Accelerating Open and FAIR Data Practices Across the Earth, Space, and Environmental Sciences

Chris Erdmann
Assistant Director, Data Leadership
American Geophysical Union
0000-0003-2554-180X
@libcce | cerdmann@agu.org

AGU
ADVANCING EARTH AND SPACE SCIENCE

CC BY

AGU ADVANCING EARTH AND SPACE SCIENCE

DATA LEADERSHIP

CONTACTS

SHELLEY STALL

Senior Director, Data Leadership | ORCID: 0000-0003-2926-8353

202.777.7307 | SSTALL@AGU.ORG

CHRISTOPHER ERDMANN

Assistant Director, Data Stewardship | ORCID: 0000-0003-2554-180X

202.777.7351 | CERDMANN@AGU.ORG

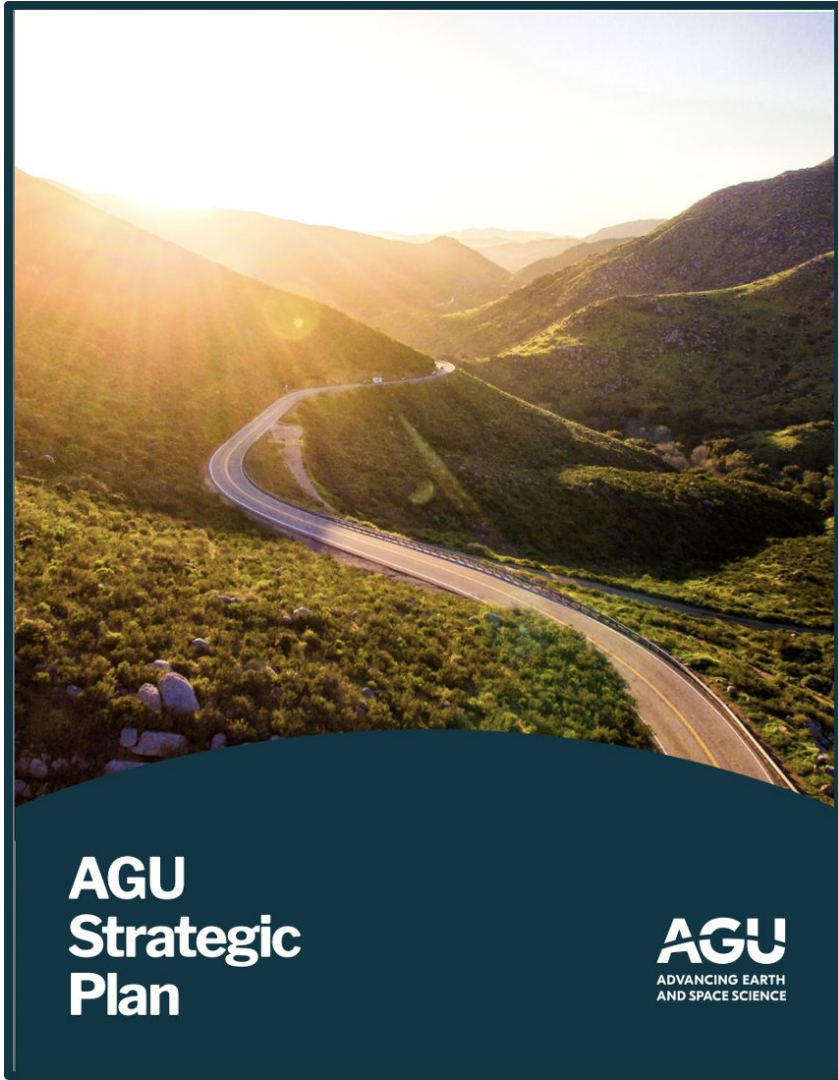
AGU’s position statement on data affirms that

“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**.”

AGU Strategic Plan: Promoting Open Science and Open Data

“Lead in open science and open data while maintaining high standards of quality and scientific integrity.”

https://www.agu.org/Learn-About-AGU/About-AGU/About/Strategic-Plan



Researchers, Journals, Data Repositories

NSF Grant 2025364

Accelerating Open and FAIR Data Practices Across the Earth, Space, and Environmental Sciences: A Pilot with the NSF to Support Public Access to Research Data (AGU NSF PAR 2.0)

Partners: Dryad, CHORUS, ESIP, Wiley (In-Kind)

2-year project aimed at implementing FAIR data practices across the Earth, space, and environmental sciences such that, by the end of the project:

- Data citations for data funded by NSF grants are captured in the NSF Public Access Repository (PAR 2.0)
- Knowledge of leading practices and workflows around data citation are well known across the AGU community.

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.

JOURNAL-SPECIFIC DATA GUIDANCE

Please note that only those journals marked with an asterisk (*) are available at this time.

AGU Advances - Open Access
Earth and Space Science - Open Access
Earth's Future - Open Access
***Geochemistry, Geophysics, Geosystems** **
GeoHealth - Open Access*
Geophysical Research Letters
Global Biogeochemical Cycles
Journal of Advances in Modeling Earth Systems (JAMES) - Open Access
Paleoceanography and Paleoclimatology
Radio Science
Reviews of Geophysics **[Invitation only]***
Space Weather - Open Access
***Tectonics** **
***Water Resources Research** **

***Journal of Geophysical Research: Atmospheres** **
***Journal of Geophysical Research: Biogeosciences** **
Journal of Geophysical Research: Earth Surface
Journal of Geophysical Research: Oceans
***Journal of Geophysical Research: Planets** **
***Journal of Geophysical Research: Solid Earth** **
Journal of Geophysical Research: Space Physics

Implementation

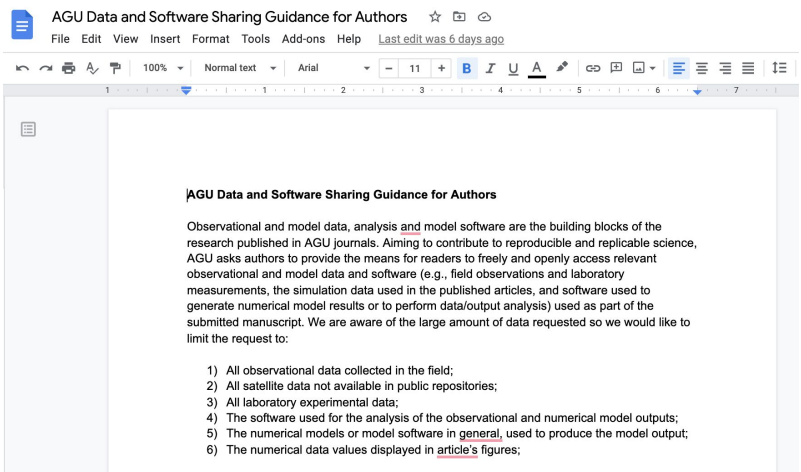
Authors, reviewers, editors, and journal staff sometimes face **gray questions on how to implement policy.**

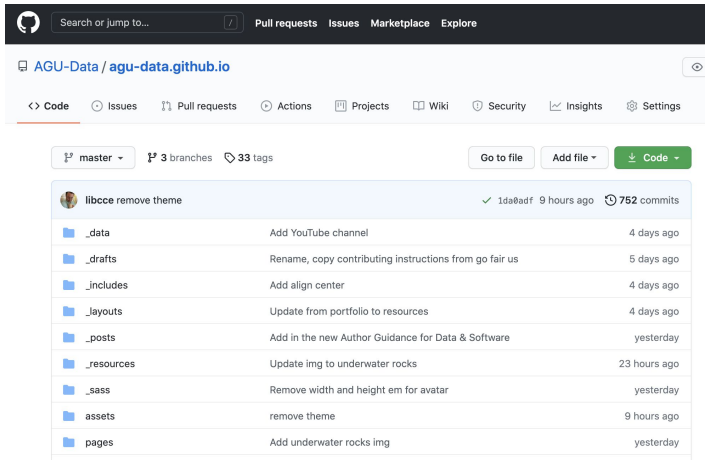
The **questions can vary** from what model information to include to navigating permissions and properly acknowledging use.

Over the span of 4 months we have responded to roughly **200 requests** that land in this gray space.

Data and software sharing guidance in process

Shift from document-based to transparent, collaborative, versioned approach.





So what are some of the challenges and opportunities we are responding to in this process?

How to select a repository?

ENABLING FAIR DATA PROJECT

OpenAIRE

Guides for Researchers

How to select a data repository?

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

FAIRSFAIR
Fostering Fair Data Practices in Europe

… and more.

Generalist Repository Comparison Chart

doi: 10.5281/zenodo.3946720

This chart is designed to assist researchers in finding a generalist repository should no domain repository be available to preserve their research data. Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. For this chart, we included a repository available to all researchers specific to clinical trials (Vivli) to bring awareness to those in this field.

https://fairsharing.org/collection/GeneralRepositoryComparison

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Brief Description	Harvard Dataverse is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data.	Open-source, community-led data curation, publishing, and preservation platform for CCO publicly available research data Dryad is an independent non-profit that works directly with: • researchers to publish datasets utlizing best practices for discovery and reuse • publishers to support the integration of data availability statements and data citations into their workflows • institutions to enable scalable campus support for research data managment best practices at low cost	A free, open access, data repository where users can make all outputs of their research available in a discoverable, reusable, and citable manner. Users can upload files of any type and are able to share diverse research products including datasets, code, multimedia files, workflows, posters, presentations, and more. With discoverable metadata supporting FAIR principles, file visualizations, and integrations, researchers can make their work more impactful and move research further faster.	Mendeley Data is a free repository specialized for research data. Search more than 20+ million datasets indexed from 1000s of data repositories and collect and share datasets with the research community following the FAIR data principles.	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices.	Vivli is an independent, non-profit organization that has developed a global data-sharing and analytics platform. Our focus is on sharing individual participant-level data from completed clinical trials to serve the international research community.	Powering Open Science, built on Open Source. Built by reserachers for researchers. Run from the CERN data centre, whose purpose is long term preservation for the High Energy Physics discipline, one of the largest scientific datasets in the world
Size limits	No byte size limit per dataset. Harvard Dataverse currently sets a file size limit of 2.5GB.	300GB/dataset	Soft limit of 20GB/file for free accounts. System limit of 5000GB/file. Unlimited storage of public data but 20GB storage for private data for free accounts. Email info@figshare.com to have upload and storage limits raised.	10GB per dataset	Projects currently have not storage limit. There is a 5GB/file upload limit for native OSF Storage. There is no limit imposed by OSF for the amount of storage used across add-ons connected to a given project.	If more than 10GB per study data, reach out to us	50GB per dataset, contact us via https://zenodo.org/support for higher limits
Storage space per researcher	1 TB per researcher	No limit	No limit	No limit	No limit	No limit	No limit
Persistent, Unique Identifier Support	DOI, Handle	DOI	DOI	DOI	DOI	DOI	DOI

Stall, Shelley, Martone, Maryann E., Chandramouliswaran, Ishwar, Crosas, Mercè, Federer, Lisa, Gautier, Julian, … Zigoni, Alberto. (2020, July 15). Generalist Repository Comparison Chart. Zenodo. http://doi.org/10.5281/zenodo.3946720

Never underestimate the power of teamwork...

To support data and software citation in publications, but also throughout the scholarly infrastructure, we've worked with a number of stakeholders including...

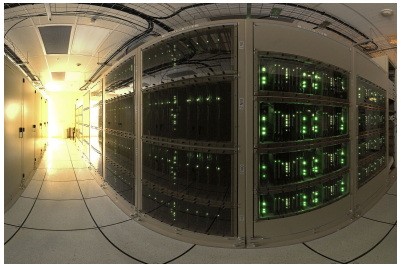
… and more.

Availability Statements

AVAILABILITY STATEMENT

An Availability Statement, located in the Open Research section of the paper, contains information about your data, software, and other research objects (e.g. notebook) and how readers can access these. The Statement should include:

1. A brief description of the type(s) of data or software
2. Repository Name(s) where they are deposited
3. Version (of software)
4. DOI, Persistent Identifier Link to Data or Software (and Identifier)
5. Link to publicly accessible development platform (in the case of Software, e.g. GitHub)
6. Access Conditions (e.g. if Registration is Required)
7. Licensing/Permissions (e.g. Creative Commons Attribution)
8. In-text citation in References (optional)



Examples: Creative Commons, HPCs

Models & Simulations

Mullendore, Gretchen, Schuster, Doug, Mayernik, Matthew, & Griffies, Stephen M. (2021, May). COPDESS Workshop: Rubric for Models and Model Data — Best Practices for Preservation and Replicability. Zenodo. http://doi.org/10.5281/zenodo.4890691

Jun 8, 2021

2nd Workshop of the Data Citation Community of Practice

A presentation of use cases identified in the 1st workshop followed by community discussion

Apr 8, 2021

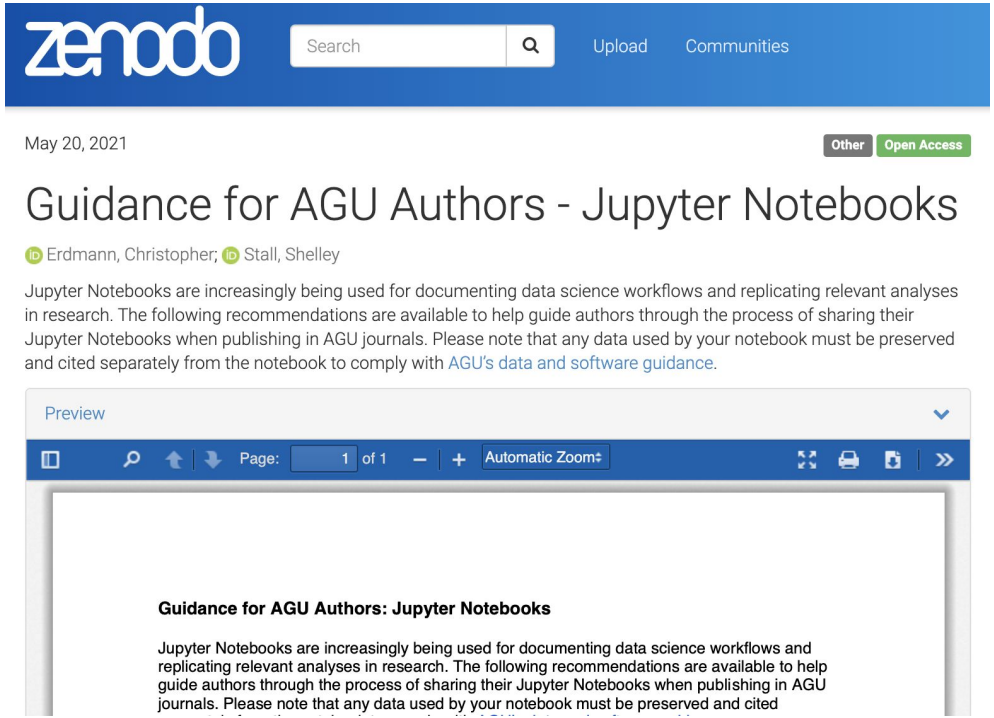
Data Citation Community of Practice Workshop for Earth, Space, and Environmental Sciences Announcement

A discussion on citing large numbers of datasets and assigning credit properly

Goal: In this workshop we want to address the use case of citing a large number of datasets such that credit for individual datasets is assigned properly. Discussion will include the concept of a **“Data Collection”** and the infrastructure and guidance still needed to fully implement so it is easier for researchers to use and receive credit when their data are cited in this manner.

Erdmann, Christopher, & Stall, Shelley. (2021, April). 5 Tips to Citing Your Research Software and Improving Discovery. Zenodo. http://doi.org/10.5281/zenodo.4706180

Guidance for Authors - Jupyter Notebooks



Erdmann, Christopher, Stall, Shelley, Gentemann, Chelle, Holdgraf, Chris, Fernandes, Filipe P. A., & Gehlen, Karsten Peters-von. (2021, May 20). Guidance for AGU Authors - Jupyter Notebooks. Zenodo. http://doi.org/10.5281/zenodo.4910038

Announcing the Launch of the AGU Data Leadership Program Website

JUNE 14, 2021

Today, June 14, 2021, we are announcing the launch of the AGU Data Leadership Program website. The website provides the Earth, space, and environmental science community with information on the latest work within AGU's Data Leadership Program as well as highlights and links from other community work around best practices that

Digital Presence - Connect your research for better discovery

JUNE 14, 2021

Connecting your research to your data, software and institution is a proven way to increase discovery of your work to potential collaborators and partners, and receive credit when others use your work.

Using persistent identifiers such as ORCIDs (for researchers), Digital Object Identifiers, and RORs (for organizations) are used to make those connections. Through the checklist and tutorial we very quickly take you through how to optimize those connections.

Get a copy of the *Digital Presence Checklist* from Zenodo: http://doi.org/10.5281/zenodo.4706119

Tutorial, 15 min (YouTube): https://youtu.be/bftzPnFdtHk

Slides for reference (Zenodo): http://doi.org/10.5281/zenodo.4706147

Thank you.

Chris Erdmann
Assistant Director, Data Leadership
American Geophysical Union
0000-0003-2554-180X
@libcce | cerdmann@agu.org

AGU
ADVANCING EARTH AND SPACE SCIENCE

CC BY

Links

1. https://www.agu.org/Learn-About-AGU/About-AGU/Data-Leadership
2. https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data
3. https://www.agu.org/Learn-About-AGU/About-AGU/About/Strategic-Plan
4. https://fromtheprow.agu.org/nsf-grant-takes-agus-data-and-publishing-practices-to-the-next-level/
5. https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors
6. https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors#guidance
7. http://www.copdess.org/enabling-fair-data-project/
8. https://www.openaire.eu/opendatapilot-repository-guide
9. https://www.re3data.org/
10. https://www.fairsfair.eu/
11. https://agu-data.github.io/DataCitationCoP/
12. https://data.agu.org/
13. https://data.agu.org/resources/digital-presence

Images

1. https://unsplash.com/photos/VUrHRFXEZ70
2. https://commons.wikimedia.org/wiki/File:Wide-angle_view_of_the_ALMA_correlator.jpg