



NATIONAL SCIENCE FOUNDATION
2415 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22314

NSF 22-055

Dear Colleague Letter: Effective Practices for Making Research Data Discoverable and Citable (Data Sharing)

March 3, 2022

Dear Colleague:

Digital data are a product of most scientific research, and as such, subject to NSF policy that requires sharing the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants (<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>). The implementation of this policy requires that proposals to the NSF contain a Data Management Plan (DMP), which includes a description of how the proposal follows best current practices on making data produced during the project's lifetime publicly accessible. The Division of Materials Research (DMR) expects its awardees to embrace NSF's policy on sharing digital data; the purpose of this Dear Colleague Letter is to help the community better understand and meet these requirements. To relay this priority, DMR recently developed guidance on preparation of DMPs specifically relevant to DMR-funded research (<https://www.nsf.gov/bfa/dias/policy/dmpdocs/dmr.pdf>). Many of the costs associated with data management are generally eligible for funding and may be included in the budget of a proposal submitted to DMR.

This Dear Colleague Letter describes and encourages effective practices for publicly sharing research data, including the use of persistent digital identifiers (PDIs).

Datasets underpinning published research findings are expected to be shared with other researchers, at no more than incremental cost and within a reasonable time. Data-sharing holds numerous benefits, from enabling broader research collaboration, through facilitating transparency and solidifying confidence in scientific research, to providing increased resources for teaching and education purposes. Recent studies found that research articles containing a link to data in a repository have markedly higher usage and visibility. Discoverable and citable data also serve to reduce barriers to entry for junior researchers, scientists in under-served communities, and researchers from underrepresented and minority groups, thus enabling improved implementation of open science principles.

The nature of digital data produced during research may vary among the different topical disciplines encompassed by the field of Materials Research. Most often, digital research data comprise one or more of the following: raw data files collected using experimental instrumentation and converted into digital format; digital files of processed experimental data; video and animation files; numerical data produced by computer simulations or computational models; computer code, scripts, software, software documentation and user manuals developed as part of the research project; digital files of theoretical models, protocols, and methods; educational, instructional, and training materials.

Open-access data sharing platforms (data repositories) comprise the most efficient way to publish and share research data¹. Moreover, as long-term data curation and preservation are core to their mission, data repositories provide a stable means for data preservation. Upon publication of a dataset, most repositories automatically generate a citation for the data, which includes identifying metadata such as the archiving repository, the data's author(s), and a PDI such as a digital object identifier (DOI). A DOI is a unique and persistent digital identifier, which, when assigned to a digital entity such as a dataset, remains unchanged over the lifetime of the object. Having a DOI (or other form of PDI) from an open-access repository renders data findable, accessible, and readily citable. Searchable global registries of data repositories provide information on indexed repositories to help researchers identify the most appropriate ones². In the case where a suitable repository is not available, researchers are strongly encouraged to use their institutional digital repositories, which typically issue DOIs to institutionally hosted content.

Optimally, publication of peer-reviewed research findings should be closely followed by sharing of the data underpinning the study. An increasing number of journals require data sharing concurrent with publication of an article. Data related to published research should generally be accessible without need for explicit or required requests from interested parties. Many repositories, additionally, allow users to set an embargo period, only after which the data will become available.

DISCLAIMER

The above guidelines are not intended to replace the guidance given in the *NSF Proposal & Award Policies & Procedures Guide* (PAPPG) and specific program solicitations. In any perceived conflict, the PAPPG or the solicitation will take precedence as appropriate for the proposal.

Sincerely,

Sean L. Jones
Assistant Director
Directorate for Mathematical and Physical Sciences

¹ Some commonly used data repositories include: NIST Materials Data Repository (<https://materialsdata.nist.gov/>); The Materials Data Facility (<https://materialsdatafacility.org/>); The Materials Project (<https://materialsproject.org/>); The Dryad Digital Repository (<https://datadryad.org/stash>).

² See, for example, Registry of Research Data Repositories (<https://www.re3data.org/>) and Designing Materials to Revolutionize our Future Software & Data (<https://dmref.org/tools>).