NIH Data Management and Sharing Policy: Compliance & Solutions

Josh Eckels CTO, LabKey





LabKey: Trusted by more than 500 Global Research Organizations







The NIH Data Management and Sharing (DMS) Policy creates a consistent minimum expectation for all research supported by the agency.

Under the DMS policy, NIH requires that new funding applications or renewals after Jan 25th 2023 will:

- Plan and budget for the managing and sharing of data
- Submit a DMS plan for review when applying for funding
- Comply with the approved DMS plan



Scope: Applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of "scientific data".

- Research Projects
- Certain Career Development Awards (Ks)
- Small Business SBIR/STTR
- Research Centers





"Scientific data" is defined 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."

Length of Time to Maintain Data:

- Grantee institutions are required to keep the data for 3 years following closeout of a grant or contract agreement. Contracts may specify different time periods.
- Grantee institutions may have additional policies and procedures regarding the custody, distribution, and required retention period for data produced under research awards.



Scientific data does not include:

- Data not necessary for or of sufficient quality to validate and replicate research findings
- Laboratory notebooks
- Preliminary analyses
- Completed case report forms
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues
- Physical objects, (e.g., laboratory specimens)



Justifiable reason for limiting sharing include:

- Informed consent will not permit or limits scope of sharing or use
- Privacy or safety of research participants would be compromised and available protections insufficient
- Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- Restrictions imposed by existing or anticipated agreements with other parties

Reasons <u>NOT</u> generally justifiable to limit data sharing:

- Data are considered too small
- Researchers anticipate data will not be widely used
- Data are not thought to have a suitable repository

Creating a DMS Plan

- Target length is two pages
- Almost all NIH-provided samples are three pages
- No specific format required
- NIH-provided template is available
 - Twelve prompts/sections
 - Public reporting burden for this collection of information is estimated to average 2 hours per response
- Many repositories and other resources are providing sample responses





Data Type:

- Summarize the types and amount of scientific data to be generated and/or used in the research. Descriptions may include the data modality, level of aggregation, and/or the degree of data processing.
- Describe which scientific data from the project will be preserved and shared. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors. The plan should provide the reasoning for these decisions.
- A brief listing of the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data



Related Tools, Software and/or Code:

Indicate whether specialized tools are needed to access or manipulate shared scientific data to support replication or reuse, and name(s) of the needed tool(s) and software. If applicable, specify how needed tools can be accessed.

Standards:

Describe what standards, if any, will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation).



Data Preservation, Access and Associated Timelines:

Give plans and timelines for data preservation and access, including:

- The name of the repository(ies) where scientific data and metadata arising from the project will be archived.
- How the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.
- When the scientific data will be made available to other users and for how long. Identify any differences in timelines for different subsets of scientific data to be shared.



Access, Distribution, or Reuse Considerations:

Applicable factors affecting access, distribution, or reuse of scientific data.

- Informed consent
- Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local laws, regulations, and policies
- Whether access to scientific data derived from humans will be controlled
- Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements
- Any other considerations that may limit the extent of data sharing.
- Any potential limitations on subsequent data use should be communicated to the individuals or entities (for example, data repository managers) that will preserve and share the scientific data.





Oversight of Data Management and Sharing:

Indicate how compliance with the DMS Plan will be monitored and managed, the frequency of oversight, and by whom (e.g., title, roles).

This element refers to oversight by the funded institution, rather than by NIH. The DMS Policy does not create any expectations about who will be responsible for Plan oversight at the institution.

Allowable costs include:

- Curating, de-identifying, annotating, and documenting data
- Local data management solutions
- Fees associated with depositing into a repository
- All costs must be incurred during performance period
- Capture in budget's appropriate cost categories, and include justification

Unallowed costs include:

- Costs already covered through institutional overhead
- Double-charging as direct and indirect costs

Proper data management helps ensure that researchers and collaborators are using data consistently and accurately.

- Carefully storing and documenting data allows more people to use the data in the future, potentially leading to more discoveries beyond the initial research.
- NIH emphasizes the importance of good data management practices and encourages data management to be reflective of practices within specific research communities.
- NIH encourages data management and sharing practices to be consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

- Encourage data management throughout the research process
- Well-established tools, support, and recommendations

Metadata and Other Associated Documentation

Metadata and other documentation associated with a dataset allow users to understand how the data was collected and how to interpret the data. Importantly, this ensures that others can use the dataset and prevents misuse, misinterpretation, and confusion.

Examples of metadata or other information that may be included with research data:

- Methodology and procedures used to collect the data
- Data labels
- Definitions of variables
- Any other information necessary to reproduce and understand the data

- Capture structured metadata about samples, assays, subjects, etc
- Use standard or custom ontologies and simple value lists



Naming Conventions

Within a project team, agreement on naming conventions for multiple objects or files—or multiple versions of files—could be useful before embarking on a project that generates large amounts of data that need names or unique identifiers.

- Add data validation and quality control flagging
- Automatically route data based on file names and other conventions
- Assign sample, subject, and other identifiers



Common Data Elements

Common data elements (CDEs) are pieces of data common to multiple datasets across different studies. NIH encourages researchers to use CDEs, which helps improve accuracy, consistency, and interoperability among datasets within various areas of health and disease research.

- Integrate disparate data sources
- Connect to existing systems like REDCap, OpenEMPI, and relational databases



Data Storage Format

There are many storage formats for different types and sizes of datasets. For instance, small and simple datasets can be managed in spreadsheets while more complicated or larger datasets may need to be managed in a database. Remember that some types of data storage incur costs, which may be part of the project budget.

- Capture spreadsheets and other simple tabular data
- Import specialized file formats in a database
- Audit log history and security controls
- Consistently analyze data via R scripts, Jupyter notebooks, SQL queries, and other coding tools



NIH encourages use of established repositories.

Depositing data in a quality repository generally improves the FAIRness of data – Findable, Accessible, Interoperable, Reusable

- For data generated from research for which no data repository is specified by NIH, researchers are encouraged to select a data repository that is appropriate for the data generated from the research project.
- Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse.





- NIH isn't endorsing specific repositories, but does have a list of examples
- Consider whether to select one or multiple
- Leverage ones you're already using (due to journal requirements, etc)
- Think about where you look for public datasets
- Don't plan to establish your own repository without a good reason

HOW LABKEY HELPS:



Targeted mass spectrometry repository for Skyline users



Clinical trial results from Immune Tolerance Network



Data management and analysis of human immunological data



Real-time Zika and Covid data sharing

- Applies to all new NIH funding application and renewals
- Plan and budget for the managing and sharing of data using FAIR standards
- Create and submit a DMS plan for review when applying for funding
- Ensure you have the systems and resources in place to comply with your plan



LabKey has helped researchers and institutions manage scientific data for more than 20 years.

Get in touch to see how our solutions can help you!



