

***Congressionally Directed Medical Research Programs:
Policy on Sharing Data and Research Resources***

SHARING DATA AND RESEARCH RESOURCES¹:

It is the intent of the Congressionally Directed Medical Research Programs (CDMRP) that data and research resources generated by CDMRP-funded research be made available to the research community (which includes both scientific and consumer advocacy communities) and to the public at large.

For the purposes of the CDMRP, expectations for sharing of data and research resources apply to all basic research, clinical studies, surveys, and other types of research funded by the CDMRP. This includes all data and research resources generated during the project's period of performance through grants, cooperative agreements, or contracts. It is critically important to share unique data and research resources that cannot be readily replicated, including but not limited to the following:

- **Unique Data²** are defined as data that cannot be readily replicated. Examples of unique data include large surveys that are expensive to replicate; studies of unique populations, such as patients to whom access is not widely available; studies conducted at unique times, such as during military conflict; studies of rare phenomena, such as rare diseases.
- **Final Research Data³** are defined as recorded factual material commonly accepted in the scientific community as necessary to document and support research findings. These are not the summary statistics or tables; rather, final research data are the data on which summary statistics and tables are based. Final research data do not include laboratory notes or notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens.
- **Research Resources⁴** include, but are not limited to, the full range of tools that scientists and technicians use in the laboratory, such as cell lines, antibodies, reagents, animal models, growth factors, combinatorial chemistry, DNA libraries, clones and cloning tools (such as PCR), methods, laboratory equipment and machines.

Data and research resources generated from CDMRP-funded research should be made as widely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data and third-party intellectual property. By sharing and leveraging data and research resources, duplication of very expensive and time-consuming efforts can be avoided, allowing for the support of more investigators with federal funds. The CDMRP believes that such sharing allows for a more expeditious translation of research results into knowledge, products, and procedures to improve human health.

¹ Adapted from the National Institutes of Health Data Sharing Policy and Implementation Guidance (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm) and Sharing Biomedical Research Resources - Principles and Guidelines for Recipients of NIH Research Grants and Contracts (<http://www.nih.gov/science/models/sharing.html>)

² Adapted from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#unique

³ Adapted from http://ott.od.nih.gov/policy/rt_guide_final.html

⁴ Adapted from http://ott.od.nih.gov/policy/rt_guide_final.html

Congressionally Directed Medical Research Programs: Policy on Sharing Data and Research Resources

Policies with respect to the sharing of data and research resources vary across institutions and countries. Investigators should familiarize themselves with their institution's policies governing the sharing of data and research resources as well as the policies of the host countries and/or institutions in which they plan to conduct research.

A. Methods for Sharing Data

There are many ways to share data, such as under the auspices of the investigator or laboratory, through use of a data archive or enclave, or through mixed mode sharing, as described below. The selected method for sharing 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.

Regardless of the mechanism used to share data, each dataset will require documentation. Proper documentation is needed to ensure that the dataset will not be misused or misinterpreted. Documentation provides information about the methodology and procedures used to collect the data, details about codes, definitions of variables, variable field locations, frequencies, and any other unique background information. The precise content of documentation will vary by scientific area, study design, type of data collected, and characteristics of the dataset.

1. Sharing under the auspices of the PI or laboratory

Investigators sharing under their own auspices may share data with the requestor via external storage media, 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 share the data by transferring them to a data archive facility to distribute more widely to interested users, maintain associated documentation, and meet reporting requirements.

Investigators sharing data under their own auspices should consider using a data-sharing agreement to impose appropriate limitations on users. Such an agreement typically indicates the criteria for data access, the conditions for research use, privacy and confidentiality standards to ensure data security at the recipient site, and prohibitions for manipulating data for the purposes of identifying subjects.

2. Data Archives

Data archives can be particularly attractive for investigators concerned about a large volume of requests, vetting inappropriate requests, or providing technical assistance for users seeking help with analyses.

***Congressionally Directed Medical Research Programs:
Policy on Sharing Data and Research Resources***

3. Data Enclaves

A data enclave provides a controlled, secure environment in which researchers are allowed to perform analyses using restricted data resources. 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, should be accessed through a data enclave.

4. Mixed Mode Sharing

Investigators may also 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.

B. Methods for Sharing Unique Research Resources

To facilitate the availability of unique or novel biological materials and resources developed with CDMRP funds, investigators may distribute the materials through their own laboratory or organization or submit them, if appropriate, to entities such as the American Type Culture Collection or other repositories or distributors. When distributing unique resources, investigators are to include pertinent information on the nature, quality, or characterization of the materials.

In addition, CDMRP-funded investigators are expected to submit unique biological information, such as DNA sequences or crystallographic coordinates, to the appropriate data banks so that they can be made available to the broad scientific community.

Organizations that believe they will be unable to comply with these requirements should promptly contact the CDMRP to discuss the circumstances, obtain information that might enable compliance, and reach an understanding in advance of an award.