

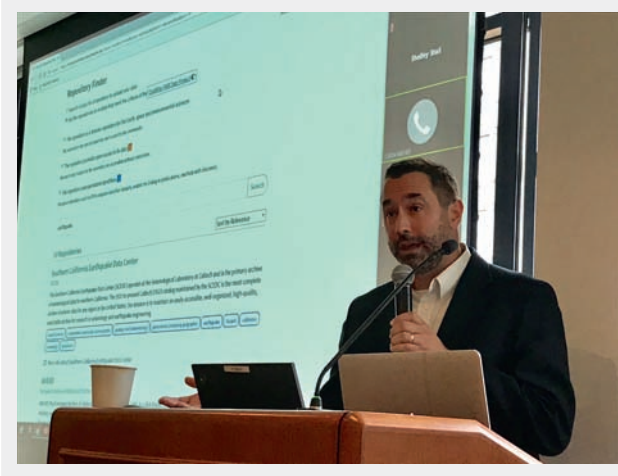
Data Sharing and Citations: New Author Guidelines Promoting Open and FAIR Data in the Earth, Space, and Environmental Sciences

Shelley Stall, Patricia Cruse, Helena Cousijn, Joel Cutcher-Gershenfeld, Anita de Waard, Brooks Hanson, Joerg Heber, Kerstin Lehnert, Mark Parsons, Erin Robinson, Michael Witt, Lesley Wyborn, and Lynn Yarmey

New author guidelines supporting open and FAIR data in scholarly publishing are being adopted throughout the Earth, space, and environmental sciences community. With the new guidelines, supporting resources are provided. These include a new tool for finding the right repository and answers to frequently asked questions. Adoption of these new guidelines requires a shift in the scientific culture around data sharing. Support for this change is needed by researchers, institutions, funders, journals, repositories, and connecting infrastructure—which will advance research across the geosciences.

Scholarly publishing today is, in many ways, all about the data. Publications increasingly describe large, complex, and diverse data sets. Preserving, making available, and ensuring the integrity of the underlying data are as important as

SHELLEY STALL and BROOKS HANSON are with the American Geophysical Union; PATRICIA CRUSE and HELENA COUSIJN are with DataCite; JOEL CUTCHER-GERSHENFELD is with the Heller School for Social Policy and Management, Brandeis University; ANITA DE WAARD is with Elsevier; JOERG HEBER is with the Public Library of Science; KERSTIN LEHNERT is with the Lamont-Doherty Earth Observatory of Columbia University; MARK PARSONS is with Tetherless World Constellation, Rensselaer Polytechnic Institute, ERIN ROBINSON is with Earth Science Information Partners; MICHAEL WITT is with Purdue University; LESLEY WYBORN is with the Australian National University; and LYNN YARMEY is with the Research Data Alliance, Rensselaer Polytechnic Institute.



Michael Witt (Purdue University) demonstrating the new Repository Finder tool developed by DataCite for the project.

developing the primary publication. This is particularly the case in the Earth, space, and environmental sciences where diverse data are critical for understanding the dynamics of our planet and solar system.

Recognizing the key value of curated, well-described data, The American Geophysical Union (AGU) first adopted a position statement¹ on the importance of Earth, space, and environmental science data in 1997. This was amended in 2015 and states:

“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. They should be preserved longterm for future use. They should be made openly available to the scientific community and the public as soon as possible. They should be accessible in usable formats with sufficient machine-readable documentation to allow informed reuse. These responsibilities are an integral part of scientific

CONTINUED



Erik Schultes, (GO FAIR International Support and Coordination Office), Natasha Simons (Australian Research Data Commons), and Jens Klump (CSIRO Mineral Resources) are participating in an assumptions wrangling exercise with Leslie Hsu (back turned, U.S. Geological Survey).

research shared by individual scientists, data stewards, research institutions, and funding organizations.”

Many other organizations have also recognized the need for well-curated data. For example, there are, at time of writing, 119 organizational endorsements² of the Joint Declaration of Data Citation Principles adopted by FORCE11.³ In the Earth, space, and environmental sciences, the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS)⁴ and the major journal publishers in these domains were signatories of a statement of commitment regarding data in an effort brought together in 2014. COPDESS galvanized a community of publishers, repositories, and infrastructure “to help translate the aspirations of open, available, and useful data from policy into practice” as stated on their website. This effort established a dialog between publishers and repositories with common goals and collaborative objectives in a way that rarely occurs.

In 2016, the FAIR Data Principles,⁵ were published as a compilation of principles arranged around into four themes specific to better stewardship for scientific data: making that data Findable, Accessible, Interoperable, and Reusable (FAIR). COPDESS embraced the FAIR principles as an opportunity to move from aspiration and agreement to actual implementation of open and FAIR data by targeting the roles of the scholarly publisher, the scientific repository, and the connecting infrastructure.

The Enabling FAIR Data Project⁶ in the Earth, space, and environmental sciences, funded by the Laura and John Arnold Foundation to the AGU, invited the participants of the COPDESS effort along with other key stakeholders to

form a new coalition to establish common policies across all publishers that require data be as open as possible and preserved in a repository that follows the FAIR data guidelines. The project objectives for the two primary stakeholder groups include:

- Scholarly publishers: to adopt common policy that data are no longer archived in the supplementary information of a manuscript, that all data are to be deposited, documented, and preserved in a FAIR-aligned repository,⁷ and cited in the manuscript with an appropriate data availability statement.
- Scientific repositories: to support authors and researchers by providing services to ensure data and software that support published research are well documented, identified with global persistent identifiers (PIDs), and have landing pages that support both machine and human readable data citation information.

Journal editors and reviewers have a pivotal role in implementing any new publisher commitments by guiding authors through the new expectations of citing data and software in their manuscript. Understanding that editors and reviewers need resources to consistently guide authors and reviewers, we have prepared the information and guidelines below.

The Enabling FAIR Data coalition has recently completed the development of a set of common data authoring policies, author guidelines,⁸ and defined expectations of each stakeholder community in a commitment statement encouraging project participants and members of these communities to become signatories and put into place the policy and practice needed to meet the criteria in the next year.

Currently, there are over 90 organization and individual signatories⁹ for the commitment statement working to meet the criteria in the next year. We invite you and your communities to also become signatories.

Journal editors and reviewers have a pivotal role in implementing any new publisher commitments by guiding authors through the new expectations of citing data and software in their manuscript. Understanding that editors and reviewers need resources to consistently guide authors and reviewers, we have prepared the information and guidelines below.

Enabling FAIR Data Author Guidelines⁸

Each publisher who has signed the Commitment Statement¹⁰ will ensure that their author guidelines include the text developed by the Enabling FAIR Data project. We

CONTINUED

are expecting the text to be incorporated into existing online content, and not necessarily copied in verbatim, understanding that each publisher has its own approach to guidance.

The guidelines include the common practices expected by authors across all the Earth, space, and environmental science journals. A paraphrased excerpt of the guidelines publishers require of authors is below with the full text available online.¹¹

1. Deposit research data in a FAIR-aligned repository,¹² with a preference for those that explicitly follow the FAIR Data Principles and demonstrate compliance with international standards for data repositories (e.g., CoreTrustSeal¹³). Supplements to articles must not be used as an archive for data.
2. Cite and link to the data in the article, following the Joint Declaration of Data Citation Principles¹⁴ and ESIP Guidelines,¹⁵ using the unique, resolvable, and persistent identifiers provided by the repository in which the data are archived.
3. Include a Data Availability Statement describing how the data underlying the findings of their article can be accessed and reused.
4. Provide unrestricted access to all data and materials underlying reported findings for which ethical or legal constraints do not apply.¹⁶

Frequently Asked Questions

To support editors and reviewers in the implementation process of the new author guidelines and Enabling FAIR Data policies, we worked with the project stakeholders to develop a list of Frequently Asked Questions¹⁷ and answers that will be kept updated.

Repository Finder Tool

An important new tool available to researchers is Repository Finder.^b Many researchers do not yet have a relationship with a repository that can provide support services. There are over 2000 repositories internationally cataloged in re3data.org¹⁸ with different criteria for the types of data they accept, and which researchers are eligible to deposit. Repositories that provide support to researchers helping them to document their data to make it more understandable by others are preferred along with those meeting the criteria defined in the Commitment Statement.¹⁹ DataCite²⁰ developed this tool on top of re3data.org, a registry of data repositories, and recently published a blog²¹ where you can learn more. It lists repositories that are open to researchers and support globally registered persistent identifiers. Additionally, a seal logo indicates a third-party certification of capabilities.

CoreTrustSeal is one of these certifications and is expected to be increasingly adopted within the Earth, space, and environmental science repository community over the next few years. Having a CoreTrustSeal certification is not required to be FAIR-aligned but does indicate that the repository meets the majority of the Enabling Fair Data project repository criteria and more.

Adoption of Open and FAIR Data Principles in other Scientific Domains

Good work is being done in domains such as chemistry with new efforts by the International Union of Pure and Applied Chemistry (IUPAC)²² to become a GO FAIR²³ Implementation Network, and the health domain with FAIR standards as part of the criteria for the new National Institute of Health Data Commons,²⁴ encouraging sharing of research data. It is difficult to get significant change to occur within and across domains unless many of the stakeholders adopt similar policies and practices in a coordinated way. Societies and communities can be a strong influence helping to bring together the wider stakeholder communities including journals, repositories, institutions, and funders for common goals. The method used by the Enabling FAIR Data project can readily be adapted to any domain. The author guidelines can be directly incorporated into the guidance provided by any scientific journal to their authors. Communities that are platforms for working groups like Earth Science Information Partners (ESIP),²⁵ through their Data Stewardship Committee,²⁶ can help create examples of data and software citations that are domain specific, building on the work by FORCE11. By having common policies, guidelines, common answers to frequent questions, and domain-specific examples the beginning of the journey to become open and FAIR is a good start. The work ahead to sustain the new author guidelines supporting open and FAIR principles is with our current culture and identifying the barriers that remain for data sharing, attribution, and credit in order to be fully integrated in the research process and valued by our institutions and funders.

Culture Change for Sharing Data through Assumptions Wrangling

The Earth, space, and environmental sciences depend, in part, on increased collaboration and sharing of data. However, such sharing runs counter to long-standing assumptions that are deeply embedded in the culture of science; assumptions that position science as a competitive enterprise centered on advancing the narrow self-interests of key stakeholders.

During the most recent multi-stakeholder workshop for the Enabling FAIR Data project in September 2018, participants used an experimental process, “assumptions

CONTINUED

wrangling,” to work towards defining concrete actions to reduce barriers to the culture change needed for embracing open and FAIR data. The process was developed by the facilitator, Joel Cutcher-Gershenfeld, and designed to guide engagement with deeply embedded cultural assumptions using a method that is also applicable in other contexts.

The workshop included an international assembly of scholarly publishers, research data facilities, public and private funders, professional societies, and nongovernmental organizations engaged in this process of assumptions wrangling, borrowing terminology from what is called “data wrangling” in science. The exercise was motivated by observing that “everyone complains about culture, here is a way to do something about it.”

The assumptions wrangling process builds on the work by Douglas McGregor, and advanced by Ed Schein, and involves four steps:

- Step 1: From/To Assumptions
- Step 2: Driving/Restraining Forces
- Step 3: Indicators
- Step 4: Personal and Ecosystem Implications

The process begins with identifying current, partly problematic embedded assumptions and alternative aspirational assumptions: What we termed the “From/To” stage. We call these “operating assumptions” since they are deeply embedded in the operating practices of the science enterprise. Table 1 contains a few edited examples from the dozens of From/To pairs that were identified by the workshop participants.

Note that these examples were also presented in a report on the assumptions wrangling process for the Winter 2019 issue of *Heller Magazine*, published by the Heller School for Social Policy and Management at Brandeis University.

Note that the “partly problematic” assumptions are also partly functional. They have various logics supporting them—shifting these assumptions is not just a matter of calling them out. In this workshop, small groups brainstormed lists of driving and restraining forces associated with selected From/To pairs, recognizing that there are restraining forces that serve the interests of some or all stakeholders. In this case, restraining forces include incentives and rewards associated with career advancement (emphasizing individual rather than collective efforts), lack of knowledge and skill in the associated data work, and funding models that do not anticipate long-term storage and reuse of data, among many other factors. Driving forces include the coordinated efforts of the key stakeholders (such as the commitment statement), changes in incentives (data sharing will be part of the selection criteria for fellows in the AGU), changes in policies (funding agencies enforcing required data management plans in proposals), and other developments.

The third step in assumptions wrangling involves identifying specific indicators that would represent evidence of change in the underlying assumptions. There are many in the geosciences, including demonstrated compliance with data management plans listed in research proposals, increased ingest of data and other research objects in data facilities, documented reuse of data from data facilities, inclusion of evidence of data reuse in tenure and promotion cases, and other behavioral indicators. However, the most important indicators are advances in the Earth, space, and environmental sciences that would not have been possible without the sharing and reuse of data. Tracking these impacts are what will be most important in a long-term shift in the underlying assumptions.

Table 1. Selected examples of “From” and “To” operating assumptions

“From” Partly Problematic Assumptions	“To” Aspirational Assumptions
As a researcher, I am in competition with my colleagues.	As a researcher, I am part of a greater community that is both cooperative and competitive. In this context, I am responsible for sharing output (data, samples, software tools, and models), with appropriate embargo periods, so as to ensure reproducibility and enable reuse.
Posting data on a website or in an attached document with an article is sufficient for reproducibility and progress in science.	Researchers submit data to appropriate repositories in formats and file types that are immediately (or easily) ingestible and interoperable. Associated metadata is complete and able to be transformed into multiple formats.
Scientific funding and other resources should follow people and organizations, not data.	Data, physical samples, and software tools and models are first-class scientific objects worthy of direct investment.
Data should only be attached to scientific articles.	Data can have unique identifiers and sometimes it is the articles that should be attached to the data.

CONTINUED

Completing the workshop process involved asking over 60 institutional leaders to indicate specific behavioral changes they would advance in their work over the next 18 months that are reflective of the “To” assumptions, as well as larger changes in the ecosystem that they see as essential. In this case, the commitments will be evident in editorials in leading scientific journals; workshops at professional meetings; new prizes and honors lifting up data sharing; collected success stories; policy changes to require data submissions with articles; tools to help researchers find relevant repositories; methods to attach unique digital identifiers to data, samples, and software; and other developments. We developed the assumptions wrangling approach with the expectation that these action commitments would be further reaching than if we just asked people to identify next steps without taking a deep dive on assumptions, and this was indeed the case.

Ultimately, the process of shifting deeply embedded operating assumptions will be an iterative one, rather than a one-time event. In this case, there are concrete plans to track the various indicators identified as a “check and adjust” on the action commitments. Progress in the case of the geosciences is important to us all, in that it is the planet Earth that is at stake and advances in the Earth, space, and ecological sciences depend on culture changes that foster increased cooperation and data sharing. Adaptation to other settings is also important since there are so many social impact domains relevant to the Heller community where culture change is needed.

Summary

As stated in the AGU’s Data Position Statement, “Earth and space sciences data are a world heritage.” Discoveries made in the near and distant future will benefit from our stewardship of data collected today. Moving our community towards better understanding of this investment in our data is critical and pivotal to future science. Datasets we create as part of our research must stand on their own for possible use and reuse by others in our own domains or possible a completely different domain. As data are easier to find and understand as a result of these policy changes and work by others, our community has an opportunity to conduct new and exciting research with higher levels of trust in good data stewardship.

References, Notes, and Links

1. <http://bit.ly/AGUDataPosition>
2. <https://www.force11.org/datacitation/endorsements>
3. Data Citation Synthesis Group. Joint declaration of data citation principles. Martone M, editor. San Diego (CA): FORCE11; 2014. <https://doi.org/10.25490/a97f-egyk>
4. <http://www.copdess.org/>
5. Wilkinson MD, Dumontier M, Aalsbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Woiten, J-W, da Silva Santos LB, Bourne PE et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>
6. <http://www.copdess.org/enabling-fair-data-project/>
7. A FAIR-aligned repository complies with the tenets in the Enabling FAIR Data Commitment Statement (<http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/>) specific to scientific repositories.
8. Enabling FAIR Data Community, de Waard A, Cousijn H, Heber J, Hanson B, Bradford M, Friedman M, Hou S, Jones P, Kavanagh E, et al. Author guidelines for enabling FAIR data in the earth, space, and environmental science. Zenodo 5 October 2018. <http://doi.org/10.5281/zenodo.1447108>
9. <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/signatories/>
10. Enabling FAIR Data Community. Commitment statement to enabling FAIR data in the earth, space, and environmental sciences. Zenodo 8 October 2018. <https://doi.org/10.5281/zenodo.1451971>
11. <http://www.copdess.org/enabling-fair-data-project/author-guidelines/>
12. A tool to assist in identifying FAIR-aligned repositories is available from DataCite and can be found at <https://repositoryfinder.datacite.org>.
13. <https://www.coretrustseal.org/>
14. <https://www.force11.org/datacitationprinciples>
15. <http://commons.esipfed.org/node/308>
16. There may be a need to restrict some access to data because of fragile environments, endangered species, geopolitical tensions, or cultural sensitivities (e.g., indigenous land rights).
17. <http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/>
18. <https://www.re3data.org/>
19. <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/>
20. <https://www.datacite.org/>
21. <https://doi.org/10.5438/wday-8958>
22. <https://iupac.org/event/supporting-fair-exchange-chemical-data-standards-development/>
23. <http://www.go-fair.org/>
24. <https://commonfund.nih.gov/commons>
25. <https://www.esipfed.org/>
26. http://wiki.esipfed.org/index.php/Preservation_and_Stewardship