

National Institutes of Health (nih.gov): NIH-GEN: Generic (Current until 2023)

Data sharing plan

How do you plan to provide access to your data?

Guidance:

The method for sharing that an investigator selects is likely to depend on several factors, including the sensitivity of the data, the size and complexity of the dataset, and the volume of requests anticipated. Investigators sharing under their own auspices may simply mail a CD with the data to the requestor, or post the data on their institutional or personal Website. Although not a condition for data access, some investigators sharing under their own auspices may form collaborations with other investigators seeking their data in order to pursue research of mutual interest. Others may simply share the data by transferring them to a data archive facility to distribute more widely to interested users, to maintain associated documentation, and to meet reporting requirements. Data archives can be particularly attractive for investigators concerned about a large volume of requests, vetting frivolous or inappropriate requests, or providing technical assistance for users seeking help with analyses.

There are several mechanisms for data sharing that investigators can use. For example, investigators sharing under their own auspices should consider using a **data-sharing agreement** to impose appropriate limitations on users. Such an agreement usually indicates the criteria for data access, whether or not there are any conditions for research use, and can incorporate privacy and confidentiality standards to ensure data security at the recipient site and prohibit manipulation of data for the purposes of identifying subjects. Many examples of data sharing agreements for specific datasets are available on the Internet, including the following:

[AHRQ National Inpatient Sample](#)

[Russian Longitudinal Monitoring Survey](#)

[Center for Medicare and Medicaid Services Data \(PDF\)](#)

Alternatively, researchers may want to add their data to a data archive or a data enclave. Datasets that cannot be distributed to the general public, for example, because of participant confidentiality concerns, third-party licensing or use agreements that prohibit redistribution, or national security considerations, can be accessed through a data enclave. A data enclave provides a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.

Investigators may also wish to develop a "mixed mode" for data sharing that allows for more than one version of the dataset and provides different levels of access depending on the version. For example, a redacted dataset could be made available for general use, but stricter controls through a data enclave would be applied if access to more sensitive data were required.

Investigators will need to determine which method of data sharing is best for their particular dataset. The Data Sharing Workbook ([PDF](#) - 75 KB) or ([MS Word](#) - 74 KB) provides information and examples of how others have shared data.

From: [NIH Data Sharing Policy and Implementation Guidance](#)

When will you make the data available?

Guidance:

It is NIH policy that the results and accomplishments of the activities that it funds should be made available to the public. PD/PIs and recipient organizations are expected to make the results and accomplishments of their activities available to the research community and to the public at large. (See also [Availability and Confidentiality of Information-Confidentiality of Information-Access to Research Data](#) in Part I for policies related to providing access to certain research data at public request.) If the outcomes of the research result in inventions, the provisions of the Bayh-Dole Act of 1980, as implemented in 37 CFR 401, apply.

[NIH Data Sharing Policies](#)

[NIH Sharing Policy Statement](#)

Which archive/repository/central database have you identified as a place to deposit data?

Guidance:

[This](#) table lists NIH-supported domain-specific data repositories that make data accessible for reuse and are open for both submitting and accessing data. Submission is typically limited to data of a certain type or related to a certain discipline. The table provides links to information about

submitting data to and accessing data from the listed repositories. Repositories in this list have current NIH funding, sustained support, open data submission and access, and open time frame for data deposit, based on information [provided by the repository](#) about funding and data availability

Will a data-sharing agreement be required?

Guidance:

There are several mechanisms for data sharing that investigators can use. For example, investigators sharing under their own auspices should consider using a **data-sharing agreement** to impose appropriate limitations on users. Such an agreement usually indicates the criteria for data access, whether or not there are any conditions for research use, and can incorporate privacy and confidentiality standards to ensure data security at the recipient site and prohibit manipulation of data for the purposes of identifying subjects. Many examples of data sharing agreements for specific datasets are available on the Internet, including the following:

[AHRQ National Inpatient Sample](#)

[Russian Longitudinal Monitoring Survey](#)

[Center for Medicare and Medicaid Services Data](#)

Taken from: [NIH Data Sharing Policy and Implementation Guidance](#)

What metadata/documentation will be submitted alongside the data?

Guidance:

Regardless of the mechanism used to share data, each dataset will require documentation. (Some fields refer to data documentation by other terms, such as metadata or codebooks). Proper documentation is needed to ensure that others can use the dataset and to prevent misuse, misinterpretation, and confusion. Documentation provides information about the methodology and procedures used to collect the data, details about codes, definitions of variables, variable field locations, frequencies, and the like. The precise content of documentation will vary by scientific area, study design, the type of data collected, and characteristics of the dataset.

It is appropriate for scientific authors to **acknowledge the source of data** upon which their manuscript is based. Many investigators include this information in the methods and/or reference sections of their manuscripts. Journals generally include an acknowledgement section, in which the authors can recognize people who helped them gain access to the data. Authors using shared data should check the policies of the journal to which they plan to submit to determine the precise location in the manuscript for such acknowledgement. Most journals now expect that DNA and amino acid sequences that appear in articles will be submitted to a sequence database before publication. From [NIH Data Sharing Policy and Implementation Guidance](#).

What file formats will you use for your data, and why?

Guidance:

Given the breadth and variety of science that NIH supports, neither the precise content for the data documentation, nor the formatting, presentation, or transport mode for data is stipulated. What is sensible in one field or one study may not work at all for others. It would be helpful for members of multiple disciplines and their professional societies to discuss data sharing, determine what standards and best practices should be proposed, and create a social environment that supports data sharing. presentation, or transport mode for data is stipulated. What is sensible in one field or one study may not work at all for others. From [NIH Data Sharing Policy and Implementation Guidance](#)

What transformations will be necessary to prepare data for preservation/data sharing?

Do you need funding for the implementation of this data sharing plan?

Guidance:

Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections. Some investigators have more experience than others in estimating costs associated with preparing the dataset and associated documentation, and providing support to data users. As investigators gain experience with the process, their ability to estimate costs will improve. Investigators working with archives can get help with data preparation and cost estimation. Investigators who are concerned about paying for data-sharing costs at the end of their grant can make prior arrangements with archives. Investigators facing considerable delays in the preparation of the final dataset for sharing should consult with the NIH program about how to manage this situation, such as requesting a no-cost extension.

(Taken from: [NIH Data Sharing Policy and Implementation Guidance](#))