

DATA CURATION NETWORK

*Data Curation Network
Based at the University Libraries
University of Minnesota
Minneapolis, MN 55455*

Ethical. Reusable. Better.

December 20, 2023

To Whom it May Concern:

On behalf of the Data Curation Network (DCN), we appreciate the opportunity to provide [input on the National Science Foundation's \(NSF\) Public Access Plan 2.0](#). The DCN has been an advocate for open and ethically shared data since our inception and are excited to see NSF's continued efforts to incentivize and fund equitable data sharing efforts. In addition to our comments submitted directly to the request for input, we offer the following additional feedback.

The DCN applauds NSF's explicit expectation to share research data affiliated with peer-reviewed publications, recognizing that open data is an important step forward in enabling a FAIR (Findable, Accessible, Interoperable, and Reusable) and equitable research environment. This requirement represents an opportunity to foster a data sharing culture that ensures data is as open as possible and as closed as necessary. Furthermore, we commend NSF for recognizing that open data, resulting from data sharing requirements, will enable researchers to gain easier access to a larger corpus of research, which will help reduce duplication of effort and enable research reproducibility.

As a member-based consortium of 17 academic and 2 non-profit data repositories, the DCN notes that disciplinary repositories are specifically highlighted in the NSF Public Access Plan 2.0 as appropriate for sharing data. The emphasis on disciplinary repositories overlooks a wide range of established and trusted repositories that could be used to share, publish, and importantly preserve research data. This includes, but is not limited to, institutional repositories as well as generalist repositories like [Dryad](#).^{*} Generalist and institutional repositories are essential research infrastructure, as there are disciplines that may not yet have established disciplinary repositories, or may not have repositories operating on stable funding models. Furthermore, from the perspective of a researcher, individuals may be unable to transfer the large amount of data generated through their project or may have to [pay to deposit their data](#). While we acknowledge the benefits of NSF investments in disciplinary repositories over the year, we strongly recommend that NSF highlight a more inclusive group of established repositories that align with the [Desirable Characteristics of Data Repositories for Federally Funded Research](#), which includes generalist and disciplinary repositories.

In addition, we recommend the NSF highlight institutional infrastructure potentially available to its funded researchers to not only share and preserve scholarly outputs, but also support end-to-end research [data management and sharing activities](#). This includes support for writing

grant proposals (including data management and sharing plan reviews and IRB support) to support for publishing that is offered at many academic institutions. We recommend that the NSF adopt language similar to that of the [National Institutes of Health \(NIH\)](#), which recognizes that while “discipline or data-type specific repositories may not exist for every type of data ... the broader repository ecosystem provides suitable data repositories to accommodate scientific data.”

Specific to sharing data associated with peer-reviewed articles, as addressed in [NSF 23-104](#), section 3.B.ii, we encourage NSF to continue to develop approaches and timelines for sharing other types of federally funded scientific data not associated with peer-reviewed scholarly publications, including sharing data that may be related to null findings or that are otherwise unpublished. The DCN encourages NSF to adopt guidelines for data sharing that extends beyond the peer-review process, as datasets can be reused in ways unanticipated by authors and researchers. Additionally, the DCN recognizes that there are valid reasons that data may not be shared, including ethical and legal restrictions; we affirm the statements listed in section 3.C.i.

Further, while the plan currently states that data “will be made available,” we encourage NSF to adopt stronger language that reflects the significant investment of time and resources required to truly publish research data. Beyond being ‘made available’ or shared, data need to be thoroughly curated, reviewed, and documented in order to be [published in a FAIR and reproducible manner](#). As we consider the [value of curation to researchers](#), we encourage NSF to adopt stronger language for researchers to publish their data, including a curatorial review, instead of simply making it available. While not every repository or institution has the capacity to support researchers in curating their research data, collaborative efforts like the DCN are helping to bridge the gap between demands and resources, which includes curating and managing data, beyond simply making it available.

Lastly, we applaud the desire to make publications accessible to assistive technology, and we invite NSF to push the envelope further by encouraging that research data, especially those associated with a publication, also be made accessible. [Creating and publishing accessible data](#) is an involved process that data curators can assist with, but it is important that NSF continue to fund efforts and technologies that will make this process easier. This way, NSF can further demonstrate its commitment to making publicly funded research accessible.

Thank you again for the opportunity to provide feedback on the NSF Public Access Plan 2.0. We would be happy to provide additional information or clarification, and would welcome a meeting to discuss further.

**Note: Dryad is a member of the DCN.*

DATA CURATION NETWORK