EGU GA 2022 - ESSI3.1

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

27 May 2022

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

Shelley Stall
Vice President, Data Leadership
American Geophysical Union
0000-0003-2926-8353
@ShelleyStall | sstall@agu.org

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)

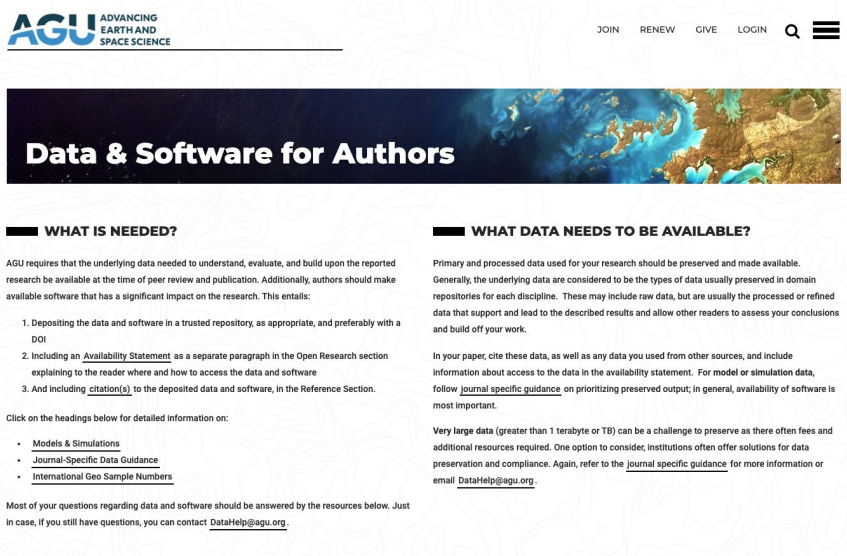
2-year project aimed at implementing FAIR data practices across the Earth, space, and environmental sciences such that, by the end of the project:

- 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.

AGU Data & Software Sharing Guidance

What is covered:

- What data needs to be available?
- Repository Selection
- Availability Statement
- Data & Software Citation
- Citation Formatter
- Models & Simulations
- Journal Specific Guidance
- International Geo Sample Numbers
- Data Help Desk



https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors

Availability Statement/Citation Paper Example

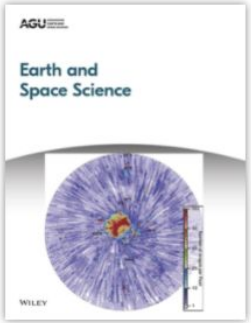
Earth and Space Science

Research Article | Open Access

Performance Assessment of Geophysical Instrumentation Through the Automated Analysis of Power Spectral Density Estimates

M. R. Koymans, J. Domingo Ballesta, E. Ruigrok, R. Sleeman, L. Trani, L. G. Evers

First published: 22 July 2021 | https://doi.org/10.1029/2021EA001675



Volume 8, Issue 9
September 2021
e2021EA001675

https://doi.org/10.1029/2021EA001675

Availability Statement/Citation Paper Example (Cont.)

Example: Koymans, M. R., Domingo Ballesta, J., Ruigrok, E., Sleeman, R., Trani, L., & Evers, L. G. (2021). Performance Assessment of Geophysical Instrumentation Through the Automated Analysis of Power Spectral Density Estimates. In Earth and Space Science (Vol. 8, Issue 9). American Geophysical Union (AGU). https://doi.org/10.1029/2021ea001675

From Open Research (Availability Statement) Section: Maps were created through PyGMT (prerelease) (Uieda et al., 2021) using Generic Mapping Tools (GMT) version 6 (Wessel et al., 2019a, 2019b) licensed under LGPL version 3 or later, available at https://www.genericmapping-tools.org/.

References:

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019a). The generic mapping tools version 6 [Software]. Zenodo. (Funded by US National Science Foundation grants OCE-1558403 and EAR-1829371). https://doi.org/10.5281/zenodo.3407866

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019b). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, 20(11), 5556– 5564. https://doi.org/10.1029/2019gc008515

AGU
ADVANCING EARTH AND SPACE SCIENCE

Help Desk Challenges

- Government Sites, Similar - Technical, Permissions
- Firewalls, Authentication - Openness, Availability, Anonymity
- Supplemental Information - Tradition, Peer Review
- FTP, Directories, Storage - Institutional, Compliant Solution
- Curation, Deposit Workflows - Service, Publication Workflows
- **Web Sharing, Dev Platforms - Citation Information**
- Databases / Dynamic Services - Direct Access, Linking
- Available Upon Request - Culture
- Citation Nothingness (Paper not the Data) - Culture
- Website Home (Parachuting) - Laziness
- English Language - Language Diversity, Translation
- Many Data Links/Citations - Tables, Supplements (See Data Citation Community of Practice)
- ...

What repository?

AGU Data Leadership

Home/Blog About Data FAIR Resources

Domain-Discipline Repositories Useful to AGU Journals

The data that supports the research reported in your paper must be deposited in a community-accepted, trusted preservation repository. Additionally, authors should make available software that has a significant impact on the research. A repository that specializes in dommain-discipline specific data and software will maximize the probability that the deposited data and software will be findable, accessible, interoperable and reusable (FAIR). Repsoitories that use persistent identifier links (e.g. DOI or digital object identifier over URLs (and not to the home page) are recommended. Note, an English language translation is necessary in order for the data/software to be accessible to the wider community. Domain-discipline repositories useful to AGU journals below may also be at different stages in supporting the FAIR principles. For any additional domain-discipline repositories recommendations, contact datahelp@agu.org or submit a GitHub issue/pull request. Otherwise, look to your institutional repository, your computing center, a general repository (e.g., Zenodo, Dryad, figshare), or search for a repository using re3data, OpenAire, or DataOne. Consult Data and Software for Authors and Data and Software Sharing Guidance for Authors Submitting to AGU Journals for more in-depth guidance.

The following is a list developed with AGU community members of useful repositories by journal:

- Multiple Journals
- GeoHealth
- JAMES

https://data.agu.org/resources/useful-domain-repositories

Preserving Large Data!

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 the time of publication by:

https://data.agu.org/2021/10/01/challenges-preserving-very-large-data.html

AGU Journals - Data/Software Citations (2021)

JOURNAL	PAPER AVAILABILITIES INTEXT CITATION COUNT	TOTAL PAPERS	PERCENTAGE
JOURNAL OF GEOPHYSICAL RESEARCH: PLANETS	209	274	76.28%
JOURNAL OF GEOPHYSICAL RESEARCH: SOLID EARTH	341	809	42.15%
GEOCHEMISTRY, GEOPHYSICS, GEOSYSTEMS	132	332	39.76%
TECTONICS	77	197	39.09%
GLOBAL BIOGEOCHEMICAL CYCLES	50	140	35.71%
JOURNAL OF GEOPHYSICAL RESEARCH: EARTH SURFACE	62	191	32.46%
REVIEWS OF GEOPHYSICS	7	23	30.43%
JOURNAL OF GEOPHYSICAL RESEARCH: BIOGEOSCIENCES	96	323	29.72%
WATER RESOURCES RESEARCH	201	716	28.07%
GEOHEALTH	26	93	27.96%
JOURNAL OF ADVANCES IN MODELING EARTH SYSTEMS	60	215	27.91%
PALEOCEANOGRAPHY AND PALEOCLIMATOLOGY	132	483	27.33%
EARTH'S FUTURE	49	181	27.07%
GEOPHYSICAL RESEARCH LETTERS	494	1856	26.62%
EARTH AND SPACE SCIENCE	86	325	26.46%
SPACE WEATHER	74	340	21.76%
JOURNAL OF GEOPHYSICAL RESEARCH: OCEANS	112	529	21.17%
AGU ADVANCES	16	76	21.05%
JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES	146	843	17.32%
JOURNAL OF GEOPHYSICAL RESEARCH: SPACE PHYSICS	135	788	17.13%
RADIO SCIENCE	15	108	13.89%
PERSPECTIVES OF EARTH AND SPACE SCIENTISTS	0	9	0.00%
Year 2021 counts	**2520**	**8851**	**28.47%**

AGU
ADVANCING EARTH AND SPACE SCIENCE

Soon to be submitted for publication…

Journal Production Guidance for Software and Data Citations - Draft

xx May 2022 - Draft Review

Journal Production Guidance for Software and Data Citations

Authors: Shelley Stall, …

Abstract

Software and data citation are emerging best practices in scientific communication. This article provides structured guidance to the academic publishing community on how to implement software and data citation in publishing workflows. These best practices support the verifiability and reproducibility of scientific results, sharing and reuse of valuable data, software tools, and resources, and credit to the originators of the data and software. They provide a basis for making both data and software FAIR (Findable, Accessible, Interoperable, and Reusable). Data citation is becoming increasingly well-established. With the current intensive use of software, including specialized tools and models for scientific research problems, the research community has begun to recognize that software, as a key research resource, requires the same level of transparency, accessibility, and disclosure as data. Software and data that support scientific results should be preserved and shared in scientific repositories for discovery, transparency, and used by other researchers. This can be achieved by citing these products in the references section of papers and effectively associating them.

AGU
ADVANCING EARTH AND SPACE SCIENCE

Thank you.

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

Shelley Stall
Vice President, Data Leadership
American Geophysical Union
0000-0003-2926-8353
@ShelleyStall | sstall@agu.org