January 10, 2020

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

**Re: NOT-OD-20-013 "Request for Public Comments on a DRAFT NIH Policy for Data
Management and Sharing and Supplemental Draft Guidance"**

The Association of American Medical Colleges (AAMC) appreciates the opportunity to respond
to NIH's request for information regarding proposed provisions for a draft data management and
sharing policy. The AAMC is a not-for-profit association representing all 154 accredited U.S.
medical schools, nearly 400 major teaching hospitals and health systems, and more than 80
academic and scientific societies. Through these institutions and organizations, the AAMC
represents nearly 173,000 faculty members, 89,000 medical students, 129,000 resident
physicians, and more than 60,000 graduate students and postdoctoral researchers in the
biomedical sciences. These comments on NIH's proposed provisions incorporate feedback
provided by AAMC-member institutions on their data sharing practices as well as broader
standards in the scientific community.

The AAMC concurs with the NIH's assertion that increased access to research data advances
biomedical research by enabling further validation of scientific results, facilitating reuse of hard-
to-generate data, catalyzing new research, and generally promoting more responsible stewardship
of federal resources. These advantages can be realized through meaningful data sharing and the
development of community-wide norms, as well as ensuring that accessed data are used for the
advancement of discovery and in furtherance of rigorous scientific discourse.

The AAMC supports NIH's efforts to integrate data management into the research review and
funding process, to increase sharing and re-use of scientific data generated through NIH-funded
research, and to develop a clearly defined policy to accomplish these objectives. In addition to
responding to the specific areas for which NIH has requested information, AAMC provides the
following high-level comments on the draft policy:

- As NIH moves forward in the policy development process, we encourage the agency to
  consider the type of policy that will lead to meaningful and positive, rather than
  compliance-based, data management and sharing practices. When deciding how

prescriptive to make the policy's requirements, NIH's focus should be on feasibility of consistent implementation and on encouraging the sharing of data that are scientifically valuable, discoverable and reusable.

- The agency can further incentivize the goal of increased data sharing through encouraging the use of persistent identifiers (PIDs) so researchers can track and receive credit for their data[1], as well as issuing funding opportunities focused on data reuse.
- It is critical to have as much as harmonization and standardization as possible across the NIH in both the policy requirements and implementation. This includes all grantees as well as consistency in evaluation of compliance and in institute-specific requirements.
- We appreciate that the draft policy does not require researchers to share all scientific data, since requiring the sharing of all data without considering its usefulness or likelihood of re-use does not contribute to scientific progress and would constitute a substantial burden on the researcher and institution.
- Given the scope of this new policy, incorporating flexibility is appreciated by the research community. However, throughout the draft policy there are many optional elements and very few requirements, which may lead to overcompliance or an ineffective or inconsistently implemented policy.
- If NIH or its Institutes, Centers, or Offices (ICOs) have specific but unstated expectations for any aspects of data management and sharing, such as what types of data should always be shared, how accessible that data should be, or a timeline for data sharing, those expectations should be included in the policy or otherwise explicitly stated.
- Successful implementation of this policy will require additional resources from both the NIH and grantee institutions. In addition to these resources, grantees will need substantial guidance from the NIH.
- We understand that NIH intends to undertake ongoing evaluation of the costs and impact of this policy as implemented. We encourage NIH to treat the implemented final policy as a robust pilot initiative and recommend that a strong and detailed statement regarding the evaluation and revision process be included in the policy itself.

General Policy Requirements

The draft policy states that researchers must submit a Data Management and Sharing Plan (hereinafter "Plan") to NIH, as well as comply with the final NIH ICO-approved Plan, leading many in the research community to the concern that over time there may be 27 different data sharing policies at the NIH for which investigators are responsible: an overarching policy and one from each ICO. While ICOs may have additional expectations for data management and sharing above the base NIH policy, these additional ICO requirements should be narrow and rare, with a priority placed on standardization across the agency whenever possible. Further, NIH should make publicly available the process (or at a minimum, basic criteria) ICOs will use to establish these requirements and explain the role of the Data Science Governance Council in

---

[1] https://www.nature.com/articles/d41586-019-01715-4

making these decisions. We understand the need for special, large-scale projects to have specific data management and sharing requirements, but stress that these should be put into place through a transparent and deliberate process.

In order for researchers to develop an effective and executable Plan, there should be clarity about the evaluation criteria and the assessment that will be used by program officers. If NIH can make public any relevant tools that program officers are using, that would be very helpful to the research community and future applicants. We also recommend that any guidance provided to program officers regarding requirements for or evaluation of the adequacy of data management and sharing plans be developed in collaboration with external experts. With the Plan submission proposed to be submitted as a Just-in-Time requirement, the Plan will no longer receive feedback from peer reviewers with expertise in the field and instead will be added to the application after the researcher initially creates the grant budget. As the policy is implemented, the agency should evaluate if this is the most effective timing for submission of the Plan. As each ICO will be responsible for communicating with the researchers as they develop and comply with their Plans, there should be clear points of contact at each ICO for questions, including where researchers can go for assistance if they are unable to receive it from their designated program officer.

The policy currently states that NIH may make data management and sharing plans publicly available. It is critical that a Plan functions to help researchers manage data and clearly lay out their obligations to the agency. Given that researchers may have hesitations about making these plans publicly available, including concerns about privacy or progress of the research, the agency should consider an embargo period or exceptions for this requirement. However, there are also clear benefits to making data management plans broadly available, in allowing researchers and the public to be able to find the data associated with a particular grant, as well as provide examples of effective Plans to NIH investigators. We recommend that the NIH find a mechanism to make the data location element and when/whether the data will be shared publicly available (e.g. one or more dataset PIDs included as part of a RePORTER listing), and secondarily that the NIH commit to creating and making available a collection of Plans that have been submitted to and approved by the agency.

Data Management and Sharing Plans

The AAMC suggests that NIH define clearly a set of minimum requirements that researchers should include in the Plan submitted to the agency. The draft supplemental guidance on elements of a Plan currently contains a number of options for researchers to include in a Plan, with no indication of the relative importance or hierarchy of these elements. While it is understandable that unique projects may have different priorities and needs for inclusion in a Plan, we recommend that the policy define minimum requirements for researchers to include in a Plan, such as data type, standards and metadata, plans for data preservation, and projected data accessibility. Researchers should also be required to indicate in the Plan whether the project will

involve data derived from human participants or specimens, and if so, include strategies for maintaining privacy, rights, and confidentiality. In the absence of sample templates and/or further guidance for the level of detail, each institution (or researcher) will create their own guidelines and tools, which may or may not meet the objectives of the policy. Providing greater guidance about the expected content of a Plan will better serve both the researchers assembling the document and the goals of the agency.

We would also recommend that the agency reconsider the currently proposed limit of 2 pages for a Plan. Many researchers who actively practice data sharing and frequently prepare DMPs have suggested to us that this length is insufficient to include all of the necessary information for the Plan to be a useful document with the appropriate level of detail. We suggest that NIH increase this limit to 4 pages and, in its ongoing evaluation of the policy's impact and effectiveness, determine whether this is an appropriate limit after the policy goes into effect.

There are a number of resources the agency will need to develop to facilitate researchers both creating and implementing a Plan. We recommend that NIH create and maintain an online clearinghouse that lists data elements and metadata for common data types for which best practices exist in the scientific community, as well as other existing resources such as DMPTool. The development of this policy presents an opportunity to amplify and disseminate efforts for good data management and sharing, particularly for certain disciplines, such as neuroimaging, or data types, such as microarrays or sequencing, which have well-defined standards and formats. This will provide a basis for standardization in the data that is submitted by researchers, and hopefully increase the usability of NIH-funded data.

We recommend that NIH identify key characteristics of suitable data repositories and additionally provide lists of accepted repositories for scientific disciplines where they have been well established (see current efforts from Springer Nature/FAIRsharing/DataCite). In order to meet the presumed expectations that most or all data from NIH-funded research will need to be stored and made available for others to use, many institutions are planning to expand and use their own repositories. Without guidance from the agency on standards for data storage and discoverability as well as some level of centralized infrastructure or coordination, holding data in such disparate platforms and systems will place a significant technical burden on anyone who wants to reuse the data, thwarting the agency's laudable goals to increase and improve data re-use.

We appreciate that the draft policy acknowledges that valuable data are not always used to support a scholarly publication—this understanding is essential to recognizing data as a first-class research object and promoting a data-centric model of research. We also agree that investigators should have the opportunity to provide a rationale for decisions about which scientific data will not be made available for sharing. In order to accommodate this flexibility and also push forward the desired result of increasing sharing, the agency could consider setting a baseline for data that should be shared, such as the minimum underlying data to replicate and

validate published findings, while still providing the researcher the ability to justify whether or not this is reasonable for a given study.

Allowable Costs for Data Management and Sharing

Data management and preservation will require significant infrastructure investment on the part of the institution; however, the allowable costs as currently defined specifically exclude infrastructure costs typically included in institutional overhead. We would recommend that if these costs are not permitted on a grant-by-grant basis, that the agency offers additional supplemental funding to institutions to develop this infrastructure.

The guidance on costs also should have additional clarity around what constitutes an "established repository," and particularly whether institutional repositories may fit this role and be included in the grant budget. While costs for deposition and storage in an established and/or commercial repository may be more well-documented, it can be difficult to define the costs for an institutional resource in the same way. The current statement that researchers can request funds for "unique and specialized information infrastructure" would benefit from examples on what this includes.

Increasing data management and sharing activities often requires significant support from personnel outside of the traditional laboratory environment, including librarians and data scientists, to provide the necessary expertise and guidance needed to comply with a data sharing policy and build good data management practices into an investigator's research process. NIH should strongly consider including these additional staff as part of the allowable costs. Again, if this is not doable, it will be necessary for the agency to provide supplemental funding to institutions in building up and maintaining services that support scientific data sharing.

Finally, the draft guidance does not instruct grantees on what happens after a grant period comes to an end and whether additional funding would be available at this juncture, when much of the data preservation and storage will take place. It is critical that the agency specify how it plans to support these costs that will occur after the normal grant period ends and indicate whether there will be additional funding available specifically for this purpose.

Considerations for Implementation, Compliance, and Enforcement

The NIH should institute an implementation timeframe that allows for researchers to fully understand the policy requirements, and for institutions to develop the necessary training, resources and infrastructure. Because this is such a wide-ranging policy that impacts the way research is conducted and includes every NIH-funded investigator and project, **we recommend a minimum implementation date of one year after the release of the final policy, with a delay in enforcement actions for at least one year after the implementation deadline.** Any determination of non-compliance should follow well-defined and transparent criteria.

Over time, if the policy is not found to be generating the desired impact and does not increase accessibility to data from NIH-funded research, the agency should reconsider aspects of the policy including the elements of a Plan, defining minimum requirements for data that should be shared, as well as a timeframe for data deposition.

In AAMC's response to the proposed key policy provisions[2], we noted that a policy alone will not be sufficient to reach the stated goal of increasing scientific data sharing, and that the agency must provide "adequate training, education, and guidance, increasing available financial resources, and leading the development of tools and infrastructure in order to enable and facilitate policy implementation." The research community has expressed concern about the lack of clarity regarding which resources NIH will provide to implement this policy, including options for data storage and additional funding mechanisms. It is important to acknowledge that the significant culture change that will be required in a move to a data sharing ecosystem will involve many factors, such as incentives and community support, in addition to any policy or mandate.

We appreciate NIH's intent to create a policy that is flexible, responsive to researcher feedback, and able to keep pace with the state of biomedical science. Plans to evaluate the impact of the policy should be described and implemented prior to its effective date to align agency and community expectations about the metrics that will be evaluated. A feedback loop between the agency and researchers, and clear communication are key, but without detailed guidance from the NIH, there will be a wide range of interpretations and policy implementation that does not necessarily serve the end goal. As NIH develops this guidance, we encourage the agency to refer to established criteria and policies from other funders and federal agencies, journals, and scientific societies, as well as consider the impact of any given requirement on how institutions are already complying with existing NIH policy. The AAMC would be happy to work with institutions to provide the agency with examples of how they are affected by and complying with varying NIH policies.

Finally, as the policy is put in place, NIH should engage specifically with working groups consisting of researchers who generate data, librarians and other data science support at institutions, and labs that have research programs based on sharing and re-using scientific data, to ensure that the policy is responsive to the needs and concerns of different stakeholders and supports the scientific community as effectively as possible while meeting its desired goals. The AAMC would be glad to assist in identifying these partners from our member medical institutions.

---

[2] https://www.aamc.org/system/files/c/1/493914-aamcsubmitscommentstonihonproposedprovisionsforadatamanagementa.pdf

The AAMC appreciates NIH soliciting feedback from stakeholders during the policy process and looks forward to continued engagement on this issue as the final policy and other guidance are developed. Please feel free to contact me or my colleagues Anurupa Dev, PhD, Lead Specialist for Science Policy (adev@aamc.org) and Heather Pierce, JD, MPH, Senior Director for Science Policy and Regulatory Counsel (hpierce@aamc.org) with any questions about these comments.

Sincerely,

Ross E. McKinney, Jr., MD
Chief Scientific Officer