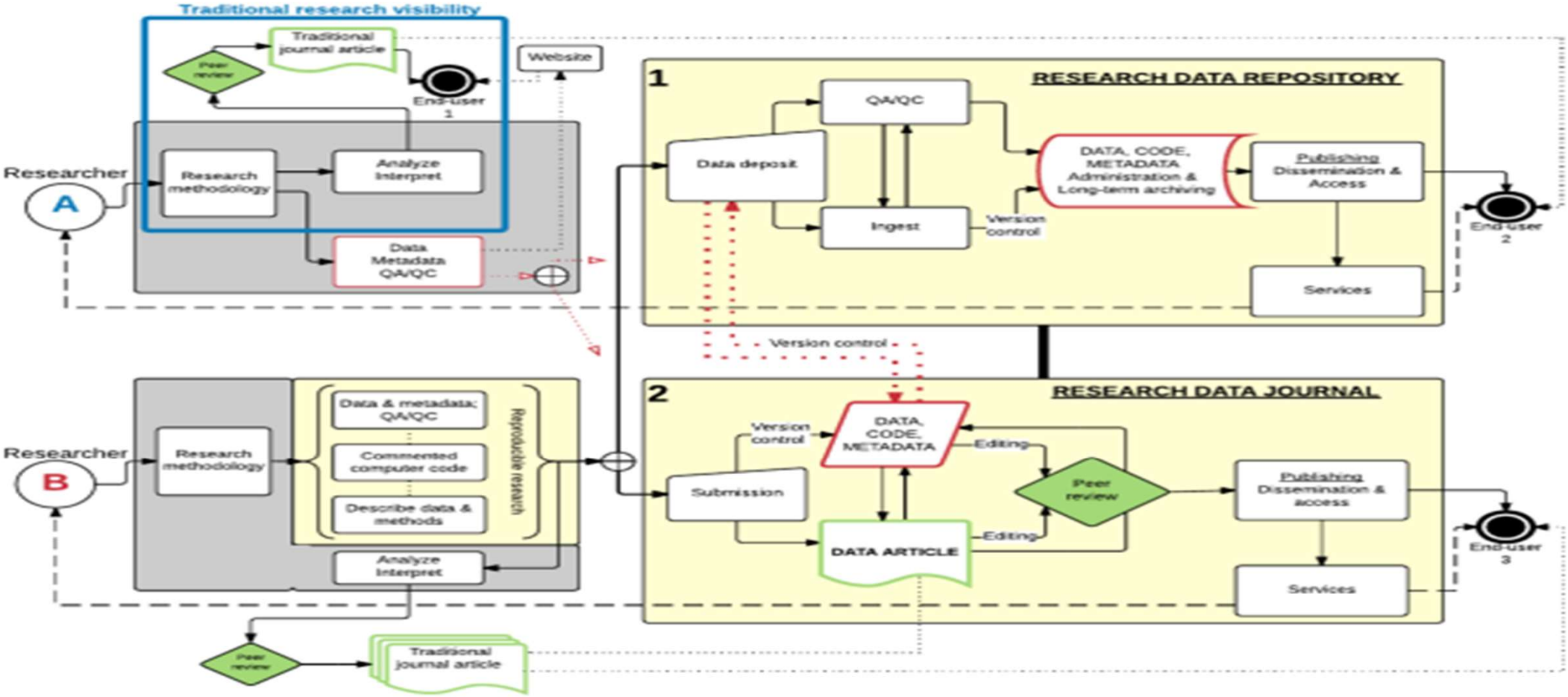




4) Callaghan (2013.04) Data Centre Workflows for publication

데이터 리포지터리와 데이터 저널 시스템 연계 워크플로우 5)

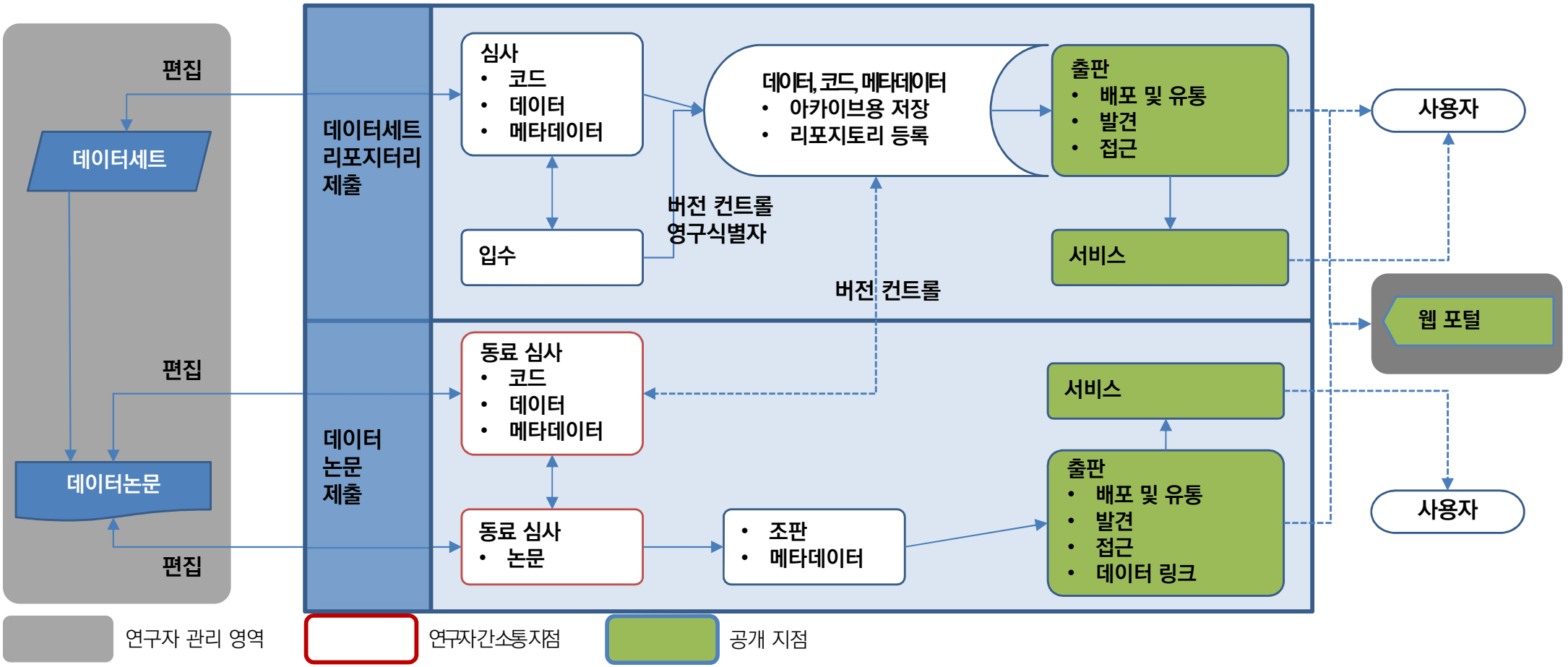
- 전통적인 연구 논문 대비 데이터 논문의 동료심사 및 출판 워크플로우 간 차이점 제시
- 데이터 및 데이터 논문 동료심사에서 각 기여자(editor, reviewer, author)간 역할 및 교환 데이터 기술 미흡



5) Bloom et al (2015)

데이터 리포지터리와 데이터 저널 시스템 연계 워크플로우 6)

- 선행 연구와 유사한 절차임. 기여자별 관리 영역 정의하였으나 상세한 역할 정의 기술 미흡
- 데이터 리포지터리와 데이터 저널 시스템 간 연계 미흡(데이터 버전 컨트롤 및 서비스단 연계만 제시)



6) Austin et al (2017)

신속한 학술 커뮤니케이션을 지원하는 PRC 모델 [7]

- 신속한 학술 커뮤니케이션 수요 증가로 Publish-Review-Curate(PRC) 모델 제안 및 도입 학술지 증가
- 기존 연구 논문의 preprint 출판 및 심사 절차만 개괄적으로 기술, 데이터 논문에 적용을 위한 상세 기술 필요

eLife의 PRC 모델 절차

- **제출**: eLife 편집자가 리뷰를 위해 제출된 논문을 선별
- **동료 심사**: 분야 전문가가 선별된 논문 심사. 해당 단계에서 저자에게 출판비가 청구됨.
- **출판**: 심사된 논문은 저널 웹사이트에 'Reviewed Preprints'로 게시됨.
- **저자 수정**: 출판 후 저자가 수정본 제출 여부를 결정함. 논문을 수정할 경우, 위와 같이 다시 심사 후 게재 절차를 거침.
- **출판본(Version of Record)**: 이 반복 프로세스 중 언제든지 저자는 논문의 현재 버전을 출판 버전이라고 선언할 수 있음. 저널 요구 사항이 완료되면 색인 및 큐레이션을 위해 해당 논문의 메타데이터는 PubMed로 전송됨.

PRC 모델 도입 학술지

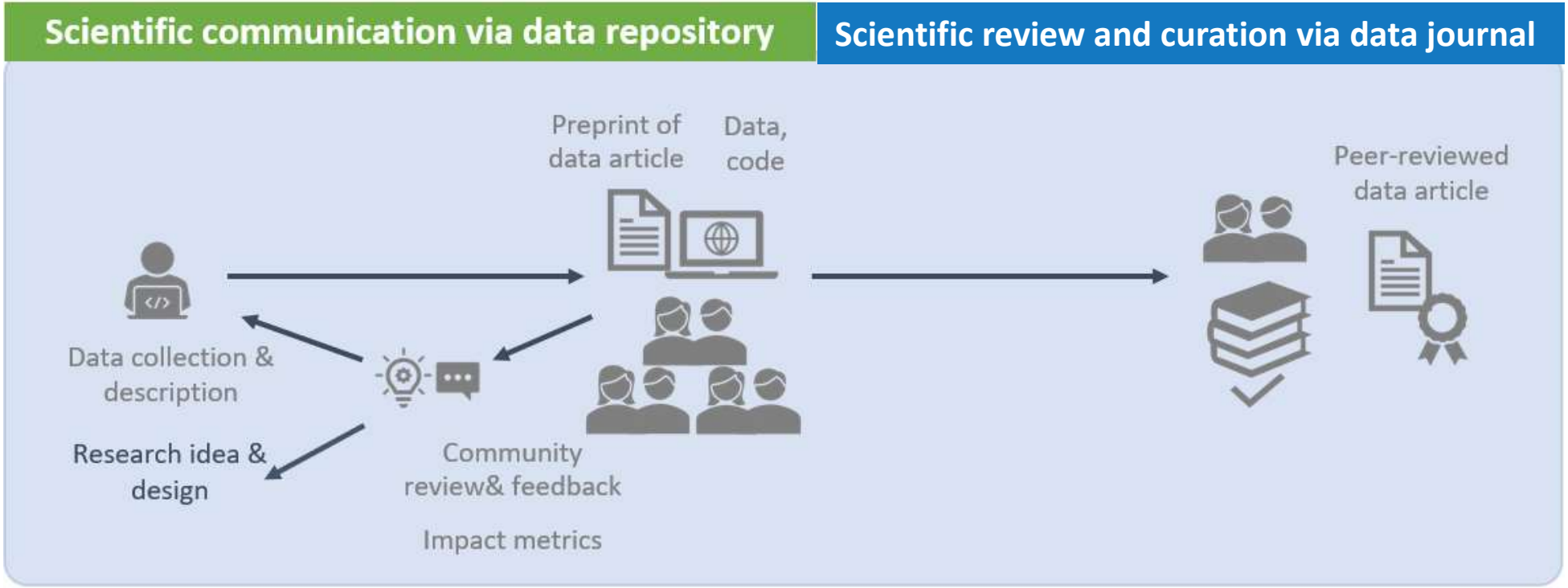
7) EisenAnna et al (2020) Peer Review: Implementing a "publish, then review" model of publishing

신속한 데이터 출판을 지원하는 PRC 모델

- 일반 연구논문과는 달리 연구데이터 기술을 목적으로 하는 데이터 논문은 프리프린트 형태의 신속한 출판 및 확산으로 인한 부작용이 적을 수 있음
- 데이터 논문의 동료심사에 주제 분야 전문 심사뿐만 아니라 데이터세트의 무결성, 형식적 일관성 등을 평가할 수 있는 데이터 전문가에 의한 기술 심사, 데이터세트의 유용성, 재사용성에 대한 공개 평가 등이 결합되면 데이터 논문에 대해 보다 전문적이고 포괄적인 동료심사가 가능
- "연구자 대상 데이터 출판에 대한 인식과 태도" 설문조사에서 대다수의 연구자들은 일반 연구논문의 동료 심사와는 다른 절차와 기준, 심사 기준, 다른 접근 방식이 데이터 논문 동료심사 및 출판에 필요함을 역설

PRC 모델을 적용한 데이터 논문 심사 과정

- ☑ 데이터 리포지터리에 데이터 및 데이터 논문 프리프린트 기탁을 통한 데이터 품질 검사, 논문의 기술적 심사와 출판 가속화
- ☑ 리포지터리 내 분야 전문가로부터 공개 피드백, 이용/인용지표 축적을 거쳐 선정된 프리프린트의 저널 투고
- ☑ 저널 내 과학적 동료 심사를 거쳐 출판본 색인 및 큐레이션 서비스

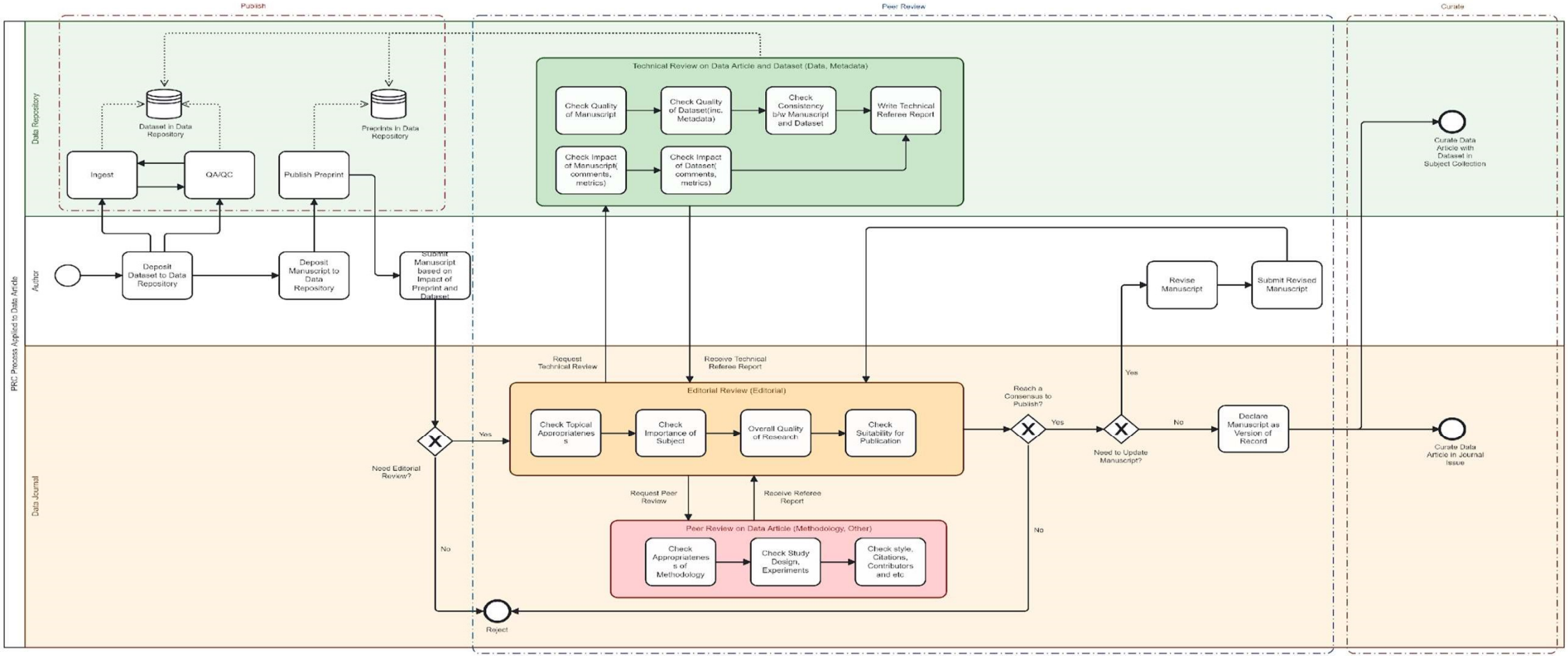


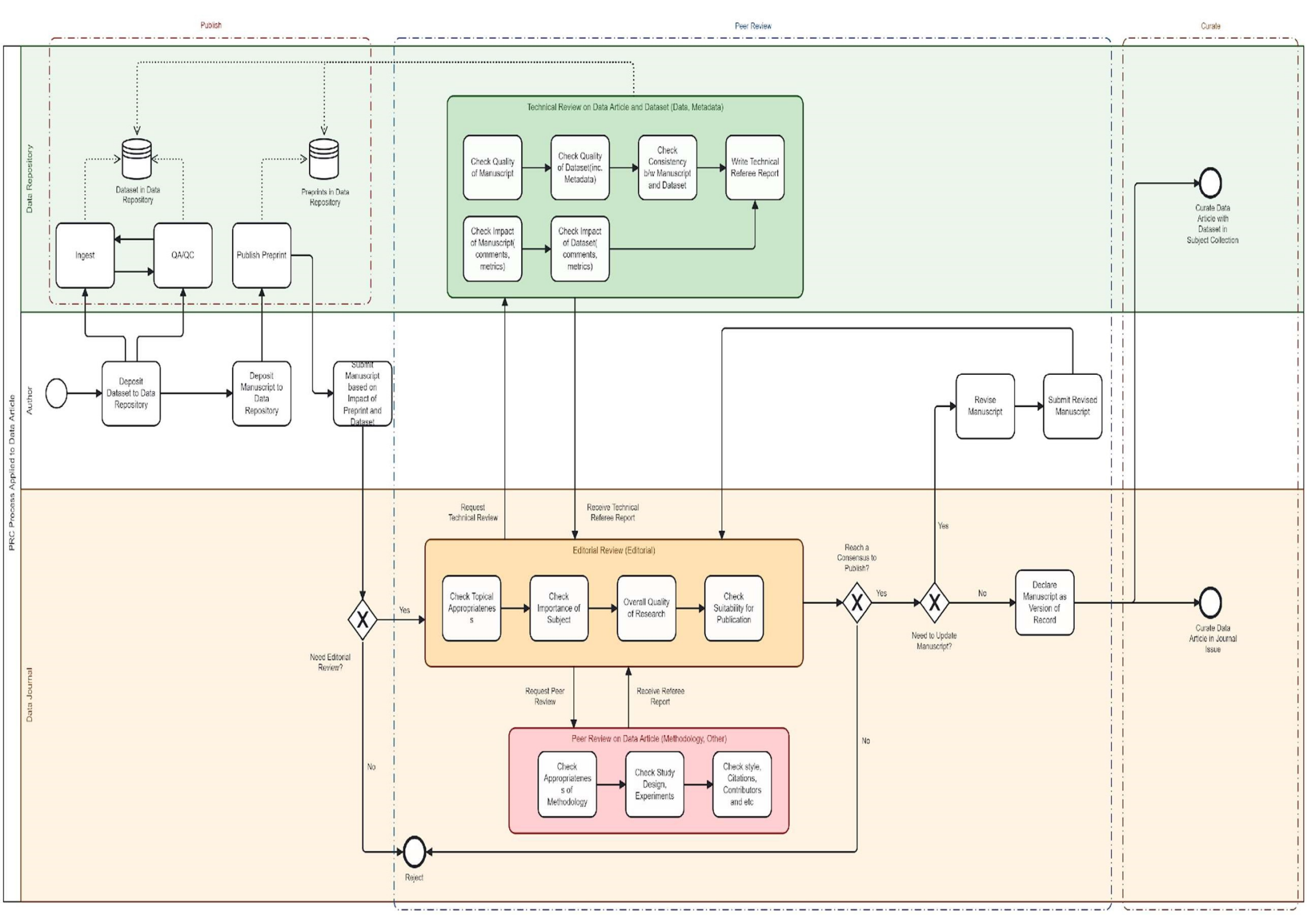
➡ 신속한 가시화 및 커뮤니케이션, 일관적이고 전문적인 기술 심사, 과학적 심사 가능

KISTI SINCE 1962

PRC 모델을 적용한 데이터 논문 심사

- ☑ 데이터와 데이터 논문의 Publish– Review–Curate 단계별 관여 기여자 역할의 명확한 정의
- ☑ 데이터 리포지터리와 데이터 저널 시스템 간 프로세스별 교환 데이터의 명확한 기술
- ☑ 중복된 과정 없이 전문적인 기술 심사, 과학적 동료심사, 편집 심사 협업 및 연계 모델링




Publish
Peer Review
Curate
PRC Process Applied to Data Article
Data Repository
Author
Data Journal
Dataset in Data Repository
Preprints in Data Repository
Ingest
QA/QC
Publish Preprint
Technical Review on Data Article and Dataset (Data, Metadata)
Check Quality of Manuscript
Check Quality of Dataset(inc. Metadata)
Check Consistency b/w Manuscript and Dataset
Write Technical Referee Report
Check Impact of Manuscript(comments, metrics)
Check Impact of Dataset(comments, metrics)
Curate Data Article with Dataset in Subject Collection
Deposit Dataset to Data Repository
Deposit Manuscript to Data Repository
Submit Manuscript based on Impact of Preprint and Dataset
Revise Manuscript
Submit Revised Manuscript
Request Technical Review
Receive Technical Referee Report
Editorial Review (Editorial)
Check Topical Appropriatenes s
Check Importance of Subject
Overall Quality of Research
Check Suitability for Publication
Reach a Consensus to Publish?
Yes
No
Need to Update Manuscript?
Declare Manuscript as Version of Record
Curate Data Article in Journal Issue
Need Editorial Review?
Yes
No
Request Peer Review
Receive Referee Report
Peer Review on Data Article (Methodology, Other)
Check Appropriatenes s of Methodology
Check Study Design, Experiments
Check style, Citations, Contributors and etc
Reject

데이터 리포지토리 연계 데이터 및 데이터 논문 심사 모델 제안

- 급격하게 증가하는 데이터 저널과 데이터 논문 출판 수요
- 일반 연구논문과 차별되는 데이터 논문의 신속한 출판, 동료심사 및 큐레이션 서비스 모델 제안함
- 데이터 리포지터리 연계 데이터 및 데이터 논문의 PRC 모델은 신속한 가시화 및 학술 커뮤니케이션, 일관적이고 전문적인 기술 심사, 과학적 심사 가능

데이터 논문 PRC 모델 적용을 위한 추가 고려사항

☑ **Policy issues** in data repository

- Ingesting/managing/versioning preprint of data articles
  - How different preprint of data article should be from the metadata of dataset or data management plan

☑ **Technical issues** in data exchange between journal and repository

- Standardized data format is necessary for data exchange
  - Article template and guideline from journal platform to data repository
  - Manuscript, comments and metrics from data repository to journal platform
  - Existing Standards such as MECA(Manuscript Exchange Common Approach) might be revised and re-used
- COUNTER, altmetrics can be utilized as standards for measuring impact of datasets and preprint of data articles

☑ **Business model issues** on who pays for technical review

- Direct author-pay model
- Journal curation service model

06 참고 문헌

- 정영임, 권오석, 김기동, 김소형, 서태설, 김선태 (2020). 생태학 분야 데이터 저널 발행 전략 연구: K 기관을 중심으로. 한국도서관. 정보학회지. 51(4): 83-100, DOI: 10.16981/kliss.51.4.202012.83
- 정영임, 조성남 (2022). 데이터 논문 출판 현황 분석, 데이터출판 세미나, 서울, 국립중앙도서관
- 황혜경, 정영임, 조성남, 서태설, 김지현 (2023). 과학기술분야 연구자들의 데이터 출판경험 및 인식 연구, 한국도서관정보학회지, 54(1), 45-68. https://doi.org/10.16981/kliss.54.1.202303.45
- Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. K., Murphy, F., Nurnberger, A., Raymond, L., Stockhause, M., Tedds, J., & Vardigan, M. (2017). Key components of data publishing: Using current best practices to develop a reference model for data publishing. International Journal on Digital Libraries, 18(2), 77–92. Available: https://doi.org/10.1007/s00799-016-0178-2
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC bioinformatics, 12, 1-12.
- Jung, Y. (2022) How to apply peer review to data publishing, SciDataCon2022, Seoul:Korea
- Jung, Y. & Hwang, H. (2023) Considerations to support data article publishing in South Korea, SciDataCon2023, Salzburg:Austria
- Michael B EisenAnna AkhmanovaTimothy E BehrensDiane M HarperDetlef WeigelMone Zaidi (2020) Peer Review: Implementing a "publish, then review" model of publishing eLife 9:e64910. https://elifesciences.org/articles/64910
- Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Smith, V., Shotton, D., ... & Edmunds, S. (2017). Strategies and guidelines for scholarly publishing of biodiversity data. Research Ideas and Outcomes, 3, e12431.

감사합니다.