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

Thank you for joining the COPE Webinar on Creating and implementing research data policies. The Webinar will begin promptly at 4.00pm (BST) and finish at 5.00pm (BST).

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](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
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

(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

The role of research institutions

• Data Sharing Policies in Scholarly Publications: Interdisciplinary Comparisons
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
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.


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 polices is difficult

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


Data source: Linda Naughton, JISC Journal Research Data Policy Bank project presentation (n = 250)
**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
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
<table>
<thead>
<tr>
<th>Research Data Helpdesk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queries are answered within two business days</td>
</tr>
<tr>
<td>Run by members of the Springer Nature Data Publishing team</td>
</tr>
<tr>
<td>Expertise in data curation and management, archiving and digital preservation, copyright and licensing, Open Access publishing</td>
</tr>
<tr>
<td>Always encourage best practices, e.g. the use of community repositories for specific data types</td>
</tr>
<tr>
<td>Email: <a href="mailto:researchdata@springernature.com">researchdata@springernature.com</a></td>
</tr>
</tbody>
</table>
Policy adoption across journals

Total journals: 1091

- Policy 1: 475
- Policy 2: 253
- Policy 3: 357
- Policy 4: 6

SPRINGER NATURE
Implementing the Type 4 Springer Nature Research Data Policy

<table>
<thead>
<tr>
<th>Policy Type</th>
<th>Policy summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>Data sharing and data citation is encouraged</td>
</tr>
<tr>
<td>Type 2</td>
<td>Data sharing and evidence of data sharing encouraged</td>
</tr>
<tr>
<td>Type 3</td>
<td>Data sharing encouraged and statements of data availability required</td>
</tr>
<tr>
<td>Type 4</td>
<td>Data sharing, evidence of data sharing and peer review of data required</td>
</tr>
</tbody>
</table>
Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping
Details of sharing via repositories is referred to in journal guide to authors (Data sharing via repositories supported mandated)

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 requirements for data access, preservation and stability. The repositories on this page may only accept data deposition. Some repositories 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 here, please refer to additional repositories.

Authors must deposit their data to a recommended data repository as part of our submission process; manuscripts will not otherwise be considered for publication. Temporary deposition of your data to a general repository for your data does not support confidential peer review. Other than the peer review process, Scientific Data does not review the deposited data.

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

www.nature.com/sdata/data-policies/repositories
Submission system/review process integrated with a journal-specific or general repository, such as figshare

• During manuscript submission
Peer reviewer guidelines and process give guidance on accessing and reviewing data files

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>Dataset Title</th>
<th>Dataset Accession Number</th>
<th>URL</th>
<th>Reviewer Passcode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Segmentation Labels</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MetaboLights</td>
<td>Classifying distinct grades of human non-alcoholic fatty liver disease employing a systems biology approach</td>
<td>MTBLS174</td>
<td><a href="http://www.ebi.ac.uk/metabolights/reviewex0Xmh45">http://www.ebi.ac.uk/metabolights/reviewex0Xmh45</a></td>
<td></td>
</tr>
</tbody>
</table>

Reviewer username / passcode allowing access to embargoed data

Reviewer links to embargoed data

Guide to referees [www.nature.com/sdata/policies/for-referees](http://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

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

![FAIRsharing.org](https://fairstandards.org)

**1171 Standards**

- Terminology Artifact
- Model/Format
- Reporting Guideline
- Identifier Schema

**View all**
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 StemCellOmicsRepository (SCOR) at http://scor.chem.wisc.edu/
Data citation permitted mandated

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

Logan T. Berner & B. E. Law

Affiliations | Contributions

Scientific Data 3, Article ID 1292 (2015)

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
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.

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**

- 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.
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
Whole Lot of Data Publishing
Going On

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

**Table 2**
Datasets published by Scientific Data Repositories.

Number of Research Data Policies grows

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:

<table>
<thead>
<tr>
<th>Policy Type</th>
<th>Policy summary</th>
<th>Example Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>Data sharing and data citation is encouraged</td>
<td>Photosynthesis Research (link, Instructions for Authors)</td>
</tr>
<tr>
<td>Type 2</td>
<td>Data sharing and evidence of data sharing encouraged</td>
<td>Plant and Soil (link, Instructions for Authors)</td>
</tr>
<tr>
<td>Type 3</td>
<td>Data sharing encouraged and statements of data availability required</td>
<td>Plasmid Communications (see Editorial policies)</td>
</tr>
<tr>
<td>Type 4</td>
<td>Data sharing, evidence of data sharing and peer review of data required</td>
<td>Scientific Data (see Data policies)</td>
</tr>
</tbody>
</table>

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 and Materials Policy

Yale University

Responsible Office: Vice Provost for Research

Policy Summary:

This policy applies to all research data materials generated with the support of Yale resources and it applies to all Yale faculty members, staff members, post-doctoral appointees, residents, students, consultants and any others involved in the design, conduct or reporting of research. The policy does not change current copyright, patent, and record retention policies, which also apply to research.

Scope:

This policy applies to all research data materials generated with the support of Yale resources, and it applies to all Yale faculty members, staff members, post-doctoral appointees, residents, students, consultants and any others involved in the design, conduct or reporting of research. The policy does not change current copyright, patent, and record retention policies, which also apply to research.

Policy Statement:

Accurate and appropriate research records are an essential component of any research project. Both the University and Yale researchers have responsibilities and rights concerning access to, use of, and maintenance of original research data.

TABLE OF CONTENTS

- Principles
- Roles and Responsibilities
- DOE Sponsoring Research Offices
- Responsibilities of DOE Sponsoring Agencies
- Use of DOE Research Funding
- Requirements and Guidance from DOE Sponsoring Offices
- Requirements
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
Institutional Identification
# Other Persistent Identifiers

<table>
<thead>
<tr>
<th>ID</th>
<th>Persistent Identifier</th>
<th>Description</th>
<th>Source</th>
<th>Domain Relevance</th>
<th>Notes/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABN</td>
<td>Australian Business Number</td>
<td>Australian business number</td>
<td><a href="https://en.wikipedia.org/wiki/Identifier">https://en.wikipedia.org/wiki/Identifier</a></td>
<td>Australian businesses, government agencies, charities, super funds</td>
<td>Issued by the Australian Business Register (ABR) which is operated by the Australian Taxation Office <a href="https://abr.gov.au">https://abr.gov.au</a></td>
</tr>
<tr>
<td>AGR</td>
<td>Agricola identifier</td>
<td>Agricola identifier</td>
<td><a href="https://members.orcid.org/api/supported-work-identifiers">https://members.orcid.org/api/supported-work-identifiers</a></td>
<td>Agriculture</td>
<td>Several identifiers are available in Agricola. See here for more information: <a href="http://agricola.nal.usda.gov/help/disphlp1.htm">http://agricola.nal.usda.gov/help/disphlp1.htm</a></td>
</tr>
<tr>
<td>ARK</td>
<td>Archival Resource Key</td>
<td>Archival Resource Key</td>
<td><a href="http://dictionary.casrai.org/Output_ID_Types/ARK">http://dictionary.casrai.org/Output_ID_Types/ARK</a></td>
<td>General use</td>
<td>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.</td>
</tr>
<tr>
<td>AAVSO ID</td>
<td>AAVSO (American Association of Variable Star Observers) Unique Identifier</td>
<td>AAVSO (American Association of Variable Star Observers) Unique Identifier</td>
<td><a href="https://www.aavso.org/aavso-unique-identifier">https://www.aavso.org/aavso-unique-identifier</a></td>
<td>Astronomical data literature</td>
<td>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.</td>
</tr>
<tr>
<td>AUID</td>
<td>Scopus author identifier</td>
<td>Scopus author identifier</td>
<td><a href="https://www.elsevier.com/solutions/scopus/support/authorprofile">https://www.elsevier.com/solutions/scopus/support/authorprofile</a></td>
<td>General</td>
<td>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.</td>
</tr>
</tbody>
</table>
Description & Discovery
Findable  Accessible  Interoperable  Reusable
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
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. metadata 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

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dublin Core</td>
<td></td>
</tr>
<tr>
<td>ISO 19115</td>
<td></td>
</tr>
<tr>
<td>DCI - Data Documentation Initiative</td>
<td></td>
</tr>
<tr>
<td>FGDC/CDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
</tr>
<tr>
<td>DataCite Metadata Schema</td>
<td></td>
</tr>
<tr>
<td>DIF - Directory Interchange Format</td>
<td></td>
</tr>
<tr>
<td>Repository-Developed Metadata Schemas</td>
<td></td>
</tr>
<tr>
<td>CF (Climate and Forecast) Metadata Conventions</td>
<td></td>
</tr>
<tr>
<td>EML - Ecological Metadata Language</td>
<td></td>
</tr>
<tr>
<td>RDF Data Cube Vocabulary</td>
<td></td>
</tr>
<tr>
<td>Darwin Core</td>
<td></td>
</tr>
<tr>
<td>ISA-Tab</td>
<td></td>
</tr>
<tr>
<td>FITS - Flexible Image Transport System</td>
<td></td>
</tr>
<tr>
<td>OA-ORE - Open Archives Initiative Object Reuse and Exchange</td>
<td></td>
</tr>
<tr>
<td>MIBBI - Minimum Information for Biological and Biomedical Investigations</td>
<td></td>
</tr>
<tr>
<td>AIBCD - Access to Biological Collection Data</td>
<td></td>
</tr>
<tr>
<td>DCAT - Data Catalog Vocabulary</td>
<td></td>
</tr>
<tr>
<td>CIF - Crystallographic Information Framework</td>
<td></td>
</tr>
<tr>
<td>International Virtual Observatory Alliance Technical Specifications</td>
<td></td>
</tr>
</tbody>
</table>
Peer Review
“It was generally agreed that data should be peer reviewed.”

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

For More:


# Five Broad Categories of Peer Review Processes for Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editorial Review</td>
<td>9</td>
</tr>
<tr>
<td>Metadata Quality</td>
<td>9</td>
</tr>
<tr>
<td>Data Quality</td>
<td>15</td>
</tr>
<tr>
<td>Methodology</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
</tbody>
</table>
## MOST-REFERENCED PEER REVIEW CRITERIA

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Included Statement</th>
<th>Lawrence et al</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editorial Review</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Metadata Quality</td>
<td>33</td>
<td>x</td>
</tr>
<tr>
<td>Public Data Sharing, Open License Requirement</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Overall Quality</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Link to Public Repository</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Descriptions of How to Access Data</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Topical Appropriateness</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Citations to Other Relevant Materials</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Suitability for Publication</td>
<td>28</td>
<td>x</td>
</tr>
<tr>
<td>Ethics of Experimentation</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Originality/Novelty - Of Science</td>
<td>27</td>
<td>x</td>
</tr>
<tr>
<td>Title/Abstract/Writing Clarity</td>
<td>25</td>
<td>x</td>
</tr>
<tr>
<td>Methodology - Appropriate</td>
<td>25</td>
<td>x</td>
</tr>
<tr>
<td>Metadata Presentation</td>
<td>24</td>
<td>x</td>
</tr>
<tr>
<td>Data Reuse</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>Included Statement</td>
<td>Lawrence et al</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Any data errors introduced by transmissions (checksums)?</td>
<td>3</td>
<td>x</td>
</tr>
<tr>
<td>Methodology - Quality Control</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Keyword Selection</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Data - Of High Quality</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Data Anomalies Documented</td>
<td>2</td>
<td>x</td>
</tr>
<tr>
<td>Methodology - Equipment Description</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Metadata Rights Information</td>
<td>1</td>
<td>x</td>
</tr>
<tr>
<td>Data - Plausibility</td>
<td>1</td>
<td>x</td>
</tr>
<tr>
<td>How are outliers identified &amp; treated?</td>
<td>1</td>
<td>x</td>
</tr>
<tr>
<td>Fairness</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Data sharing - Platform Agnostic</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Provenance</td>
<td>0</td>
<td>x</td>
</tr>
</tbody>
</table>
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
Scholarly Electronic Publishing Bibliography

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


Amsterdam Manifesto on Data Citation

1. Data should be regarded as a central product of research.
2. Unique and persistent identifiers (DataCite, recommended)
3. Data citation should support meaningful attribution of credit to all contributors.
4. Sufficient metadata to make data citable.
5. Resolve to open data.
6. Data should support version of the data set.
7. Ensure data citation should cite datasets that are expected to be persistent.
8. Date citation in a publication should be resolvable in bibliographic citation.
9. There should be a method to go back to other versions of the data.
10. If a publication depends on data, the data should be cited in the publication cited.
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.
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

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

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
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.”

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?

publicationethics.org
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)

publicationethics.org
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
  o 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
  o 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

publicationethics.org