



**BILLING CODE 4140-01-P**

**DEPARTMENT OF HEALTH AND HUMAN SERVICES**

**National Institutes of Health**

**Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance**

**AGENCY:** National Institutes of Health, HHS.

**ACTION:** Request for Comments.

**SUMMARY:** The National Institutes of Health (NIH) is seeking public comments on a DRAFT NIH Policy for Data Management and Sharing and supplemental DRAFT guidance. The purpose of this DRAFT Policy and supplemental DRAFT guidance is to promote effective and efficient data management and sharing to further NIH's commitment to making the results and accomplishments of the research it funds and conducts available to the public.

**DATES:** To ensure that your comments will be considered, please submit your response to this Request for Comments no later than January 10, 2020.

**ADDRESSES:** Comments may be submitted online at: <https://osp.od.nih.gov/draft-data-sharing-and-management>.

**FOR FURTHER INFORMATION CONTACT:** Andrea Jackson-Dipina, Dr.PH,  
Director of the Division of Scientific Data Sharing Policy, Office of Science Policy, NIH,  
6705 Rockledge Drive, Suite 750, Bethesda, MD 20892, 301-496-9838,  
jacksondipinaac@od.nih.gov.

**SUPPLEMENTARY INFORMATION:**

**Background**

In compliance with the requirement of Section 3506(c)(2)(A) of the Paperwork Reduction Act of 1995, which provides opportunity for public comment on proposed projects, we do not see this information collection as sensitive or controversial in nature, as the information collection will enable continued Policy for Data Management and Sharing allowing the research community to more effectively continue their research and serve the public. NIH's mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability. Sharing scientific data advances this mission by enhancing NIH's stewardship of taxpayer funds and maximizing research participants' contributions. Moreover, increasing access to scientific data resulting from NIH-funded or conducted research advances biomedical research by enabling the validation of scientific results, allowing analyses to be strengthened by combining data, facilitating reuse of hard-to-generate data, and accelerating future research.

NIH has a long history of making the products of Federally-funded research available to the public. For example, in 2003, NIH released its first *NIH Data Sharing Policy* to set

the expectation that final research data would be shared from awards requesting \$500,000 or more in direct costs in any single year. The *NIH Public Access Policy*, which applies to manuscripts accepted for publication after April 7, 2008, ensures that the public has access to the published results of NIH-funded or conducted research by requiring NIH researchers to submit final peer-reviewed journal manuscripts to PubMed Central. NIH also has implemented policies to facilitate sharing of certain high-value data-types, such as the 2007 *NIH Genome-Wide Association Studies Policy* and the 2014 *NIH Genomic Data Sharing Policy*, establishing expectations for sharing large-scale genomic data resulting from NIH-funded or conducted studies. To maximize critical investments in clinical research, NIH has established policies specific to sharing clinical research data. Most recently in 2016, NIH issued the *NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information* which sets forth the expectation that NIH-funded or conducted clinical trials will be registered and have summary results information submitted to *ClinicalTrials.gov*, complementing the *HHS Final Rule for Clinical Trials Registration and Results Information Submission*.

Through this Notice, NIH is seeking public input on a trans-NIH data management and sharing policy proposal that further advances the Agency's commitment to responsible data management and sharing. Of note, NIH first announced its intent to encourage broad data sharing in 2015 with the release of the *NIH Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research* and further stakeholder input was sought via the 2017 *Request for Information: Strategies for NIH Data Management, Sharing, and Citation*. NIH shared its initial proposed policy

provisions for a future draft data management and sharing policy in 2018 through its *Request for Information on Proposed Provisions of a Draft Data Management and Sharing Policy for NIH Funded or Supported Research*. In response to the 2018 Request for Information, NIH received a total of 183 submissions from both national and international stakeholders, the majority of whom described themselves as scientific researchers or institutional officials from a variety of organizational affiliations and areas of research interest. Most respondents strongly supported data sharing and the concept of defining “scientific data” as encompassing the data and metadata needed to replicate and validate research findings. Additionally, respondents generally agreed that researchers should prospectively outline where, when, and how scientific data resulting from NIH-funded or conducted research will be managed and shared while allowing for data sharing exceptions, when justified. Many respondents expressed concerns about varying expectations across diverse scientific domains, the NIH Institutes, Centers, and Offices (ICOs), and Federal agencies, in addition to concerns of potential burden on the research community.

Public comments received from these Requests for Information, coupled with engagement efforts and lessons learned from other Federal agencies’ data sharing policies, were considered in crafting an NIH-wide data management and sharing policy proposal. After thorough review and consideration of stakeholder input, NIH developed the current DRAFT NIH Policy for Data Management and Sharing (herein referred to as “DRAFT Policy”) for public input which, when finalized and effective, would apply to all NIH-funded or conducted research generating scientific data, regardless of data type,

size, or the requested amount of funding. NIH recognizes that while all scientific data need to be managed, not all data generated in the course of research may be necessary to validate and replicate research findings. Therefore, this DRAFT Policy proposes that applicants **submit a plan** outlining how scientific data are to be managed and shared. Importantly, the proposed DRAFT Policy allows for flexibility across various scientific domains by outlining minimum expectations for NIH-wide Data Management and Sharing Plans (Plans), on which NIH ICOs may build. This DRAFT Policy also proposes that Plans could be submitted at “Just-In-Time” and reviewed by NIH program staff, which reduces applicant burden because only those applicants likely to be funded would submit Plans. This approach may facilitate consistent evaluation across NIH ICOs as well as throughout the lifetime of the award, during which updates to Plans may be made.

Paramount to this DRAFT Policy is the incorporation of principles that respect the autonomy and privacy of research participants and protection of confidential data. Thus, in the Data Management and Sharing Plan, researchers can describe practices for responsible management and sharing of sensitive scientific data, such as those from human participants (i.e., through de-identification or other protective measures), including when there should be exceptions to sharing or only limited sharing of data. These considerations are particularly germane when working with small or underserved populations. For instance, NIH recognizes that sovereign Tribal Nations may have unique data sharing concerns and the Agency has engaged these communities through Tribal Consultation sessions across the U.S. to consider their potential needs in the formation of this DRAFT Policy. NIH intends to continue conversations with Tribal

Nations to develop culturally sensitive data management and sharing resources for researchers seeking to collaborate with Tribal Nations. NIH encourages comments on specific strategies for promoting responsible data management and sharing in these types of research settings, including identification of areas in which further guidance may be needed.

NIH recognizes that the deliberate flexibility of its DRAFT Policy may require additional implementation guidance. It is important to acknowledge that NIH recognizes that expectations for robust data management and sharing practices will need to be met with investments in and evolution of accompanying data infrastructure. As indicated in the *NIH Strategic Plan for Data Science*, NIH's policy development efforts are being considered in tandem with its efforts to modernize the data infrastructure ecosystem. Thus, NIH also seeks feedback on proposals for supplemental DRAFT guidance documents intended to help researchers prospectively integrate Data Management and Sharing Plans into routine research practices. The supplemental DRAFT guidance: *Allowable Costs for Data Management and Sharing* (see below) proposes the types of costs that could be considered for inclusion in a research proposal to support data sharing activities. The supplemental DRAFT guidance: *Elements of An NIH Data Management and Sharing Plan* (see below) proposes a framework by which applicants could structure Data Management and Sharing Plans, including descriptions of elements such as the data type(s), standards employed, and timelines for data sharing. NIH encourages feedback on the utility of these supplemental DRAFT guidance documents and welcomes suggestions for any additional guidance that may be helpful to the community.

Substantive input is needed to ensure future policy decisions facilitate tangible and effective data management and sharing strategies. In this Request for Comment, NIH seeks public input on its proposed DRAFT NIH Policy for Data Management and Sharing and supplemental DRAFT guidance documents, including ways to promote access to research findings while minimizing burden on the research community. Feedback obtained through this Notice and other outreach efforts will help inform a final NIH Policy for Data Management and Sharing, which upon the effective date, would replace the 2003 *NIH Data Sharing Policy*.

**Request for Comments:**

NIH encourages the public to provide comments on any aspect of the DRAFT Policy and supplemental DRAFT guidance, described below.

- I. DRAFT NIH Policy for Data Management and Sharing,
- II. Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing, and
- III. Supplemental DRAFT Guidance: Elements of An NIH Data Management and Sharing Plan.

**Submitting a Response:**

Comments should be submitted electronically to the following webpage: <https://osp.od.nih.gov/draft-data-sharing-and-management> by January 10, 2020. Unedited comments will be compiled and may be posted, along with the

submitter's name and affiliation, on the NIH Office of Science Policy web site after the public comment period closes. Submitted comments are considered public information. Please do not include any proprietary, classified, confidential, or sensitive information in your response.

## **DRAFT NIH Policy for Data Management and Sharing**

### **I. Purpose**

The NIH Policy for Data Management and Sharing (herein referred to as the Policy) reinforces NIH's longstanding commitment to making the results and outputs of the research that it funds and conducts available to the public. Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery. In addition, NIH emphasizes the importance of good data management practices, which provide the foundation for effective data sharing and improve the reproducibility and reliability of research findings. NIH encourages data management and data sharing practices consistent with the NIH Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research and the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

To promote effective and efficient data management and data sharing, NIH expects researchers to manage scientific data resulting from NIH-funded or conducted research and prospectively plan for which scientific data will be preserved and shared. Under this Policy, individuals and entities would be required to provide a Data Management and

Sharing Plan (Plan) describing how scientific data will be managed, including when and where the scientific data will be preserved and shared, prior to initiating the research study. Shared data should be made accessible in a timely manner for use by the research community and the broader public. This Policy is intended to establish expectations for Data Management and Sharing Plans upon which other NIH Institutes, Centers and Offices (ICO) may supplement as appropriate.

## II. Definitions

For the purposes of this Policy, terms are defined as follows:

- **Data Management and Sharing Plan (Plan):** A plan describing how scientific data will be managed, preserved, and shared with others (e.g., researchers, institutions, the broader public), as appropriate.
- **Data Management:** The process of validating, organizing, securing, maintaining, and processing scientific data, and of determining which scientific data to preserve.
- **Data Sharing:** The act of making scientific data available for use by others (e.g., researchers, institutions, the broader public).
- **Metadata:** Data describing scientific data that provide additional information to make such scientific data more understandable (e.g., date, 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,

regardless of whether the data are 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. NIH expects that reasonable efforts will be made to digitize all scientific data.

### **III. Scope**

This Policy applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data. This includes research funded or conducted by extramural grants, contracts, intramural research projects, or other funding agreements regardless of NIH funding level or funding mechanism.

### **IV. Effective Date (s)**

The effective date of this Policy and subsequent implementation deadlines are dependent upon feedback on this proposal. This Policy is proposed to be effective for NIH-funded or conducted research, including:

- Competing grant applications that are submitted to NIH for a future receipt date or subsequent receipt dates (date yet to be determined);
- Proposals for contracts that are submitted to NIH on or after a future date (date yet to be determined);
- NIH Intramural research conducted on or after a future date (date yet to be determined); and

- Other funding agreements (e.g., Other Transactions) that are executed on or after a future date (date yet to be determined), unless otherwise stipulated by NIH.

## **V. Requirements**

This Policy would require:

- Submission of a Data Management and Sharing Plan (Plan) outlining how scientific data will be managed and shared, taking into account any potential restrictions or limitations.
- Compliance with the NIH ICO-approved Plan, prospectively describing effective management and timely sharing of scientific data (as appropriate) and accompanying metadata resulting from NIH-funded or conducted research.

The funding NIH ICO may request additional or specific information to be included within the Plan in order to meet expectations for data management and data sharing in support of programmatic priorities or to expand the utility of the scientific data generated from the research. Costs associated with data management and data sharing may be allowable under the budget for the proposed project (see below, Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing).

## **VI. Data Management and Sharing Plans**

Researchers with NIH-funded or conducted research projects resulting in the generation of scientific data are required to submit a Plan to the funding NIH ICO as part of Just-in-

Time for extramural awards, as part of the technical evaluation for contracts, as part of the NIH Intramural Annual Report, or prior to release of funds for other funding agreements. Plans should explain how scientific data generated by a research study will be managed and which of these scientific data will be shared. Plans may be updated by researchers (with appropriate NIH ICO approval) during regular reporting intervals if changes are necessary or at the request of the NIH ICO to reflect changes in the previously documented approach to data management and data sharing throughout the research project, as appropriate. NIH encourages shared scientific data to be made available as long as it is deemed useful to the research community or the public. Plans should also identify strategies or approaches to ensure data security and compliance with privacy protections are in place throughout the life of the scientific data. NIH may make Plans publicly available.

NIH prioritizes the responsible management and sharing of scientific data derived from human participants. Applicable Federal, Tribal, state, and local laws, regulations, statutes, guidance, and institutional policies dictate how research involving human participants should be conducted and how the scientific data derived from human participants should be used. Researchers proposing to generate scientific data derived from human participants should outline in their Plans how human participants' privacy, rights, and confidentiality will be protected, i.e., through de-identification or other protective measures. NIH recognizes that certain factors (e.g., legal, ethical, technical) may limit the ability to preserve and share data. Plans should include consideration of these factors, when applicable, in describing the approach to data management and data

sharing. NIH encourages the use of established repositories for preserving and sharing scientific data.

*Plan Elements:* Consider addressing specific elements outlined in DRAFT guidance (see below, Supplemental DRAFT Guidance: Elements of An NIH Data Management and Sharing Plan).

*Plan Assessment:* The funding NIH ICO will assess the Plan, through the following processes:

- Extramural Awards: Plans will undergo a programmatic assessment by NIH staff within the proposed funding NIH ICO. NIH encourages potential awardees to work with NIH staff to address any potential concerns regarding the Plan prior to submission.
- Contracts: Plans will be included as part of the technical evaluation performed by NIH staff.
- Intramural Research Projects: Plans will be assessed by the Scientific Director (or designee) or Clinical Director (or designee) of the researcher's funding NIH ICO.
- Other funding agreements: Plans will be assessed in the context of other funding agreement mechanisms (e.g., Other Transactions).

## **VII. Compliance and Enforcement**

*During the Funding or Support Period:*

During the funding period, compliance with the Plan will be determined by the funding NIH ICO. Compliance with the Plan, including any Plan updates, will be reviewed

during regular reporting intervals (e.g., at the time of annual Research Performance Progress Reports (RPPRs)) at a minimum.

- **Extramural Awards:** The Plan will become a Term and Condition of the Notice of Award. Failure to comply with the Terms and Conditions may result in an enforcement action, including additional special terms and conditions or termination of the award, and may affect future funding decisions.
- **Contracts:** The Plan will become a Term and Condition of the Award, and compliance with and enforcement of the Plan will be consistent with the award and the Federal Acquisition Regulations (FAR), as applicable.
- **Intramural Research Projects:** Compliance with and enforcement of the Plan will be consistent with applicable NIH policies established by the NIH Office of Intramural Research and the applicable NIH ICO.
- **Other funding agreements:** Compliance with and enforcement of the Plan will be consistent with applicable NIH policies.

*Post Funding or Support Period*

After the end of the funding period, non-compliance with the NIH ICO-approved Plan may be taken into account by the funding NIH ICO for future funding decisions for the recipient institution (e.g., as authorized in the NIH Grants Policy Statement, Section 8.5, Special Award Conditions, and Remedies for Noncompliance (Special Award Conditions and Enforcement Actions)).

**Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing**

NIH recognizes that making data accessible and reusable for other users, while integral to the research process, may require costs above and beyond the routine costs of conducting research. To assist individuals and entities who may be subject to a future NIH Policy for Data Management and Sharing, NIH is proposing supplemental DRAFT guidance regarding potential categories of allowable NIH costs associated with data management and sharing for public comment. NIH is proposing that reasonable, allowable costs may be included in NIH budget requests when associated with:

- 1. Curating data and developing supporting documentation**, include formatting data according to accepted community standards; de-identifying data; attaching metadata to foster discoverability, interpretation, and reuse; and formatting data for transmission and storage at a selected repository for long-term preservation and access.
- 2. Preserving and sharing data through established repositories**, such as data deposit fees and charges necessary for making data available and accessible. When proposing to use a repository that charges recurring fees, budgets may include costs that would be incurred for preserving and sharing data. If the Plan proposes use of multiple repositories, consider including costs associated with use of each proposed repository.
- 3. Local data management considerations**, such as unique and specialized information infrastructure necessary to provide local management, preservation, and access to data, (e.g., before deposit into an established repository). Budget estimates should not include infrastructure costs typically included in institutional overhead (e.g., Facilities and Administrative costs), nor costs associated with the routine conduct of

research. Costs associated with collecting or otherwise gaining access to research data (e.g., data access fees) are considered costs of doing research and should not be included in budgets.

### **Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan (Plan)**

To assist those who may be subject to a future NIH Policy for Data Management and Sharing, NIH is proposing supplemental DRAFT guidance regarding elements of a Data Management and Sharing Plan (Plan) for public comment. A Plan should describe in two pages or less the proposed approach to data management and sharing that the specific research will employ. If certain elements of a Plan have not been determined at the time of submission, an entry of “to be determined” may be acceptable if a justification is provided along with a timeline or appropriate milestone at which a determination will be made. Note, NIH does not expect researchers to share all scientific data generated in a study. Elements of a Plan should consider:

1. **Data Type:** A description of the types and estimated amount of scientific data that will result from NIH-funded or conducted research, which scientific data will be preserved and shared, and the rationale for these decisions. Descriptions may include any additional metadata, information, or documentation about the scientific data that will be made publicly available (e.g., study protocols, data collection instruments). In describing the data types to be managed, preserved, and shared, consider:
  - Describing data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., exome

sequences of 20 to 30 gene variants from an estimated 800 cases and fMRI data from ~100 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

- Providing a rationale for decisions about which scientific data are to be preserved and made available for sharing, taking into consideration scientific utility, validation of results, availability of suitable data repositories, privacy and confidentiality, cost, consistency with community practices, and data security.
  - Identifying metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) which will be made accessible to facilitate interpretation of the scientific data.
  - For scientific data derived from human participants or specimens, outlining plans for providing appropriate protections of privacy and confidentiality (i.e., through de-identification or other protective measures) that are consistent with applicable federal, tribal, state, and local laws, regulations, statutes, guidance, and institutional policies.
2. **Related Tools, Software and/or Code:** An indication of whether specialized tools are needed to access or manipulate shared data to support replication or reuse, and name(s) of the needed tool(s) and software. Consider specifying how needed tools 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).

3. **Standards:** An indication of what standards, if any, will be applied to the scientific data and associated metadata to be collected, including data formats, data identifiers, definitions, unique identifiers, and other data documentation. While many scientific fields have developed and adopted common data standards, others have not. In such cases, the Plan may indicate that no appropriate data standards exist for the data to be collected, preserved, and shared. Provide the name of any data standards or metadata standards proposed for use, considering:
- Use of existing, widely adopted standards for scientific data and associated metadata. Some examples include: Clinical Data Interchange Standards Consortium, Minimum Information About a Microarray Experiment, Minimum Information about a high-throughput SEQUencing Experiment, and the Office of the National Coordinator for Health Information Technology Interoperability Standards Advisory.
  - 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 (CDE) Resource Portal.
4. **Data Preservation, Access, and Associated Timelines:** An indication of the timelines for data preservation and access, considering:
- Where scientific data will be archived to ensure long-term preservation (i.e., which repository(ies)). If scientific data will be archived in an existing data repository(ies), consider providing the name and URL web address of the

repository(ies). If an existing data repository(ies) will not be used, consider indicating why not and how scientific data will be preserved and shared.

- How the scientific data will be findable and whether a persistent unique identifier or other standard indexing tools will be used, and any provisions for maintaining the security and integrity of the scientific data (e.g., encryption and backups).
- Whether additional considerations are needed to implement the Plan, (e.g., whether permission needs to be sought to use a specific data repository, and from whom).
- Whether scientific data generated from humans or human biospecimens will be available through unrestricted (made publicly available to anyone) or restricted access (made available only after the requestor has received approval to use the requested scientific data). If the scientific data will be shared through a restricted access mechanism, consider describing the general terms of access for the data.
- Anticipated timeframes for preserving scientific data, describing if different timelines will apply to different subsets of scientific data, and when the scientific data will be submitted to specified data repositories.
- When the scientific data will be made available to other users (e.g., researchers and the broader public). In general, scientific data should be made available as soon as practicable, independent of award period and publication schedule. If applicable, consider indicating when scientific data will no longer be available to other users.

5. **Data Sharing Agreements, Licenses, and Other Use Limitations:** NIH encourages the broadest use of scientific data resulting from NIH-funded or conducted research,

consistent with privacy, security, informed consent, and proprietary issues. In describing proposed plans for managing data sharing agreements and other types of arrangements, consider indicating:

- A description of any restrictions imposed by existing agreements that would limit the ability to broadly share scientific data, as well as a summarizing what those limitations on sharing or reuse are.
  - Whether the applicant anticipates entering into any agreements that could limit the ability to broadly share scientific data and describe those agreements.
  - Any other considerations that may result in limitations on the ability to broadly share scientific data.
  - How relevant limitations to sharing are consistent with community expectations, and how scientific data will be shared to the maximum extent possible while honoring these limitations.
6. **Oversight of Data Management:** An indication of the individual(s) who will be responsible for executing various components (e.g., data collection, data analysis, data submission) of the Plan over the course of the research project and the roles of the individual(s) in data management, and a description of the appropriate expertise for oversight.

Dated: October 30, 2019.

Lawrence A. Tabak,  
Principal Deputy Director,  
National Institutes of Health.

