

## Proposed Provisions for a Draft NIH Data Management and Sharing Policy

NIH has a longstanding commitment to making the results and accomplishments of the research that it funds and conducts available to the public.<sup>1</sup> Increasing access to scientific data resulting from NIH funding or support offers many benefits, and reflects NIH's responsibility to maintain stewardship over taxpayer funds. Specifically, sharing of scientific data and results enables researchers to more vigorously test the validity of research findings, strengthen analyses by combining data sets, access hard-to-generate data, and explore new frontiers. Data sharing also informs future research pathways, increases the return on investment of scientific research funding, and accelerates the translation of research results into knowledge, products, and procedures to improve health and prevent disease. Effective data sharing practices rely upon appropriate identification, adoption, and crediting of good data management and sharing practices, thus, NIH encourages data sharing consistent with the FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles.<sup>2</sup>

NIH is invested in ensuring that scientific data are managed, preserved, and made accessible in a timely manner for appropriate use by the research community and the public, as outlined in the NIH Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research.<sup>3</sup> In accordance with this plan, NIH is implementing measures to update its 2003 Data Sharing Policy,<sup>4</sup> consistent with Administration priorities<sup>5</sup> and other authorities granted to NIH, such as under the 21<sup>st</sup> Century Cures Act.<sup>6</sup> Taking these factors into consideration, this draft proposal outlines key provisions that could serve as a foundation for a new NIH data management and sharing policy, including potential scope and proposed review requirements for NIH-supported research.

### Proposed Provisions

#### I. Definitions

The following definitions related to data management and sharing are proposed for consideration:

- **Data Management and Sharing Plan:** A plan describing how scientific data will be managed, preserved, shared, and made accessible to others (e.g., other researchers and the broader public), as appropriate.
- **Data Sharing:** To make scientific data accessible for use by others (e.g., other researchers and the broader public) in a manner that is consistent with the FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles.
- **Metadata:** Data that provide additional information to make data more usable (e.g., independent sample and variable description, outcome measures, and any intermediate, descriptive, or phenotypic observational variables).
- **Scientific Data:** The recorded factual material commonly accepted in the scientific community as necessary to validate and replicate research findings including, but not limited to, data used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens. For the purposes of a possible Policy, **scientific data** may include certain individual

level and summary or aggregate data, as well as **metadata**.<sup>7</sup> NIH expects that reasonable efforts should be made to digitize all scientific data.

## II. Purpose

**Scientific data** are a valuable output of NIH-funded or -supported research, and they should be managed, preserved, and made accessible in a timely manner for appropriate use by the research community and the broader public. Any new NIH policy for managing and sharing data would establish:

- Requirements for responsible management and sharing of **scientific data** resulting from NIH-funded or -supported research; and
- Expectations for other NIH Policies related to the management and sharing of **scientific data** such as the NIH Genomic Data Sharing Policy,<sup>8</sup> the NIH Policy on the Dissemination of NIH-funded Clinical Trial Information,<sup>12</sup> and the Intramural Research Program Human Data Sharing (HDS) Policy.<sup>9</sup>

Any new NIH policy would be evaluated at appropriate intervals and adjusted as needed to ensure the stated purpose is achieved.

## III. Scope and Requirements

A new NIH data management and sharing policy would apply to all intramural and extramural research, funded or supported in whole or in part by NIH, that results in **scientific data**, regardless of NIH funding level or mechanism. This includes research funded or supported by grants (including cooperative agreements), contracts, Other Transactions, intramural support mechanisms, Cooperative Research and Development Agreements (CRADAs), or other funding agreements. Components could include requirements for:

- Submission of a **Data Management and Sharing Plan (Plan)** as part of the funding/support application process, proposal, CRADA, or other funding agreement, or intramural research report. If there are perceived barriers to sharing **scientific data** (e.g., sharing includes specific restrictions or sharing all data is not possible), the **Plan** would outline how **scientific data** will be managed and preserved and include an explanation of the perceived barriers; and
- Sharing and managing the **scientific data** resulting from NIH-funded or -supported research according to the NIH Institute or Center-approved **data management and sharing plan**.

The funding or supporting NIH Institute or Center (IC) could request that additional data be included within the Plan, e.g., establish certain minimum requirements for data management and sharing to support programmatic priorities or expand the utility of the **scientific data** generated in the research. Reasonable costs associated with data management and sharing could be requested under the budget for the proposed project.<sup>10</sup>

#### IV. Requirements for Data Management and Sharing Plans

Applications and proposals for NIH-funded or -supported research projects that result in **scientific data** would be required to include a **Plan**. If perceived barriers to sharing **scientific data** exist (e.g., sharing includes specific restrictions or sharing is not possible), the **Plan** would be required to outline how **scientific data** will be managed and preserved and include an explanation of the perceived barriers. The **Plan** would also need to identify strategies or approaches to ensure adequate data security and compliance with privacy protections.

*Plan Review and Evaluation:* The funding or supporting NIH IC would consider the evaluation and determine the acceptability of the **Plan**, which could be implemented in a variety of ways, including:

- Extramural Grants: **Plans** could be evaluated as an Additional Review Consideration,<sup>11</sup> i.e., evaluated as acceptable or unacceptable by reviewers, but not be factored into the overall impact score through the peer review process. This allows for NIH staff to work with potential awardees to ensure that any reviewer concerns regarding the Plan could be addressed for meritorious applications as a contingency of NIH funding. Plan compliance would be integrated into terms and conditions as appropriate.
- Contracts: **Plans** could be included as part of the technical evaluation performed by NIH staff and incorporated in the subsequent terms of the contract.
- NIH Intramural Research Projects: **Plans** could be reviewed by the Scientific Director (or designee) or Clinical Director (or designee) of the researcher's funding IC and integrated into approval conditions as appropriate.
- Other funding/support agreements: **Plans** could be evaluated in the context of other funding/support agreement mechanisms, e.g., CRADA, Other Transactions, and integrated into the terms and conditions as appropriate.

*Plan Elements:* **Plans** could have a two-page limit and address the following research elements: (i) data types, (ii) related tools and software, (iii) data standards, (iv) data preservation, access (including timelines) and discoverability, (v) terms for re-use and redistribution, (vi) limitations on access, and (vii) oversight of data management. Examples of guidance about how these **Plans** could be implemented are included below:

1. Data Type: Indicate the types and estimated amount of **scientific data** that will result from NIH-funded or -supported research and indicate the rationale for which **scientific data** will be preserved and shared.
  - 1.1. Describe in general terms the type and amount of **scientific data** to be collected and used in the project (e.g., exome sequences of 22 gene variants from 790 cases and fMRI data from 100 research participants). Describe the data modality (e.g., imaging, genomic, mobile, and survey) and whether the **scientific data** will be individual, aggregated, or summarized, and how raw or processed the data will be.

- 1.2. Describe any other information that is anticipated to be shared along with the **scientific data**, such as relevant associated data, and any other information necessary to interpret the data (e.g., study protocols and data collection instruments).
  - 1.3. For **scientific data** derived from human participants or specimens, if appropriate, describe plans for protecting privacy and confidentiality, e.g., through de-identification or data aggregation prior to sharing.
2. Related Tools, Software and/or Code: Indicate what software/computer code will be used to process or analyze the **scientific data** (the inclusion of scripts may be helpful), why the software/code was chosen, and whether it is free and open source. If software/code that is not free and open source is needed to access or further analyze the **scientific data**, briefly describe why this particular software/code is needed. Describe whether there is an alternative free and open source software/code that may be used to further analyze the **scientific data**.
3. Standards: Indicate what standards, if any, apply to the **scientific data** to be collected, including data formats, data identifiers, definitions, and other data documentation, including terms of use. NIH encourages the use of existing data standards, such as standards for collecting and representing **scientific data** and information describing the **scientific data**. NIH encourages the use of common data elements (CDEs) to facilitate broader and more effective use of **scientific data** and to advance research across studies. For assistance in identifying NIH-supported CDEs, the NIH has established a Common Data Element Resource Portal.<sup>12</sup>
4. Data Preservation and Access: Data preservation must be consistent with the NIH Strategic Plan for Data Science.<sup>13</sup> Additionally, examples of NIH-supported data repositories that make data accessible for reuse are available at the NIH Trans-NIH BioMedical Informatics Coordinating Committee website.<sup>14</sup>
  - 4.1. Indicate where **scientific data** will be archived to ensure its long-term preservation. If **scientific data** will be stored in an existing repository, provide the name and URL web address of the repository. If an existing repository will not be used, indicate why not and how **scientific data** preservation will be assured (e.g., in a newly created repository or by the investigator's organization).
  - 4.2. Indicate how the **scientific data** will be made discoverable and whether a persistent unique identifier or other standard indexing tools will be used.
  - 4.3. Describe any provisions for maintaining the security and integrity of the **scientific data** (e.g., encryption and backups).
  - 4.4. Describe alternative plans for maintaining, preserving, and providing access to **scientific data** should the original **Plan** not be achieved.

- 4.5. If perceived barriers to sharing **scientific data** exist (e.g., sharing includes specific restrictions or sharing is not possible), outline how **scientific data** will be managed and preserved and include an explanation of the perceived barriers.
  - 4.6. Indicate whether additional considerations are needed to implement the **Plan** (e.g., prior permission to use a specific repository).
  - 4.7. Indicate whether **scientific data** generated from humans or human biospecimens will be available through unrestricted (made publicly available to anyone) or restricted access (made available after the requestor has received approval to use the requested **scientific data** for a particular project or projects). If the **scientific data** will be shared through a restricted access mechanism, describe the terms of access for the data.
5. Data Preservation and Access Timeline: Provide information, if available, on the anticipated timeframes for **scientific data** storage and accessibility, and criteria for how decisions affecting **scientific data** storage and accessibility will be made throughout the course of the study. Timelines should include all of the **scientific data** types listed in element 1 (Data Type) and may be described in chronologic terms (e.g., Year 1) and/or in terms relating to the conduct of the research (e.g., relative to completion of particular specific aims or publications, value to other researchers). Timing may vary, based on a variety of factors, including – but not limited to – different **scientific data** types, specific programmatic priorities, and expected utility of the **scientific data**.
    - 5.1. Describe when the **scientific data** will be made available to secondary data users (e.g., researchers and the broader public).
    - 5.2. Describe the time period the **scientific data** will be made available to secondary data users with approval to use the requested data for a particular project (e.g., researchers and the broader public).
6. Data Sharing Agreements, Licensing, and Intellectual Property: Describe any provisions that will govern the appropriate reuse and redistribution of the **scientific data** while ensuring broadest sharing, as appropriate. NIH encourages terms that provide for the broadest use of data resulting from NIH-funded or -supported research, consistent with privacy, security, informed consent, and proprietary issues.
    - 6.1. Describe any relevant data sharing agreement(s), outlining the responsibilities of each party, as well as how **scientific data** can and cannot be used.
    - 6.2. Describe any general licensing terms, and any limitations on the **scientific data** use and reuse based on these terms. Describe whether the licensing is imposed by the applicant institution or whether it comes from any existing agreement(s).
    - 6.3. If applicable, indicate how intellectual property, including invention or other proprietary rights, will be managed in a way to maximize sharing of **scientific data**. Include any information relevant to the intellectual property rights associated with the

**scientific data**, such as whether the intellectual property stems from an existing agreement or is anticipated to arise from the proposed research project itself.

7. Oversight of Data Management: Indicate the individual(s) who will execute various components of the **Plan** over the course of the research project (e.g., data collection, data analysis, data management, data submission, data sharing, data distribution, and project completion).

Additionally, other considerations for any **Plan**, could include topics such as:

- **Scientific Data Archiving.** Investigators would be encouraged to consider using repositories that make **scientific data** available at no cost for extended periods of use. Timing may vary based on, e.g., different data types, specific programmatic priorities, and expected utility of the data. Investigators should include information, if available, on the timeframe of storage and accessibility of the **scientific data** and on how decisions will be made to stop storing the **scientific data** or change its level of accessibility.
- **Sharing of Scientific Data Resulting from Human Specimens or Samples.** Sharing **scientific data** derived from human specimens warrants special attention and should be done as broadly as possible, consistent with the consent of the individual study participants. Data may be shared across institutions and repositories to maximize utility, and informed consent should permit broad sharing wherever possible. Informed consent for collecting and sharing **scientific data** derived from human specimens must be consistent with Federal Human Subjects Regulations (e.g., the “Common Rule”<sup>15</sup>) and applicable NIH policies, such as human subject protection program policies.<sup>16</sup>

## V. Compliance and Enforcement

NIH encourages the sharing of data for as long as it is useful to the scientific community. Many repositories offer data preservation and access for free. The NIH could consider intramural- and extramural-funded or -supported institutions to have satisfied their obligations under a new policy if:

- The data sharing repository meets community-based standards at the time of deposition, OR
- The **scientific data** is posted to a federal government repository(ies) (i.e., .gov repositories), or a repository(ies) specified by NIH in funding opportunity announcements.

Responsibilities for monitoring and enforcing compliance would be undertaken by the NIH-funding or supporting IC, with consideration taken to encourage consistency and coordination across NIH as appropriate. Possible strategies could include:

### *During the Funding or Support Period*

Compliance with the **Plan** would be determined by the funding or -supporting NIH IC and reviewed at a minimum, annually [e.g., at the time of annual Research Performance Progress Reports (RPPRs)].

- Extramural Grants: The **Plan** would become a Term and Condition of the Notice of Award. Failure to comply with the award Terms and Conditions may result in an enforcement action, including additional special terms and conditions or termination of the award.<sup>17</sup>
- Contracts: The **Plan** would become a Term and Condition of the Award, and compliance with and enforcement of the **Plan** would be consistent with the award and the Federal Acquisition Regulations (FAR),<sup>18</sup> as applicable.
- NIH Intramural Research Projects: Compliance with and enforcement of the **Plan** would be consistent with applicable NIH policies established by the NIH Deputy Director for Intramural Research and the applicable NIH ICs.

*Post Funding or Support Period*

Non-compliance with the NIH-approved **Plan** would be taken into account by the funding IC for future funding or support decisions (e.g., as authorized in the NIH Grants Policy Statement, Section 8.5, Special Award Conditions and Enforcement Actions).

- 
- <sup>1</sup> See, for example: NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources <http://grants.nih.gov/grants/sharing.htm>; the NIH Intramural Research Program Human Data Sharing (HDS) Policy <https://oma1.od.nih.gov/manualchapters/intramural/3016/index.html>; the NIH Trans-NIH BioMedical Informatics Coordinating Committee website [http://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html)
- <sup>2</sup> Wilkinson, M., Dumontier, M. et al, The FAIR Guiding Principles for scientific data management and stewardship (2016), available at <https://www.nature.com/articles/sdata201618>
- <sup>3</sup> National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (February 2015) <https://grants.nih.gov/grants/nih-public-access-plan.pdf>
- <sup>4</sup> NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources [https://grants.nih.gov/grants/policy/data\\_sharing/](https://grants.nih.gov/grants/policy/data_sharing/)
- <sup>5</sup> Office of Science and Technology Policy, National Science and Technology Council: <https://www.whitehouse.gov/ostp/nstc/>
- <sup>6</sup> Public Law 14-255, 21<sup>st</sup> Century Cures Act (December 13, 2016) <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf>
- <sup>7</sup> NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information (September 16, 2016) <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-149.html>
- <sup>8</sup> National Institutes of Health Genomic Data Sharing Policy (August 27, 2014) [https://osp.od.nih.gov/wp-content/uploads/NIH\\_GDS\\_Policy.pdf](https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf)
- <sup>9</sup> The NIH Intramural Research Program Human Data Sharing (HDS) Policy: <https://policymanual.nih.gov/3016>
- <sup>10</sup> NIH Grants Policy Statement, Section 7 Cost Consideration: [https://grants.nih.gov/grants/policy/nihgps/HTML5/section\\_7/7.1\\_general.htm](https://grants.nih.gov/grants/policy/nihgps/HTML5/section_7/7.1_general.htm); NIH Data Sharing Policy and Implementation Guidance [http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm#inc](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#inc)
- <sup>11</sup> NIH Peer Review *Review Criteria at a Glance* (March 9, 2018) [https://grants.nih.gov/grants/peer/guidelines\\_general/Review\\_Criteria\\_at\\_a\\_glance.pdf](https://grants.nih.gov/grants/peer/guidelines_general/Review_Criteria_at_a_glance.pdf)
- <sup>12</sup> NIH Common Data Element (CDE) Resource Portal: <https://www.nlm.nih.gov/cde/>
- <sup>13</sup> NIH Strategic Plan for Data Science: [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf)
- <sup>14</sup> NIH Trans-NIH BioMedical Informatics Coordinating Committee website: [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html)
- <sup>15</sup> Title 45 – Public Welfare Part 46 Protection of Human Subjects (October 1, 2013): <https://www.gpo.gov/fdsys/pkg/CFR-2013-title45-vol1/xml/CFR-2013-title45-vol1-part46.xml>
- <sup>16</sup> NIH Resources on Research Involving Human Subjects: <https://humansubjects.nih.gov/>
- <sup>17</sup> NIH Grants Policy Statement, Administrative Requirements (October 2017) [https://grants.nih.gov/grants/policy/nihgps/HTML5/section\\_8/8.5\\_special\\_award\\_conditions\\_and\\_enforcement\\_actions.htm](https://grants.nih.gov/grants/policy/nihgps/HTML5/section_8/8.5_special_award_conditions_and_enforcement_actions.htm)
- <sup>18</sup> Federal Acquisition Regulation (August 22, 2018) <https://www.acquisition.gov/?q=browsefar>