The New NIH Policy for Data Management and Sharing and What it Means for You at the University of Cincinnati

Tiffany Grant and Amy Koshoffer

Co-leaders UC Libraries Research & Data Services

Table of Contents

Introduction

"Data without context are inert, but data within contexts become information, knowledge[1]."

Researchers submitting for funding through the National Institutes of Health (NIH) on or after January 25, 2023, should be aware of the requirement to submit a Data Management and Sharing Plan (DMSP) for any NIH-funded or conducted research that will generate scientific data. Previously, the NIH only required grants with funding of $500,000/year or greater in direct costs to provide a short explanation of how and when data resulting from the grant would be publicly shared. However, this new mandate requires all grant applications or renewals to include a detailed plan for data management and sharing for the funded period. This requirement is mandated through the Final NIH Policy for Data Management and Sharing that emphasizes the importance of good data management practices and establishes the expectation for maximizing the appropriate sharing of scientific data generated from NIH-funded or conducted research. The NIH defines scientific data as the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. The NIH has long championed the proper management and sharing of scientific data to accelerate biomedical discovery through the promotion of data reuse for future research studies.

The NIH encourages data management and sharing that is consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles. The FAIR Data Principles are a concise set of principles designed by representatives from academia, industry, funding agencies, and publishers, that serve to support and enhance reuse of data[2]. In the first formal paper documenting the FAIR Principles, the authors suggest that good data management is critical not only for knowledge, discovery, and innovation, but also for the integration and reuse of data post-publication. The FAIR principles refer very specifically to data that is "open". Open data is simply defined as "data that anyone can access, use, and share[3]". The NIH has a long-standing commitment to open data to increase the utility of data produced by federal funding and has done so through mandating data management and sharing initiatives. Proper management and sharing of research data have numerous benefits to researchers. Authors found that articles that include statements that link to data in a repository were associated with an up to 25% higher citation impact[4]. In another study, the authors showed a 69% increase in citations when data was made publicly available, and this increase was independent of impact factor, publication date, or the author's country of origin[5]. Citations are a type of currency in the scholarly community, as they can be directly tied to research funding, promotion, and notoriety in the respective field by facilitating increased visibility of the author's works. Moreover, allowing for greater access to data can foster collaboration opportunities, increase transparency in research, and maximize the reuse of data all while meeting funder and publisher requirements.

This document will serve as a single resource for researchers at the University of Cincinnati to learn about the new NIH Data Management and Sharing Policy that went into effect on January 25, 2023. Researchers can use this document and the embedded links to find information on what is required of them as they prepare to submit NIH grant proposals and what resources UC has available to them to facilitate the process. Throughout this document, researchers will find links to information and tools that will aid them as they prepare Data Management and Sharing Plans as well as information about available data repositories for data sharing.

---

[1] Sven Birkerts, *Changing the Subject: Art and Attention in the Internet Age* (Graywolf Press, 2015).

[2] Mark D. Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (March 15, 2016): 160018, https://doi.org/10.1038/sdata.2016.18.

[3] "What Makes Data Open?," accessed January 26, 2023, https://www.theodi.org/article/what-makes-data-open/.

[4] Giovanni Colavizza et al., "The Citation Advantage of Linking Publications to Research Data," *PloS One* 15, no. 4 (2020): e0230416, https://doi.org/10.1371/journal.pone.0230416.

[5] Heather A. Piwowar, Roger S. Day, and Douglas B. Fridsma, "Sharing Detailed Research Data Is Associated with Increased Citation Rate," *PLOS ONE* 2, no. 3 (March 21, 2007): e308, https://doi.org/10.1371/journal.pone.0000308.

The Policy in a Nutshell

At the time of this writing, the policy is effective for *all* grant proposals submitted to the NIH.

NIH has issued the Data Management and Sharing Policy to promote the sharing of scientific data. The policy serves as a mandate for researchers to prospectively plan for how scientific data will be managed and ultimately shared. To comply with the policy, researchers must:

- Determine if their proposed research is subject to the DMS policy.
- Identify appropriate methods/approaches and repositories for managing and sharing scientific data.
- Develop a Plan for managing and sharing scientific data and include it in the application or proposal. If subject to Genomic Data Sharing Policy, submit a plan that specifically addresses genomic data considerations.
- Estimate and request funds for data management and sharing activities (if not already covered by the institution or other sources.)

Applicants planning to generate scientific data must submit their data management and sharing plan to the NIH as part of the funding application or proposal. If awarded funding, researchers are expected to carry out data management and sharing as outlined in the approved plan and as a condition of the award.

Required Elements of Your NIH Data Management and Sharing Plan

All of the information in this section is part of a [checklist](#)[6] that was created by the NIH DMSP Guidance for Data Support Services Working Group. The document can be accessed [here](#)[7]. Additionally, the NIH has developed an optional DMS Plan format page that aligns with the recommended elements of a DMS Plan. This blank form can be accessed [here](#). More detailed Information regarding the elements that researchers should include in their plans can be found below.

1. **Data Type** **--Summarize the scientific data necessary to validate your findings.**
    a. List or create a table to describe the datasets that will be generated in the project including:
        i. Data type, format, size, and number of files
        ii. Which datasets will be shared
        iii. The level of aggregation, de-identification, or processing/cleaning that will be done before sharing
        iv. The source of any secondary data (previously collected data reused in the project)
    b. Scientific data that will be preserved and shared, and the rationale for doing so.
    c. List the metadata and other documentation (e.g. a README file) that will be shared with your data to facilitate interpretation.

2. **Related Tools, Software, and/or Code** **-- Identify tools, software, and/or code necessary to access or manipulate the shared data.**
    a. State whether or not specialized tools are needed, and for each tool that is necessary list the following:
        i. Version number and operating system,
        ii. How they can be accessed (i.e., open source and freely available, generally available for a fee in the marketplace, or available only from the research team or some other source),
        iii. How long they will be available (if known).

3. **Standards** **-- List the standards that will be used for sharing the data and metadata.**

---

[6] "Data Management and Sharing Plan Checklist for Researchers.Docx," October 13, 2022, https://osf.io/https://osf.io/awypt.
[7] Hao Ye et al., "Working Group on NIH DMSP Guidance," June 28, 2022, https://doi.org/10.17605/OSF.IO/UADXR.

    a. State whether or not there are data standards for your field that apply to your project. Typical data standards include:
  - i. Metadata schemas
  - ii. Standard Terminologies (Controlled Vocabulary and Ontologies)
  - iii. Content/ Encoding Standards
  - iv. Common Data Elements
  - v. Identifiers (PIDs)

4. Data Preservation, Access, and Associated Timelines **-- Provide details and timelines for sharing and preserving data for long-term usability.**
   a. Name the repository(ies) where data will be archived.
      - i. If a particular metadata standard is required, list it in the standards section.
      - ii. A specific NIH repository *may* be required in the funding opportunity announcement.
         1. A later section of this document will detail repository options available to you at UC.
   b. Specify which type of unique identifier is used by the repository (DOI, handle, ID number, accession number). Please note that an identifier is not required at the time of submission of your data management plan.
   c. Revisit your data list from section 1 (Data Types) and state when the data will be made available (portions of the data may be released at different times). Timelines required by the policy are:
      - i. Data will be made available when the work is published or the award/support period ends (whichever comes first) OR
      - ii. Data will be made available earlier
   d. State the minimum number of years data will be available, based on repository policies.
      - i. Note that per UC policy, data must be kept for a minimum of 5 years per UC Board of Trustees Code 10-43-18. Thus, if any repository has a minimum duration of under 5 years, it should be stated that the data will be retained for a minimum of 5 years to remain in compliance with UC policy.

5. Access, Distribution, or Reuse Considerations **-- Describe how sharing will be maximized while respecting restrictions.**
   a. Describe any considerations that may affect the extent of data sharing, including legal, technical, and/or ethical.
   b. Consider whether data can be shared with access controls or, if there are intellectual property concerns, an embargo period, rather than refraining from sharing altogether
   c. If you have human subject data, describe how you will protect the privacy, rights, and confidentiality of study participants (de-identification, etc.).

6. Oversight of Data Management and Sharing **– Identify who will be responsible for plan compliance and oversight.**
   a. List names and titles/roles of everyone who will be responsible for monitoring compliance with the data management plan and updating it as needed.
   b. State how often compliance with the data management plan will be verified (e.g. every ___ months, on the first of each month, etc.).

University of Cincinnati Support Teams, Tools and Resources

University of Cincinnati Libraries Research & Data Services Unit

The UCL Research & Data Services (RDS) group inspires the creation of knowledge and enhances research productivity across the UC research community through the development and implementation of interdisciplinary research data services that enable research and promotes synergistic collaborations between UCL and UC researchers. UC Libraries provides access to a wide range of Research Data and GIS services and resources for the UC research community. Members of the unit are available across the campus (both East and West) to assist researchers in managing and preserving research data, finding, and acquiring external data, and in utilizing GIS techniques and software. The library also provides a variety of computing and collaboration spaces to support researchers. More information about the team can be found on this website and a more detailed overview of the team's accomplishments can be found here. To submit a question to the team, email askdata@uc.edu.

The RDS unit also provides many services related to data management. These data services align with researchers at the time of proposal planning and continue through the course of the project. These services include:

- Data Management Workshops and Laboratory Consultation
- Data Management Consultation (best practices, project setup, data storage, archiving options, etc.)
- Data Management Plan Advice and Consultation
- Project Management with Open Science Framework
- Assistance with Data Sharing and Archiving
- Clinical Data Management using REDCap

Data Management Plan LibGuide

The Data Management Plan LibGuide is a guide created by the RDS unit offering the most up-to-date information on best practices for managing research data. The guide provides practical tips, tools, and resources necessary for researchers to understand the importance of good data management and efficient approaches to proper data management.

DMPTool

The DMPTool is a free, open-source, online application that helps researchers create data management plans (DMPs). The DMPTool provides a click-through wizard for creating a DMP that complies with funder requirements. It also has direct links to funder websites, help text for answering questions, and data management best practices resources. Templates for data management plans are based on the specific requirements listed in funder policy documents. The DMPTool maintains these templates, however, researchers should always consult the program officers and policy documents directly for authoritative guidance. Sample plans are provided by a funder or another trusted party. Here is an example template provided by the DMPTool.

Access DMPTool from this link. Once there enter your UC email address to be directed to the UC single sign-on access.

University of Cincinnati Tools and Resources (Data Repositories)

Because the new NIH Data Management and sharing policy (and some publishers) require sharing of scientific data, researchers need to understand the available options and requirements to do so. In general, the NIH does not endorse or require sharing data in any particular repository, although some initiatives and funding opportunities do have individual requirements. Overall, NIH encourages researchers to select the repository that is most appropriate for their data type and discipline. However, for *some* programs and types of data, NIH and/or Institute, Center, Office (ICO) policy(ies) and Funding Opportunity Announcements (FOAs) identify particular data repositories (or sets of repositories) to be used to preserve and share data. For data generated from research subject to such policies or funded under such FOAs, researchers should use the designated data repository(ies). For data generated from research for which no data repository is specified by the NIH, researchers are encouraged to select a data repository that is appropriate for the data generated from the research project.

To ensure compliance, researchers must understand the data-sharing requirements of the opportunity for which they are applying. This information is often specified within the FOA, and/or through the Institute, Center, or Office of the specific funding body. Some have requirements that the researcher must comply with. For instance, all studies generating human genomic data that fall within the scope of the NIH Genomic Data Sharing policy and must first be registered in the Database of Genotypes and Phenotypes (DbGaP—NIH's central repository for human genomic and associated phenotypic data) — even if the data will be submitted elsewhere. In this instance, although no specific repository is required, any studies that generate human genomic data must register with DbGaP. However, because this data likely contains sensitive information that must be protected, researchers should ensure deposition into a repository where proper data security measures are in place.

When no specific repository is designated by the NIH, researchers will want to carefully consider their target audience and submit to a repository that is targeted to this audience. This ensures that the data is findable to those that have the greatest need for it. When searching for a data repository, researchers should consider the following[8].

- Does the repository provide globally unique and persistent identifiers (e.g., DOIs and global unique identifiers, or GUIDs) so data can be cited?
- Can the repository be publicly searchable so data can be found?
- Does the repository allow for licenses that clarify how data can be reused?
- Does the repository allow for rich metadata descriptions so data are understandable and reusable?

When researchers are deposition of datasets into a repository, each submission should at minimum include the following:

- **Complete Metadata** - Metadata aids Discovery and Reuse. Scholar@UC requires metadata such as the title of the work submitted, creator name, data submitted, and description. All text in Scholar is searchable. Researchers should write a detailed descriptions to increase the discoverability of their content. Additional metadata fields can be found under the "Show Additional Description" link. There, researchers can add additional fields such as subject terms for search enhancement. If a discipline-specific schema/standard is being followed, this should be indicated in the description or the next document – the README file.

- **README.txt File** - This is a text document that provides relevant information such as the purpose of the project and the organizational structure or relationship of the files. It explains terms that are unique to the dataset, keywords, omissions, and errors. If you are using a file naming convention, you can explain it in the readme file. It is also the place to put additional details that were not included in the metadata, such as additional information about external storage of the data, metadata schema followed, and researcher contact information. An example readme file can be found [here][9].

- **Data Dictionary/Codebook** - This document explains all the variables and abbreviations associated with the dataset.

Below are listed data repositories available to UC researchers at no cost through their institutional affiliation. In general, these repositories provide mechanisms to publicly share research data, providing ways to meet data-sharing requirements.

ICPSR (Discipline focused Data Archive and Repository)

Inter-university Consortium for Political and Social Research (ICPSR) provides leadership and training in data access, curation, and methods of analysis for the social science research community. It is an international consortium of more

---

[8] Courtney K. Soderberg, "Using OSF to Share Data: A Step-by-Step Guide," *Advances in Methods and Practices in Psychological Science* 1, no. 1 (March 1, 2018): 115–20, https://doi.org/10.1177/2515245918757689.

[9] "README.Txt » Data Ab Initio," accessed January 27, 2023, http://dataabinitio.com/?p=378.

than 750 academic institutions and research organizations that maintains a data archive of more than 250,000 files of research in the social and behavioral sciences. ICPSR hosts 21 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields. ICPSR collaborates with several funders, including U.S. statistical agencies and foundations, to create thematic data collections and data stewardship and research projects.

ICPSR also operates the Health and Medical Care Archive (HMCA) which is the exclusive data archive of the Robert Wood Johnson Foundation (RWJF), the largest philanthropy devoted exclusively to health and health care in the United States. HMCA preserves and disseminates data collected by selected research projects funded by RWJF and facilitates secondary analyses of the data. The data collections in HMCA primarily include large-scale surveys of the American public about public health, attitudes towards health reform, and access to medical care; surveys of health care professionals and organizations, public health professionals, and nurses; evaluations of innovative programs for the delivery of health care, and many other topics and populations of interest.

The University of Cincinnati is an ICPSR member institution, and everyone affiliated with a member institution receives full access to the data archive of more than 4,000 research studies. In addition, ICPSR data users get:

- Convenience in the form of powerful search tools, useful metadata, easy access, and curation services all in one place.
- Data enhancements, including SAS, SPSS, Stata, and R files to facilitate analysis.
- Online analysis tools for hundreds of studies.
- Technical help in using data.
- Teaching tools featuring Teaching & Learning with ICPSR provide dozens of online Data-Driven Learning Guides to enhance the teaching of core concepts in the social sciences.

As a data repository, ICPSR offers two modes of data deposition. The first is self-publication via openICPSR which is very similar to what researchers get with Scholar@UC. OpenICPSR is a no-cost, self-publishing repository for social, behavioral, and health sciences research data that is particularly well-suited for researchers who need to publish data associated with a journal article to advance scientific transparency, meet funder requirements for data sharing, and allows other researchers to replicate their findings. Projects published are available immediately with both a citation and persistent identifier. More information about OpenICPSR can be found here.

The other mode of ICPSR data deposition comes as more of a service to the UC researcher. It too is free (as an institutional member), but data would have to be accepted into the archive to undergo the curation process. Once curated, the researcher receives a secure, curated data record. To access the service through ICPSR, researchers would need to submit a form that will help to inform the consortium of the amount, types, and file formats of the data contained in the dataset. If the dataset is selected by ICPSR, it will be cleaned, indexed, saved in interoperable file formats, and maintained in a safe, compliant manner by those at ICPSR. More information about the data curation services provided by ICPSR can be found here. The curation process greatly facilitates the reuse and reproducibility of data. However, both openICPSR and a curated data deposit via ICPSR will help you make data publicly available and meet funder requirements.

Open Science Framework (A General Repository Option)

Open Science Framework (OSF) is a free and open-source project management tool that supports researchers throughout their entire project lifecycle and collaborative research workflows. OSF helps research teams work on projects privately or make the entire project publicly accessible for broad dissemination. OSF allows researchers to manage files, data, code, and protocols in one centralized location and easily build custom organizations for their project. As a collaboration tool, researchers can manage which parts of their project they would like to make public, making it easy to collaborate and share with the community or just a research team. As a project management tool, OSF supplies a project dashboard, project logs to document work on the project, version control, and project analytics that help researchers measure the impact of their work by providing data on how many people are accessing and downloading their research materials. OSF is designed to help researchers collaboratively manage, store, and share their research process and files related to their

research. Although many repositories are simply archival in purpose, OSF also allows researchers to store and interact with files during the research process and to preregister their work and upload preprints if they so desire.

Explore the Open Science Framework here.

Scholar@UC (An Institutional option - UC's own Digital Repository)

Scholar@UC is a digital repository that enables the University of Cincinnati community to share research and scholarly work with a worldwide audience. Faculty and staff can use Scholar@UC to collect work in one location and create a durable and citeable record of papers, presentations, publications, datasets, or other scholarly creations. Students, through an approved process, may contribute capstone projects such as senior design projects, theses, and dissertations. Scholar is an open-source, agile development project that is supported by the University of Cincinnati Libraries and UCit.

The mission of Scholar@UC is to preserve the permanent intellectual output of UC, to advance discovery and innovation, to foster scholarship and learning through the transformation of data into knowledge, to collect a corpus of works that can be used for teaching, and to inspire derivative works, and to enhance discoverability and access to these resources. Scholar@UC allows researchers to publish their work in a single location and make unpublished output available. When archiving with Scholar, data will no longer reside solely on inaccessible computers or services, but it will become searchable and accessible to the world, enhancing the global impact and recognition of scholarly work while also contributing to the intellectual output of the University of Cincinnati.

Because Scholar is findable by harvesters and search engines, deposition into Scholar may help to meet grant and publisher requirements on data sharing. Scholar can ingest content including, pre-, post-, and published versions of manuscripts, senior design projects, theses/dissertations, posters, datasets, presentations, and other types of creative works. Content can be in many different forms, including text, image, video, audio, or mixed-media in nature. With Scholar, researchers can make their research private until they are ready to publish, embargo content until a specific date, limit access to a group of specific UC users, the UC community, or make their content available worldwide. Once in Scholar, users can have their submission assigned a globally unique and persistent identifier (DOI). Scholar@UC has a full-service API, search indexing, and Linked Open Data metadata that makes content findable by harvesters and search engines. Scholar does have some limitations, specifically, Scholar is not approved for protected data. Therefore, researchers needing to store and archive data containing personal identifiers should not be housed in the repository. Also, Scholar has a 3GB file size limitation, so if researchers have larger files, a member of the Scholar team can work with them to deposit the data into Scholar.

Scholar@UC can be accessed via this link, and members of the Scholar Team can be reached via this email address: scholar@ucmail.uc.edu.

Conclusions

Although the NIH Data Management and Sharing Policy may have taken some researchers by surprise, researchers should not feel ill-prepared for their grant submissions. In this document we have provided a summary and detailed information on the new plan and resources at UC and beyond to assist with compliance. Although this document is intended to be comprehensive in nature, it is not exhaustive. We encourage all researchers with further questions to contact RDS (askdata@uc.edu) or their program officer for more detailed information on their specific submission.