

Creating and implementing research data policies

COPE Webinar

2 July 2018

Guest speakers:

- **Rebecca Grant** (Research Data Manager) & **Varsha Khodiyar** (Data Curation Manager), Open Research Group, Springer Nature, Rebecca.Grant@springernature.com, Varsha.Khodiyar@nature.com
- **Todd Carpenter**, Executive Director, National Information Standards Organization, tcarpenter@niso.org

Moderator:

- **Trevor Lane**, Education & Engagement Consultant, Edanz Group and COPE Council Member, tlane@edanzgroup.com

COPE Webinar: Creating and implementing research data policies



Trevor Lane

Education &
Engagement
Consultant,
Edanz Group

COPE Council
Member
(Chair, Member
Services)
tlane@edanzgroup.com



Rebecca Grant

Research Data Manager,
Rebecca.Grant@springernature.com

Varsha Khodiyar

Data Curation Manager,
Varsha.Khodiyar@nature.com

Open Research Group,
Springer Nature

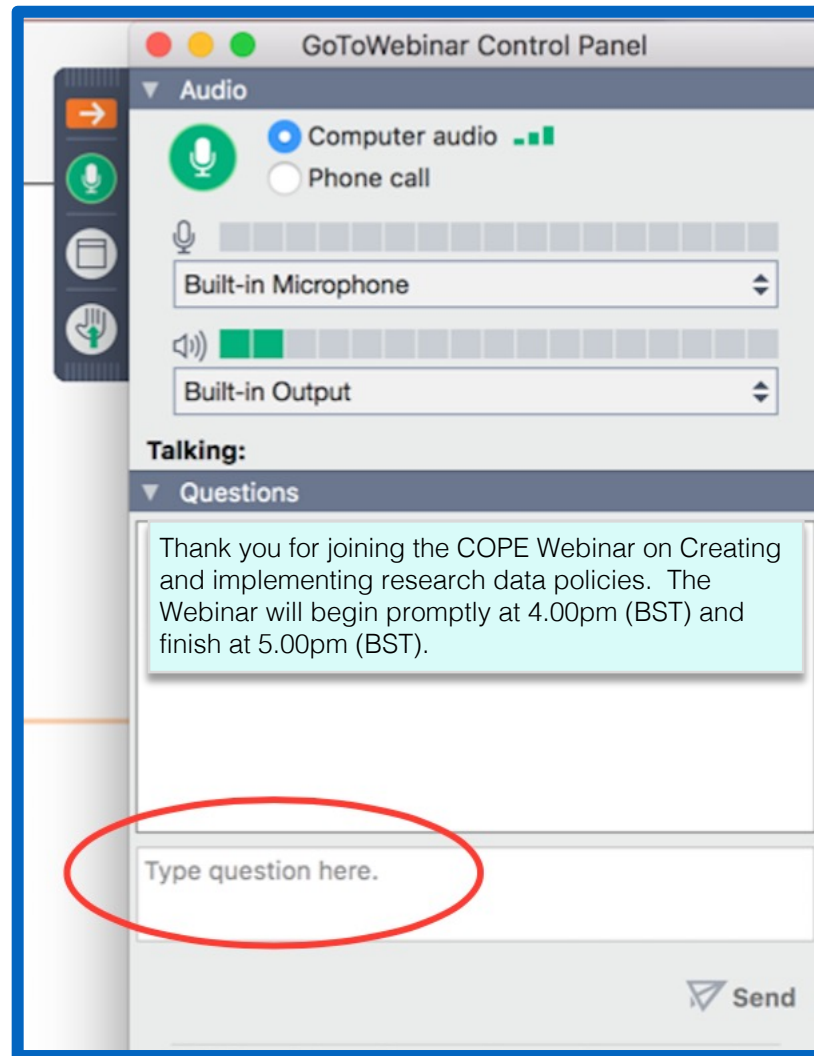
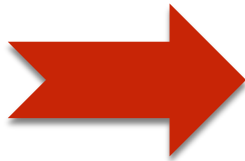


Todd Carpenter

Executive
Director, NISO
National
Information
Standards
Organization
tcarpenter@niso.org

COPE Webinar: Creating and implementing research data policies

**Please type your
questions in the
Question Box**



Note:
***Recording &
summary report
will be uploaded
to COPE website***

Creating and implementing research data policies

COPE Webinar

2 July 2018

Lightning Poll 1

COPE Webinar: Creating and implementing research data policies

COPE resources

- **Core Practice 5**
<https://publicationethics.org/data>



5. Data and reproducibility

Journals should include policies on data availability and encourage the use of reporting guidelines and registration of clinical trials and other study designs according to standard practice in their discipline

[View all Data and reproducibility resources](#)

COPE Webinar: Creating and implementing research data policies

COPE resources

- **Digest Jan 2018: Research integrity and how to buy a Persian carpet: Transparency and Openness Promotion (TOP) Guidelines, part 2**
<https://publicationethics.org/news/research-integrity-and-how-buy-persian-carpet-top-guidelines-part-2>

(TOP Guidelines 2015: <https://cos.io/our-services/top-guidelines/>)
- **COPE Forum discussion notes on data sharing, 12 Feb 2016**
https://publicationethics.org/files/u661/Notes%20from%20Forum%20Discussion%20Topic_12_February%202016_DATA_SHARING_final.pdf
- **Digest Feb 2018: Data and reproducibility — Focus from COPE Education Subcommittee**
<https://publicationethics.org/news/data-and-reproducibility-focus-cope-education-subcommittee>
The role of research institutions
<https://publicationethics.org/news/data-and-reproducibility-role-research-institutions>
- **Data Sharing Policies in Scholarly Publications: Interdisciplinary Comparisons**
https://publicationethics.org/files/u7140/Data%20sharing%20poster_2017.pdf

COPE Webinar: Creating and implementing research data policies

International Committee of Medical Journal Editors (ICMJE)

<http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html#two>

- “As of 1 July 2018 manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement...”
(eg, what data will be shared, when, for how long, where, and with whom?)
- “Clinical trials that begin enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial's registration....If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.”

Creating and implementing research data policies

COPE Webinar

2 July 2018

Lightning Poll 2

COPE webinar: Creating and implementing research data policies

Rebecca Grant, Research Data Manager and Varsha Khodiyar, Data Curation Manager, Springer Nature

ADVANCING
DISCOVERY



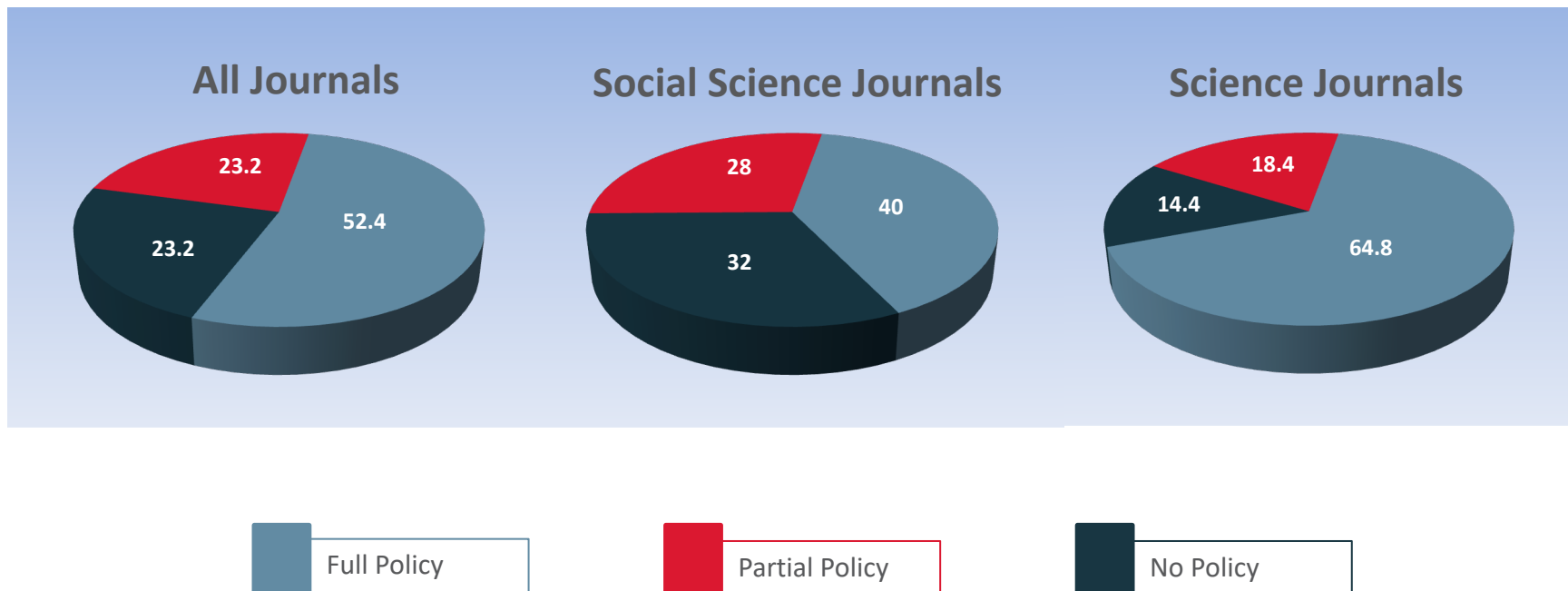
Springer Nature is a leading research, educational and professional publisher, providing quality content to our communities through a range of innovative platforms, products and services.

Home to brands including Springer, Nature Research, BioMed Central, Palgrave Macmillan and Scientific American.

As the leading open access publisher, we see the rise of open research in all its manifestations as one of the major forces reshaping the way that researchers communicate and collaborate to advance the pace and quality of discovery.

Since 2016 we have been rolling out standard data policies across all Springer Nature journals.

Understanding journal policies is difficult



Data source: Linda Naughton, JISC Journal Research Data Policy Bank project presentation (n = 250)

“The evidence shows that the current research data policy ecosystem is in critical need of standardization and harmonization”

Policy Types

Type 1

Data sharing and data citation is encouraged but not required

Type 2

Data sharing and evidence of data sharing encouraged

Type 3

Data sharing encouraged and statements of data availability required

Type 4

Data sharing, evidence of data sharing and peer review of data required

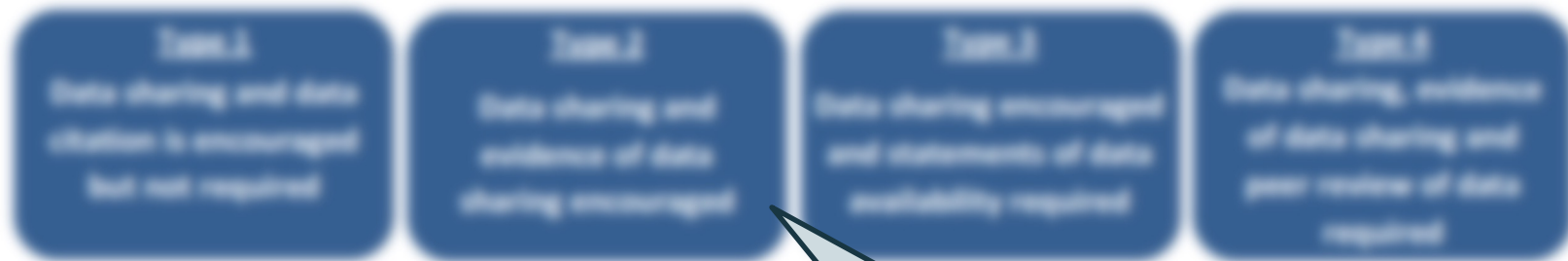
Process

1. Identify and agree the most relevant policy type for individual journal
2. Implement standardised text and processes into relevant journal guides and publishing workflows
3. Provide a consistent and easy-to-follow journal data policy for authors, researchers and peer reviewers

Standardising and harmonising research data policy in scholarly publishing

Iain Hrynaszkiewicz, Aliaksandr Birukou, Mathias Astell, Sowmya Swaminathan, AmyeKenall, Varsha Khodiyar
bioRxiv 122929; doi: <https://doi.org/10.1101/122929>

Policy Types



Process

1. Identify and agree the most relevant policy type

2. Implement standardised text and processes into

3. Provide a consistent and easy-to-follow journal d

All policy types:

- Preference **sharing of data via repositories** (rather than ESM/SI)
- **Allow citation of public datasets** in reference lists/bibliographies
- **Encourage use of publisher helpdesk** to ensure compliance with funder mandates

Research Data Helpdesk

Queries are answered within two business days

Run by members of the Springer Nature Data Publishing team

Expertise in data curation and management, archiving and digital preservation, copyright and licensing, Open Access publishing

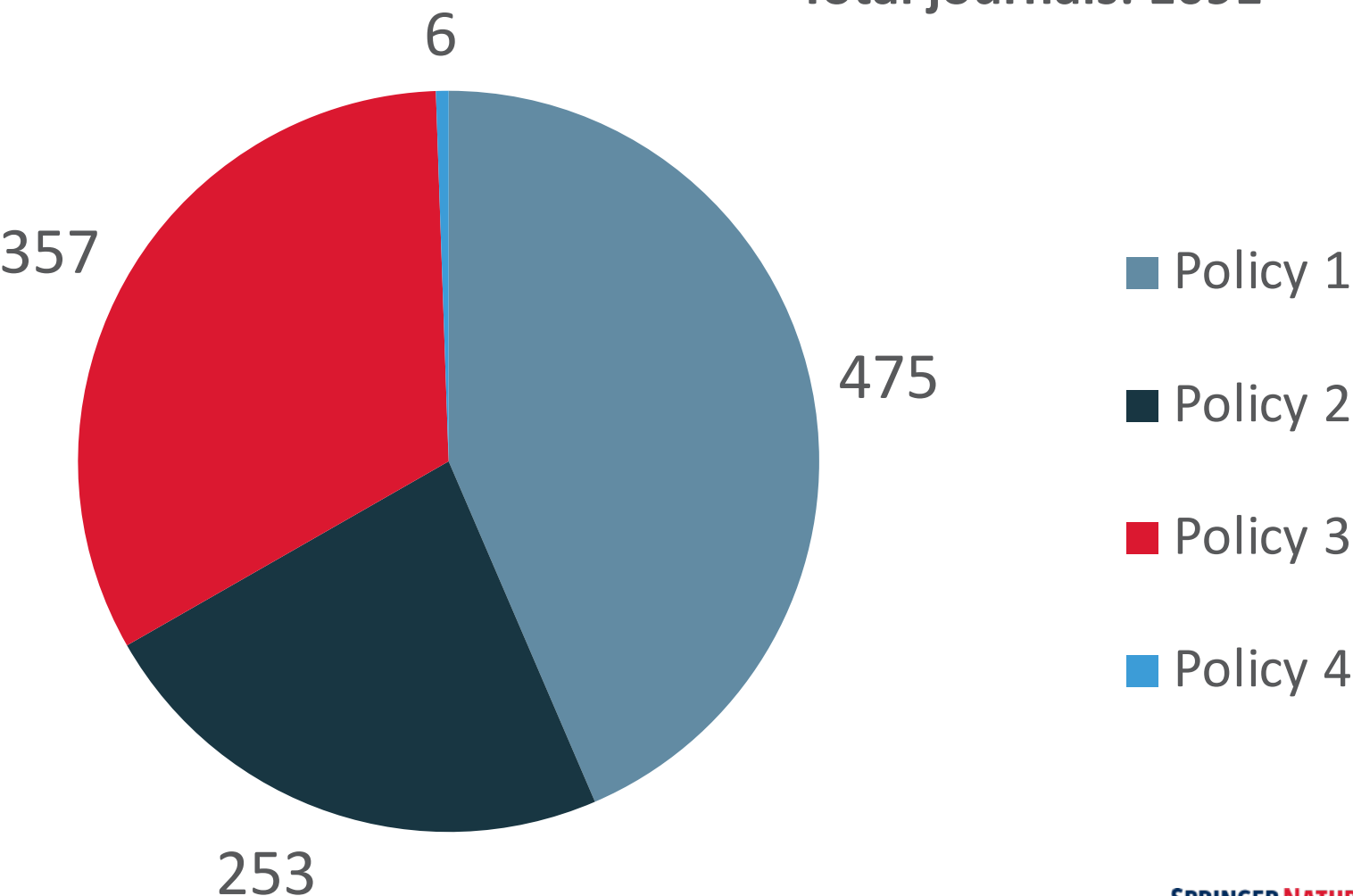
Always encourage best practices, e.g. the use of community repositories for specific data types

Email: researchdata@springernature.com



Policy adoption across journals

Total journals: 1091



Implementing the Type 4 Springer Nature Research Data Policy

Policy Type	Policy summary
Type 1	Data sharing and data citation is encouraged
Type 2	Data sharing and evidence of data sharing encouraged
Type 3	Data sharing encouraged and statements of data availability required
Type 4	Data sharing, evidence of data sharing and peer review of data required

MENU

SCIENTIFIC DATA



Search



E-alert



Submit



Login

Data Descriptor | 12 June 2018 | OPEN

Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping

Comment | 19 June 2018 | OPEN

MEG-BIDS, the brain imaging data structure extended to magnetoencephalography



Guiomar Niso, Krzysztof J. Gorgolewski [...] Sylvain Baillet

Article | 19 June 2018 | OPEN

Sustainable data and metadata management at the BD2K-LINCS Data Coordination and Integration Center



Vasileios Stathias, Amar Koleti [...] Stephan C. Schürer

Data Descriptor | 19 June 2018 | OPEN

Seamless bathymetry and topography datasets for New South Wales, Australia



Kaya M. Wilson & Hannah E. Power

Search Scientific Data

All Subjects



Find out more about Scientific Data



Read our Aims
& Scope

SCIENTIFIC DATA

Discover data associated with our content



isaexplorer
a data discovery tool

Find the right repository for your data

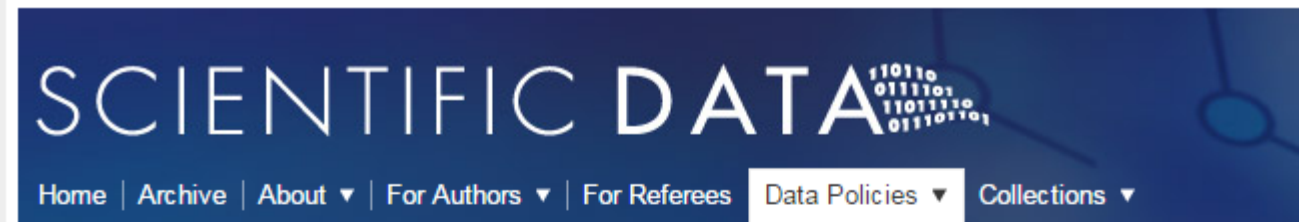


Access our Recommended
Repositories list

SCIENTIFIC DATA

Latest Research

Details of sharing via repositories is referred to in journal guide to authors (Data sharing via repositories supported mandated)

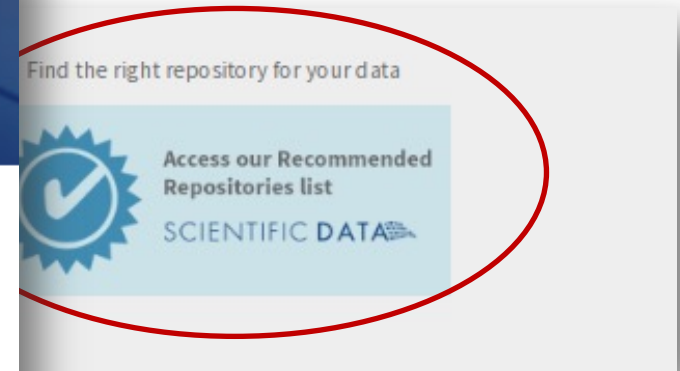


[Home](#) ► [Data Policies](#) ► Recommended Repositories

Recommended Data Repositories

Scientific Data mandates the release of datasets accompanying our manuscripts, but we do not ourselves host data. Instead, we encourage submission of datasets to community-recognized repositories where possible, or to general-science repositories if no community resource is available. Repositories included on this page have been selected based on their compliance with requirements for data access, preservation and stability. Some repositories on this page may only accept data deposition in certain formats or may charge for deposition of data. Please ensure you are aware of the policies of your chosen repository. If your repository of choice is not listed on this page, please contact us for additional repositories.

Authors must deposit their data to a recommended data repository as part of the submission process; manuscripts will not otherwise be considered for publication. Temporary deposition of your data to a general repository is not acceptable. If your repository of choice does not support confidential peer review, please contact us for additional repositories.



View data repositories

- [Biological sciences:](#)
 - [nucleic acid sequence](#); [protein sequence](#); [molecular & supramolecular structure](#); [neuroscience](#); [omics](#); [taxonomy & species diversity](#); [mathematical & modelling resources](#); [cytometry](#); [organism-focused resources](#)
- [Health sciences](#)
- [Chemistry & chemical biology](#)
- [Earth and environmental sciences](#)
- [Physics, astrophysics & astronomy](#)
- [Social sciences](#)
- [Generalist repositories](#)
- [Institutional or project-specific repositories](#)

Submission system/review process integrated with a journal-specific or general repository, such as figshare

The screenshot shows a web interface for manuscript submission. At the top, there are tabs: 'Manuscript Information', 'Files' (highlighted with a red circle), 'Validate', and 'Submit'. Below these are several buttons: 'Upload Files', 'Replace Files', 'File Type', 'File Description', 'File Order', 'Dataset Deposition' (highlighted with a red circle), and 'Dataset Upload'. The 'Dataset Deposition' section contains a question: 'Have you already uploaded ALL of your primary data files to a publicly accessible repository? *'. Below this question are two radio buttons: 'Yes' (selected) and 'No'. A red circle highlights the entire 'Dataset Deposition' section. Below the radio buttons is a text prompt: 'Please enter all details for datasets uploaded to publicly accessible repositories in the fields below'. This is followed by five input fields: 'Repository Name *', 'Title *', 'Accession Number *', 'Url *', and 'Password'. A 'Clear' button is located to the right of the 'Password' field. Below these fields is a button labeled 'Additional Data Files'. At the bottom of the section is a text prompt: 'Were the data collected and reported according to any specific community reporting standards, such as those described in the [EQUATOR Network](#) for health-related research, the [ARRIVE guidelines](#) for animal research, or the [MIBBI](#) checklists, available via the BioSharing portal? *'. Below this prompt are two radio buttons: 'Yes' and 'No'.

Manuscript Information | **Files** | Validate | Submit

Upload Files | Replace Files | File Type | File Description | File Order | **Dataset Deposition** | Dataset Upload

Dataset Deposition * ? Have you already uploaded ALL of your primary data files to a publicly accessible repository? *

☒ Yes ☐ No

Please enter all details for datasets uploaded to publicly accessible repositories in the fields below

Repository Name *	Title *	Accession Number *	Url *	Password
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Clear

Additional Data Files

Were the data collected and reported according to any specific community reporting standards, such as those described in the [EQUATOR Network](#) for health-related research, the [ARRIVE guidelines](#) for animal research, or the [MIBBI](#) checklists, available via the BioSharing portal? *

☐ Yes ☐ No

- During manuscript submission

Peer reviewer guidelines and process give guidance on accessing and reviewing data files

Repository Name	Dataset Title	Dataset Accession Number	URL	Reviewer Passcode
Cancer Imaging Archive	Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection	10.7937/K9/TCIA.2017.KLXWJJ1Q	https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q	Username: Reviewer1 Password: NSDanon1!
	Segmentation Labels			

Reviewer username / passcode allowing access to embargoed data

Repository Name	Dataset Title	Dataset Accession Number	URL	Reviewer Passcode
Gene Expression Omnibus	Classifying distinct grades of human non-alcoholic fatty liver disease employing a systems biology approach	GSE46300	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=yvfoeiqhnczbel&acc=GSE46300	
MetaboLights	Classifying distinct grades of human non-alcoholic fatty liver disease employing a systems biology approach	MTBLS174	http://www.ebi.ac.uk/metabolights/reviewerxpXmh418	

Reviewer links to embargoed data

figshare Fileset http://npg.figshare.com/content/129_35053/be235cd8be46cb3ea

Editorial office available to help reviewers in case of issues accessing data



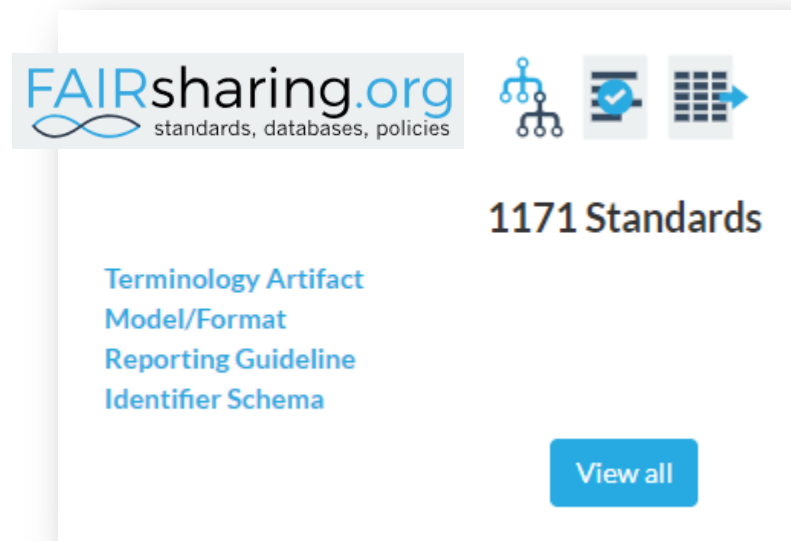
Guide to referees www.nature.com/sdata/policies/for-referees

Peer Reviewers are asked to consider

- Experimental Rigor and Technical Data Quality
- Completeness of the Description
- Integrity of the Data Files and Repository Record

Data deposition checked as part of the publishing process where there is an established research community mandate

- Peer reviewers are specifically asked
 - Consistent with relevant minimum reporting standards?
 - Are data archived to the most appropriate repository?
- Editor & Editorial Board Member will also advise authors on use of appropriate repositories and reporting standards
- Implementation example – publication did not pass peer review until data were moved to a community mandated repository.



Statement in published articles explaining how supporting data can be accessed

Data Descriptor article sections:

- Title
- Abstract
- Background & Summary
- Methods
- **Data Records** ←
- Technical Validation
- Usage Notes
- Figures & Tables
- References
- Data Citations

Data Records

All the samples used in this study are summarized in Table 1. Consistent identifiers are used in Tables 2 and 3 to allow mapping between the proteomic and transcriptomic data outputs.

Data Record 1

The raw data, peaklists (.mgf), ProteomeDiscoverer result files (.msf) and ProteomeDiscoverer workflow files (.xml) have been uploaded to ProteomeXchange (<http://www.proteomexchange.org/>) with the following accession number PXD000134 (ref. 67; Table 2).

Data Record 2

Microarray data are available at the NCBI Gene Expression Omnibus (GEO) database under the accession numbers GSE26451 (ref. 68) and GSE26453 (ref. 69; Table 3).

Data Record 3

The peptide and protein identification data sets have been annotated by The Global Proteome Machine at <http://gpmdb.thegpm.org/>

Data Record 4

The peptide and protein identification data sets have been annotated by the StemCellOmicsRepository (SCOR) at <http://scor.chem.wisc.edu/>

Data citation ~~permitted~~ mandated

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN



Plant traits, productivity, biomass and soil properties from forest sites in the Pacific Northwest, 1999–2014

Logan T. Berner & B...

Affiliations | Contrib...

Scientific Data 3, Article...

The dataset (*NACP TERRA-PNW: Forest Plant Traits, NPP, Biomass, and Soil Properties, 1999–2014*) is hosted with other contributions from the North American Carbon Program (NACP) by the Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics (Data Citation 1) Oak Ridge National Laboratory Distributed Active Archive Center

Data Citations

[Abstract](#) • [Background & Summary](#) • [Methods](#) • [Data Records](#) • [Technical Validation](#) • [Additional Information](#) • [References](#) • [Data Citations](#) • [Acknowledgements](#) • [Author information](#)

1. Law, B. E., & Berner, L. T. *Oak Ridge National Laboratory Distributed Active Archive Center* <http://dx.doi.org/10.3334/ORNLDAAAC/1292> (2015).

```
<ref-list content-type="data-citations">
```

```
<ref id="d1">
```

```
<element-citation>
```

```
<source>Oak Ridge National Laboratory Distributed Active Archive Center</source>
```

```
<ext-link ext-link-type="dummy" specific-use="url"
```

```
xlink:href="http://dx.doi.org/10.3334/ORNLDAAAC/1292">http://dx.doi.org/10.3334/ORNLDAAAC/1292</ext-link>
```

[Abstract](#) • [Background & Summary](#) • [Methods](#) • [Data Records](#) • [Technical Validation](#) • [Additional Information](#) • [References](#) • [Data Citations](#) • [Acknowledgements](#)

Plant trait measurements are needed for evaluating ecosystem conditions and for ecosystem process model development.

Relevant dataset citations in reference lists provided and verified

Data Curation Editors ensure the capture of key metadata about the dataset(s) being described in each Data Descriptor.

Source	Sample	Characteristics	Protocol	Data
Where did the sample originate?	What is your sample?	What do users need to know about your source and samples to fully understand your work?	How did the samples become data?	Where is the data?

During the metadata curation process

- Manuscript re-read
- Data archive checked
- Minor issues with the data and/or manuscript often identified
- Data Citations checked for formatting and accuracy

Data publication at *Scientific Data*

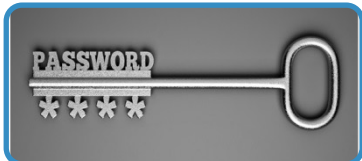


Access our Recommended
Repositories list
SCIENTIFIC DATA

Advice on data repositories available & some repositories integrated with submission system



Editors can advise on best practice for sharing sensitive, human derived data



Reviewers are given access to data during peer review, along with specific guidance



Authors mandated to include data availability and access statements in their manuscripts



Data citation content, format & links checked by specialist editors, editorial office & production team

Implementing data policies across Springer Nature journals: lessons learned

It's important to acknowledge disciplinary differences and how ready different research communities are to share.

Implementing any kind of research data policy will assist a journal in beginning its data sharing journey.

The process must be easy for editors, and a journal-by-journal approach to implementation may be necessary.

To support stronger policies, tools, services and resources are needed.

Thank you

For more information on Research Data Support and other data-related activities at Springer Nature:

Email: researchdata@springernature.com

Website: <http://go.nature.com/ResearchDataServices>

Community portal:
<https://researchdata.springernature.com/>

The story behind the image



Chien Shiung Wu (1912–1997)

Chien Shiung Wu was a Chinese American experimental physicist best known for conducting The Wu experiment that bears her name. This experiment showed that the conservation of parity was violated by a weak interaction and it was possible to distinguish between a mirrored variation of the world and the mirror image of the current world. This discovery earned Wu the Wolf Prize in Physics in 1978.

Creating and implementing research data policies

COPE Webinar

2 July 2018

Lightning Poll 3

How to Implement a Good Research Data Policy? Build It Around Standards.

Todd Carpenter, Executive Director, NISO
COPE Webinar: Creating & implementing research data policies
July 2, 2018

Science data isn't what it used to be



Image: [Walters Art Museum](#)

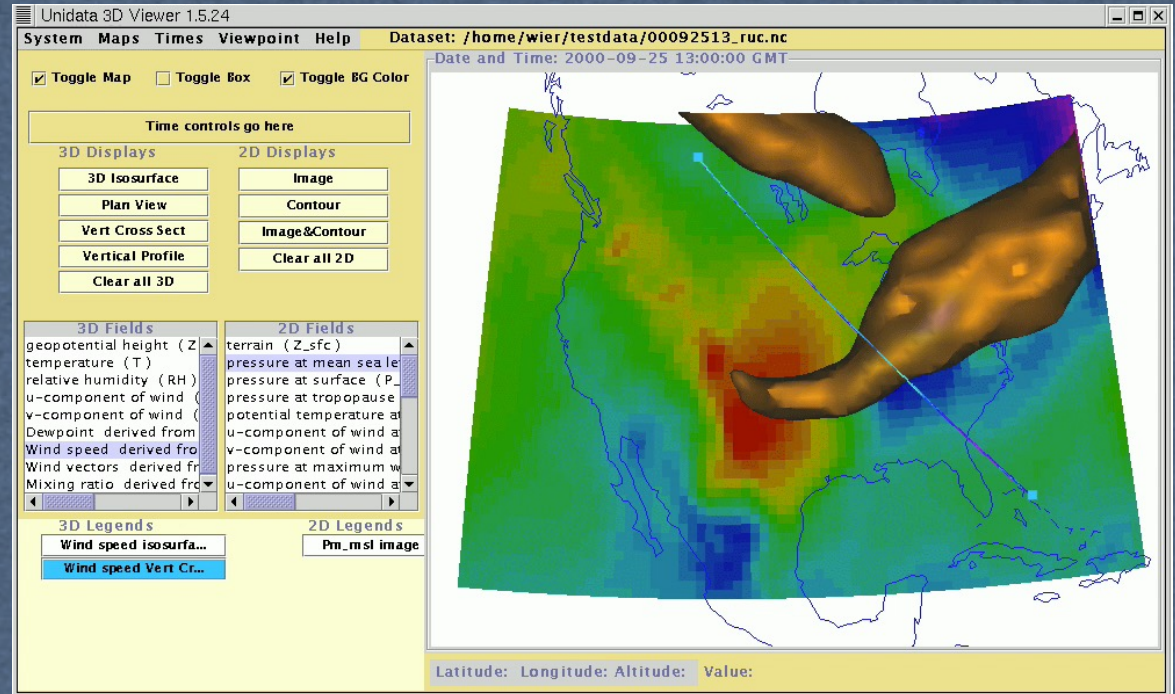


Image: [Domenico Caron, Davis, et al.](#)

Whole Lot of Data Publishing Going On

	up to 2010	2011	2012	2013	2014	2015	Total
3TU.Datacentrum	1692	446	379	345	371	296	3529
CSIRO DAP	0	46	62	438	454	418	1418
Dryad	493	773	1309	1990	2687	2424	9676
Figshare	0	16,929	28,224	108,221	94,223	72,818	320,415
Zenodo	99	24	68	43	268	1107	1609
Total	2284	18,218	30,042	111,037	98,003	77,063	336,647

Table 2

 Excel | CSV

Datasets published by Scientific Data Repositories.

Source: Assante, M. et al., (2016). Are Scientific Data Repositories Coping with Research Data Publishing?.
Data Science Journal. 15, p.6. DOI: <http://dx.doi.org/10.5334/dsj-2016-006>

Number of Research Data Policies grows



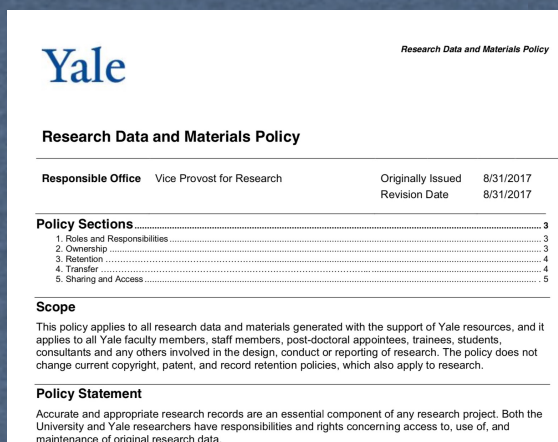
Data Availability



The following policy applies to all PLOS journals, unless otherwise noted.

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.



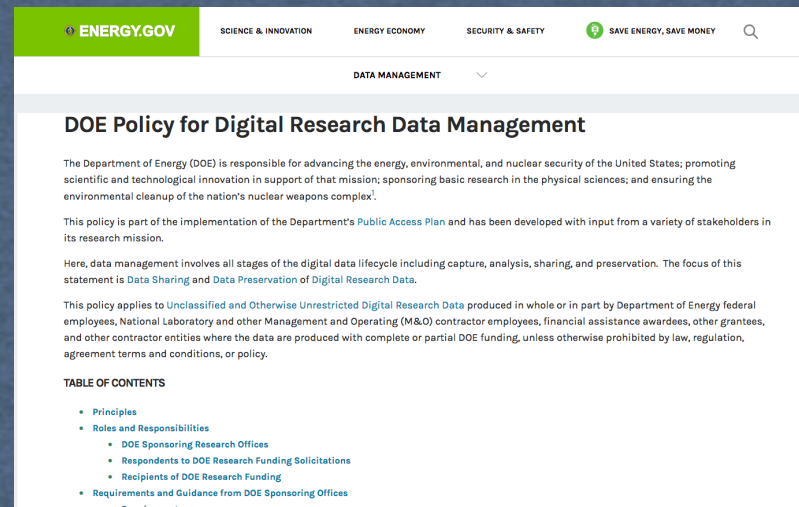
Research Data Policy Types **SPRINGER NATURE**

The 4 types of research data policy are provided in full below. These policy texts are templates and journals may make minor changes to fit with their journal scope and website style. See FAQs for a summary of the requirements of each policy type.

Springer Nature has made the research data policy texts, unless otherwise stated, available for reuse by the research data community under a [Creative Commons attribution license](#).

Here are examples of journals that support each policy type:

Policy Type	Policy summary	Example Journal
Type 1	Data sharing and data citation is encouraged	<i>Photosynthesis Research</i> (click, 'Instructions for Authors')
Type 2	Data sharing and evidence of data sharing encouraged	<i>Plant and Soil</i> (click, 'Instructions for Authors')
Type 3	Data sharing encouraged and statements of data availability required	<i>Palgrave Communications</i> (see Editorial policies)
Type 4	Data sharing, evidence of data sharing and peer review of data required	<i>Scientific Data</i> (see Data policies)



Our standards need to keep up



“The nice thing about standards is that there are so many of them to choose from...”

Andrew S. Tanenbaum

About



- Non-profit industry trade association accredited by American National Standards Institute (ANSI)
- Mission of developing and maintaining technical standards related to information, documentation, discovery and distribution of published materials and media
- Volunteer driven organization: 500+ participants spread out across the world
- Responsible for standards such as DOI, Dublin Core Metadata, SUSHI, MARC, ISBN, ISSN

What roles do standards play in Data Management Policies?

Identification

Description/Discovery

Peer review

Citation

Identification



[ORCID.ORG](https://orcid.org)

Institutional Identification





DataCite

Helping you to find,
access, and reuse data

www.datacite.org

Other Persistent Identifiers

Sold	Persistent identifier	Description	Source	Schema Information, link to project site	Domain relevance	Notes/Description
	ABN	Australian Business Number	https://en.wikipedia.org/wiki/Identifier	http://downloads.icbglobal.org/au/ATO/ATO_Format_of_the_ABN_NAT2956.pdf	Australian businesses, government agencies, charities, super funds	Issued by the Australian Business Register (ABR) which is operated by the Australian Taxation Office https://abr.gov.au/
	AGR	Agricola identifier	https://members.orcid.org/api/supported-work-identifiers		Agriculture	Several identifiers are available in Agricola. See here for more information: http://agricola.nal.usda.gov/help/disphlp1.htm
	ANSI/NISO Number	ANSI/NISO number	http://id.loc.gov/vocabulary/identifiers.html	http://id.loc.gov/vocabulary/identifiers/ansi.html	General use	American National Standards Institute and National Information Standards Organization number for an ANSI or ANSI/NISO standard
	ARK	Archival Resource Key	http://dictionary.casrai.org/Output_ID_Types/ARK	https://wiki.ucop.edu/display/Curation/ARK	General use	Actionable identifiers that can connect to three things: the object itself, a metadata record, and a commitment statement. ARKs are championed by the California Digital Library and there are no fees for assigning or using them.
	arXiv	Eprints identifier for published works	http://dictionary.casrai.org/Output_ID_Types/ARXIV	https://arxiv.org/help/arxiv_identifier	Published works	ArXiv offers a community-based sustainability model for scientific papers in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, and statistics.
	AAVSO ID	AAVSO (American Association of Variable Star Observers) Unique Identifier	https://www.aavso.org/aavso-unique-identifier	https://www.aavso.org/aavso-unique-identifier	Astronomical data literature	Replaces use of the Harvard Designation, which identified stars by their position in the sky. This system supports variable stars, and scales to identify the hundreds of thousands expected to be discovered as modern surveys come online.
	AUID (Elsevier's Scopus)	Scopus author identifier	https://www.elsevier.com/solutions/scopus/support/authorprofile		General	The Scopus author identifier distinguishes authors from one another by assigning each a unique number and then grouping all documents by the same person together.

Description & Discovery

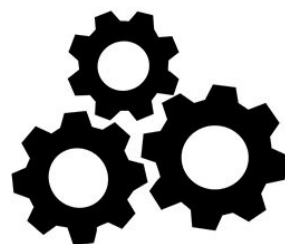
F_{indable}



A_{ccessible}



I_{nteroperable}



R_{eusable}



FAIR Principles Detail

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

FAIR Principles Detail

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

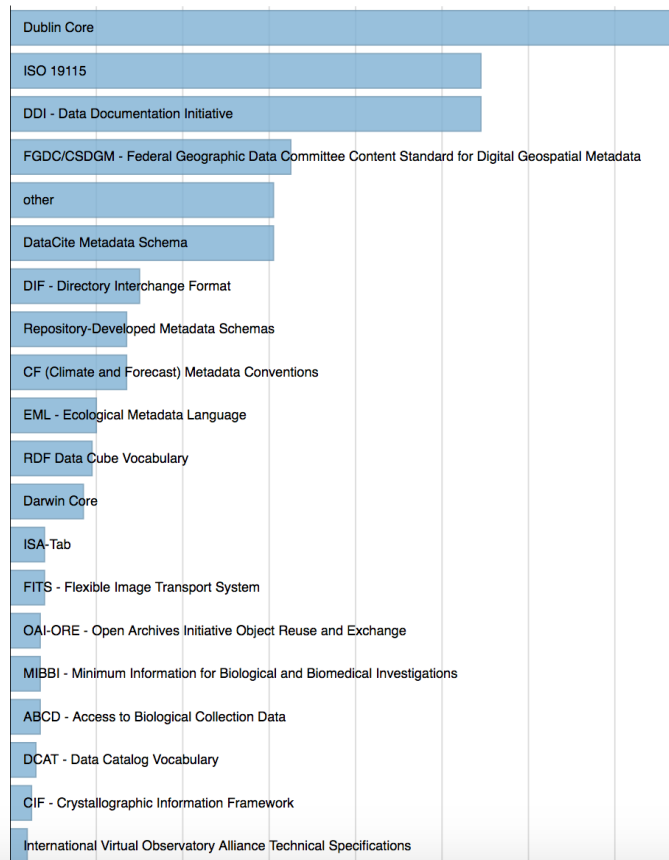
To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Metadata standards



Peer Review

Just like All Other Scholarly Content

“It was generally agreed that data should be peer reviewed.”

- 2014 Study of 4,000 researchers by David Nicholas et al.

Nicholas, David, et al. "Peer review: still king in the digital age." *Learned Publishing* 28.1 (2015): 15-21. Also CICS/CIBER. *Trust and Authority in Scholarly Communications in the Digital Era* (2013). Available at: http://ciber-research.eu/download/20140115-Trust_Final_Report.pdf

WHAT CONSTITUTES PEER REVIEW OF RESEARCH DATA – A STUDY OF CURRENT PRACTICE

TODD A. CARPENTER
EXECUTIVE DIRECTOR, NISO
CO-CHAIR, RDA DATA PUBLISHING IG

1 of 34

Carpenter: What Constitutes Peer Review of Data Study

542 views

Share

Like

Download



National Information Standards Organization (NISO)

Follow



Cornell University
Library

We gratefully acknowledge support from
the Simons Foundation
and member institutions

arXiv.org > cs > arXiv:1704.02236

Search or Article ID inside arXiv: All papers
(Help | Advanced search)

Broaden your search using: Semantic Scholar

Computer Science > Digital Libraries

What Constitutes Peer Review of Data: A survey of published peer review guidelines

Todd A Carpenter

(Submitted on 7 Apr 2017)

Since a number of journals specifically focus on the review and publication of data sets, reviewing their policies seems an appropriate place to start in assessing what existing practice looks like in the 'real world' of reviewing and publishing data. This article outlines a study of the publicly available peer review policies of 39 scientific publications that publish data papers to discern which criteria are most and least frequently referenced. It also compares current practice with proposed criteria published in 2012.

Comments: 17 pages, 7 tables

Subjects: Digital Libraries (cs.DL)

Cite as: arXiv:1704.02236 [cs.DL]

(or arXiv:1704.02236v1 [cs.DL] for this version)

Submission history

From: Todd Carpenter [view email]

[v1] Fri, 7 Apr 2017 14:07:57 GMT (132kb)

Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)

Download:

• PDF only

Current browse context:

cs.DL

< prev | next >

new | recent | 1704

Change to browse by:

cs

References & Citations

• NASA ADS

DBLP – CS Bibliography

listing | bibtex

Todd A. Carpenter

Bookmark

For More:

Article Preprint: <https://arxiv.org/abs/1704.02236>

PPT: <https://www.slideshare.net/BaltimoreNISO/carpenter-what-constitutes-peer-review-of-data-study>

Scholarly Kitchen: <https://scholarlykitchen.sspnet.org/2017/04/11/what-constitutes-peer-review-research-data/>

Five Broad Categories of Peer Review Processes For Data

Editorial Review (9 Criteria)

Metadata Quality (9 Criteria)

Data Quality (15 Criteria)

Methodology (12 Criteria)

Other – (11 Criteria)

MOST-REFERENCED PEER REVIEW CRITERIA

<u>Attribute</u>	<u>Included Statement</u>	<u>Lawrence et al</u>
Editorial Review	36	
Metadata Quality	33	x
Public Data Sharing, Open License Requirement	33	
Overall Quality	31	
Link to Public Repository	31	
Descriptions of How to Access Data	30	
Topical Appropriateness	29	
Citations to Other Relevant Materials	29	
Suitability for Publication	28	x
Ethics of Experimentation	28	
Originality/Novelty - Of Science	27	x
Title/Abstract/Writing Clarity	25	
Methodology - Appropriate	25	x
Metadata Presentation	24	x
Data Reuse	23	

LEAST-REFERENCED PEER REVIEW CRITERIA

<u>Attribute</u>	<u>Included Statement</u>	<u>Lawrence et al</u>
Any data errors introduced by transmissions (checksums)?	3	x
Methodology - Quality Control	3	
Keyword Selection	2	
Data - Of High Quality	2	
Data Anomalies Documented	2	x
Methodology - Equipment Description	2	
Metadata Rights Information	1	x
Data - Plausibility	1	x
How are outliers identified & treated?	1	x
Fairness	1	
Data sharing - Platform Agnostic	1	
Provenance	0	x

WHAT CONSTITUTES ROBUST PEER REVIEW

- Editorial purpose and cohesion is important, even in data publication
- Significant push for openness of data & reuse, with recognition of anonymization or access control if appropriate
- Links to public repositories and details on gaining access
- Ethical concerns regarding data collection need to be documented
- Overall quality of metadata, with specifics depending on domain
- Focus on replicability

Citation

Scholarly Electronic Publishing Bibliography

4.1 General Works: Research (Multiple-Types of Electronic Works)

Al-Aufi, Ali, and Paul Genoni. "An Investigation of Digital Scholarship and Disciplinary Culture in Oman." *Library Hi Tech* 28, no. 3 (2010): 414-432.

Asefeh Asemi, and Nosrat Riyahiniya. "Awareness and Use of Digital Resources in the Libraries of Isfahan University of Medical Sciences, Iran." *The Electronic Library* 25, no. 3 (2007): 316-327.

Atakan, Cemal, Dogan Atilgan, Ölem Bayram, and Sacit Arslantekin. "An Evaluation of the Second Survey on Electronic Databases Usage at Ankara University Digital Library." *The Electronic Library* 26, no. 2 (2008): 249-259.

Bhatt, R. K. "Use of UGC-Infonet Digital Library Consortium Resources by Research Scholars and Faculty Members of the University of Delhi in History and Political Science: A Study." *Library Management* 31, no. 4/5 (2010): 319-343.

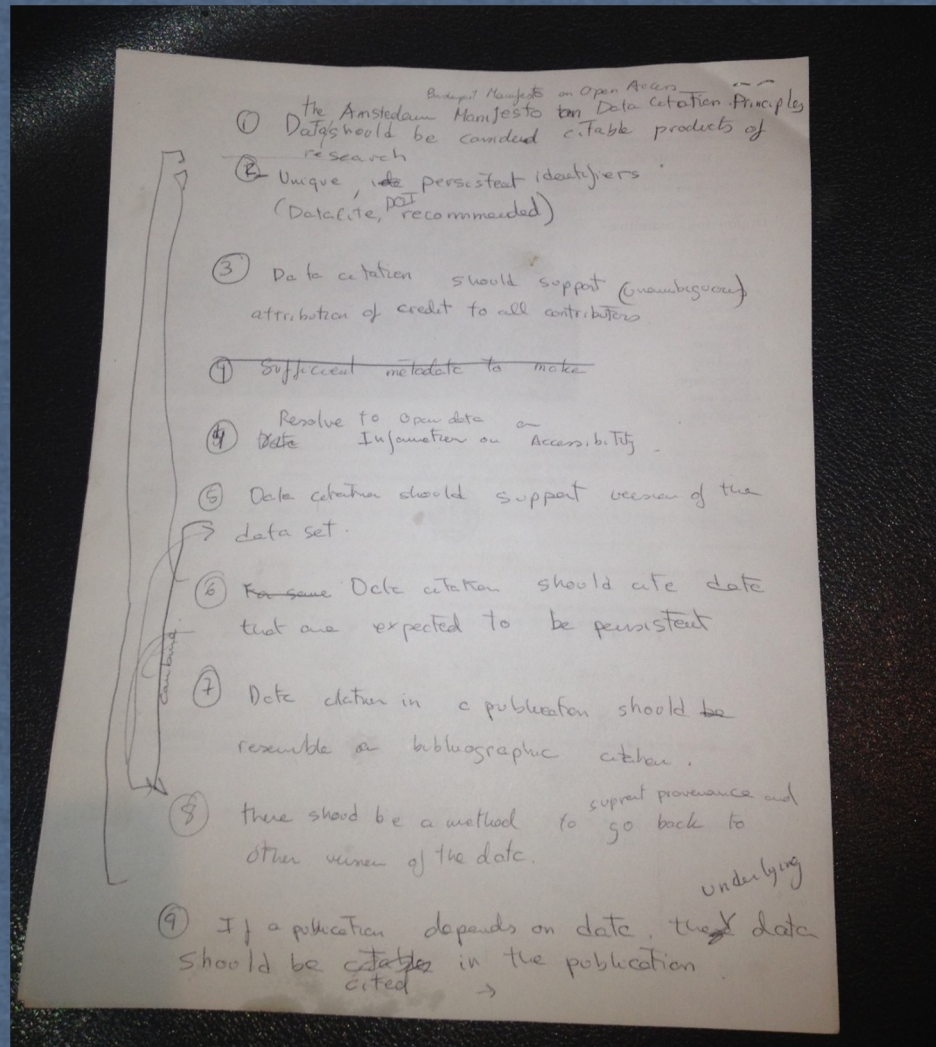
Björk, Bo-Christer, and Ziga Turk. "How Scientists Retrieve Publications: An Empirical Study of How the Internet is Overtaking Paper Media." *The Journal of Electronic Publishing* 6 (December 2000). <http://hdl.handle.net/2027/spo.3336451.0006.202>

Boumarafi, Behdja. "Electronic Resources at the University of Sharjah Medical Library: An Investigation of Students' Information-Seeking Behavior." *Medical Reference Services Quarterly* 29, no. 4 (2010): 349-362.

Crawford, John. "The Use of Electronic Information Services and Information Literacy: A Glasgow Caledonian University Study." *Journal of Librarianship and Information Science* 38, no. 1 (2006): 33-44.

Cryer, Emma, and Karen S. Grigg. "Consortia and Journal Package Renewal: Evolving Trends in the 'Big Package Deal'?" *Journal of Electronic Resources in Medical Libraries* 8, no. 1 (2011): 22-34.

Amsterdam Manifesto on Data Citation



Joint Declaration of Data Citation Principles

JOINT DECLARATION OF DATA CITATION PRINCIPLES - FINAL



When citing please use: Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [[datacitation](#)].

ENDORSEMENT LIST

PREAMBLE

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.

In support of this assertion, and to encourage good practice, we offer a set of guiding principles for data within scholarly literature, another dataset, or any other research object.

These principles are the synthesis of work by a [number of groups](#). As we move into the next phase, we welcome your participation and endorsement of these principles.

DC¹
Data Citation Principles

Advancing Standards for Citations

Revision of ISO 690 – Bibliographic References

Revision to include more robust coverage of data,
datasets and other material forms

Revision of NISO Z39.29 - Bibliographic References

Similar issues as above

Many Bibliographies are also revising their practices
with regard to coverage and inclusion of data
citations

TOP GUIDELINES

TRANSPARENCY AND OPENNESS PROMOTION

THE GUIDELINES

Transparency, open sharing, and reproducibility are core values of science, but not always part of daily practice. Journals, funders, and societies can increase reproducibility of research by adopting the TOP Guidelines and helping them evolve to meet the needs of researchers and publishers while pursuing the most transparent practices.

8 MODULAR STANDARDS

Citation Standards Describes citation of data	Data Transparency Describes availability and sharing of data
Analytical Methods Transparency Describes analytical code accessibility	Research Materials Transparency Describes research materials accessibility
Design and Analysis Transparency Sets standards for research design disclosures	Preregistration of Studies Specification of study details before data collection
Preregistration of Analysis Plans Specification of analytical details before data collection	Replication Encourages publication of replication studies

ACROSS 3 TIERS

1 DISCLOSURE:
the final research output must disclose if the work satisfies the standard

2 REQUIREMENT:
the final research output must satisfy the standard

3 VERIFICATION:
third party must verify that the standard is being met

WHAT DOES YOUR SIGNATURE MEAN?

A statement of approval for the principle of rewarding transparency in research.

The organization commits to review the standards within one year and determine which are appropriate for their field.

OVER 5,000 JOURNAL SIGNATORIES

LEARN MORE AT [COS.IO/TOP](https://cos.io/top)



The Center for Open Science is a non-profit company with the mission of improving the openness, integrity, and reproducibility of scientific research.

COS: cos.io | OSF: osf.io | Email: contact@cos.io

TOP2 Implementation

Meeting in Charlottesville, VA in September 2017

Goal is to create implementable approach to quality measures based on the openness and replicability of science

Not judging science on citations, but how well it is being communicated.

Transparency and Openness Promotion (TOP) Statement Guidelines

“We propose that, going forward, authors of all scientific articles disclose the availability and location of all research items, including data, materials, and code, related to their published articles in what we will refer to as a TOP Statement.”

<http://bit.ly/top2preprint>

[The authors are currently seeking feedback on this preprint. Please email david@cos.io or comment directly on a working draft here: <http://bit.ly/top2preprint>]

Making Science Transparent By Default; Introducing the TOP Statement

The TOP Statement Working Group¹

In order to increase the replicability of scientific work, the scientific community has called for practices designed to increase the transparency of research (McNutt, 2014; Nosek et al., 2015). The validity of a scientific claim depends not on the reputation of those making the claim, the venue in which the claim is made, or the novelty of the result, but rather on the empirical evidence provided by the underlying data and methods. Proper evaluation of the merits of scientific findings requires availability of the methods, materials, and data and the reasoned argument that serve as the basis for the published conclusions (Clairbout and Karrenbach 1992; Donoho, et al 2009; Stodden, et al 2013; Borwein, et al 2013; Munafò, et al, 2017). Wide and growing support for these principles (see, for example, signatories to Declaration on Research Assessment, DORA, <https://sf.dora.org/>), and the Transparency and Openness Promotion Guidelines <https://cos.io/our-services/top-guidelines/>) must be coupled with

¹ Authors listed alphabetically: [Ulfarsson](#), [Jan Aalbersberg](#), Elsevier; Tom Appleyard, [BioMed Central](#), Springer Nature; Sarah Brookhart, Association for Psychological Science; Todd Carpenter, National Information Standards Organization; Michael Clarke, Clarke & Company, STM Advisers; Stephen Curry, Imperial College; Josh Dahl, [Clarivate Analytics](#); Alex [DeHaven](#), Center for Open Science; Eric [Eich](#), University of British Columbia; [Macros](#), Franko, Health Research Alliance; Len Freedman, Global Biological Standards Institute; Chris Graf, Wiley and COPE; Sean Grant, RAND Corporation; Brooks Hanson, American Geophysical Union; Heather Joseph, Scholarly Publishing and Academic Resources Coalition; Veronique [Kucenas](#), Public Library of Science; Bianca Kramer, Utrecht University; Alan Kraut, Center for Open Science Board Member; Roshan Kumar [Kato](#), Open Access Nepal; Carole Lee, University of Washington; Ais MacFarlane, Wellcome Trust; Maryann [Madison](#), UC San Diego; Evan Mayo-Wilson, Johns Hopkins Bloomberg School of Public Health; Marcia McNutt, National Academy of Sciences; Meredith McPhail, Laura and John Arnold Foundation; David Mellor, Center for Open Science; David Moher, University of Ottawa, Equator Network; Alison [Mudditt](#), Public Library of Science; Brian [Nosek](#), Center for Open Science; Belinda Orland, American Heart Association; Tim Parker, Whitman College; Mark Parsons, Rensselaer Polytechnic Institute; Mark Patterson, [eLife](#); Solange Santos, Scientific Electronic Library Online; Carolyn Shore, National Academy of Sciences; Dan Simons, Association for Psychological Science; Bobbie Spellman, University of Virginia; Jeff Spies, Center for Open Science; Matt Spitzer, Center for Open Science; Victoria [Stodden](#), University of Illinois; [Sowmya Swaminathan](#), Springer Nature; Deborah Sweet, Cell Press; Anne [Tsui](#), Notre Dame, [Simone Vazira](#), UC Davis

TOP2 Implementation

The disclosure would address a simple set of questions:

Is the item publicly available in a persistent location?
("Yes", "Not applicable", or "See explanation")

If Yes, provide unique, persistent identifier(s) and applicable license information.

Otherwise, provide a brief explanation.

TOP Statements are a mandate for disclosure, not a mandate for access

Thank you!

Todd Carpenter, Executive Director

tcarpenter@niso.org

National Information Standards Organization (NISO)

3600 Clipper Mill Rd, Suite 302

Baltimore, MD 21211 USA

+1 (301) 654-2512

www.niso.org

Creating and implementing research data policies

COPE Webinar

2 July 2018

Questions?

COPE Webinar: Creating and implementing research data policies



Trevor Lane

Education &
Engagement
Consultant,
Edanz Group

COPE Council
Member
(Chair, Member
Services)



Rebecca Grant

Research Data
Manager,

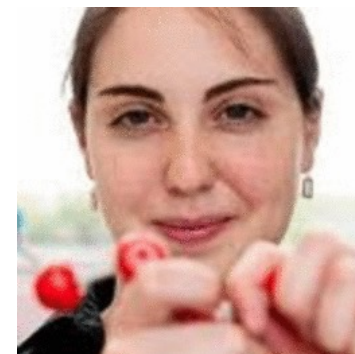
Varsha Khodiyar

Data Curation
Manager,
Open Research
Group, Springer
Nature



Todd Carpenter

Executive
Director, NISO
National
Information
Standards
Organization



Heather Tierney

Managing Editor
Journals and
Ethics Policy,
American
Chemical Society

COPE Council
Member
(Chair,
Education)

Next steps

- **COPE Forum discussion notes on data sharing, 12 Feb 2016**
https://publicationethics.org/files/u661/Notes%20from%20Forum%20Discussion%20Topic_12_February%202016_DATA_SHARING_final.pdf
 - If you have any comments, please email the Executive Officer, Natalie Ridgeway, cope_execofficer@publicationethics.org
- **Digest Jan 2018: Research integrity and how to buy a Persian carpet: Transparency and Openness Promotion (TOP) Guidelines, part 2**
<https://publicationethics.org/news/research-integrity-and-how-buy-persian-carpet-top-guidelines-part-2>
 - Comment on the TOP Part 2 G-doc
- Please give us your **feedback** by responding to the email we will send you after this webinar

Creating and implementing research data policies

COPE Webinar

2 July 2018

Guest speakers:

- **Rebecca Grant** (Research Data Manager) & **Varsha Khodiyar** (Data Curation Manager), Open Research Group, Springer Nature, Rebecca.Grant@springernature.com, Varsha.Khodiyar@nature.com
- **Todd Carpenter**, Executive Director, National Information Standards Organization, tcarpenter@niso.org

Moderator:

- **Trevor Lane**, Education & Engagement Consultant, Edanz Group and COPE Council Member, tlane@edanzgroup.com

Creating and implementing research data policies

COPE Webinar

2 July 2018

Thank you!

Special thanks to:

COPE Education Subcommittee and

Linda Gough, Sarah Gilmore, Sabah Moran, Natalie Ridgeway

