Open and FAIR Data Sharing: The Enabling FAIR Data project



6 June 2019

Shelley Stall, AGU Sr. Director, Data Leadership sstall@agu.org @ShelleyStall https://orcid.org/0000-0003-2926-8353



The National Academies of SCIENCES • ENGINEERING • MEDICINE

CONSENSUS STUDY REPORT

OPEN SCIENCE BY DESIGN

Realizing a Vision for 21st Century Research



National Academies of Sciences, Engineering, and Medicine. 2018. Open Science by Design: Realizing a Vision for 21st Century Research. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116.

Open Science by Design is aimed at overcoming barriers and **moving** toward open science as the default approach across the research enterprise. This report explores specific examples of open science and discusses a range of challenges, focusing on stakeholder perspectives. It is meant to provide guidance to the research enterprise and its stakeholders as they build strategies for achieving open science and take the next steps. Report released July 2018



Definition of Open Science

Open science aims to ensure the open availability and usability:

- scholarly publications
- the data that result from scholarly research
- and the methodologies, including code or algorithms, that were used to generate those data.



Why Open Science?

Openness and sharing of information are fundamental to the progress of science and to the effective functioning of the research enterprise.

Specific benefits:

- Rigor and reliability
- Ability to address new questions
- Faster and more inclusive dissemination of knowledge
- Broader participation in research
- Effective use of resources
- Improved performance of research tasks
- Open publication for public benefit



Barriers and Limitations

- Costs and infrastructure
- Structure of scholarly communications
- Lack of supportive culture, incentives and training
- Privacy, security and proprietary barriers to sharing
- Disciplinary differences



We are at an inflection point

- New IT tools/services hold potential to revolutionize scientific practice
- Research funders have introduced mandates and support systems to ensure that the results of the research they sponsor are open
- Publishers are adopting open frameworks and strengthening data/methods openness requirements
- US federal policies support openess



1. Building a Supportive Culture

Finding: Continued effort by stakeholders, working <u>internationally</u> and across disciplinary boundaries, is needed to change evaluation practices and introduce other incentives.

Recommendation One: Research institutions should work to create a culture that actively supports Open Science by Design by better *rewarding and supporting researchers engaged in open science practices.*



2. Training for Open Science by Design

Finding: There is little formal training and education in the principles and practices of open science. The emergence of data science as a recognized interdisciplinary field has highlighted the <u>need for new educational</u> <u>content and approaches related to data</u>.

Recommendation Two: Research institutions and professional societies should *train students and other researchers to implement open science practices effectively* and should support the development of educational programs that foster Open Science by Design.



3. Ensuring long term preservation and stewardship Finding: <u>Developing and sustaining the infrastructure</u> required for longterm stewardship of research products will present a continuing challenge. Recommendation Three: Research funders and research institutions should develop the policies and procedures to identify the data, code,

specimens, and other research products that should be preserved for long-term public availability, and they should *provide the resources necessary* for the long-term preservation and stewardship of those research products.



4. Facilitating data discovery and reuse

Finding: As progress toward opens science by design continues, it is important that the community adhere to the ultimate goal of <u>achieving</u> <u>the availability of research products under open principles</u>.

Recommendation Four: Funders that support the development of research archives should work to ensure that these are *designed and implemented according to the FAIR data principles*. Researchers should seek to ensure that their research products are made available according to the FAIR principles and state with specificity any exceptions based on legal and ethical considerations.



Five Findings and Recommendations 5. Developing new approaches to fostering Open Science by Design

Finding: Public and private funders have made significant

contributions to fostering open science to this point. They

should continue to support initiatives that accelerate progress,

and evaluate and revise their policies as needed.

Recommendation Five: *The research community should work together* to realize Open Science by Design to advance science and help science better serve the needs of society.



The National Academies of SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

Reproducibility and Replicability in Science

National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

Reproducibility and Replicability in Science defines reproducibility and replicability and examines the factors that may lead to non-reproducibility and nonreplicability in research. Unlike the typical expectation of reproducibility between two computations, expectations about replicability are more nuanced, and in some cases a lack of replicability can aid the process of scientific discovery. This report provides recommendations to researchers, academic institutions, journals, and funders on steps they can take to improve reproducibility and replicability in science. Released 2019

AND SPACE SCIENC



AGU: Open and FAIR Data



AGU's position statement on data affirms that

"Earth and space sciences data are a world heritage. Properly documented, credited, and preserved, they will help future scientists understand the Earth, planetary, and heliophysics systems."





Reluctance to share data or models Lack of computational and numerical analysis skills Lack of awareness of relevant and potential data sources Poor programming style or data analysis workflow Lack of well-maintained libraries and codes Inability to process large amounts of data Overuse of spreadsheets Lack of multi- and interdisciplinary collaborations Poor compliance with data management requirements Inefficient numerical analysis, no awareness of advances Reliance on old fashioned languages or legacy systems Other

Data Management Skills Gap Analysis, April 7, 2017 http://bfe-inf.org/document/skills-gap-analysis

FAIR Guiding Principles

FAIR is... Findable Accessible Interoperable Reusable

Article in Nature journal *Scientific Data*: Wilkinson,
M. D. *et al.* The FAIR Guiding Principles for
scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).



FAIR Data Principles (applies to software too)

• Findable

 Assign persistent IDs (PIDs), provide rich metadata, register in a searchable resource, ...

• Accessible

 Retrievable by their ID using a standard protocol, metadata remain accessible even when data are no longer available...

Interoperable

 Use formal, broadly applicable languages, use standard vocabularies, qualified references...

Reusable

 Rich, accurate metadata, clear licenses, provenance, use of community standards...

Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

Evolving Support for Open Science and Data



Research Data Ecosystem



Other Roles:

- Research Labs
- Service providers to the ecosystem (e.g. PID providers like DataCite, github/Zenodo, CrossRef, CHORUS, Scholix)
- Research offices -- not at institutions (e.g. Ronin)
- International Efforts
- Societies
- Cyberinfrastructure
- IRBs
- Industry

Authors: Shelley Stall, AGU Director of Data Program and Erin Robinson, Executive Director, Earth Science Information Partners

https://copdess.org/enabling-fair-data-project/

COPDESS

COPDESS V Enabling FAIR Data V Data FAIRs V

Q

Coalition for Publishing Data in the Earth and Space Sciences



enable PAR (initiable, accessible, interoperable, and reusable) data on a large scale. This project will accelerate scien

transparency, and reproducibility of this data.

Commitment to Enabling FAIR Data in the Earth, Space, and Environmental Sciences

Publication of scholarly articles in the Earth space and snyiropmental science computity is conditional upon the consurrent

READ MORE

Enabling FAIR Data Project - Objectives

- FAIR-aligned data repositories add value to research data, provide metadata and landing pages for discoverability, and support researchers with documentation guidance, citation support, and curation.
- FAIR-aligned Earth, space, and environmental science publishers align their policies to establish a similar experience for researchers. Data, software, technology will be available through citations that resolve to repository landing pages. Availability statements are provided.

Data are not placed in the supplemental information.



Current Publisher Signatories...

(as of 11 June 2019)

- American Geophysical Union
- Copernicus Publications
- Ubiquity Press
- California Digital Library CDL
- Wiley

Over 100 signatories to date!

- PANGAEA, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI), Center for Marine Environmental Sciences, University of Bremen (MARUM)
- WDC Climate, Deutsches Klimarechenzentrum (DKRZ)
- Science
- Science Advances
- PLOS
- Elsevier
- F1000
- Nature
- Scientific Data
- Taylor & Francis Group
- Hindawi Ltd.

https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earthspace-and-environmental-sciences/signatories/

FAIR-Aligned: Researcher Commitment

- Locating trustworthy, community-accepted, FAIR-aligned repositories that support:
 - Documenting data and software (and other research outputs as is possible) to agreed community standards that describe provenance and enable discovery, assessment of reliability, and reuse
 - Persistent identifiers for data and software (and other research outputs as is possible)
 - Licenses for data and software (and other research outputs as is possible) that is as open as possible to enable the widest potential reuse.
- Citing data, software, physical samples, and other research products
- Developing data availability statements
- Preparing and managing data management plans. Make them living documents.

Data Sharing and Citations: New Author Guidelines

Stall, S., et al. (2018), Data sharing and citations: New author guidelines promoting open and FAIR data in the Earth, space, and environmental sciences, *Sci.*

Editor, https://www.csescienceeditor.org/article/data-sharing-and-citations-new-authorguidelines-promoting-open-and-fair-data-inthe-earth-space-and-environmentalsciences/.





Author Guidelines for Data

• Deposit research data in a FAIR-aligned repository.

-A tool to assist: https://repositoryfinder.datacite.org .

- Cite and link to the data using the unique, resolvable, and persistent identifiers.
- Include a Data Availability Statement.
- Provide unrestricted access to all data and materials underlying reported findings for which ethical or legal constraints do not apply.
 - Some data may need restricted access because of fragile environments, endangered species, geopolitical tensions or cultural sensitivities (e.g. indigenous land rights).

Frequently Asked Questions

https://copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/

Table of Contents

- 1. Selecting a Repository
- 2. Data Deposition and Sharing
- 3. Data Availability Statement and Data Citation
- 4. Software Citation and Curation
- 5. Physical Samples
- 6. Enabling FAIR Data Project Questions and additional resources

Editorial Board for incorporating new updates about to be established. All Commitment Signatories and Stakeholders are invited to participate.

Developed by: DataCite/re3data https://repositoryfinder.datacite.org Repository Finder Find a repository to upload your data.

Repository Finder, a pilot project of the Enabling FAIR Data Project led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

Search re3data for a repository to upload your data



or



See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project.



Author, Reviewer, Editor – Open and FAIR – New Data Guidelines – One-Page Reference





- 1. Summary of objectives for open and FAIR data
- 2. Resources for Editors, Reviewers, and Authors
- 3. Information and description of the tools for editors, reviewers, and researchers
- 4. Data Citation Training, Guidelines, and Examples
- 5. Software Citation Guidelines and Examples
- 6. Enabling FAIR Data Project Overview
- 7. FAIR Guiding Principles

http://bit.ly/FAIROnePage



Beyond the Earth, Space, and Environmental Sciences





National Science Foundation WHERE DISCOVERIES BEGIN

🖶 Print 🦰 Share Email

QE

NSF 19-069 Dear Colleague Letter: Effective Practices for Data

May 20, 2019

Dear Colleague:

Open science principles are increasingly being adopted by industry, government, and academia. Open science gives rise to public benefits by offering broader access to publication, data, and other research materials; broader access enables broader circulation of scientific knowledge, greater return on investments in research data, and more opportunities for replicating and building upon scientific findings.

NSF's open science policy is articulated in the Foundation's Public Access Plan (NSF 15-052) and formally implemented in the NSF Proposal and Award Policies and Procedures Guide and in the Award Terms and Conditions that accompany each award that NSF makes. Implications of this policy are further clarified in an actively-maintained set of Frequently Asked Questions (NSF 18-041).

The purpose of this Dear Colleague Letter (DCL) is to describe — and encourage — effective practices for managing research data¹, including the use of persistent identifiers (IDs) for data and machine-readable data management plans (DMPs).

NSF's DMP requirement, as stated in NSF 15-052, expands on NSF's long-standing data-sharing policy. The DMP requirement specifies that every proposal submitted to NSF must include a supplementary document of no more than two pages, titled "Data Management Plan." This document should describe how activities described in the grant proposal will conform to NSF policy on the dissemination and sharing of research results.

https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

National Science Foundation

Dear Colleague Letter: **Effective Practices for Data**

Released 20 May 2019

References the AGU's **Enabling FAIR Data project** as example of guidelines from journals on requiring data citations and data availability statements.

UD SPACE SCIEN

Published 4 June 2019

Shelley Stall, American Geophysical Union Lynn Yarmey, Research Data Alliance Joel Cutcher-Gershenfeld, Brandeis University **Brooks Hanson,** American Geophysical Union Kerstin Lehnert, Lamont–Doherty Earth *Observatory* **Brian Nosek,** Center for Open Science, University of Virginia Mark Parsons Tetherless World Constellation, *Rensselaer Polytechnic Institute* **Erin Robinson**, *Earth Science Information* Partners (ESIP)

Lesley Wyborn, National Computational Infrastructure Australia, Australian National University



Nature News & Comment 🤣

Findable, accessible, interoperable and reusable: The benefits of the FAIR data initiative are enormous.



Make scientific data FAIR

All disciplines should follow the geosciences and demand best practice for publishing and sharing data, argue Shelley Stall and colleagues.

Stall, S, et al. (2019), Make scientific data FAIR, Nature 570, 27-29 (2019) doi: 10.1038/d41586-019-01720-7

Follow

Meeting of Societies – NSF Funded (OAC)

Enabling FAIR Data, Sharing our Method, Encouraging Broad Adoption

- June 19-20, 2019
- AGU Building 2000 Florida Ave, NW, Washington DC
- Target: Natural Sciences and friends

Belmont Forum

Collaborative Research Action (CRA), Science-driven e-Infrastructure Innovation (SEI) for the Enhancement and of Transnational, Interdisciplinary and Transdisciplinary Data Use in Environmental Change

PARSEC - Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas





PARSEC : Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

Consortium Leaders: Nicolas Mouquet, David Mouillot, Alison Specht and Shelley Stall.

Objectives

(a) Determine the influence of natural protected areas (PAs) on the socio-economy of local communities;

(b) Develop metrics for prediction and mitigation of adverse effects of PAs;

(c) Improve linkage between data, publications and researchers;

(d) Improve recommendations for the research data workflow and skills for research teams;

(e) Improve future environmental decision-making;

(f) Increase the number of citations to data sets and better attribute them to the data creator;

(g) Promote the incentive of credit for open and FAIR data management and preservation for data reuse;

(h) Provide tools for researchers to view how the data they have deposited is cited and reused.



FUNDING: 1258K€

DURATION: 48 months

FRANCE: Foundation for Research on Biodiversity - ANR (N. Mouquet) USA: American Geophysical Union - NSF (S. Stall) JAPAN: National Institute of Information & Communications Technology - JST (Y. Murayama) BRAZIL: University of São Paulo - FAPESP (P. Pizzigatti Corrêa) plus postdoc and technical support (FAPESP) Cooperating partners: NCI, Australia (L. Wyborn), BGS, UK (H. Glaves). Associates: DataCite, ORCID, ESIP, RDA, EDI

USA Team Members and International Partners

- Shelley Stall Country Team lead and Lead PI for data-science. —Synthesis-science
 - Jamie Trammell (country team leader): natural resource planning for conservation.
 - Robin Chazdon: biodiversity and human well-being in tropical forested areas.
 - David Kramer: on-ground community engagement in developing countries.
 - To be determined technician/intern

-Data-science

- Shelley Stall, (country team leader and Lead PI for Data-science): data sharing, reuse, management, attribution, and process improvement. FAIR data expert. Scholarly publishing and scientific society.
- Margaret O'Brien: semantics and management of heterogeneous data.

Partners

• DataCite, ORCID, Earth Science Information Partners (ESIP), Research Data Alliance (RDA)



Things to do:

Want information on the author guidelines for data? https://www.csescienceeditor.org/article/data-sharing-and-citations-new-authorguidelines-promoting-open-and-fair-data-in-the-earth-space-and-environmentalsciences/.

Want to be on the email list for updates on Enabling FAIR Data and/or COPDESS? Email <u>sstall@agu.org</u>. [at the moment it is a manual process – keeps costs low]

Want to be a signatory of the Enabling FAIR Data Commitment Statement? <u>https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/</u>.

Want to learn more about data citation, CrossRef schemas, and making sure they are coded correctly? <u>https://scholarlykitchen.sspnet.org/2018/12/12/making-magic-happen-implementing-and-contributing-data-citations-in-support-of-todays-scholarship/</u>.



References

Stall, S, et al. (2019), Make scientific data FAIR, *Nature* **570**, 27-29 (2019) *doi:* <u>10.1038/d41586-019-01720-7</u>

Stall, S., et al. (2018), Data sharing and citations: New author guidelines promoting open and FAIR data in the Earth, space, and environmental sciences, *Sci. Editor*, <u>https://www.csescienceeditor.org/article/data-sharing-and-citations-new-author-</u> guidelines-promoting-open-and-fair-data-in-the-earth-space-and-environmental-sciences/.

Hanson, Brooks, et al. (2018), Making Magic Happen: Implementing and Contributing Data Citations in Support of Today's Scholarship, Scholarly Kitchen, <u>https://scholarlykitchen.sspnet.org/2018/12/12/making-magic-happen-implementing-and-contributing-data-citations-in-support-of-todays-scholarship/</u>. Published 12 December 2018.

Stall, S., et al. (2018), Advancing FAIR data in Earth, space, and environmental science, Eos, 99, <u>https://doi.org/10.1029/2018EO109301</u>. Published on 05 November 2018.





Open and FAIR data are the building blocks to Open Science.

An international effort of broad collaboration is an critical element to making progress towards open and FAIR data.



Questions?

Shelley Stall AGU Senior Director, Data Leadership <u>sstall@agu.org</u> @ShelleyStall https://orcid.org/0000-0003-2926-8353



AGU

