



Billing Code 4140-01-P

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Final NIH Policy for Data Management and Sharing and Supplemental Information

Agency: National Institutes of Health, HHS.

Action: Notice of final Policy.

SUMMARY: The National Institutes of Health (NIH) is issuing this final NIH Policy for Data Management and Sharing (DMS Policy) to promote the management and sharing of scientific data generated from NIH-funded or conducted research. This Policy establishes the requirements of submission of Data Management and Sharing Plans (hereinafter Plans) and compliance with NIH Institute, Center, or Office (ICO)-approved Plans. It also 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, with justified limitations or exceptions. This Policy applies to research funded or conducted by NIH that results in the generation of scientific data.

DATES: This final Policy is effective January 25, 2023.

FOR FURTHER INFORMATION CONTACT: If you have questions, or require additional background information about the DMS Policy, please contact Dr. Lyric Jorgenson, by e-mail at (sciencepolicy@od.nih.gov), or telephone at 301-496-9838.

SUPPLEMENTARY INFORMATION: Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing

accessibility to high-value datasets, and promoting data reuse for future research studies.¹ As a steward of the nation's investment in biomedical research, and in accordance with 42 U.S.C. 282, of the Public Health Service Act, as amended, NIH has long championed policies that make research available to the public to achieve these goals. For example, the 2003 NIH Data Sharing Policy reinforced NIH's commitment to data sharing by requiring investigators to address data sharing in applications for large research awards. NIH's 2014 Genomic Data Sharing (GDS) Policy, initially preceded by the 2008 Genome-Wide Association Studies Policy, set the expectation that researchers share large-scale genomic data, regardless of species, to enable the combination of large and information-rich datasets. In 2016, the NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information (Clinical Trials Policy) further reinforced NIH's commitment to research participants and the research community by making the results of clinical trials accessible in a timely fashion.

NIH recognizes that its data sharing policy efforts must flexibly evolve to keep pace with scientific and technological opportunities and notes that researchers' ability to generate, store, share, and combine data has never been greater. To capitalize on these advancements, NIH initiated the development of a more comprehensive data sharing policy alongside its efforts to modernize data sharing infrastructure in its 2015 Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research. With policy and infrastructure modernization efforts working in

¹ See also NIH Rigor and Reproducibility efforts at <https://www.nih.gov/research-training/rigor-reproducibility>

tandem, NIH initiated a stepwise process for seeking feedback from the community to develop a robust data sharing policy capable of reflecting the diversity of its community's data sharing needs. In 2016, NIH requested public comments on data management and sharing strategies and priorities (NOT-OD-17-015). In 2018, NIH solicited public input on proposed key provisions that could serve as a foundation for a future NIH policy for data management and sharing (NOT-OD-19-014). Using public feedback to inform its thinking, in 2019 NIH released a draft proposal for a future data management and sharing policy in the Federal Register (84 FR 60398).

Along with the Draft Policy proposal, NIH sought feedback on supplemental materials that could help researchers integrate effective data management and sharing practices into research, including “Elements of an NIH Data Management and Sharing Plan” and “Allowable Costs for Data Management and Sharing.” We note that a third document, “Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research,” was developed in response to public comments received on both the Draft Policy and the “Request for Public Comments on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research,” which was released for public comment by the White House Office of Science and Technology Policy (OSTP) to promote consistency across federal agencies and reduce researcher burden (85 FR 3085).

In respect and recognition of Tribal sovereignty, NIH also initiated Tribal Consultation on its Draft Policy proposal, in accordance with the HHS Tribal Consultation Policy and the NIH Guidance on the Implementation of the HHS Tribal Consultation Policy. The NIH Tribal Consultation Report – NIH Draft Policy for Data Management and Sharing provides more detail on the Tribal Consultation process relative to the development of the final DMS Policy and NIH's response. Briefly, three themes emerged from Tribal Nations' input: 1) Strengthen engagement built on trust between researchers and Tribal Nations; 2) Train researchers to responsibly and respectfully manage and share American Indian and Alaska Native (AI/AN) data; and 3) Ensure research practices are aligned with the laws, policies, and preferences of AI/AN community partners. NIH intends to continue discussions to ensure appropriate implementation of the DMS Policy as it relates to these communities, and details about some of the implementation planning follows in the discussion below.

Overview of Public Comments

NIH incorporated feedback over the course of several years to develop a data management and sharing policy proposal and released its Request for Comments on the Draft NIH Policy for Data Management and Sharing and Draft Supplemental Guidance on November 8, 2019 (84 FR 60398, comment period closing on January 10, 2020). NIH held a public webinar on December 16, 2019, with over 580 people participating. In response to the Draft Policy, NIH received 203 responses from both domestic and

international stakeholders, and the comments are publicly available.² The largest group of respondents reported affiliation with universities, followed by nonprofit research organizations, professional associations (tied with “other”), as well as small percentages of respondents affiliated with government agencies, healthcare delivery organizations, and patient advocacy organizations. Respondents typically identified themselves as scientific researchers, while another sizeable section self-identified as “other.” Remaining respondents identified as institutional officials, with smaller percentages self-identified as bioethicists or social science researchers, government officials, patient advocates, and members of the public. NIH considered all feedback in the development of the final DMS Policy, and a discussion of the public comments on topics follows below.

Discussion of Public Comments on the Draft NIH Policy for Data Management and Sharing

Clarifying Expectations for Sharing Scientific Data

Draft Policy: The Draft Policy did not explicitly set a default expectation of data sharing. Rather, it focused on requiring submission of and compliance with a Data Management and Sharing Plan (Plan) that outlines how data will be managed and shared. The Draft

² Compiled Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance (February 2020) https://osp.od.nih.gov/wp-content/uploads/RFI_Final_Report_Feb2020.pdf

Policy also included recognition of that fact that certain factors (i.e., legal, ethical, or technical) may limit the ability to preserve and share data.

Public Comments: While commenters were generally supportive of the overall scope of the Draft Policy, many requested NIH make an explicitly stronger commitment to expecting data sharing from the research community. Suggestions included requiring data sharing and indicating that data sharing should be the default, with well justified exceptions being permitted.

Final Policy: The final DMS Policy does not create a uniform requirement to share all scientific data. Unlike a requirement for submission of Plans, which can be implemented across various funding mechanisms and types of research with little variation, appropriate data sharing is likely to be varied and contextual. Through the requirement to submit a Plan, researchers are prospectively planning for data sharing, which we anticipate will increasingly lead researchers to integrate data sharing into the routine conduct of research. Accordingly, we have included in the final DMS Policy an expectation that researchers will maximize appropriate data sharing when developing Plans. The final DMS Policy retains the Draft Policy's factors (i.e., ethical, legal, or technical) that may necessitate variations in the extent of scientific data preservation and sharing, and researchers should convey such factors in their Plans. The final DMS Policy has also been modified to clarify these factors are not limited to data derived from human research participants. We believe this will provide the necessary flexibility for researchers to

accommodate the substantial variety in research fields, projects, and data types that this expectation will encompass.

Definition of “Scientific Data”

Draft Policy: The scope of which data will be shared relies on the definition of “scientific data.” This term was defined in the Draft Policy as: “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.”

Public Comments: Commenters focused on a variety of aspects of the definition of “scientific data.” They suggested that the concept of data quality be included, as data that may otherwise meet the definition but, if uninterpretable, are not of value. Commenters also suggested the definition address null or negative findings (and indicate that these data should be shared). Commenters requested clarification about the sentence that NIH expects reasonable efforts will be made to digitize all scientific data, including whether NIH would cover costs to digitize data that are not collected in digital form.

Final Policy: The final DMS Policy 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. 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.” We agree that data quality is an important concept to convey to ensure that scientific data are useful and to prevent data sharing from becoming a perfunctory administrative requirement, but rather one that should be done with the understanding that these data are intended to be used by others. Therefore, we have added to the definition that the data should be of sufficient quality to validate and replicate research findings. Even those scientific data not used to support a publication are considered scientific data and within the final DMS Policy’s scope. We understand that a lack of publication does not necessarily mean that the findings are null or negative; however, indicating that scientific data are defined independent of publication is sufficient to cover data underlying null or negative findings.

We also note that while the final DMS Policy states that scientific data are those as of sufficient quality to “validate and replicate,” we anticipate that shared scientific data will be used for a variety of purposes (consistent with applicable laws, policies, and limitations) including subsequent analyses, as suggested in the Purpose section of the final DMS Policy. Therefore, the concepts of validation and replication provide a

standard for determining what constitutes scientific data and are not intended to limit uses of shared data.

Finally, we have removed the expectation for digitizing scientific data. We encourage reasonable efforts to digitize data, recognizing that digitizing data may be a technical factor that may limit the sharing of data.

Timing of Submission of Data Management and Sharing Plans

Draft Policy: The Draft Policy proposed the submission of Plans at Just-in-Time for grants.

Public Comments: While we received a range of comments about timing of Plan submission, the majority were opposed to or requested further clarification about Just-in-Time Plan submission. Commenters were concerned about not having sufficient time to develop Plans and expressed concerns about the Plan revision process leading to delays in issuing awards. Others indicated that institutions would want to review Plans because they would ultimately be responsible for compliance, but a Just-in-Time Plan submission would not afford institutions sufficient time. A key practical concern with Just-in-Time Plan submission was difficulty submitting a budget at application that included requests for allowable data management and sharing costs prior to actually drafting the Plan. Commenters who favored submitting Plans at Just-in-Time frequently cited decreased

burden on applicants, because with Just-in-Time, only those applicants likely to be funded would be required to submit Plans, rather than all applicants.

Final Policy: The final DMS Policy requires submission of a Plan for extramural grants at application. This approach is more conducive to achieving NIH's goal of promoting a culture in which data management and sharing are recognized to be an integral component of a biomedical research project, rather than an administrative or additive one. While NIH is aware that this approach places the requirement on the general pool of grant applicants rather than on those likely to be funded, it is precisely this approach of prospective planning for data management and sharing that NIH hopes to promote and that a number of commenters suggested is crucial for ensuring more regular planning for data management and sharing. We were swayed by the logistical concerns expressed in comments, namely how applicants could submit budgets appropriately reflective of data management and sharing when not yet required to submit the Plan that is intended to help them consider these issues. In addition, the concerns about institutions having sufficient time to review Plans and potential logistical challenges in issuing timely awards was persuasive. This approach is also consistent with the 2018 Request for Information on Proposed Provisions of a Draft Data Management and Sharing Policy for NIH Funded or Supported Research, which proposed Plans be submitted with extramural grant applications. The responses to that proposal generally favored Plan submission at the time of application.

Assessment of Plans

Draft Policy: The Draft Policy proposed that NIH Program Staff in the funding NIH ICO assess Plans from extramural grants.

Public Comments: Many commenters supported peer review of Plans, noting their skill and that peer review of Plans would promote a cultural shift in favor of data sharing. Commenters also suggested that NIH Program Staff review may lead to more consistent Plan assessment and decrease peer reviewer burden.

Final Policy: The final DMS Policy maintains NIH Program Staff assessments of Plans' merits. However, peer reviewers may comment on the proposed budget for data management and sharing, although these comments will not impact the overall score. This approach balances the benefit of consistency afforded by NIH Program Staff review of Plans, review of updates, and compliance monitoring, with the opportunity for peer reviewers to comment on the requests for data management and sharing costs. Over time, and through these reviews, we hope to learn more about what constitutes reasonable costs for various data management and sharing activities across the NIH portfolio of research.

NIH ICO Consistency of Data Sharing Expectations

Draft Policy: The Draft Policy noted that NIH ICOs may supplement the Policy's expectations for Plans with their own complementary requirements to further advance their specific program or research goals. In addition, the Draft Policy stated the funding

NIH ICOs may request additional or specific information to be included within Plans to meet expectations for data management and sharing in support of programmatic priorities or to expand the utility of the scientific data generated from the research.

Public Comments: In light of various existing NIH ICO data sharing policies, commenters expressed confusion around having potentially varying expectations in data sharing policy implementation across NIH. There were concerns about insufficient direction to NIH ICO and around a potentially uncoordinated variety of approaches. Commenters suggested guidance to facilitate NIH ICO consistency and suggested that NIH provide a centralized location of NIH ICO-specific expectations to help researchers navigate variations, particularly when subject to more than one NIH ICO's data sharing policies.

Final Policy: While the final DMS Policy's language on this issue has not substantively changed from that of the Draft Policy, we have heard the concerns and intend to address them during the period of implementation planning prior to the DMS Policy's Effective Date. NIH ICOs can, within certain bounds, meet their scientific, policy, and programmatic goals in different ways. As such, this Policy affords NIH ICOs the opportunity to meet the goals of this Policy in ways that enhance their respective science. However, we intend to promote consistency on some key tenets of the final DMS Policy, such as the requirement for submission of Plans and the timing of their submission. The DMS Policy represents the minimum requirements for the NIH, but NIH ICOs may expect more specificity in Plans. For example, NIH ICOs and Programs may wish to

promote, via specific Funding Opportunity Announcements (FOAs) or across their research portfolios, the use of particular standards to enable interoperability of datasets and resources. We are appreciative of the suggestion about how to organize NIH ICO-specific expectations and will be working to ensure clear implementation materials for applicants and awardees.

Data Derived from Human Participants

Draft Policy: The Draft Policy acknowledged the applicability of laws, regulations, guidance, and policies that govern the conduct of research with human participants and how data derived from human participants should be used. It also described that Plans should indicate how human participants and data derived from them would be protected. Finally, the Draft Policy acknowledged that certain factors may limit the ability to share data and proposed that these factors be described in the Plan. Importantly, the Draft Policy did not propose any new expectations for the conduct of research with human participants.

Public Comments: Commenters expressed concerns about how to safeguard participant privacy and confidentiality when sharing data, with some requesting information on de-identification practices. Commenters also requested guidance on best practices in communicating data sharing in informed consent. They also stressed the importance of data sharing to maximize the contributions of those who volunteer to participate in NIH-funded studies. Some pointed to special populations with preferences on data sharing

issues, such as AI/AN populations, and asked how sharing of data from these participant populations is expected to be handled.

In addition to the public comments submitted during the comment period, NIH received input from the Secretary's Advisory Committee on Human Research Protections (SACHRP).³ SACHRP provided a set of recommendations relating to applying the DMS Policy to research with human participants, some of which we have incorporated into the final DMS Policy and are discussed below.

AI/AN communities provided input through various channels, including through letters sent to NIH as part of government-to-government communications. The Tribal Consultation process also led to valuable input that is informing NIH's implementation efforts, described further below.

Final Policy: As with the Draft Policy, the final DMS Policy does not introduce new requirements for protections for research with human participants. Existing laws (e.g., Certificates of Confidentiality), regulations (e.g., the Common Rule), and policies (e.g., the NIH Genomic Data Sharing Policy) continue to apply. However, through this Policy and associated supplemental information and other activities, NIH promotes thoughtful practices regarding the treatment of data derived from human participants.

³ Attachment A – NIH Data Sharing Policy (September 2020) <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/august-12-2020-attachment-a-nih-data-sharing-policy/index.html>

In response to public comments and SACHRP's recommendations on the Draft Policy, we have included in the final DMS Policy three concepts that we believe are important to emphasize for investigators as they think through how to engage prospective participants regarding what is expected to happen with the data they contribute and, downstream, how best to respect these contributions. First, we encourage investigators to consider, while developing their Plans, how to address data management and sharing in the informed consent process, such that prospective participants will understand what is expected to happen with their data. This planning will serve investigators as they develop their Plans, because some of the Plan elements prompt investigators to outline anticipated factors that might affect the ability to share and preserve scientific data, such as any limitations arising from the informed consent process. NIH also intends to develop resources to help researchers and institutions in communicating the intent to share data with prospective research participants. Second, we note that any limitations on subsequent use of data (which may apply to non-human data as well) should be communicated to those individuals or entities preserving and sharing the scientific data. This ensures that factors that may affect subsequent use of data are properly communicated and will travel with the data. Finally, we highlight the importance of researchers considering whether, in choosing where and how to make their data available (if not already specified by an FOA or funding NIH ICO expectation), access to scientific data derived from humans should be controlled, even if de-identified and lacking explicit limitations on subsequent use.

We note that data carrying explicit limitations on subsequent use require access controls to manage such limitations. This approach honors the wishes and autonomy of the

participants who contributed their data and is important to uphold, even if the data are de-identified. In addition, investigators should consider whether access to data even without such limitations should be controlled. SACHRP identified concerns regarding re-identification of otherwise de-identified data, and indeed technological advances and increasing interoperability among data resources, while providing opportunities for new analyses, present identifiability concerns that are widely acknowledged. In response to concerns expressed in public comments and by SACHRP, NIH may support development of resources to assist researchers and institutions in determining how to appropriately de-identify data from human participants, as well as for communicating data sharing in informed consent.

The final DMS Policy does not preclude the open sharing of data from human participants in ways that are consistent with consent practices, established norms, and applicable law. For example, open sharing of a compilation of a population's genotype at a particular locus may be an acceptable and established practice if consistent with informed consent. And importantly, we are aware that some patient communities prioritize openness to speed scientific progress and discovery. Nothing in the final DMS Policy is intended to prevent these approaches, as long as participants are appropriately informed and prospectively agree to them.

We emphasize that respecting participant autonomy and maintaining privacy of participants and confidentiality of their data can be consistent with data sharing. Through the final DMS Policy, we outline a balance that accommodates various responsible

approaches that meet data sharing expectations and honor appropriate limitations in sharing. In addition, while the DMS Policy sets the expectation that, through their Plans, researchers maximize the appropriate sharing of scientific data (acknowledging factors that may limit such sharing, as discussed above), the DMS Policy does not expect that the informed consent given by participants to be obtained in any particular way, such as through broad consent.

In response to input from Tribal Nations, the final DMS Policy clarifies agency respect for Tribal sovereignty in the absence of written Tribal laws or polices. To address some of the other themes and comments we heard from both AI/AN communities as well as public commenters who expressed interest in agency efforts to promote responsible and respectful engagement of AI/AN populations, we are developing supplemental information for researchers who wish to work with AI/AN communities. Such guidance is expected to encourage researchers to (among other topics): thoughtfully consider the unique data sharing concerns of AI/AN communities; respectfully negotiate agreements for data use with Tribal Nations; and enhance researcher awareness of processes Tribal Nations use to review prospective research. NIH will seek input from AI/AN communities on the development of the guidance, to ensure it serves the goals of guiding researchers while taking into account Tribal preferences and values.

When Data Are Expected To Be Shared

Draft Policy: The Draft Policy proposed that shared scientific data should be made accessible in a timely manner for use by the research community and the broader public.

Public Comments: While commenters appreciated the flexibility afforded by this approach, they also expressed concern about its ambiguity. Some suggested timing of data sharing be connected to publication. Commenters also suggested NIH should specify outer bounds for timing of data sharing in the absence of a publication. Overall, commenters expressed the desire for more clarity.

Final Policy: The final DMS Policy states that “[s]hared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of the award/support period, whichever comes first.” This statement provides more clarity than the Draft Policy through outer bounds to guide researchers in when to make the scientific data available. It clarifies that publication triggers release of the data that underlie that publication (indeed, publishers often require the same). But it also recognizes that research does not always lead to a publication that would itself trigger the release of data. Importantly, the final DMS Policy is designed to increase the sharing of scientific data, regardless of whether a publication is produced. Important research may never be published for a variety of reasons, not least of which because the results did not prove the hypothesis. However, we believe the scientific data underlying all NIH-funded research to be of importance, particularly to serve the purposes of accountability and transparency. Data that do not form the basis of a publication produced during the award period should be shared by the end of the award period. A single research project may

take advantage of both approaches. Namely, researchers may share data underlying publication during the period of award but may share other data that have not yet led to a publication by the end of the award period.

How Long Data Should Be Available

Draft Policy: The Draft Policy stated that “NIH encourages shared scientific data to be made available as long as it is deemed useful to the research community or the public.”

Public Comments: Commenters expressed uncertainty about how the concept of usefulness would be determined, and who would determine usefulness.

Final Policy: We have indicated a framework for helping researchers think through a minimum time period for data availability. Providing this framework is anticipated to help researchers both develop Plans and also budget accordingly for data management and sharing costs, when needed. Existing requirements and expectations set forth through, for example, applicable record retention requirements, repository policies, and journal policies may guide researchers as they seek to define minimal periods for data availability. However, we encourage researchers to propose longer time periods that may be informed by other factors, such as anticipated value of the dataset for the scientific community and the public.

Where to Share Scientific Data

Draft Policy: The Draft Policy stated that “NIH encourages the use of established repositories for preserving and sharing scientific data.”

Public Comments: Commenters supported the use of established repositories for preserving and sharing scientific data.

Final Policy: The final DMS Policy strongly encourages the use of established repositories to the extent possible. This reflects NIH’s preference that scientific data be shared and preserved through repositories, rather than kept only by the researcher or institution and provided on request, with the recognition that this is not always a practical or even a preferred approach. For example, we recognize and respect that AI/AN communities, in particular, may wish to manage, preserve, and share their own data. We support efforts that enable AI/AN communities to prioritize research opportunities and to ensure sufficient protections on scientific data generated from such research. In addition, we have released the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research, which will aid researchers as they choose suitable repositories for the preservation and sharing of data. This supplemental information is discussed in more detail below.

Discussion of Public Comments on the Draft Supplemental Information: Elements of an NIH Data Management and Sharing Plan

Page Limit and Template for Plans

Draft Supplemental Information: The Draft Supplemental Information suggested a limit for Plan length of two pages or less. It did not indicate whether template Plans would be provided.

Public Comments: Commenters expressed that two pages is insufficient to describe approaches for data management and sharing, particularly for larger, more complicated projects, such as those involving consortia. In addition, commenters suggested that NIH provide a template for Plans, with Plans being machine-readable.

Final Supplemental Information: We understand the concern about describing plans for data management and sharing in two pages. In the final supplemental information, we have noted the elements to be addressed in two pages or less, indicating that these descriptions need not be long narratives. In addition, short Plans are anticipated to limit researcher burden.

The Acceptability of “To Be Determined” as a Response to Plan Elements

Draft Supplemental Information: The Draft Supplemental Information proposed that if certain elements of a Plan have not been determined by the time of Plan 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.

Public Comments: Commenters disagreed with allowing responses of “to be determined” at initial Plan submission.

Final Supplemental Information: The final Supplemental Information eliminates the language that a response of “to be determined” is acceptable. We do not expect researchers to necessarily have all details at the application stage, but we encourage researchers to fill out Plans to the best of their knowledge and ability, so the Plans may be appropriately assessed. We also note that adherence with NIH ICO-approved Plans is a requirement of the final DMS Policy. As indicated in the final DMS Policy, researchers will have opportunities to update their Plans throughout the course of their awards, subject to NIH ICO approval.

The Use of Persistent Unique Identifiers (PIDs)

Draft Supplemental Information: The Draft Supplemental Information asked for researchers to indicate how data will be findable and whether a persistent unique identifier or other standard indexing tools will be used.

Public Comments: Commenters expressed support for PIDs, explaining that researchers are incentivized to use PIDs because they enable effective citation. They also noted PIDs are a way to track data sharing compliance.

Final Supplemental Information: The final Supplemental Information asks researchers to describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools. This wording change is meant to highlight the importance of using a PID or other standard indexing tool so the data are findable, which is a key component of the FAIR (Findable, Accessible, Interoperable, and Re-usable) Principles. PIDs are also listed as a desirable characteristic of data repositories in the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research.

Data Security

Draft Supplemental Information: The Draft Supplemental Information proposed that researchers address provisions for maintaining the security and integrity of the scientific data, such as through encryption and back-ups. It also noted that data sharing should be consistent with security as well as other factors.

Public Comments: Commenters emphasized the importance of data security.

Final Supplemental Information: We have removed the prompt for researchers to address provisions related to the security of scientific data. While we agree with the importance of appropriate data security measures, we believe that technical provisions regarding data security are more appropriately addressed by the institutions and repositories preserving and sharing the scientific data. The Supplemental Information to the NIH Policy for Data

Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research (discussed in more detail below) outlines characteristics of suitable repositories, and we do not wish to burden the funded community with describing in-depth the data security processes of the data repositories preserving and sharing the data generated by their research. While data may remain with an institution prior to submission to a data repository, the DMS Policy is not designed to set any new standards for institutional data security practices.

Discussion of Public Comments on the Draft Supplemental Information: Allowable Costs for Data Management and Sharing

Timelines for Using Funds for Data Management and Sharing Activities

Draft Supplemental Information: The Draft Guidance noted that budget requests to the NIH may include costs for preserving and sharing data through repositories that charge recurring fees, however it did not specify timelines by which funds allotted for data management and sharing must be spent or how to account for paying fees to data repositories storing data after the end of the performance period.

Public Comments: Commenters generally supported the proposal but sought clarification on whether funds may be used to pre-pay fees for long-term data availability. Commenters also asked whether these funds could cover personnel expenses.

Final Supplemental Information: Personnel costs required to perform the types of data management and sharing activities described in the final Supplemental Information are allowable. Regarding the availability of data beyond the end of the project, which is crucial to achieving the goals of the DMS Policy, the final Supplemental Information clarifies that fees for long-term data preservation and sharing are allowable, but funds for these activities must be spent during the performance period, even for scientific data and metadata preserved and shared beyond the award period. NIH funds cannot legally be spent after the award period.

Discussion of Requests for Additional Guidance and Information

Public commenters requested more clarity not only on information in provided materials, but about issues key to implementation. One common theme was a request for guidance about how to choose a data repository, with some requesting a list of suitable repositories. NIH does not intend to provide a comprehensive list of suitable repositories outside of those supported or stewarded by NIH.⁴ However, NIH recognizes the need for providing a way to help researchers determine what characteristics make for a suitable repository for the preservation and sharing of data from NIH-funded research. As such, we are releasing the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research. This document stems in part from an interagency effort led by the White House OSTP to

⁴ For an example of NIH-supported or -stewarded repositories see Open Domain-Specific Data Sharing Repositories (September 2020) https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html

outline desirable characteristics of preserving and sharing data from federally funded research, released as the Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research (85 FR 3085). The purpose was also to promote consistency across federal agencies to reduce researcher burden. The public comments on this document also informed the development of the Supplemental Information.

The Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research includes a process to help researchers determine suitable repositories by providing relevant characteristics, noting that NIH ICOs may have identified preferred repositories in FOAs or through other announcements.

Concluding Points

As the DMS Policy is released, the world is in the midst of the COVID-19 pandemic. The recognition that more open sharing can lead to faster advances and treatments has led to an unprecedented worldwide effort to openly share publications and data related to both SARS-CoV-2 (the novel coronavirus that causes COVID-19) and coronaviruses more generally. While this is a specific example of an urgent public health need, patients, families, and patient advocacy groups consider the diseases and conditions that affect them to be of equal urgency, as do those who research these diseases and conditions and treat affected patients. With public input, NIH has worked to develop and

refine this DMS Policy, the goal of which is to increase the sharing of scientific data generated from NIH-funded research to ultimately enhance health, lengthen life, and reduce illness and disability.

In addition to the Supplemental Information discussed here, we intend to provide frequently asked questions and other information to aid in implementation, prior to the DMS Policy's Effective Date. We recognize that some fields and researchers plan for sharing and prepare data for preservation and sharing as a regular practice. For others, these activities may be new. We anticipate a period of learning and an evolution of implementation practices. Further, 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. We look forward to working with applicants and the funded community as they prepare to meet the DMS Policy's requirements and expectations, as we all move toward a future in which data sharing is a community norm.

The final DMS policy is set forth below. Upon its Effective Date, the DMS Policy replaces the 2003 NIH Data Sharing Policy.

NIH Policy for Data Management and Sharing

Section I. Purpose

The National Institutes of Health (NIH) Policy for Data Management and Sharing (herein referred to as the DMS Policy) reinforces NIH's longstanding commitment to making the results and outputs of NIH-funded research available to the public through effective and efficient data management and data sharing practices. Data sharing enables researchers to rigorously test the validity of research findings,⁵ 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 FAIR data principles.⁶

Under the DMS Policy, NIH requires researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing Plan (Plan). Upon NIH approval of a Plan, NIH expects researchers and institutions to implement data management and sharing practices as described. The DMS Policy is intended to establish expectations for Data Management and Sharing Plans, which applicable NIH Institutes, Centers and Offices (ICO) may supplement as appropriate.

⁵ NIH Rigor and Reproducibility <https://www.nih.gov/research-training/rigor-reproducibility>

⁶ Wilkinson, M., Dumontier, M. et al, The FAIR Guiding Principles for Scientific Data Management and Stewardship (March 2016) <https://www.nature.com/articles/sdata201618>

Section II. Definitions

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

Scientific Data: 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. 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.

Data Management: The process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data for its users.

Data Sharing: The act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example, via an established repository.

Metadata: Data that provide additional information intended to make scientific data interpretable and reusable (e.g., date, independent sample and variable construction and description, methodology, data provenance, data transformations, any intermediate or descriptive observational variables).

Data Management and Sharing Plan (Plan): A plan describing the data management, preservation, and sharing of scientific data and accompanying metadata.

Section III. Scope

The DMS 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. The DMS Policy does not apply to research and other activities that do not generate scientific data, including training, infrastructure development, and non-research activities.

Section IV. Effective Date(s)

The effective date of the DMS Policy is January 25, 2023, including for:

- Competing grant applications that are submitted to NIH for the January 25, 2023 and subsequent receipt dates;
- Proposals for contracts that are submitted to NIH on or after January 25, 2023;
- NIH Intramural Research Projects conducted on or after January 25, 2023; and
- Other funding agreements (e.g., Other Transactions) that are executed on or after January 25, 2023, unless otherwise stipulated by NIH.

Section V. Requirements

The DMS Policy requires:

- Submission of a Data Management and Sharing Plan outlining how scientific data and any accompanying metadata will be managed and shared, taking into account any potential restrictions or limitations.
- Compliance with the awardee's plan as approved by the NIH ICO.

The 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 Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing).

Section VI. Data Management and Sharing Plans

Researchers planning to generate scientific data are required to submit a Plan to the funding NIH ICO as part of the Budget Justification section of the application for extramural awards, as part of the technical evaluation for contracts, as determined by the Intramural Research Program for Intramural Research Projects consistent with the objectives of this Policy, or prior to release of funds for other funding agreements. Plans should explain how scientific data generated by research projects will be

managed and which of these scientific data and accompanying metadata will be shared. If Plan revisions are necessary (e.g., new scientific direction, a different data repository, or a timeline revision), Plans should be updated by researchers and reviewed by the NIH ICO during regular reporting intervals or sooner. Plans from NIH-funded or conducted research may be made publicly available and should not include proprietary or private information.⁷

Plan Elements: NIH has developed Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan that describes recommended elements to address in Plans.

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

- Extramural Awards: Plans will undergo programmatic assessment by NIH as determined by the proposed NIH ICO. NIH encourages potential awardees to work with NIH staff to address any potential questions regarding Plan development 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 in a manner determined to be appropriate by the Intramural Research Program.

⁷ NIH Grants Policy Statement 2.3.11 Availability and Confidentiality of Information (October 2019)
https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.11_availability_and_confidentiality_of_information.htm

- Other funding agreements: Plans will be assessed in the context of other funding agreement mechanisms (e.g., Other Transactions).

Section VII. Managing and Sharing Scientific Data

NIH expects that in drafting Plans, researchers will maximize the appropriate sharing of scientific data, acknowledging certain factors (i.e., legal, ethical, or technical) that may affect the extent to which scientific data are preserved and shared. Any potential limitations on subsequent data use should be communicated to individuals or entities (e.g., data repository managers) that will preserve and share the scientific data. The NIH ICO will assess whether Plans appropriately consider and describe these factors.

Considerations for Scientific Data Derived from Human Participants: 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 govern research involving human participants and the sharing and use of scientific data derived from human participants. NIH also respects Tribal sovereignty in the absence of written Tribal laws or polices. The DMS Policy is consistent with federal regulations for the protection of human research participants and other NIH expectations for the use and sharing of scientific data derived from human participants, including the NIH's 2014 Genomic Data Sharing (GDS) Policy, 2015 Intramural Research Program Human Data Sharing Policy, and 45 CFR 46. Researchers proposing to generate scientific data derived

from human participants should outline in their Plans how privacy, rights, and confidentiality of human research participants will be protected (i.e., through de-identification, Certificates of Confidentiality, and other protective measures).

NIH strongly encourages researchers to plan for how data management and sharing will be addressed in the informed consent process, including communicating with prospective participants how their scientific data are expected to be used and shared. Researchers should consider whether access to scientific data derived from humans, even if de-identified and lacking explicit limitations on subsequent use, should be controlled.

Data Repository Selection: NIH strongly encourages the use of established repositories to the extent possible for preserving and sharing scientific data.⁸ The Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research assists researchers in selecting a suitable data repository(ies) or cloud-computing platform.

Data Preservation and Sharing Timelines: Shared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of performance period, whichever comes first. Researchers are encouraged to consider relevant requirements and expectations (e.g., data

⁸ NIH Strategic Plan for Data Science (June 2018)
https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

repository policies, award record retention requirements, journal policies) as guidance for the minimum time frame that scientific data should be made available, which researchers may extend.

Section VIII. Compliance and Enforcement

During the Funding or Support Period

During the funding period, compliance with the Plan will be determined by the NIH ICO. Compliance with the Plan, including any Plan updates, may be reviewed during regular reporting intervals (e.g., at the time of annual Research Performance Progress Reports (RPPRs)).

- **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, 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 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 NIH 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 Information to the NIH Policy for Data Management and Sharing:
Elements of an NIH Data Management and Sharing Plan**

The final NIH Policy for Data Management and Sharing requires applicants to submit a Data Management and Sharing Plan (Plan) for any NIH-funded or conducted research that will generate scientific data. This supplemental information outlines the elements to be addressed in a Plan within two pages or less. A Plan should reflect the proposed approach to data management and sharing at the time it is prepared and be updated during the course of the award/support period to reflect any changes in the management and sharing of scientific data (e.g., new scientific direction, new repository option, timeline revision). For some programs and data types, NIH and/or NIH ICOs have developed specific data sharing expectations (e.g., scientific data to share, relevant standards, repository selection, timelines) that apply and should be reflected in a Plan. When no additional NIH and/or NIH ICO data sharing expectations apply, researchers should propose their own approaches to data management and sharing in a Plan. NIH

encourages data management and sharing practices to be consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and reflective of practices within specific research communities. NIH recommends addressing all elements described below.

Data Type: Briefly describe the scientific data to be managed, preserved, and shared, including:

- A general summary of the types and estimated amount of scientific data to be generated and/or used in the research. Describe 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., 256-channel EEG data and fMRI images from ~50 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).
- A description of which scientific data from the project will be preserved and shared. NIH does not anticipate that researchers will preserve and share all scientific data generated in a study. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors that may affect the extent to which scientific data are preserved and shared. Provide the rationale 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: An indication of 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, (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team) and, if known, whether such tools are likely to remain available for as long as the scientific data remain available.

Standards: An indication of what standards 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). While many scientific fields have developed and adopted common data standards, others have not. In such cases, the Plan may indicate that no consensus data standards exist for the scientific data and metadata to be generated, preserved, and shared.

Data Preservation, Access, and Associated Timelines: 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. NIH has provided additional information to assist in selecting suitable repositories for scientific data

resulting from funded research.

- 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 (i.e., the larger research community, institutions, and/or the broader public) and for how long. NIH encourages scientific data be shared as soon as possible, and no later than time of an associated publication or end of the performance period, whichever comes first. Researchers are encouraged to consider relevant requirements and expectations (e.g., data repository policies, award record retention requirements, journal policies) as guidance for the minimum time frame scientific data should be made available. NIH encourages researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, institutions, and/or the broader public. Identify any differences in timelines for different subsets of scientific data to be shared.

Access, Distribution, or Reuse Considerations: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data generated from NIH-funded or conducted research, consistent with privacy, security, informed consent, and proprietary issues. Describe any applicable factors affecting subsequent access, distribution, or reuse of scientific data related to:

- Informed consent (e.g., disease-specific limitations, particular communities’

concerns).

- Privacy and confidentiality protections (i.e., de-identification, Certificates of Confidentiality, and other protective measures) consistent with applicable federal, Tribal, state, and local laws, regulations, and policies.
- Whether access to scientific data derived from humans will be controlled (i.e., made available by a data repository only after approval).
- Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements (e.g., with third party funders, with partners, with Health Insurance Portability and Accountability Act (HIPAA) covered entities that provide Protected Health Information under a data use agreement, through licensing limitations attached to materials needed to conduct the research).
- Any other considerations that may limit the extent of data sharing.

Oversight of Data Management and Sharing: Indicate how compliance with the Plan will be monitored and managed, frequency of oversight, and by whom (e.g., titles, roles).

Supplemental Information to the NIH Policy for Data Management and Sharing:

Allowable Costs for Data Management and Sharing

NIH recognizes that making data accessible and reusable for other users may incur costs. To assist individuals and entities subject to the final NIH Policy for Data

Management and Sharing, this supplemental information outlines categories of allowable NIH costs associated with data management and sharing.

All allowable costs submitted in budget requests must be incurred (e.g., curation fees, data repository fees) during the performance period, even for scientific data and metadata preserved and shared beyond the award period. Consistent with 45 CFR 75.403 and the NIH Grants Policy Statement Section 7.4, budget requests must not include infrastructure costs that are included in institutional overhead (e.g., Facilities and Administrative costs) or 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 scientific data management and sharing budgets. Costs may not be double charged or inconsistently charged as both direct and indirect costs.

Reasonable, allowable costs may be included in NIH budget requests when associated with:

- 1. Curating data and developing supporting documentation**, including formatting data according to accepted community standards; de-identifying data; preparing metadata to foster discoverability, interpretation, and reuse; and formatting data for transmission to and storage at a selected repository for long-term preservation and access.
- 2. Local data management considerations**, such as unique and specialized information infrastructure necessary to provide local management and

preservation (e.g., before deposit into an established repository).

- 3. Preserving and sharing data through established repositories**, such as data deposit fees necessary for making data available and accessible. For example, if a Data Management and Sharing Plan proposes preserving and sharing scientific data for 10 years in an established repository with a deposition fee, the cost for the entire 10-year period must be paid prior to the end of the period of performance. If the Plan proposes deposition to multiple repositories, costs associated with each proposed repository may be included.

Supplemental Information to the NIH Policy for Data Management and Sharing:

Selecting a Repository for Data Resulting from NIH-Supported Research

This supplemental information is intended to help researchers choose data repositories suitable for the preservation and sharing of data (i.e., scientific data and metadata) resulting from NIH-funded and conducted research. NIH promotes the use of established data repositories because deposit in a quality data repository generally improves the FAIRness (Findable, Accessible, Interoperable, and Re-usable) of the data.

While NIH supports many data repositories, it will not necessarily provide data repositories to preserve and share all data resulting from the research it funds. The broader repository ecosystem for biomedical data includes data repositories supported by other organizations, both public and private. NIH anticipates that the broader repository

ecosystem will continue to evolve over time, providing different options for researchers as their data sharing needs continue to evolve.

Similarly, while discipline or data-type specific repositories may not exist for every type of data resulting from NIH-funded or conducted research, the broader repository ecosystem provides suitable data repositories to accommodate scientific data generated from all of NIH's funded or conducted research projects. Researchers may wish to consult experts in their own institutions (e.g., librarians, data managers) for assistance in selecting among data repositories.

NIH encourages researchers to select data repositories that exemplify the desired characteristics (see lists I. and II. below relating to data repository characteristics), including when a data repository is supported or provided by a cloud-computing or high-performance computing platform. These desired characteristics aim to ensure that data are managed and shared in ways that are consistent with FAIR data principles.

Selecting a Data Repository

1. For some programs and types of data, NIH and/or NIH ICO policy(ies) and 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).
2. For data generated from research for which no data repository is specified by NIH or the NIH ICO (as described above), researchers are encouraged to select a data

repository that is appropriate for the data generated from the research project and is in accordance with the desired characteristics, taking into consideration the following guidance:

- A. Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse. NIH makes a list of such data repositories available (see https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html).
- B. If no appropriate discipline or data-type specific repository is available, researchers should consider a variety of other potentially suitable data sharing options:
 - i. Small datasets (up to 2 GB in size) may be included as supplementary material to accompany articles submitted to PubMed Central (see <https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#suppm>).
 - ii. Data repositories, including generalist repositories (see https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html) or institutional repositories, that make data available to the larger research community, institutions, or the broader public.
 - iii. Large datasets may benefit from cloud-based data repositories for data access, preservation, and sharing.

I. Desirable Characteristics for All Data Repositories

The characteristics in this section are relevant to all repositories that manage and share data resulting from Federally funded research:

- A. Unique Persistent Identifiers:** Assigns datasets a citable, unique persistent identifier (PID), such as a digital object identifier (DOI) or accession number, to support data discovery, reporting (e.g., of research progress), and research assessment (e.g., identifying the outputs of federally funded research). The unique PID points to a persistent landing page that remains accessible even if the dataset is de-accessioned or no longer available.
- B. Long-Term Sustainability:** Has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; building on a stable technical infrastructure and funding plans; and having contingency plans to ensure data are available and maintained during and after unforeseen events.
- C. Metadata:** Ensures datasets are accompanied by metadata to enable discovery, reuse, and citation of datasets, using schema that are appropriate to, and ideally widely used across, the community(ies) the repository serves. Domain-specific repositories would generally have more detailed metadata than generalist repositories.
- D. Curation and Quality Assurance:** Provides, or has a mechanism for others to provide, expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata.
- E. Free and Easy Access:** Provides broad, equitable, and maximally open access to datasets and their metadata free of charge in a timely manner after submission, consistent with legal and ethical limits required to maintain privacy and confidentiality, Tribal sovereignty, and protection of other sensitive data.

- F. Broad and Measured Reuse:** Makes datasets and their metadata available with broadest possible terms of reuse; and provides the ability to measure attribution, citation, and reuse of data (e.g., through assignment of adequate metadata, unique PIDs).
- G. Clear Use Guidance:** Provides accompanying documentation describing terms of dataset access and use (e.g., particular licenses, need for approval by a data use committee).
- H. Security and Integrity:** Has documented measures in place to meet generally accepted criteria for preventing unauthorized access to, modification of, or release of data, with levels of security that are appropriate to the sensitivity of data.
- I. Confidentiality:** Has documented capabilities for ensuring that administrative, technical, and physical safeguards are employed to comply with applicable confidentiality, risk management, and continuous monitoring requirements for sensitive data.
- J. Common Format:** Allows datasets and metadata downloaded, accessed, or exported from the repository to be in widely used, preferably non-proprietary, formats consistent with those used in the community(ies) the repository serves.
- K. Provenance:** Has mechanisms in place to record the origin, chain of custody, and any modifications to submitted datasets and metadata.
- L. Retention Policy:** Provides documentation on policies for data retention within the repository.

II. Additional Considerations for Repositories Storing Human Data (even if de-identified)

The additional characteristics outlined in this section are intended for repositories storing human data, which are also expected to exhibit the characteristics outlined in Section I, particularly with respect to confidentiality, security, and integrity. These characteristics also apply to repositories that store only de-identified human data, as preventing re-identification is often not possible, thus requiring additional considerations to protect privacy and security.

- A. Fidelity to Consent:** Employs documented procedures to restrict dataset access and use to those that are consistent with participant consent (such as for use only within the context of research on a specific disease or condition) and changes in consent.

- B. Restricted Use Compliant:** Employs documented procedures to communicate and enforce data use restrictions, such as preventing reidentification or redistribution to unauthorized users.

- C. Privacy:** Implements and provides documentation of appropriate approaches (e.g., tiered access, credentialing of data users, security safeguards against potential breaches) to protect human subjects' data from inappropriate access.

- D. Plan for Breach:** Has security measures that include a response plan for detected data breaches.

E. Download Control: Controls and audits access to and download of datasets (if download is permitted).

F. Violations: Has procedures for addressing violations of terms-of-use by users and data mismanagement by the repository.

G. Request Review: Makes use of an established and transparent process for reviewing data access requests.

Dated: October 19, 2020.

Lawrence A. Tabak,

Principal Deputy Director,

National Institutes of Health.

[FR Doc. 2020-23674 Filed: 10/29/2020 8:45 am; Publication Date: 10/30/2020]