KCSE
교육연수위원회 워크숍

서태설
한국과학기술정보연구원

2022년 5월 28일(토)

연구 데이터의 재사용과 검증을 위한
데이터 출판

목차

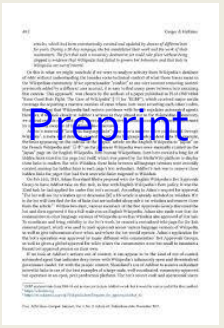
프롤로그

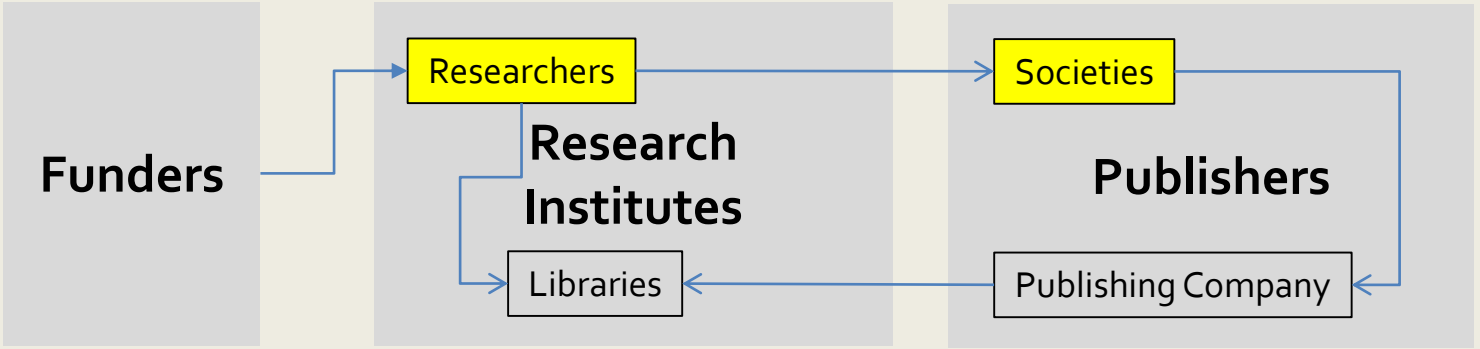
프롤로그

학술 출판 범위의 확대

- Articles
- Preprints
- Reports
- Peer-reviews
- Data
- Codes
- Protocols

Data
Code
Protocol
:





프롤로그

연구 윤리 문제와 검증을 위한 연구 데이터 공개

정량 평가 위주
성과 검증 미비

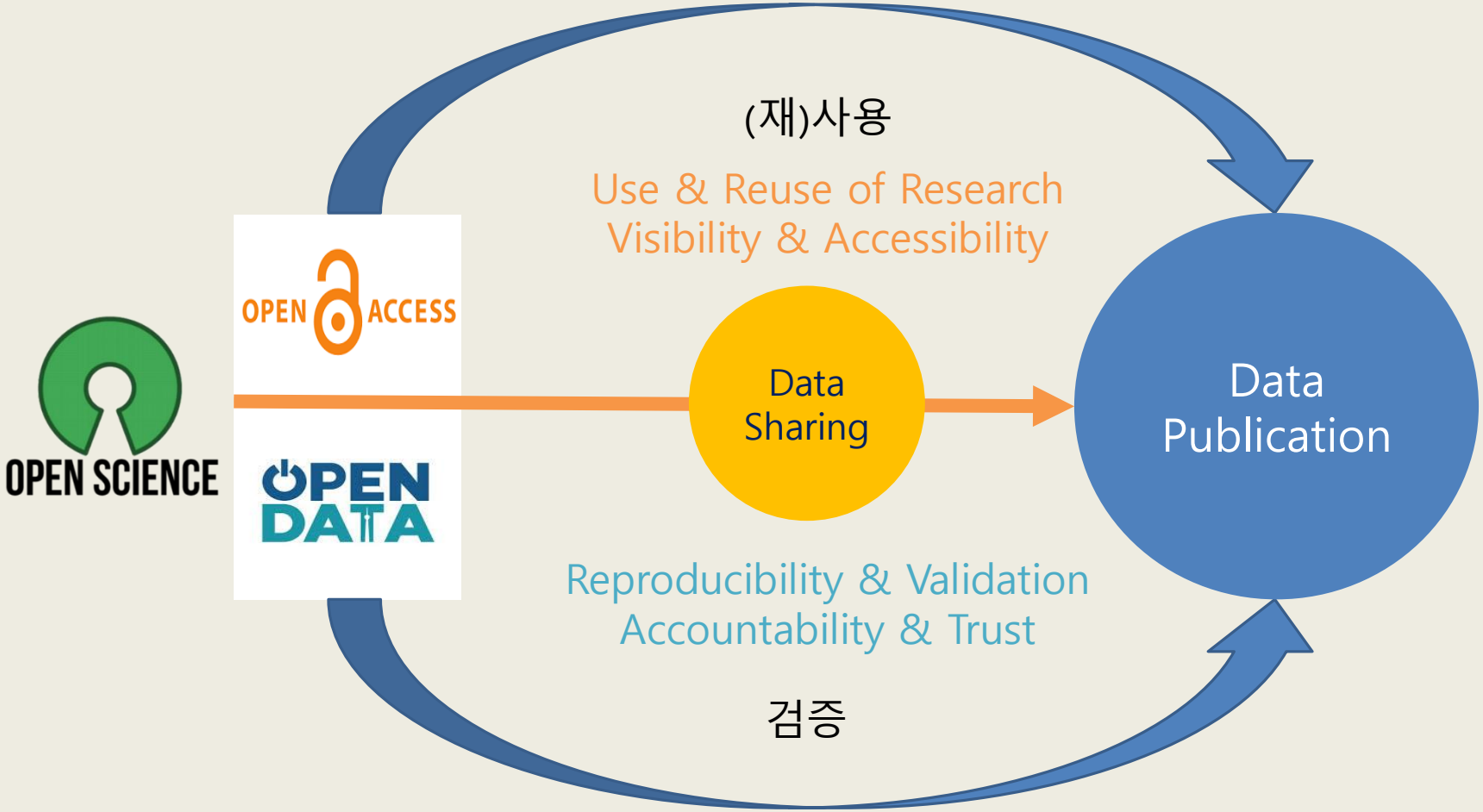
표절
이중 게재
데이터 조작
논문 쪼개기
거짓 인용
거짓 저자
:

부실 심사
부실 편집
인용 조장
학술지 쪼개기
:

데이터 출판의 필요성

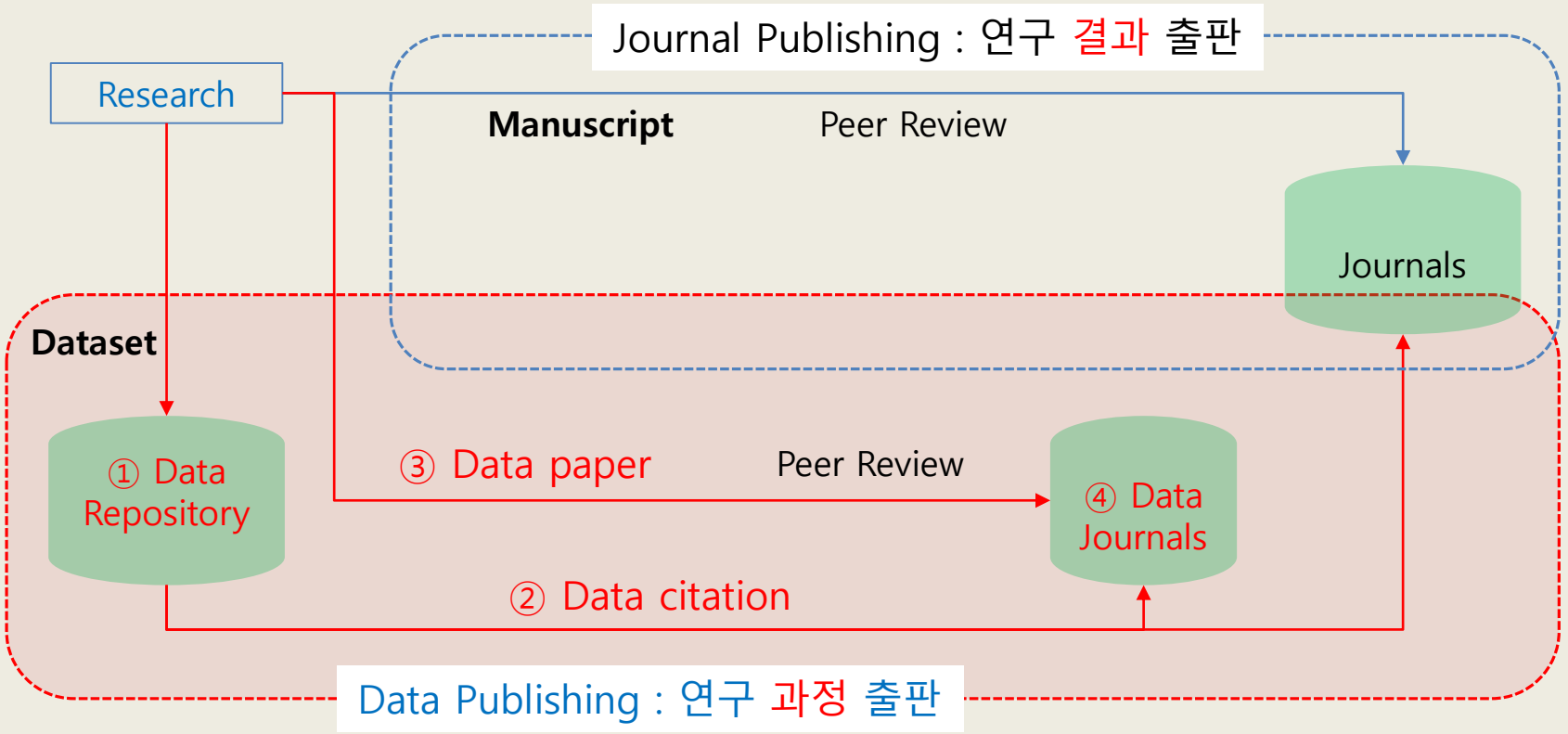
데이터 출판의 필요성

데이터 출판의 필요성



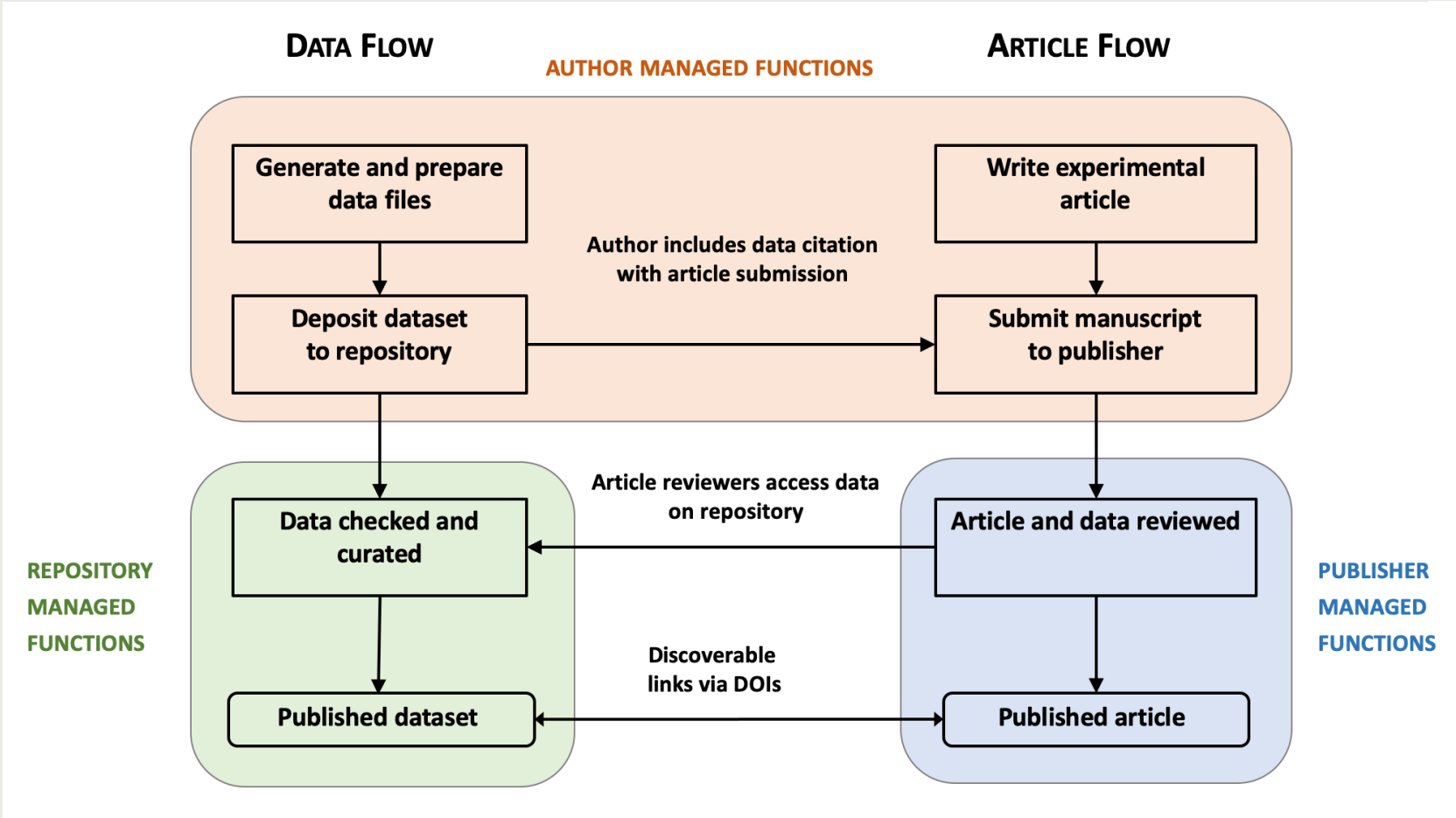
데이터 출판의 필요성

데이터 출판의 개념



데이터 출판의 필요성

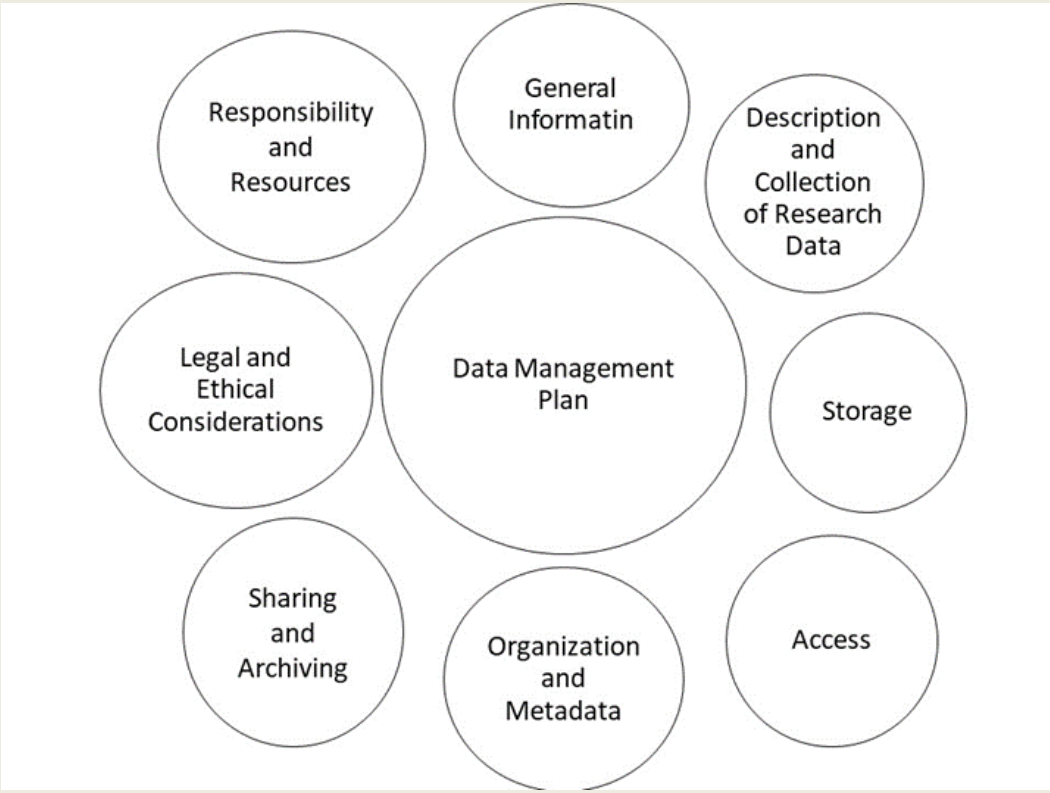
학술지 논문 출판을 위한 데이터 관리 프로세스 변화



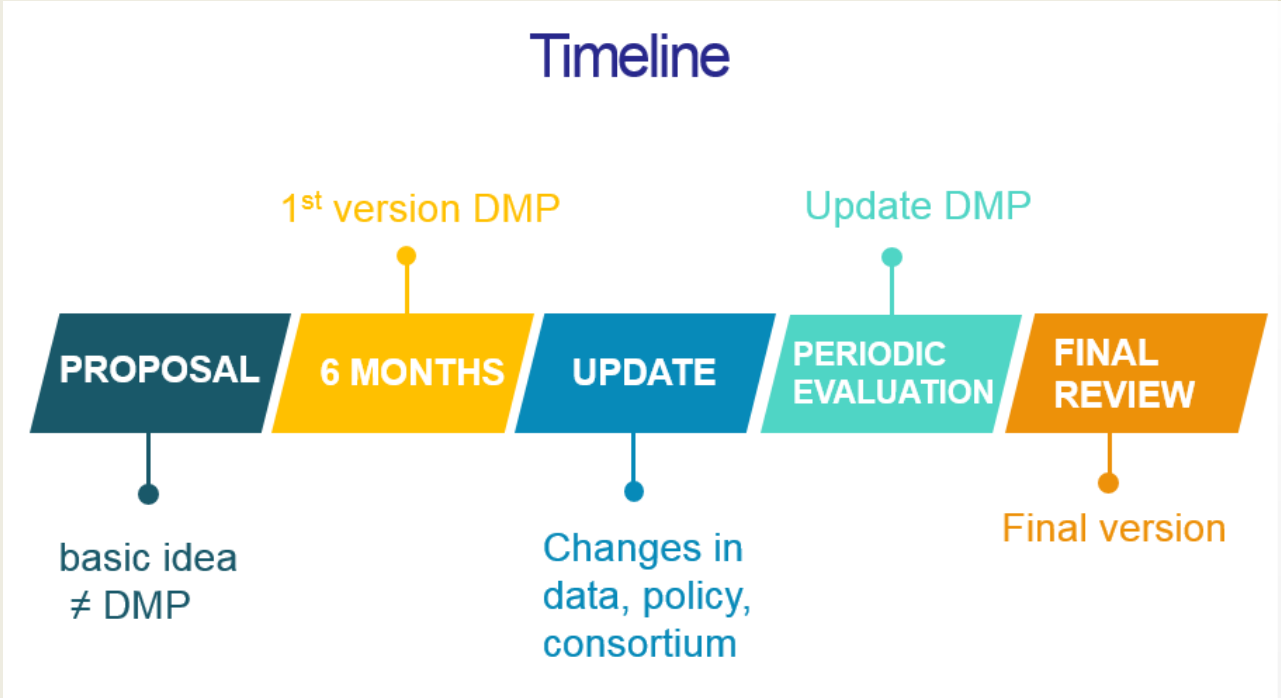
https://data.research.cornell.edu/content/preparing-fair-data-reuse-and-reproducibility

데이터 출판의 필요성

DMP (Data Management Plan) 작성 의무화



CC-BY Margaret Louise Fotland, UiO



https://www.openaire.eu/when-do-i-have-to-create-a-data-management-plan

데이터 출판이란?

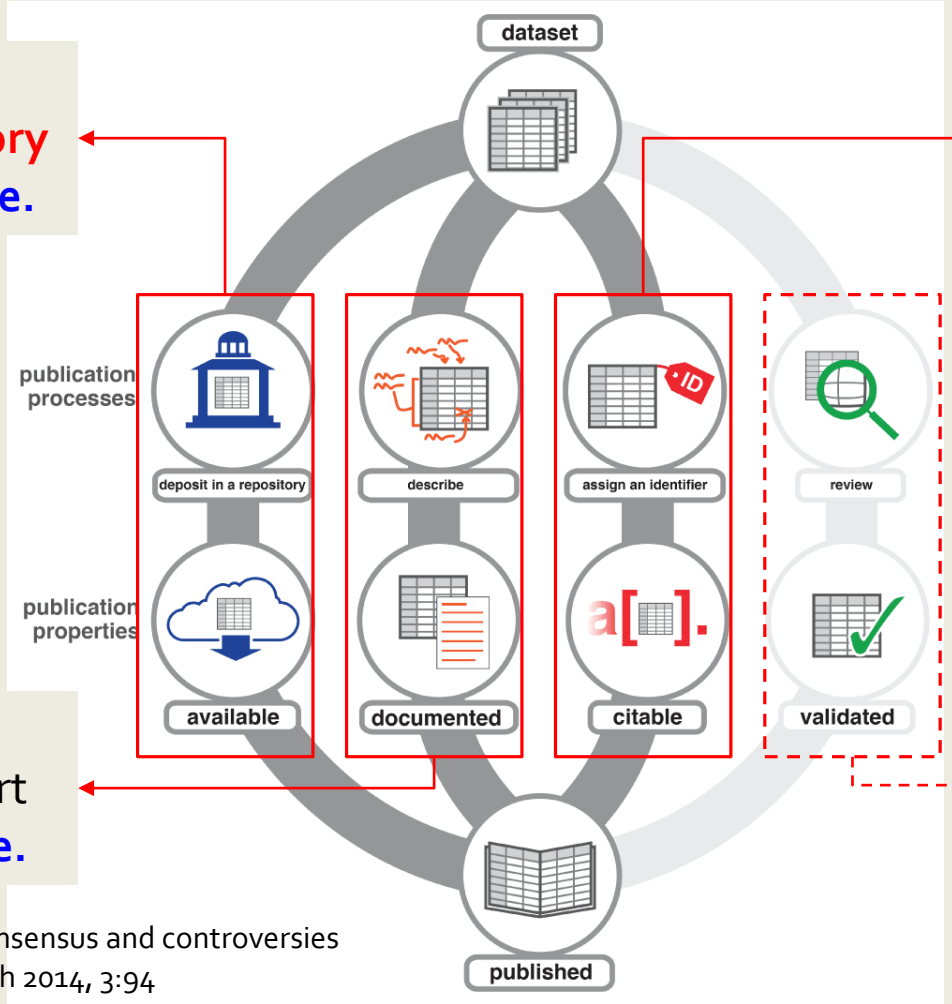

open health data
Journal of
open psychology data
Journal of
open archaeology data
RMetS
Geoscience
Data Journal

데이터출판이란?

데이터 출판의 개념

Datasets are typically deposited in a **repository** to make them **available.**

Datasets are assigned an **identifier** to facilitate **citation**



Datasets are **documented** to support reproduction and **reuse.**

Publishers **review** datasets to **validate** them.

Source: Kratz J and Strasser C. Data publication consensus and controversies [version 3; peer review: 3 approved]. F1000Research 2014, 3:94

데이터출판이란?

데이터를
어떻게 출판할까?

a) Data paper published in a journal

b) Dataset + Curation

Data Repository

Dataset
+
Curated Metadata

데이터출판이란?

데이터 저널 현황

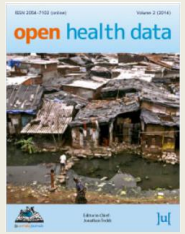
1992년에 설립된 **Pensoft 출판사**는 Data Papers, Forum Papers, Software Descriptions, Short Communications 등 기존 유형의 논문 형식이 아닌 다양한 논문 유형을 수용하는 학술지를 발간

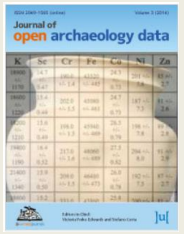
2012년에 런던정경대학의 연구자들에 의해 설립된 **Ubiquity Press**는 연구데이터와 소프트웨어의 가치를 인식하고 전문 데이터 리포지토리와 협력하여 데이터 저널을 출판함.

Wiley는 Geoscience Data Journal을 비롯한 데이터 저널 4종을 발간중임.

Nature Publishing Group(NPG)은 2014년 Scientific Data를 창간함.

Elsevier는 2014년에 다학제적 데이터 저널인 Data in Brief를 창간함






데이터출판이란?

학술지 논문과 데이터 논문의 차이

Characteristics	Journal Articles	Data Papers
Purpose of Publishing	Sharing credible **knowledge**	Sharing credible **dataset**
Aims of Peer Review	To check quality, novelty and validity of the **theories, experiments and observation**	To check completeness and collecting method of the **dataset**
Core Article Type	Original paper Review paper	**Data paper**
Composition of Manuscript	Methods Results Discussion Conclusions	Data description Metadata Dataset
Data Sharing Policy	Data sharing (Type 1~3)	Data sharing & **Peer review** (Type 4)
Deposition of data	No limitation	**Repositories**
DOI	Article	Article & **Data**

데이터출판이란?

데이터 논문에 대한 동료 심사 내용

Review I – Data description document

1. Is the method used to create the data of a high scientific standard?
2. Is enough information provided to enable the data to be re-used?
3. Comprehensive description of all the data
4. Does the data make an important and unique contribution to the geosciences?
5. Range of applications to geosciences
6. Are all contributors and existing work acknowledged?
7. Sufficient citation information of the dataset, eg dataset DOI, name of data centre etc.

Review II - metadata

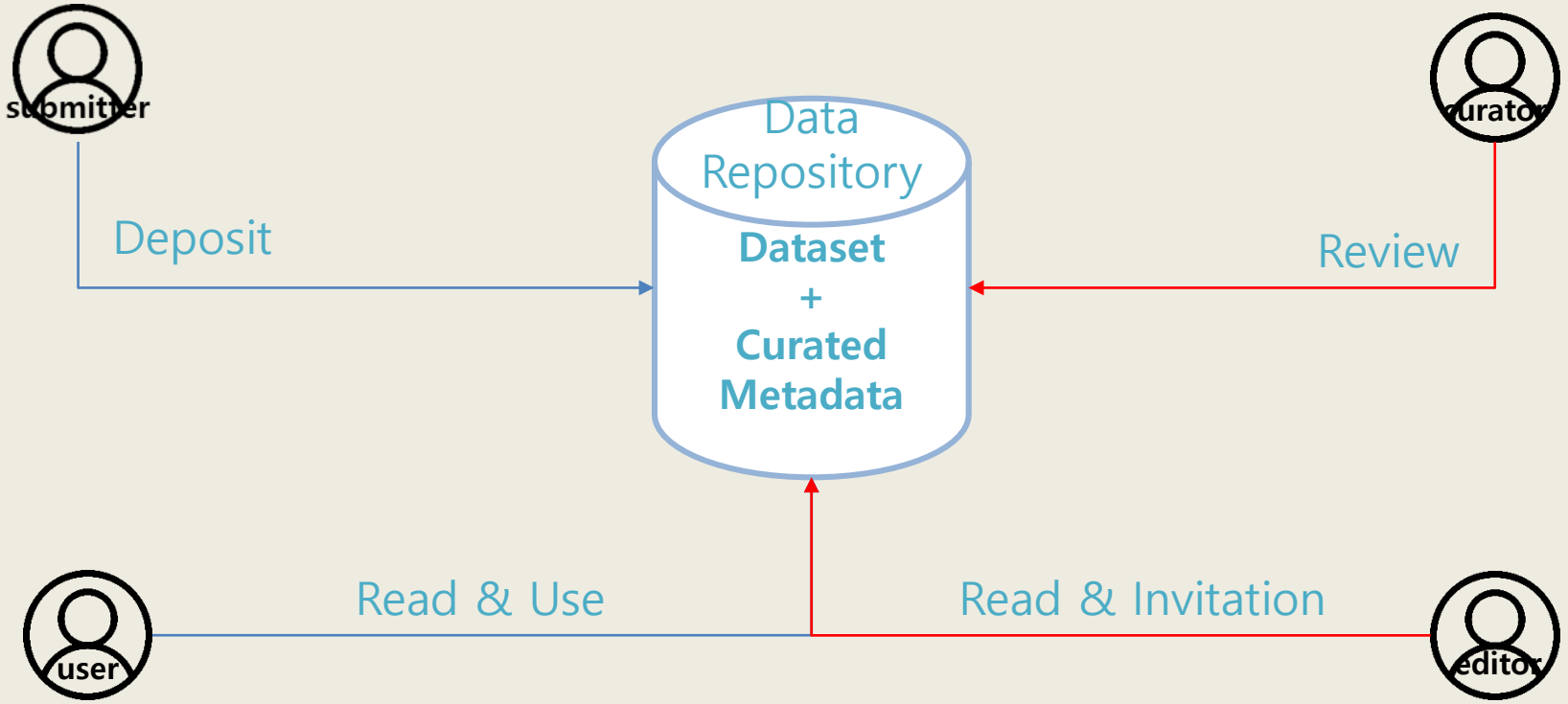
8. Does the metadata establish the ownership of the data fairly?
9. Is enough information provided to enable the data to be re-used?
10. Are the data present as described, and accessible from a registered repository using the software provided?

Review III – the data themselves

11. Readability, E.g. do they use standard or community formats?
12. Quality e.g. are error limits and quality statements adequate to assess fitness for purpose, is spatial or temporal coverage good enough to make the data useable?
13. Are the data values physically possible and plausible?
14. Are there missing data that might compromise its usefulness?

데이터출판이란?

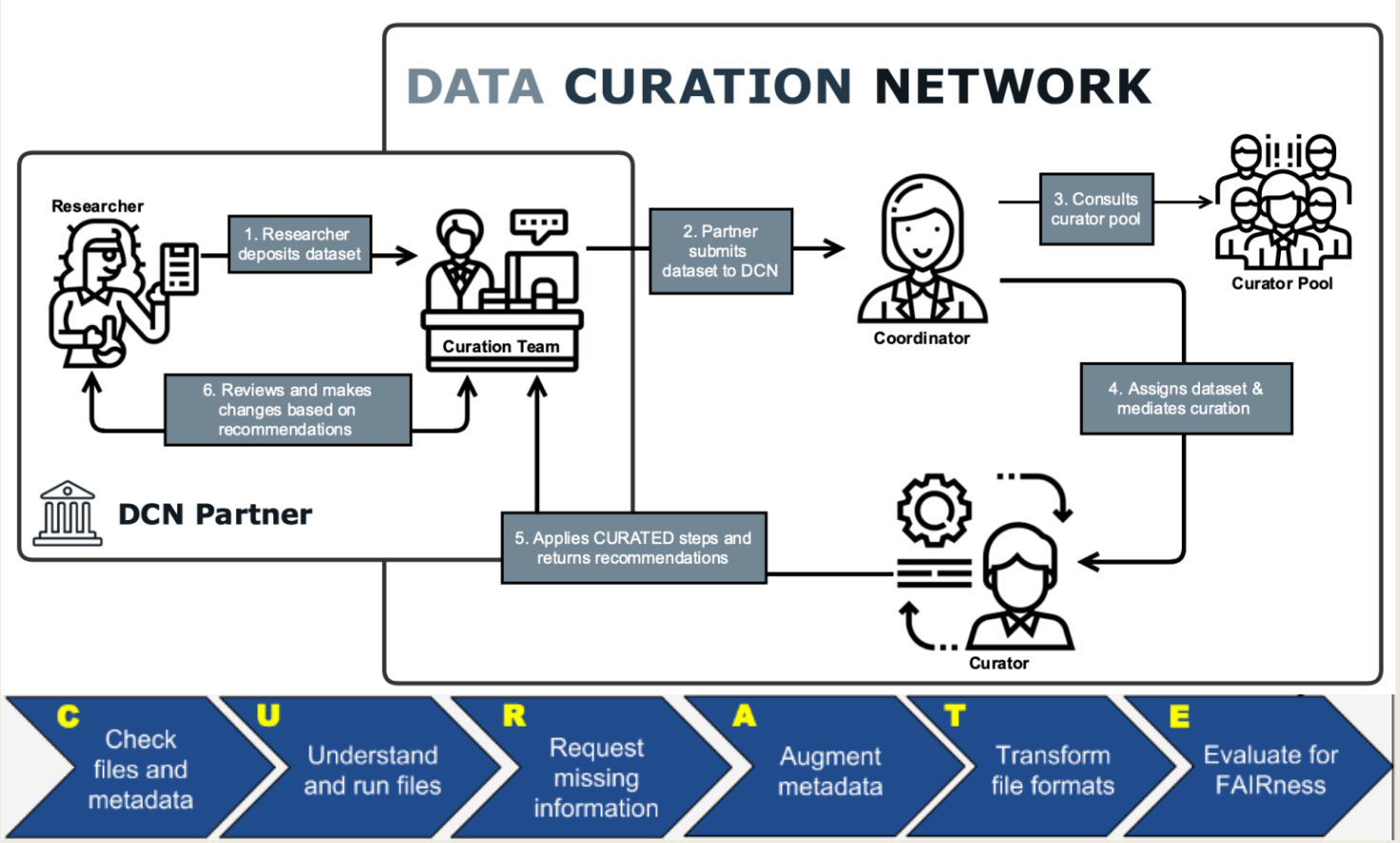
Dataset + Review Report

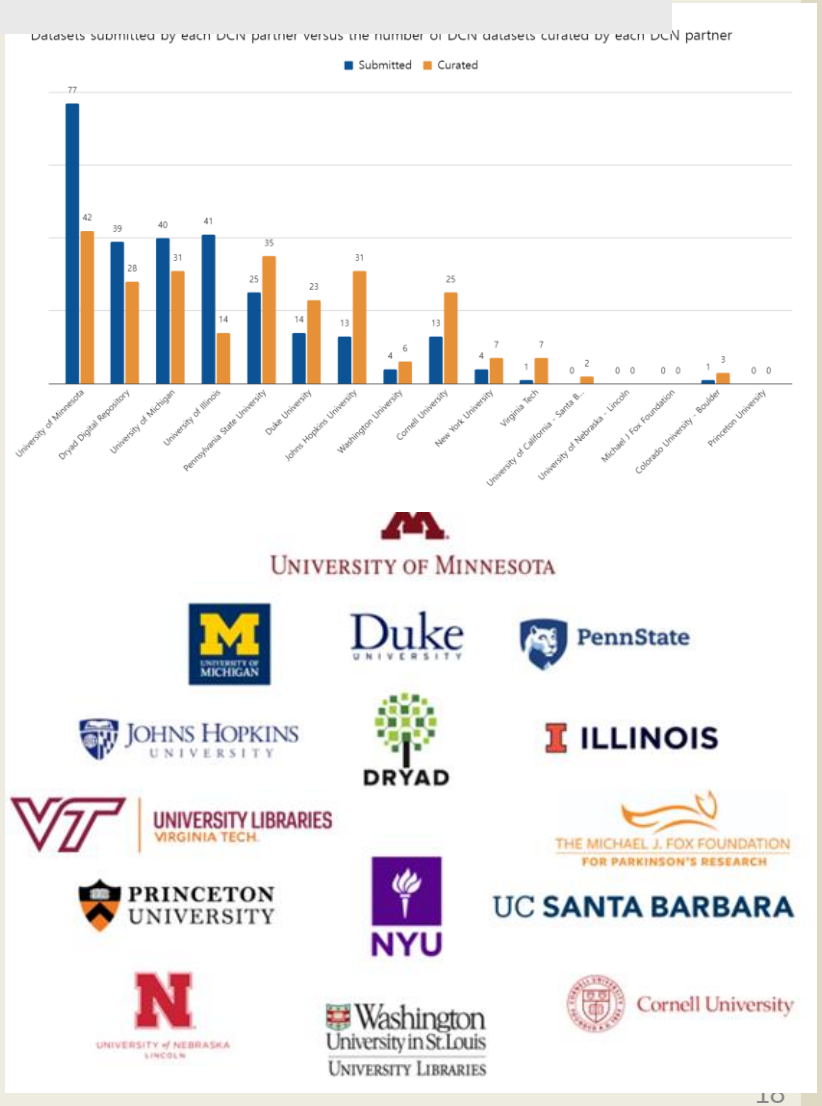


데이터출판이란?

DCN의 큐레이션 서비스

Empowering researchers to publish high quality data in an ethical and FAIR way.





데이터출판이란?

DCN의 큐레이션 서비스

DATA CURATION NETWORK
Ethical. Reusable. Better.

Home | About | Data Sets | Community | Education | Research | News | Contact

Directory of Datasets Curated by the DCN

Browse DCN-curated datasets by partner, subject area, discipline, data type, or software language (applicable to code and some statistical datasets).

All Submitting Partners | All Disciplines
All Curating Partners | All Data Types
All Subjects | All Software Languages

DCN-1: Entropically-driven macrolide polymerizations...
DCN-2: Estimating densities of zebra mussels...
DCN-3: Microfluidic long DNA sample preparation...
DCN-4: RV Atlantis Ship Data
DCN-6: TreeMerge: A new method...

DCN-4: RV Atlantis Ship Data

Dataset citation

"RV Atlantis 2019 Ship Data" available at the Data Repository for the University of Minnesota, https://doi.org/10.13020/1s80-qg61.

Curation Actions

Arrangement and Description	Contextualize
De-Identification	Documentation
Evaluate FAIRness	File Format Transformation
File Updates	Metadata
Persistent Identifier	Quality Assurance
Risk Management	

Dataset Information

Submitting Partner(s): University of Minnesota
Curating Partner(s): Johns Hopkins University
Curators: Dave Fearon, Valerie Collins
Subjects: Life Sciences
Disciplines: Biology/Integrated Biology/Integrated Biomedical Sciences
Data Types: Excel/Tabular

University Digital Conservancy Home / University of Minnesota / Data Repository for U of M (DRUM) / View Item

RV Atlantis 2019 Ship Data
Walter, Hannah; Li, Ruixuan; Stoffregen, Thomas; Wagman, Jeffrey (2019)

apaL

Persistent link to this item
https://doi.org/10.13020/1s80-qg61
https://hdl.handle.net/11299/201587

Services
Full Metadata (xml)
View Usage Statistics

Collection period
2017-06 to 2017-06

Title
RV Atlantis 2019 Ship Data

Published Date
2019-01-25

Authors
Walter, Hannah
Li, Ruixuan
Stoffregen, Thomas
Wagman, Jeffrey

Author Contact
Walter, Hannah, J. (walte666@umn.edu)

Type
Dataset

Abstract
Ocean waves cause oscillatory motion of ships. Oscillatory ship motion typically is greater in roll (i.e., the ship rolling from side to side) than in pitch (i.e., tipping from front to back). Affordances for walking on a ship at sea should be differentially influenced by ship motion in roll and pitch. When roll exceeds pitch, the maximum walkable distance within a defined path should be greater when walking along the ship's short, or athwart axis than when walking along its long, or fore-aft axis. When pitch exceeds roll, this relation should be reversed. We asked whether such changes in ship motion would be reflected in judgments of direction-specific affordances for walking. Participants (experienced maritime crewmembers) judged how far they could walk along a narrow path on the ship deck. On different days, sailing conditions were such that the relative magnitude of pitch and roll was reversed. Judgments of direction-specific affordances for walking mirrored these changes in ship motion. The accuracy of judgments was consistent across directions, and across changes in ship motion. We conclude that experienced maritime crewmembers were sensitive to dynamic variations in affordances for walking that were, themselves, a function of dynamic properties of the animal-environment system.

Description
1. File List A. Filename: RV Atlantis_APAL2017_DATA Short description: Dataset of all collected data during the experiment. Includes calculations of performance, judgments, and accuracy. B. Filename: June 16th Ship Motion Data Short description: Zipped folder of hourly data for the labeled date. Each file is raw data collected directly from the bridge of the R/V Atlantis. C. Filename: June 25th Ship Motion Data Short description: Zipped folder of hourly data for the labeled date. Each file is raw data collected directly from the bridge of the R/V Atlantis.

Referenced by
Adaptive perception of changes in affordances for walking on a ship at sea. Human Movement Science. https://doi.org/10.1016/j.humov.2019.01.002

License
CC0 1.0 Universal

Suggested Citation
Walter, Hannah; Li, Ruixuan; Stoffregen, Thomas; Wagman, Jeffrey. (2019). RV Atlantis 2019 Ship Data. Retrieved from the Data Repository for the University of Minnesota, https://doi.org/10.13020/1s80-qg61.

View/Download file

File View/Open	Description	Size	Format
RV Atlantis_APAL2017_DATA.xlsx	All Collected Data from Experiment	26.93Kb	Microsoft Excel 2007
June 16th Ship Movement Data.zip	June 16, 2017 Ship Movement Data	4.850Mb	application/zip
June 25th Ship Movement Data.zip	June 25, 2017 Ship Movement Data	5.208Mb	application/zip
NAV_IMU.txt	Metadata for Ship Movement Data	1.891Kb	Text file
RVAtlantisDataArchivalVersion.zip	Archival version of spreadsheet and figures	45.98Kb	application/zip
RV Atlantis_APAL2019_README.pdf	Readme Description of the Dataset	144.2Kb	PDF

학술지의 데이터 공유 정책

학술지의 데이터 공유 정책

Wiley's Data Sharing Policy

Journals Data Journals

Confirms data presence | Links to data | Quality of data

Policy Type	Data availability statement is published[1]	Data has been shared[2]	Data has been peer reviewed[3]
<Type 1> Encourages Data Sharing	Optional	Optional	Optional
<Type 2> Expects Data Sharing	Required	Optional	Optional
<Type 3> Mandates Data Sharing	Required	Required	Optional
<Type 4> Mandates Data Sharing and Peer Reviews Data	Required	Required	Required

Source: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html

데이터출판과 인용

데이터 저널 사례

Geoscience Data Journal provides an OA platform where scientific data can be formally published, in a way that includes **scientific peer-review**.
An online-only journal, GDJ publishes short data papers cross-linked to – and citing – datasets that have been deposited in approved **data repositories/centers** and awarded PIDs such as **DOIs**.
The journal will also accept articles on data services, and articles which support and inform data publishing best practices.

SCIE
Scopus
CAS
DOAJ

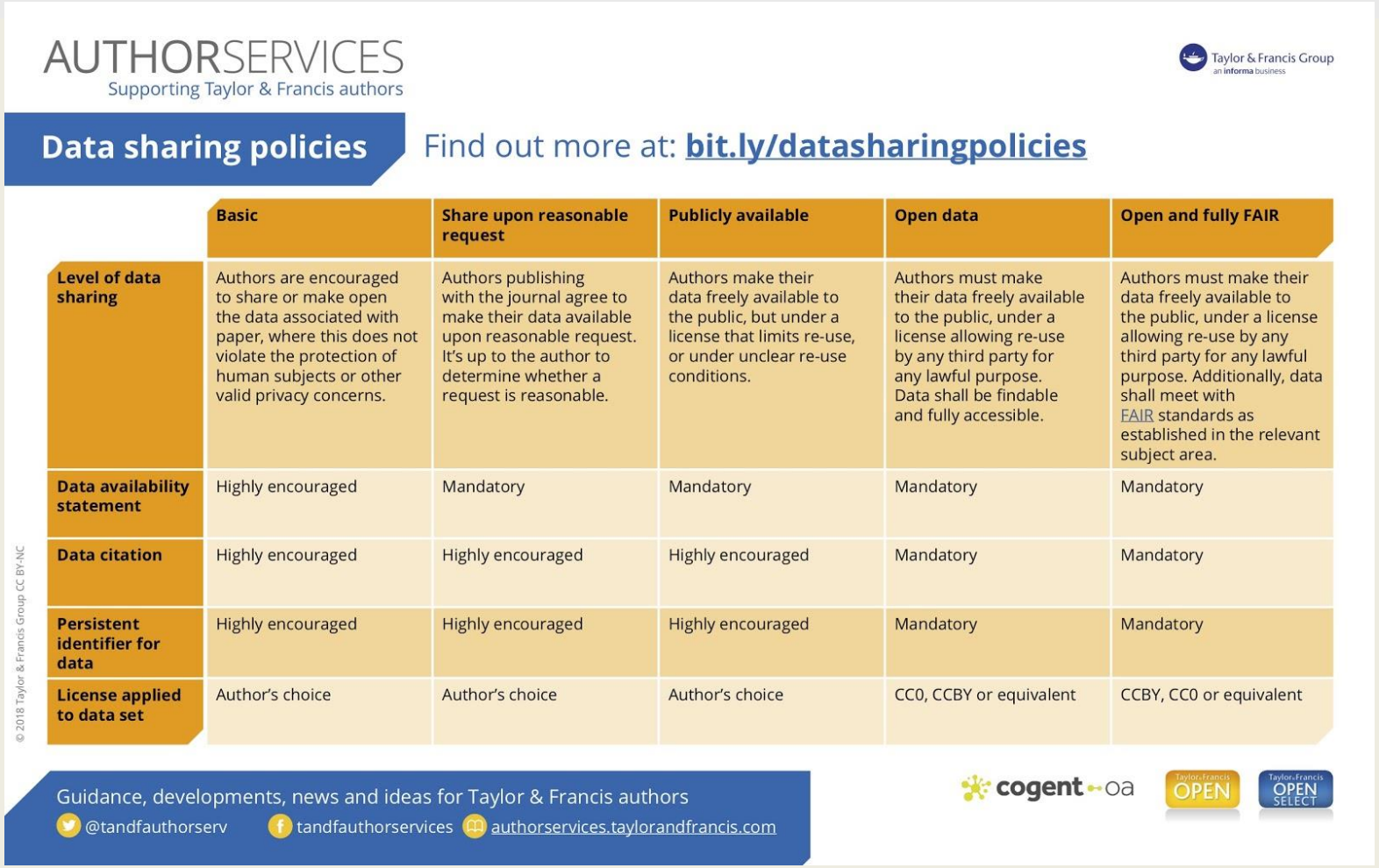
<Type 4>
Mandates Data Sharing and Peer Reviews Data

A data article describes a dataset, giving details of its collection, processing, file formats etc.
The data description document – describe how the data were created and provide information on the importance, uniqueness and applicability to other purposes of the data. **The metadata** – to clearly identify and describe the data. **The data themselves** – usability, accessibility through the repository.

Source: https://rmets.onlinelibrary.wiley.com/journal/20496060

학술지의 데이터공유정책

Data Sharing Policy of Taylor & Francis



AUTHORSERVICES
Supporting Taylor & Francis authors

Taylor & Francis Group
an informa business

Data sharing policies

Find out more at: bit.ly/datasharingpolicies

	Basic	Share upon reasonable request	Publicly available	Open data	Open and fully FAIR
Level of data sharing	Authors are encouraged to share or make open the data associated with paper, where this does not violate the protection of human subjects or other valid privacy concerns.	Authors publishing with the journal agree to make their data available upon reasonable request. It's up to the author to determine whether a request is reasonable.	Authors make their data freely available to the public, but under a license that limits re-use, or under unclear re-use conditions.	Authors must make their data freely available to the public, under a license allowing re-use by any third party for any lawful purpose. Data shall be findable and fully accessible.	Authors must make their data freely available to the public, under a license allowing re-use by any third party for any lawful purpose. Additionally, data shall meet with FAIR standards as established in the relevant subject area.
Data availability statement	Highly encouraged	Mandatory	Mandatory	Mandatory	Mandatory
Data citation	Highly encouraged	Highly encouraged	Highly encouraged	Mandatory	Mandatory
Persistent identifier for data	Highly encouraged	Highly encouraged	Highly encouraged	Mandatory	Mandatory
License applied to data set	Author's choice	Author's choice	Author's choice	CC0, CCBY or equivalent	CCBY, CC0 or equivalent

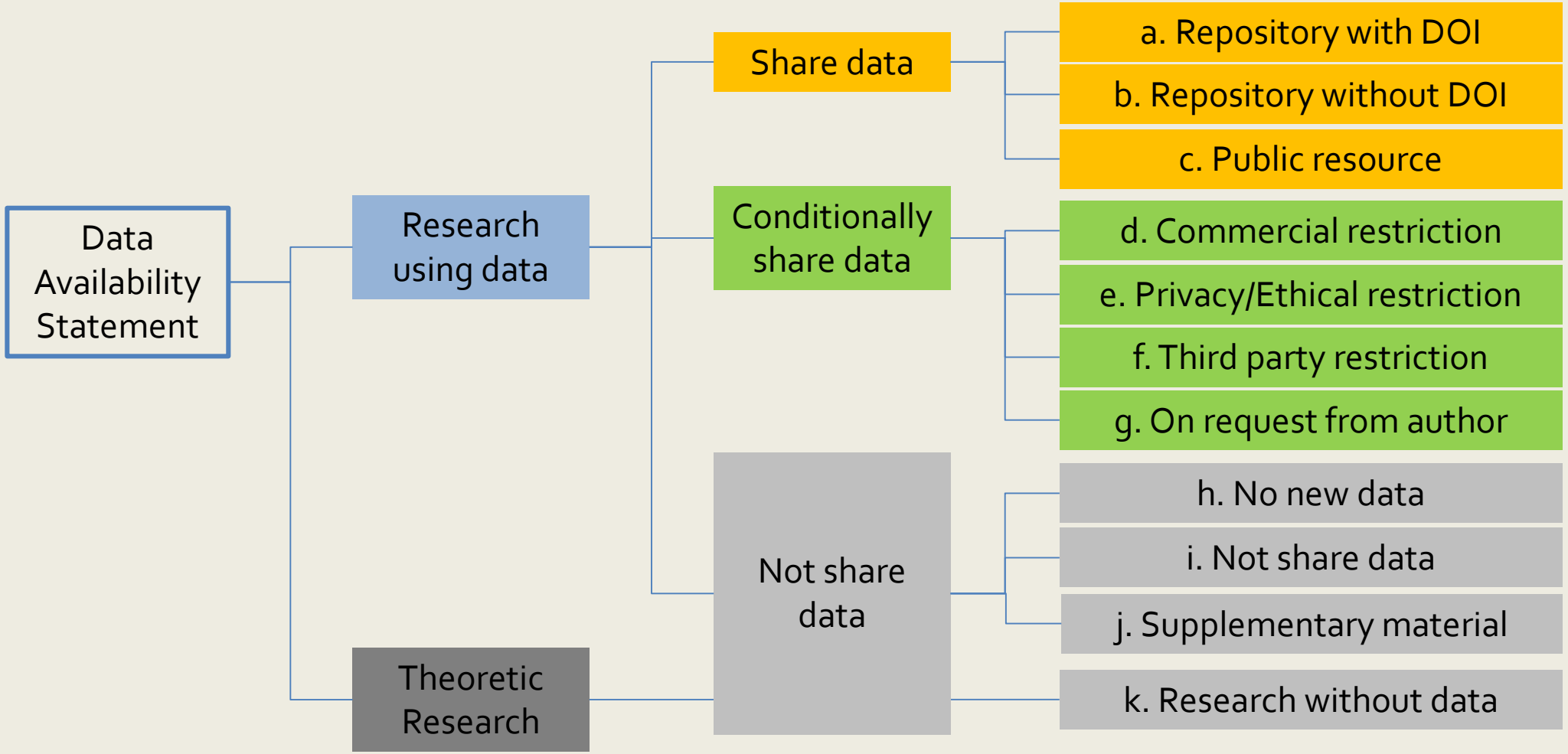
Guidance, developments, news and ideas for Taylor & Francis authors
@tandfauthorserv tandfauthorservices authorservices.taylorandfrancis.com

cogent oa Taylor & Francis OPEN Taylor & Francis OPEN SELECT

Source: https://authorservices.taylorandfrancis.com/data-sharing-policies/

학술지의 데이터 공유 정책

Data Availability Statement

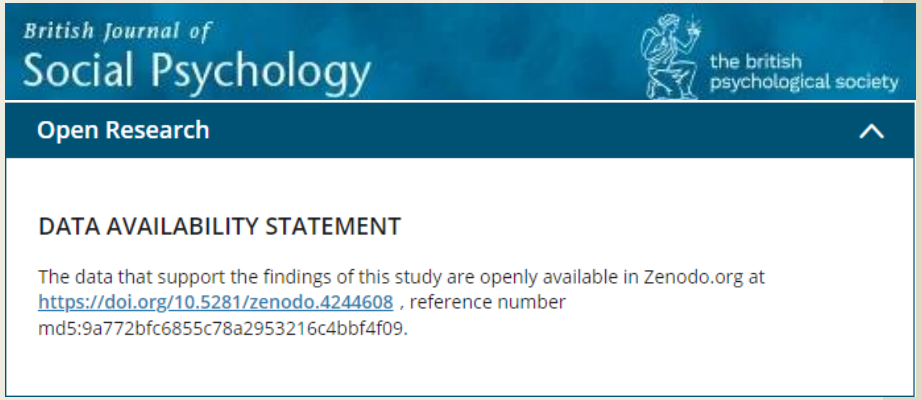


학술지의 데이터 공유 정책

Data Availability Statement : Share Data

a. Repository with DOI

The data that support the findings of this study are openly available in [repository name] at http://doi.org/[doi], reference number [reference number].

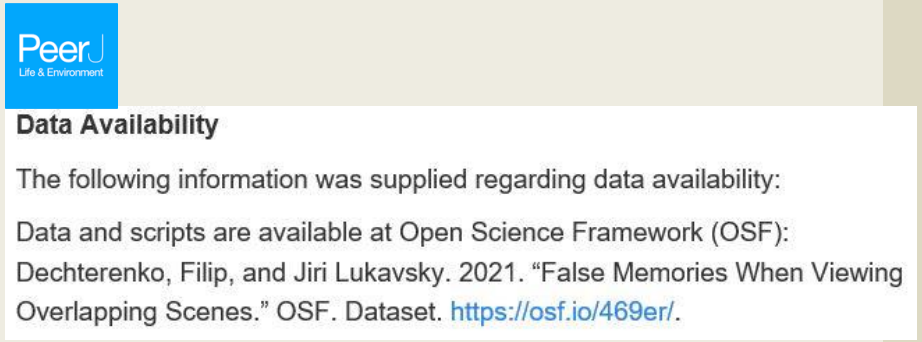

British Journal of Social Psychology
the british psychological society
Open Research
DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in Zenodo.org at https://doi.org/10.5281/zenodo.4244608 , reference number md5:9a772bfc6855c78a2953216c4bbf4f09.

b. Repository without DOI

The data that support the findings of this study are openly available in [repository name] at [URL], reference number [reference number].

c. Public resource

The data that support the findings of this study are available in [repository name] at [URL/DOI], reference number [reference number]. These data were derived from the following resources available in the public domain: [list resources and URLs]


PeerJ Life & Environment
Data Availability
The following information was supplied regarding data availability:
Data and scripts are available at Open Science Framework (OSF): Dechterenko, Filip, and Jiri Lukavsky. 2021. "False Memories When Viewing Overlapping Scenes." OSF. Dataset. https://osf.io/469er/.

학술지의 데이터 공유 정책

Data Availability Statement : Conditionally Share Data

d. Commercial restriction	The data that support the findings will be available in [repository name] at [URL / DOI link] following an embargo from the date of publication to allow for commercialization of research findings.
e. Privacy/Ethical restriction	The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
f. Third party restriction	The data that support the findings of this study are available from [third party]. Restrictions apply to the availability of these data, which were used under license for this study. Data are available [from the authors / at URL] with the permission of [third party].
g. On request from author	The data that support the findings of this study are available from the corresponding author upon reasonable request.

Data availability

The histology images supporting Fig. 2 and Figs. 4–8, are publicly available in the figshare repository, as part of this record: https://doi.org/10.6084/m9.figshare.11907768[47]. Data supporting Fig. 3, Tables 1–5 and Supplementary Tables 1–3 are not publicly available in order to protect patient privacy. These datasets can be accessed on request from Dr. Roberto Salgado, upon the completion of a Data Usage Agreement, according to policies from the German Breast Group and NSABP, as described in the data record above. Figure 9 and supplementary figures 1–8, were generated using the publicly available prognosis tool at www.tilsinbreastcancer.org/, which utilises datasets from a pooled analysis of 9 phase 3 breast cancer trials, including BIG 02-98, ECOG 1199, ECOG 2197, FinHER, GR, IBCSG 22-00, IEO, PACS01 and PACS04 (https://doi.org/10.1200/JCO.18.01010). This paper is intended to serve as a practical reference for practicing pathologists.

Respiratory Research BMC

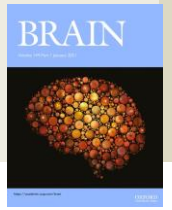
Availability of data and materials

The data underlying this article are available in the article and in its Additional file 1. Further inquiries can be directed to the corresponding author.

학술지의 데이터 공유 정책

Data Availability Statement : Not Share Data

h. No new data	Data sharing is not applicable to this article as no new data were created or analyzed in this study.
i. Not share data	Research data are not shared.
j. Supplementary material	The data that supports the findings of this study are available in the supplementary material of this article.
k. Research without data	Data sharing not applicable to this article as no datasets were generated or analysed during the current study

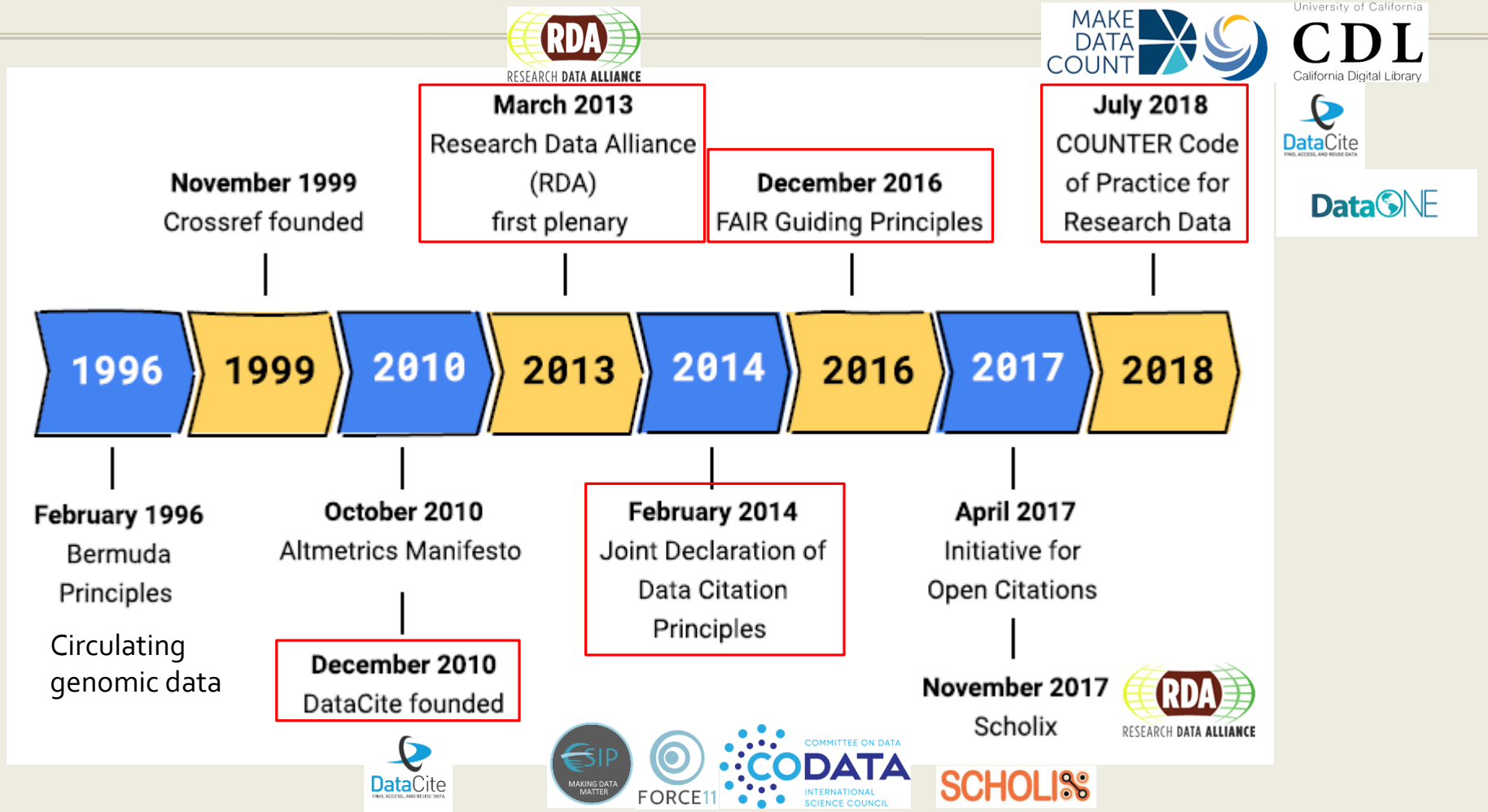


Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary material. Raw data that support the findings of this study are available from the corresponding author, upon reasonable request.

데이터 인용

데이터 인용



Source: Lowenberg, Daniella, Chodacki, John, Fenner, Martin, Kemp, Jennifer, Jones, Matthew B., 2019. Open Data Metrics: Lighting the Fire. Zenodo. https://doi.org/10.5281/zenodo.3525349

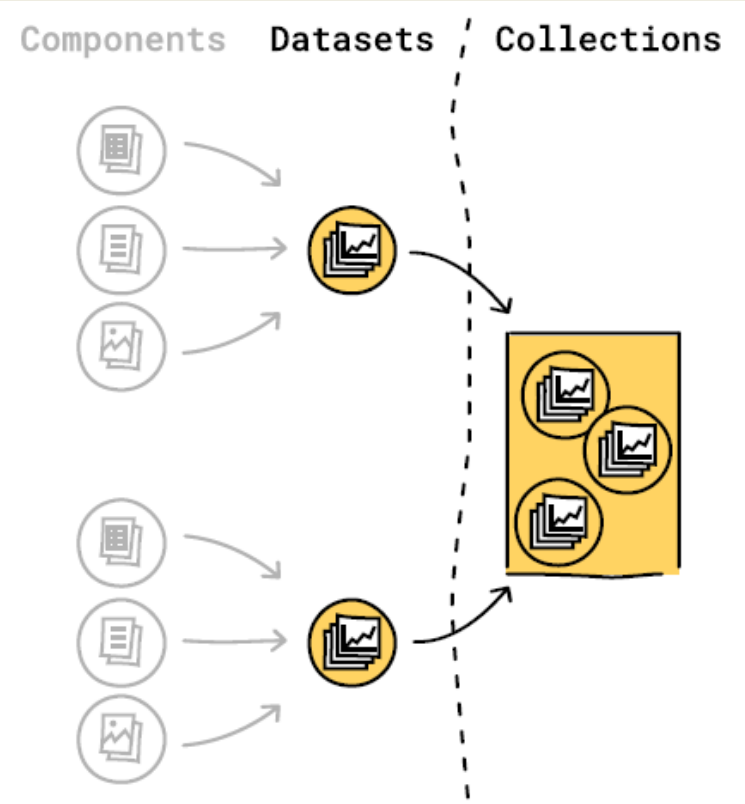
데이터 인용

데이터 평가 노력

MAKE DATA COUNT

데이터 인용

데이터셋 인용의 대상범위



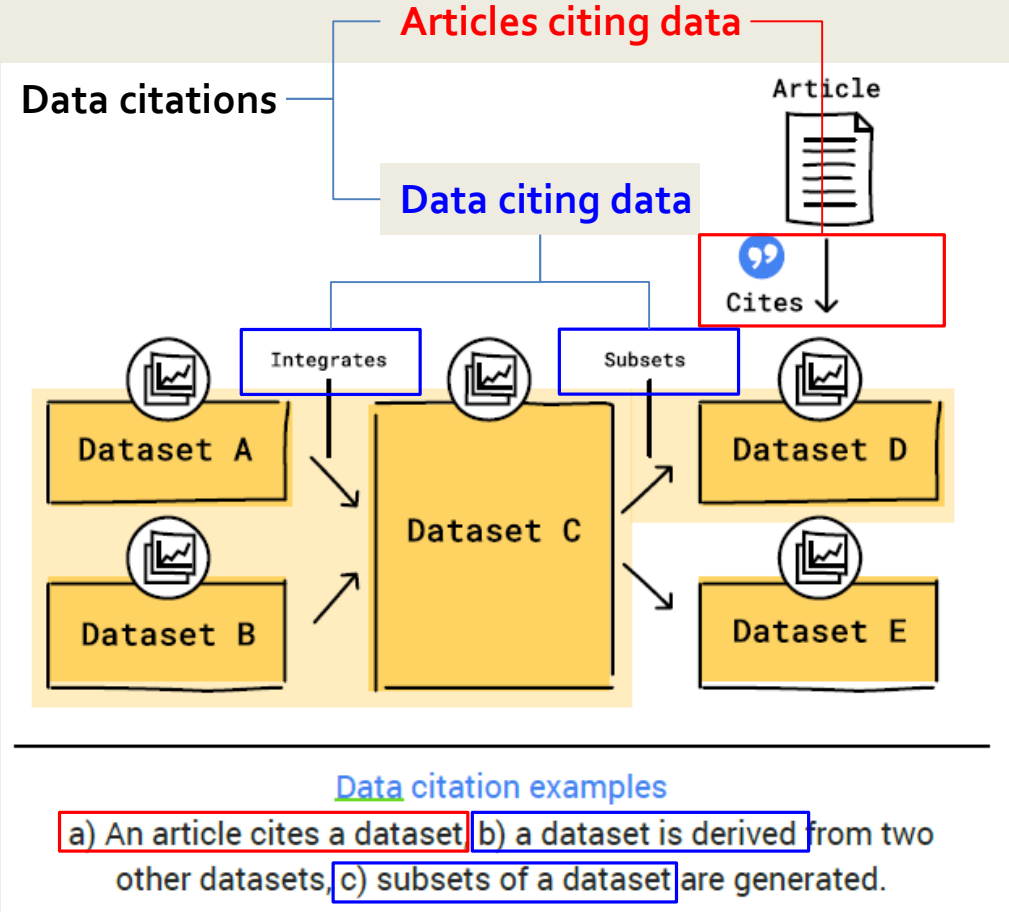
COUNTER Code of Practice for Research Data

- **Datasets**, defined as "an aggregation of data, published or curated by a single agent, and available for access or download in one or more formats, with accompanying metadata."
- **Components** each of which is "part of the data available for a dataset that can be accessed or downloaded individually."
- **Versions** which represent "significant changes to the content and/or metadata, associated with changes in one or more components, and that would result in changes to fixity attributes of the components."

Source: Lowenberg, Daniella, Chodacki, John, Fenner, Martin, Kemp, Jennifer, Jones, Matthew B., 2019. Open Data Metrics: Lighting the Fire. Zenodo. https://doi.org/10.5281/zenodo.3525349

데이터 인용

데이터셋의 인용과 평가



FORCE11 Joint Declaration of Data Citation Principles

Principle 2: Credit and attribution
Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data.

Principle 7: Specificity and verifiability
Data citations should facilitate identification of, access to, and verification of the specific data. Citations or citation metadata should include information, sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.

Source: Lowenberg, Daniella, Chodacki, John, Fenner, Martin, Kemp, Jennifer, Jones, Matthew B., 2019. Open Data Metrics: Lighting the Fire. Zenodo. https://doi.org/10.5281/zenodo.3525349

데이터 인용

데이터 인용 지침 사례: DataCite

- The preferred format for rendering a DataCite citation:

Creator (PublicationYear): Title. Version. Publisher. (resourceTypeGeneral). Identifier

- Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V. 2.1. Geological Institute, University of Tokyo. (dataset). https://doi.org/10.1594/PANGAEA.726855
- Geofon operator (2009): GEFON event gfz2009kciu (NW Balkan Region). GeoForschungsZentrum Potsdam (GFZ). (dataset). https://doi.org/10.1594/GFZ.GEOFON.gfz2009kciu
- Denhard, Michael (2009): dphase_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project. World Data Center for Climate. (dataset.) https://doi.org/10.1594/WDCC/dphase_mpeps

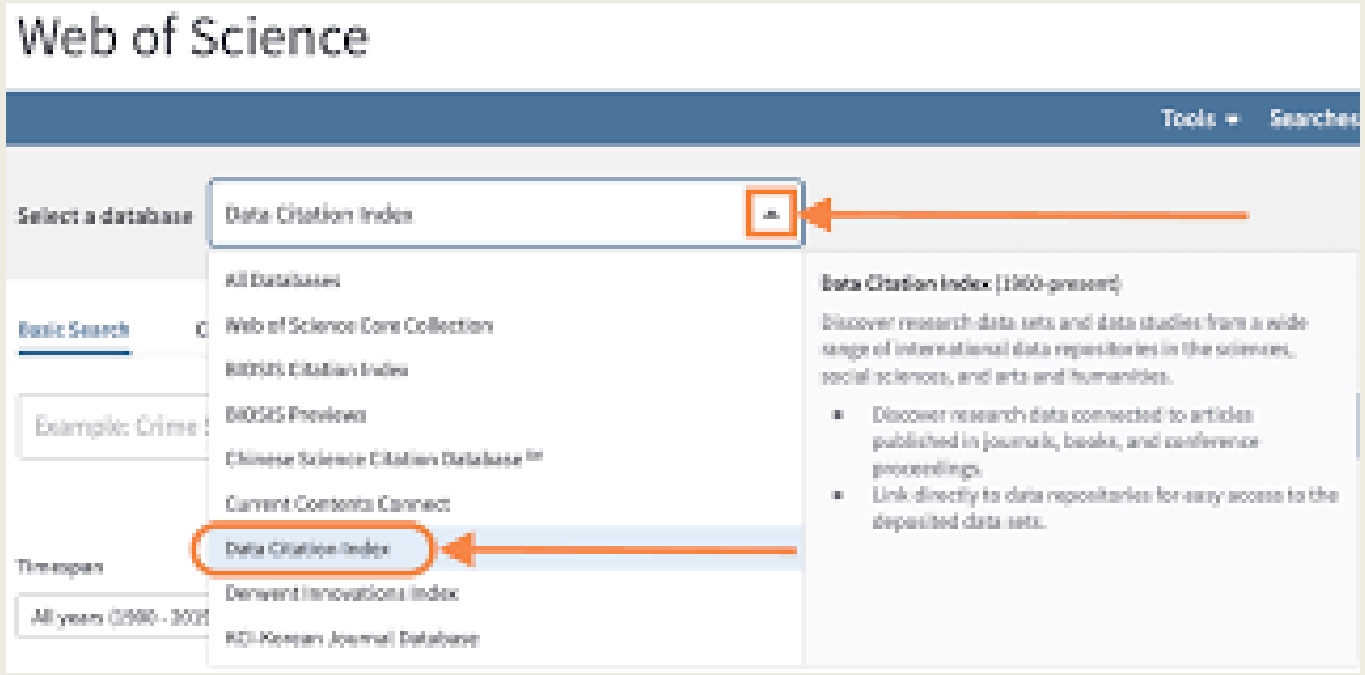
데이터 인용

Data Citation Index

- 2012년 11월부터 이터의 색인과 인용정보를 제공
- DataCite 방식의 표준화된 인용정보를 제공하고

As of 2 March 2022

- Data from 443 repositories
- 12,357,621 datasets
- 1,434,939 data studies
- 264,773 software



데이터출판과 인용

데이터 인용 지침 사례: Linguistics

The template for a **minimal bibliographic reference** (i.e. in the bibliography section of a piece of academic writing) to a dataset resource is:
Author, Date, Title, Publisher, Locator.

The template for an **expanded bibliographic reference** to a dataset resource, including *conditional elements* (i.e. required in certain cases depending on resource characteristics) is:
Author, *Other Attribution (Roles),* **Date, Title, Publisher, Locator,** *Version, Date accessed.*

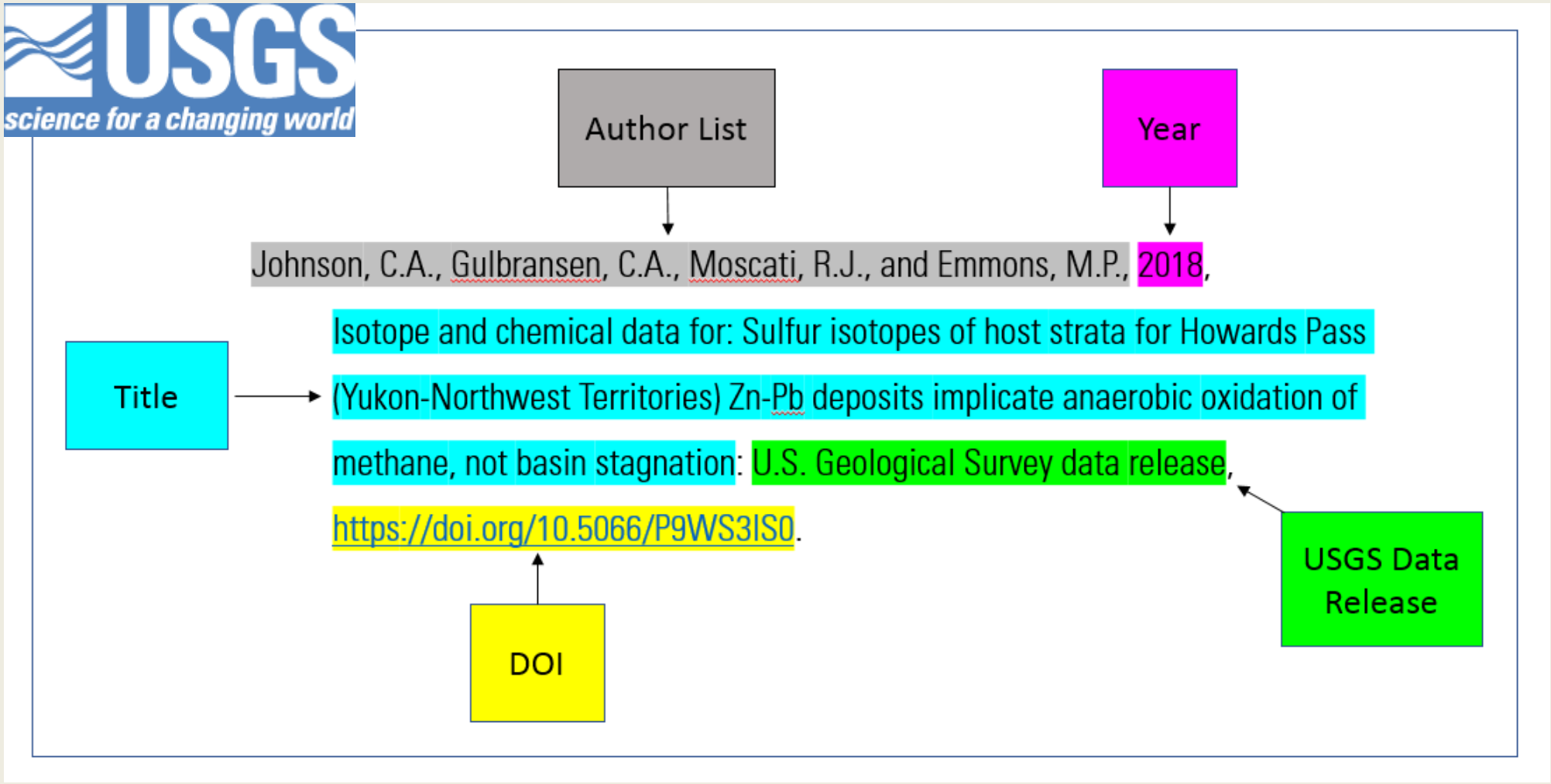
In-text (or in-line) citations must point to a bibliographic reference in the bibliography section of the published work. The template for a **minimal in-text citation** is:
Author, Date.

The template for an **expanded in-text citation** including additional potential information is:
Author, Date, *Locator, Subset, Other Attribution (Roles).*

Source: Andreassen, H. N., Berez-Kroeker, A. L., Collister, L., Conzett, P., Cox, C., Smedt, K. D., … Research Data Alliance Linguistic Data Interest Group. (2019). Tromsø recommendations for citation of research data in linguistics (Version 1). *Research Data Alliance*. DOI: 10.15497/RDA00040

데이터출판과 인용

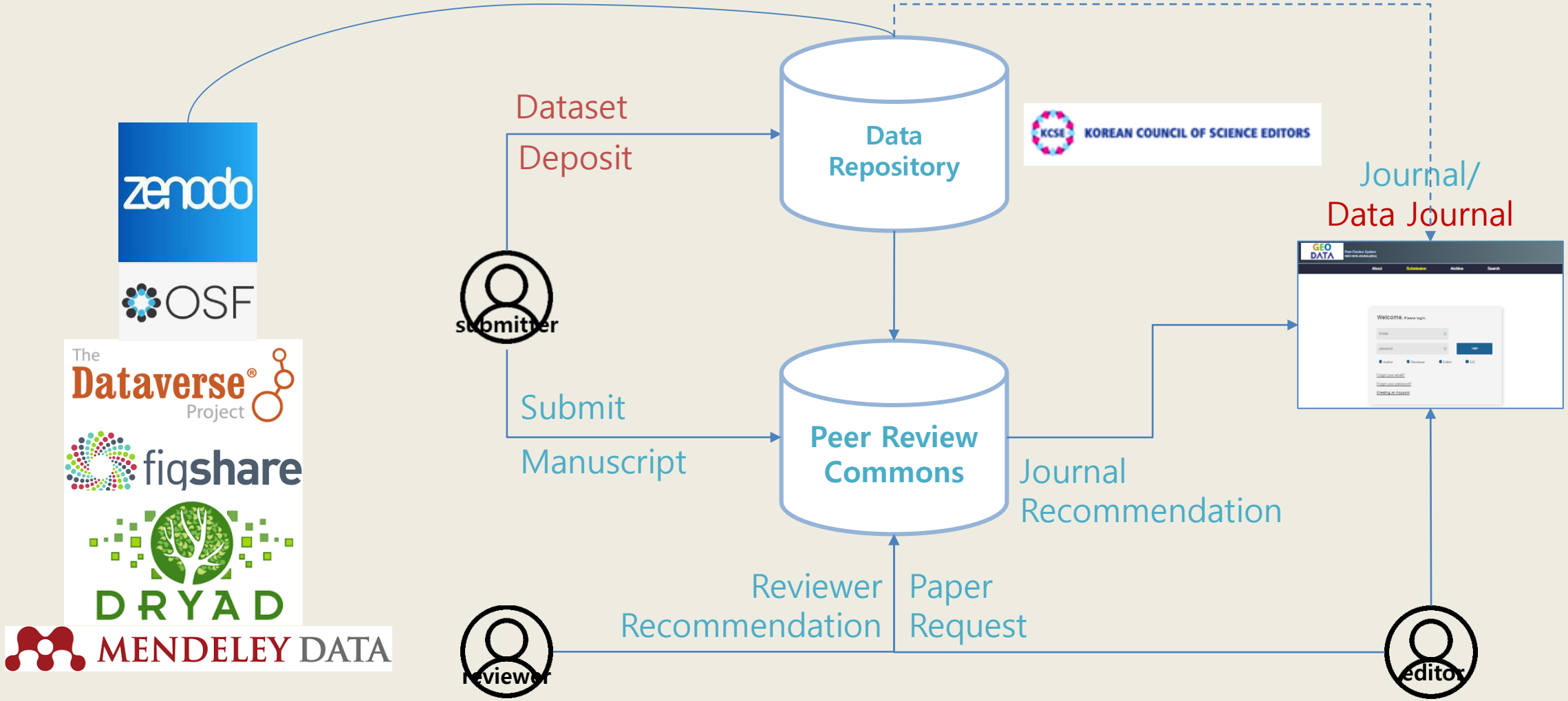
데이터 인용 지침 사례: U.S. Geological Survey



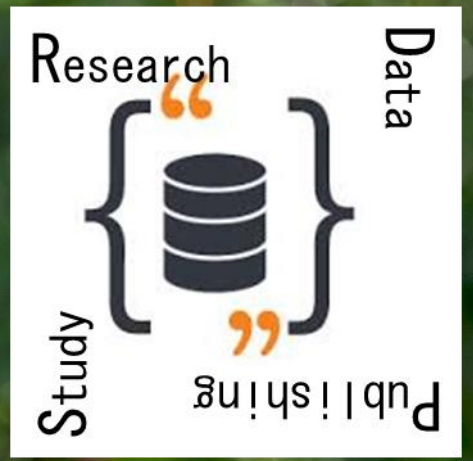
에필로그

결론

데이터 출판 생태계


zenodo
OSF
The Dataverse® Project
figshare
DRYAD
MENDELEY DATA
Dataset
Deposit
Data Repository
KOREAN COUNCIL OF SCIENCE EDITORS
Journal/
Data Journal
submitter
Submit
Manuscript
Peer Review Commons
Journal
Recommendation
Reviewer
Recommendation
Paper
Request
reviewer
editor

에필로그



투명하고 효과적인
연구 커뮤니케이션 확립