Request for Information (RFI) on NSF Public Access Plan 2.0: Ensuring Open, Immediate, and Equitable Access to National Science Foundation Funded Research

Kristina Vrouwenvelder[1], Shelley Stall[1], and Matthew Giampoala[1]

[1]Affiliation not available

January 23, 2024

December 27, 2023

Dr Martin Halbert
National Science Foundation
2415 Eisenhower Ave.,
Alexandria, VA
22314.

RE: Request for Information (RFI) on NSF Public Access Plan 2.0: Ensuring Open, Immediate, and Equitable Access to National Science Foundation Funded Research

Dear Dr Martin Halbert,

Thank you for the opportunity to participate in this Request for Information (RFI) on NSF Public Access Plan 2.0: Ensuring Open, Immediate, and Equitable Access to National Science Foundation Funded Research. The National Science Foundation, as a major funder of scientific research in the U.S., is a major leader of policy and standards that will set best practices for scientific research and innovation. The American Geophysical Union (AGU), a professional society, bases our response to the RFI on our experience working with Earth, space, and environmental science researchers as well as computer and information scientists and data and software experts in the work that we are doing promoting FAIR and open data sharing in AGU journals and in the wider community, including international and domestic work with researchers and repositories (both general and disciplinary). These efforts include providing guidance on best practices for data sharing upon publication and working with AGU authors and editors to find appropriate data and software repositories. Our response to this RFI incorporates this experience with researchers.

Sincerely,

Kristina Vrouwenvelder
Program Manager,
AGU Open Science Leadership
ORCID: 0000-0002-5862-2502

Shelley Stall
Vice President,
AGU Open Science Leadership
ORCID: 0000-0003-2926-8353

Matt Giampoala
Vice President,
AGU Publications
ORCID: 0000-0002-0208-2738

Challenges and Barriers to Immediate Access to Scholarly Peer-Reviewed Publications

3.A Peer-Reviewed Scholarly Publications

AGU is committed to open, accessible publication of scholarly peer-reviewed publications and is currently transitioning all 24 AGU journals to an open access model. AGU journals will permit deposit into NSF-PAR; however, this policy increases the likelihood that journals will need to shift away from subscription business models. As publishers adopt new business models, this policy increases the likelihood that more authors will encounter open access fees under the new business models. These open access fees present barriers for authors who are not funded by grants that cover open access publication costs or for the many authors who publish work associated with a grant after the funding period for that work has already completed and represent significant barriers to equity and access in scientific research. A recent analysis of AGU's publications demonstrates our commitment to open access publishing through both hybrid and fully open access journals, but a significant fraction of AGU authors still choose to publish articles that are not immediately open access. (Wooden et al., 2023)

3.A.ii Maximizing Public Access to Scholarly Publications With Machine-Readability

AGU recognizes and supports the value of maximizing public access to scholarly publications by including machine-readable XML versions of peer-reviewed publications in PAR. We recommend that as part of the implementation of this effort, the NSF ensure that an open, well-documented API is included for easy, machine-operable access to publications preserved in the PAR and their metadata.

Extending Access to Research Outputs beyond Peer-Reviewed Publications and Data

AGU recognizes and supports the value of research outputs beyond the peer-reviewed publication. We have elevated data and software outputs as part of scholarly research and publication, asking authors to share the data and software that underly their AGU publications. AGU is also piloting an 'open books' program to ensure that access to AGU-published books is as open as possible. As NSF considers extending PAR to books and non-peer-reviewed outcomes, we recommend the inclusion of software, books, and all conference reports from NSF-funded workshops and conferences, whether juried or not. These requirements will help ensure open, equitable access to all outputs of federally funded research.

Challenges and Barriers to Sharing Data Underlying NSF-funded Publications

3.B Scientific Data - FAIR *data in* NSF PAR 2.0

For effective data sharing and data reuse, FAIR (Findable, Accessible, Interoperable, and Reusable) data are a necessity. We recognize and support NSF's focus on enabling FAIR data and data sharing. (American Geophysical Union, 2019) However, FAIR data are not a single target; achieving FAIR-er data encompasses a spectrum of good to better to best data practices. To achieve FAIRness, NSF must ensure that data management and data curation are built into the data lifecycle from project start to data sharing and publication.

As this update to the NSF Public Access Plan acknowledges, research data are extremely heterogeneous. Data have different levels of investment, different potentials for reuse, and different opportunity costs to future scientific endeavor if FAIR and open principles are not embraced (for example, some datasets can only be collected once, while other experiments can easily be reproduced). Often, if not always, these heterogeneities are discipline- and domain-specific and should be captured with the help of guidance from the appropriate research communities and discipline data experts. Achieving FAIR-er data will look different across disciplines and data types. We urge NSF and the appropriate directorate, division, or

program levels to adopt policies on what data needs to be FAIR and what FAIR compliance looks like for that data, versus what data can be more lightly managed. Adopting such policies will help ensure that vital data for reuse is FAIR-compliant and will smooth adoption and improve researcher efficiency by applying lighter management principles to other types of data as needed. Some of the specific data types requiring different management approaches that we recognize in the Earth, space, and environmental sciences include observational data, experimental data, physical samples, and model and simulation outputs.

3.B.ii. Scientific Data Not Associated With Peer-Reviewed Scholarly Publications

If data are a primary research output, as the NSF has elevated in this update, then NSF should consider adding a recommendation or requirement that even datasets not associated with peer-reviewed scholarly publication be shared openly in NSF PAR. We further recommend that the NSF add guidance to researchers on the importance of sharing data in a timely fashion, e.g. during the period of grant funding and research, to ensure that knowledge, skills, and attention are given to this important research output.

3.B.iii. Digital Repositories for Scientific Data

Disciplinary repositories can offer important domain-specific data management & curation guidance and services to researchers, improving the FAIR-ness of data and ensuring that it is analysis-ready; aligning to discipline best practices, enhancing reuse. AGU recommends that NSF explicitly elevate these added services that disciplinary repositories provide to researchers. AGU supports data sharing by our authors in community-accepted, trusted repositories; however, since implementing our data sharing policy for authors, we have seen authors increasingly sharing data through generalist repositories lacking data curation or discipline specific metadata services. This trend presents challenges for maximizing FAIR data and enabling data reuse. (Hanson et al., 2023) Some disciplines are served by a disciplinary repository, but not all disciplines have one. Given that NSF is advising the use of disciplinary repositories, we recommend that NSF ensure that all researchers have a discipline-specific repository available for their data. Disciplinary repositories that do exist need more sustainable funding paradigms for repository operations. In particular, disciplinary repositories must employ domain experts to provide the important data and metadata curation and data discovery services that disciplinary repositories offer. These services require consistent funding.

3.C.i. Legal, Privacy, Ethical, Technical, Intellectual Property, and Security Limitations on Access to Scientific Data Arising From NSF Awards

When NSF develops best practices for data sharing, including use cases where legal, privacy, ethical, technical, intellectual property, or security limitations may arise, we would recommend that NSF include community guidance and feedback from other groups with experience in this sphere, including digital repositories and standards organizations and other federal agencies where applicable. Data are commonly used in interdisciplinary research, and standards must be set across disciplines, with all experiences and expertise in mind.

3.C.iii. Specific Online Digital Repositories -- *Data Management and Sharing Plans (DMSP)*

At the AGU, we recognize the value of data as a research output. Data sharing makes scientific research more reproducible and transparent and enables future data reuse. We support the inclusion of data sharing in data management plans for federally-funded research, just as we have supported and mandated the sharing of data underlying scientific publications for some years now. However, based on

our work with authors publishing with the AGU and with researchers more broadly, we believe that more support for researchers and for grant reviewers is needed to effectively implement the NSF's planned DMSPs. Researchers will need effective guidance and training on writing DMSPs, in particular to allow researchers to understand the benefits and drawbacks of different data repository solutions. (Stall et al., 2023) Disciplinary repositories offer many important resources to researchers, including curation services and increased FAIR-ness of data, but their timelines, costs, and other advantages and disadvantages differ. Researchers may need direction to the most appropriate repository for their data type and support preparing their data for submission. Timelines for data preservation in discipline repositories offering curation services are not immediate or even quick, as the curation process takes time. These timelines for data preservation must align with timelines for publishing peer-reviewed publications and for annual grant reporting to the NSF, or researchers will not be able to comply with publisher and funder data sharing requirements. Importantly, reviewers of DMSPs for the NSF will also need support here in evaluating effective DMSPs and offering feedback to researchers. We suggest that this could take the form of a rubric for evaluating DMSPs.

Section 4: Ensuring Scientific and Research Integrity

AGU supports the use of PIDs to identify preserved research outputs, grants, and senior personnel. We recommend that NSF work across federal agencies to ensure PID requirements and guidance are unified and avoid proliferating PID standards on the national level. Research, and researchers, work across discipline boundaries and need unified guidance and a national strategy on PID use.

Section 5: Public Access Plan Coordination Among Federal Agencies

We recommend that the NSF work across federal agencies to understand and implement the CARE principles, in particular with consideration for data sovereignty. (Carroll et al., 2020)

We recommend that under "Roles and Responsibilities", the NSF recognize the important role of disciplinary repositories in implementing and enacting this guidance and supporting FAIR, open data.

Opportunities and Benefits from Sharing Data Underlying NSF-funded Publications

This is an exciting time for scientific discovery and research, as new, powerful computational tools and methods, including AI and ML methods, enable new discovery and analysis of big data. To maximize the potential of these tools, FAIR and analysis-ready data are essential. Data management and data and metadata curation will ensure that scientific datasets are ready for reuse. Data sharing enhances reproducibility and transparency across science, reducing duplication of effort and shoring up public trust in science. At the AGU, we recognize data as an important scientific research output and our publication policies, which ask authors to share data and software underlying their scientific research in their peer-reviewed publications, reflect this. (American Geophysical Union, 2019) Since enacting these policies, we have seen the uptake of author sharing of data and software double. As we encourage data sharing, an ongoing impact analysis of the effect on scientific research is needed. The effect of our efforts towards increased transparency and reproducibility should be analyzed and an ongoing impact analysis of the use and reuse of data to enable scientific discovery is needed. These efforts will increase trust in science and help effect a culture change across research by demonstrating the value of FAIR data and open science practices.

References

American Geophysical Union. (2019, November). *AGU Position Statement on Earth and Space Science Data.* https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/Position_Data

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), Article 1. https://doi.org/10.5334/dsj-2020-043

Hanson, B., Stall, S., Cutcher-Gershenfeld, J., Vrouwenvelder, K., Wirz, C., Rao, Y. (Douglas), & Peng, G. (2023). Garbage in, garbage out: Mitigating risks and maximizing benefits of AI in research. *Nature*, 623(7985), 28–31. https://doi.org/10.1038/d41586-023-03316-8

Stall, S., Specht, A., Corrêa, P. L. P., David, R., Edmunds, R., Mabile, L., Machicao, J., O'Brien, M., Wyborn, L., Vellenich, D. F., Miyairi, N., & Murayama, Y. (2023). *PARSEC DDOMP Workbook Checklist.* https://doi.org/10.5281/zenodo.8239982

Wooden, P., Lunn, J., Vrouwenvelder, K., Moerchen, M., Ricci, M., Giampoala, M., & Sedora, B. (2023). AGU Publications Opens Science: Making Science Accessible and Equitable. *Eos*, 104. https://doi.org/10.1029/2023EO235035