Data Sharing within Scientific Societies

How is the journey so far?

Are we there yet?

Shelley Stall
Vice President, Open Science Leadership
sstall@agu.org ORCID: 0000-0003-2926-8353

Breaking it down…

Membership
- Position statement on data (… includes software)
- AGU Strategic Plan, Open Science focus
- Open Science Journey Checklist and Guidance
- Open Science and Data Help Desk with Earth Science Information Partners (ESIP)

Journal Guidance
- Data and software sharing policy
- Guidance for authors, publishing staff, editors and reviewers

Publication
- Citation quality to support automated attribution and credit

Society Community
- Data Sharing Seminar Series

Four Challenges (and solutions)

AGU
ADVANCING EARTH AND SPACE SCIENCE

AGU's position statement on data affirms that

"Earth and space science data are a world heritage, and an essential part of the science ecosystem. All players in the science ecosystem—researchers, repositories, publishers, funders, institutions, etc.—should work to ensure that relevant scientific evidence is processed, shared, and used ethically, and is available, preserved, documented, and fairly credited."

https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data

AGU's position statement on data affirms that

"Earth and space science data are a **world heritage,** and an essential part of the science ecosystem. All players in the science ecosystem—researchers, repositories, publishers, funders, institutions, etc.—should work to **ensure that relevant scientific evidence is processed, shared, and used ethically, and is available, preserved, documented, and fairly credited.**"

https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data

Your Open Science Journey

You, the researcher:

how you and your work are discovered, made visible and gets recognition

Your Research Team/Lab:

how you can work openly

Your Community:

improve interoperability, sharing, and reuse beyond your team

Beyond Your Community:

preparing for cross-domain challenges

Take the first steps on your journey to Open Science

Useful checklists for you, the researcher

Your Digital Presence

Data Documentation and Citation Checklist

Software Documentation and Citation Checklist

Guidelines for your research team

Open Science Practices for Teams

Resources and Guides for Teams

Digital Objects Preservation Checklist for Teams

PARSEC

AGU
ADVANCING EARTH AND SPACE SCIENCE

Open Science & Data Help Desk

Exhibit Hall Booth #1901

Help Desk / Platform and Tool Demos / Workshops / Presentations / Editor's Nook

Visit us!

Monday 15:00-18:00
Tuesday 10:00-18:00
Wednesday 10:00-18:00
Thursday 10:00-13:00

Data Help Desk

ESIP

AGU ADVANCING EARTH AND SPACE SCIENCE

Scan for Help Desk schedule

Open Science & Data Help Desk

Presentation Schedule

The Open Science and Data Help Desk is an in-person resource with experts across the Earth, space, and environmental disciplines to provide information and answer your questions on data, software, and Open Science topics. **Visit us at Exhibit Hall booth #1901.**

Monday 12 December

15:30 **Open Means Understandable: Writing An Effective Plain Language Summary About Your Article** *Jenny Lunn*
16:00 **The Ethos Lifecycle: Operationalizing ethics in data science** *Micaela Parker*
16:30 **Navigating Open Data, Open Access, and Funding Options** *Mia Ricci*

Tuesday 13 December

10:00 **Data Citation Primer — A Must-have if You are Publishing Soon!!** *Shelley Stall*
10:30 **How and why to cite data and software (in AGU journals)** *Mark Parsons and James Gallagher*
11:30 **The Ethos Lifecycle: Operationalizing ethics in data science** *Micaela Parker*
13:00 **Data Management and my Career** *Pranoti Asher*
14:00 **At the intersection of open data and community science** *Julia Parrish*
14:30 **Getting started with reproducible documents in R** *Ben Bond-Lamberty*
15:00 **Effective, efficient, and fair peer reviewing** *Ben Bond-Lamberty*
15:30 **AGU Pubs - Diversity, Equity, Inclusion, and Accessibility** *Mia Ricci*

Wednesday 14 December

10:00 **Software Citation Primer — Get Prepared for Publication!!** *Shelley Stall*
10:30 **How and why to cite data and software (in AGU journals)** *Mark Parsons and James Gallagher*
12:00 **Physical Samples: Identifiers, Citation, Metadata and more with SESAR and the IGSN** *Kerstin Lehnert and Saebyul Choe*
13:00 **Diversity, Equity, and Inclusion** *Billy Williams*
14:30 **Open science meta-analysis: easier than you think** *Kendal Morris*
16:00 **Working Openly as a Team** *Shelley Stall*

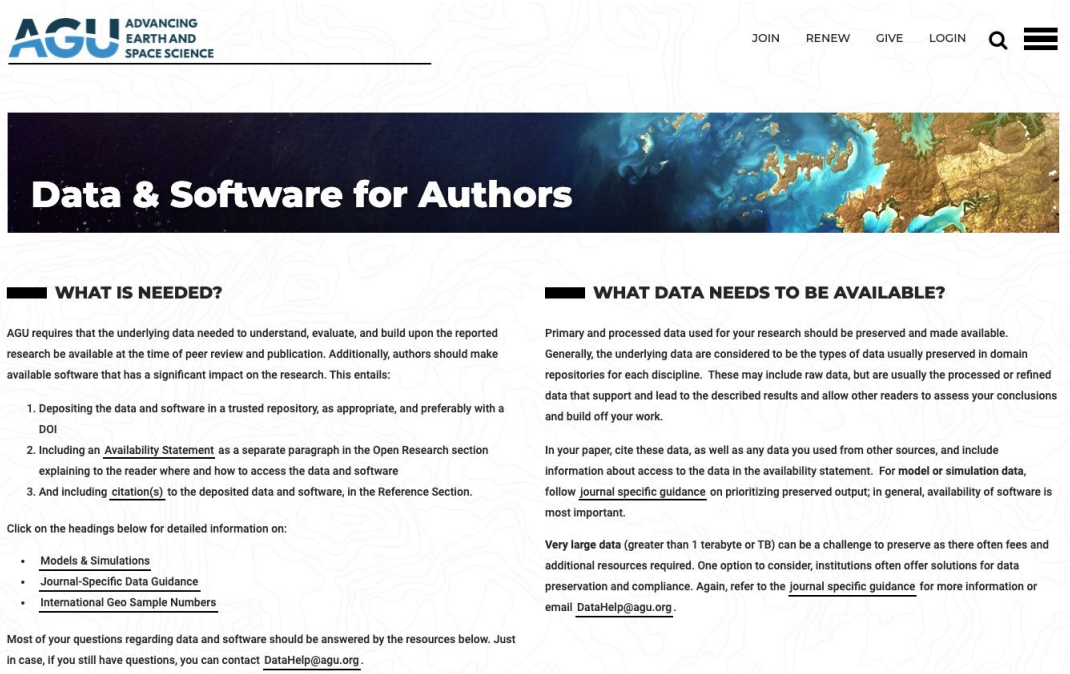
Thursday 15 December

10:00 **Digital Presence** *Shelley Stall*
11:00 **Open Means Understandable: Writing An Effective Plain Language Summary About Your Article** *Jenny Lunn*

AGU Data & Software Sharing Guidance (v. 2.0)

What is covered:

- What data needs to be available?
- Repository Selection
- Availability Statement
- Data & Software Citation
- Citation Formatter
- Models & Simulations
- Expanded Guidance
- International General Sample Numbers



AGU ADVANCING EARTH AND SPACE SCIENCE

JOIN RENEW GIVE LOGIN

Data & Software for Authors

WHAT IS NEEDED?

AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

1. Depositing the data and software in a trusted repository, as appropriate, and preferably with a DOI
2. Including an Availability Statement as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software
3. And including citation(s) to the deposited data and software, in the Reference Section.

Click on the headings below for detailed information on:

- Models & Simulations
- Journal-Specific Data Guidance
- International Geo Sample Numbers

Most of your questions regarding data and software should be answered by the resources below. Just in case, if you still have questions, you can contact DataHelp@agu.org.

WHAT DATA NEEDS TO BE AVAILABLE?

Primary and processed data used for your research should be preserved and made available. Generally, the underlying data are considered to be the types of data usually preserved in domain repositories for each discipline. These may include raw data, but are usually the processed or refined data that support and lead to the described results and allow other readers to assess your conclusions and build off your work.

In your paper, cite these data, as well as any data you used from other sources, and include information about access to the data in the availability statement. For **model or simulation data**, follow journal specific guidance on prioritizing preserved output; in general, availability of software is most important.

Very large data (greater than 1 terabyte or TB) can be a challenge to preserve as there often fees and additional resources required. One option to consider, institutions often offer solutions for data preservation and compliance. Again, refer to the journal specific guidance for more information or email DataHelp@agu.org.

We continue to iterate and improve the guidance based on feedback from editors, reviewers, authors, and staff.

The next improvement will aim at adding frequently asked questions (FAQs) and more examples.

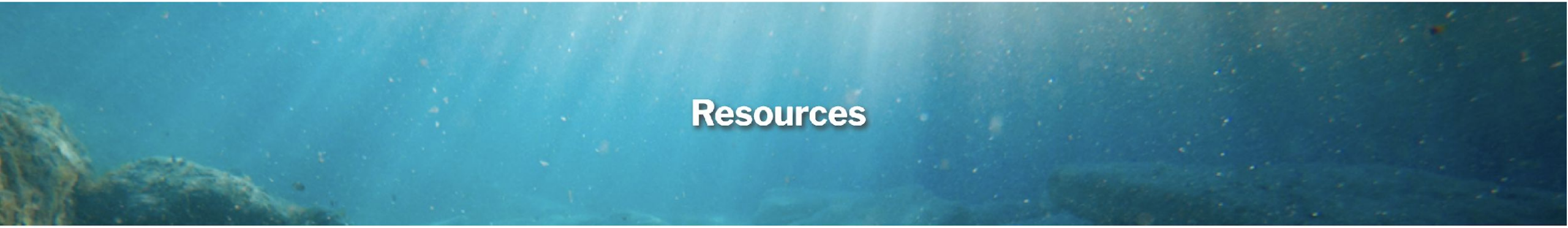
https://doi.org/10.5281/zenodo.5121488

AGU ADVANCING EARTH AND SPACE SCIENCE

AGU Data Leadership

https://data.agu.org

Home/Blog About Data FAIR Resources



Resources

Manage your Digital Objects – Research Team Member Checklist

Digital Presence - Connect your research for better discovery

Guidance for AGU Authors - Jupyter Notebooks

Software Citation - 5 Tips

CoreTrustSeal Cohort - AGU Enabling FAIR Data Project

Data and Software Sharing Guidance for Authors Submitting to AGU Journals

Data and Software Availability and Citation Checklist

Domain-Discipline Repositories Useful to AGU Journals

AGU's Paleoceanography and Paleoclimatology - Data Submission Quick Guide

Preserving very large data is a challenge. Spoilers, there are no easy answers!

OCTOBER 01, 2021

When it comes to large datasets, we are often asked by authors and editors how they should preserve the data. These questions come via datahelp@agu.org and our data and software guidance discussions. Spoilers, there are no easy answers, yet! Here we offer our experience, share the current limitations, and the approaches we recommend with what is possible right now.

AGU requires that primary and processed data used for your research should be preserved and made available. This can range from observational data to the data used to generate your figures. The raw data may be needed, but usually, the processed or refined data that support and lead to the described results and allow other readers to assess your conclusions and build off your work should be preserved.

For data that is large, over 1 Terabyte (TB), authors run into the challenge of finding a suitable repository. Many repositories have file size limitations but also costs associated with deposits over certain limits. This generalist repository comparison chart provides an overview of the limitations. Discipline-specific and institutional repositories are often a place to turn to for assistance with preserving large data but they also have limitations and potential costs. This emphasizes the importance of avoiding surprises at

Your Digital Presence - Taking advantage of the auto-update feature for your ORCID

SEPTEMBER 21, 2021

Your Digital Presence is how other researchers and the international scholarly community discover your research online with commonly used open search tools. It is also an important element of Open Science. By making discovery easy for others you help to:

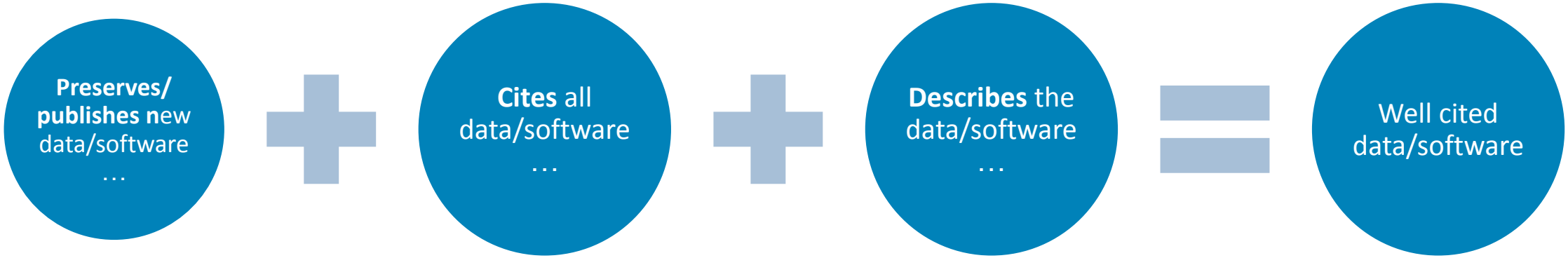
- Increase **Citations** of your work
- Build your network to explore possible future **Collaborators**
- Improve your **Open Science** practices by linking to your research

What can you do?

1. Learn more about Digital Presence using this checklist and helpful 15-min tutorial on our resources page. Info includes: registering for an ORCID and using your ORCID to link your scholarly research objects.
2. Turn on Auto Updates for your ORCID. This significantly reduces work to keep your ORCID up-to-date using trusted scholarly

The Formula for Data and Software Citations

The author…



…in a community-accepted trusted repository

…used for the research in the Reference Section; use bracketed descriptions

…in the Open Research Section – even when a citation is not possible

AGU
ADVANCING EARTH AND SPACE SCIENCE

Journal Production Guidance for Software and Data Citations

It is important to note the guidance provided by Crossref in their blog (Lin, J., 2017) **has since been amended.** If a journal implemented this guidance previously, there is a high probability that a correction is needed.

In short, if there is a DOI for a data citation, include the DOI XML tag in the file to Crossref. This means you need to know it is a dataset citation, and maintain the machine-readable integrity all the way through production and content hosting.

Reference This Paper for Details

Shelley Stall, Geoffrey Bilder, Matthew Cannon, et al. Journal Production Guidance for Software and Data Citations. *ESS Open Archive* . December 31, 2022.
DOI: 10.22541/essoar.167252601.17695321/v1

AGU
ADVANCING EARTH AND SPACE SCIENCE

Data Sharing Seminar Series for Societies

https://wesharedata.org

WELCOME

Societies have a unique role in bringing awareness of developing practices and supporting the necessary discussions within disciplines to bring their voice to the larger community.

These seminars are monthly through January 2022 held on the first Friday of each month at 10am ET (14:00 UTC). There will be 2-3 speakers with ~30 minutes of Q&A and discussion specific to society engagement to help with data sharing, credit, transparency and more. Recordings will be made available here following the session.

This series is supported and guided by these collaborating societies and federations:

Science JOURNALS AAAS

AAS

ACS Chemistry for Life®

AGU ADVANCING EARTH AND SPACE SCIENCE

AMS American Meteorological Society Board on Data Stewardship

ECOLOGICAL SOCIETY OF AMERICA esa

BRITISH ECOLOGICAL SOCIETY

CSSP council of scientific society presidents

FASEB Federation of American Societies for Experimental Biology

FABBS Federation of Associations in Behavioral & Brain Sciences

IUPAC INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

Next Webinar

Data Sharing for Societies: Pinnacle Seminar

20 April 2022, 10am ET (1400 UTC)

REGISTER

All slides, recording, and resources mentioned during the presentation: https://zenodo.org/communities/data-sharing-seminar-series-for-societies/

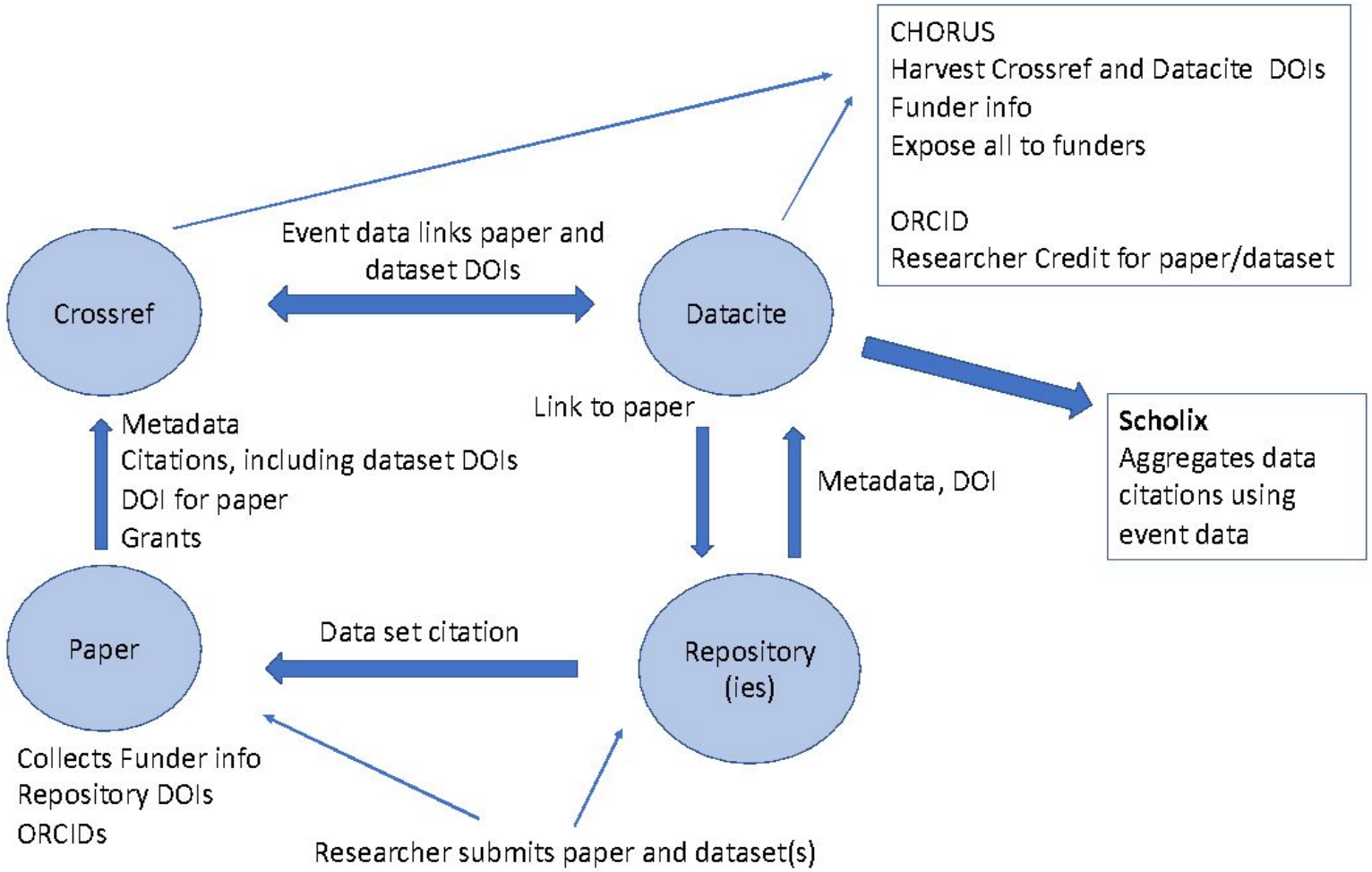
Four challenges (and solutions) for realizing the future of open science for the benefit of humanity

Requires engagement by all key stakeholders

1. Means and solutions are ready, we need the will and some actions
2. All require and involve culture change and collective action
3. We are excited by how close we are and energy around these

1. Work together to finally empower data and software citations

Publishers, Repositories, Universities, Funders



All that's needed:

- Repositories: Support PID's and reference guidance
- Publishers: Implement Markup
- Crossref et al. Enable Scholix fully
- **Publishers and Repositories: work *together* on guidance and best practices**
- Don't let perfect be a barrier

AGU
ADVANCING EARTH AND SPACE SCIENCE

2. Elevate importance of data, software, etc. in grants

Funders and all stakeholders

Major funders (e.g., NSF, NASA, NIH, ERC, UKRI) focus grants and grant review on intellectual merit.

- They in effect diminish the value of data, software and other outputs needed today for scientific and societal impact relative to the ideas that depend on them.
- DMP's are the last section of proposals and data *et al.* are not central.

An Alternative Approach:

- **Outcomes Impact Statement:** Explain how the outcomes of this research—including new data, software, materials, samples, and products—will have scientific and societal value and impact. Indicate how these outcomes will be shared using leading practices, referencing the DMP as appropriate.

AGU
ADVANCING EARTH AND SPACE SCIENCE

3. Work together to support the research repository infrastructure sensibly

All Stakeholders, Universities, Repositories, Funders

- Big Danger: General and Institutional repositories become a new supplemental material quagmire (huge non-standard dumping ground for files w/o metadata).
- Institutions have need, capacity, and resources to support future data infrastructure.
- Domain repositories have expertise; leading practices; need support
- Several examples of at least partial success already (IRIS, UNAVCO, CDL-Dryad partnership, CERN)

AGU
ADVANCING EARTH AND SPACE SCIENCE

4. Reward and Award Open Science, Open Data, and Sharing Science

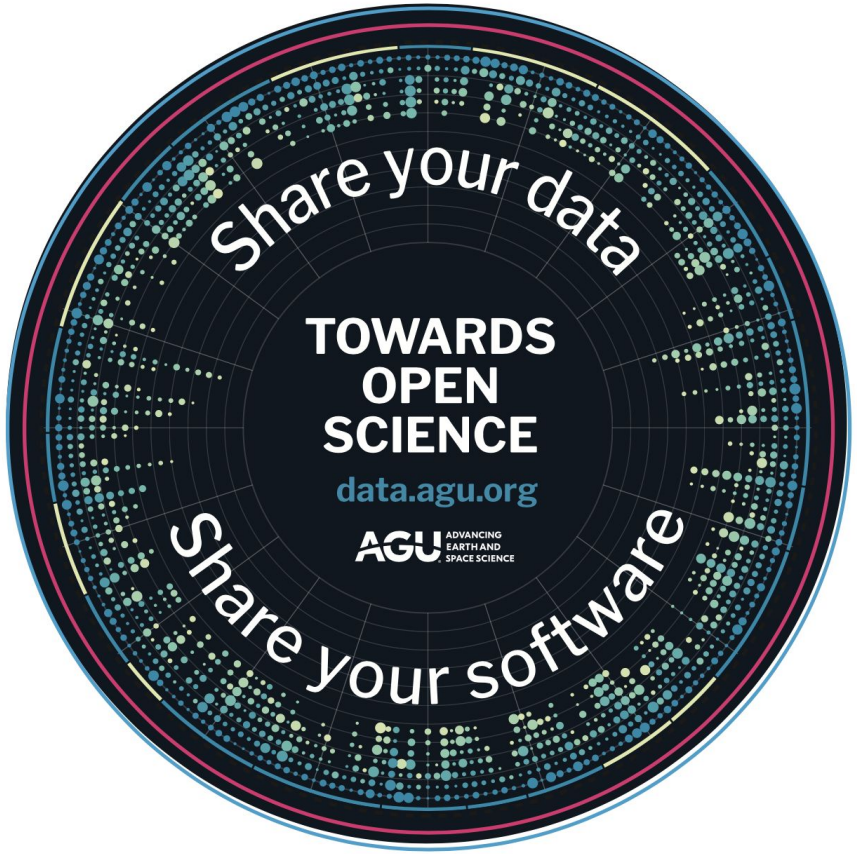
All Stakeholders

- Societies and Academies: Include open science, open data, JEDI, and outreach in awards.
- Universities: Adjust hiring, tenure and promotion guidance and practice
- Researchers: Include these values in nomination letters and reviews
- All: Reinforce each other in these practices; training

AGU
ADVANCING EARTH AND SPACE SCIENCE

Thank you

Shelley Stall
AGU Vice President, Open Science Leadership
sstall@agu.org
@ShelleyStall
https://orcid.org/0000-0003-2926-8353



This work is generously funded through grants from:

PARSEC

BELMONT FORUM

NSF

FAPESP

FUNDAÇÃO DE AMPARO À PESQUISA

AGENCE NATIONALE DE LA RECHERCHE ANR

JST