**MLA/AAHSL Comments on Draft NIH Policy for Data Management and Sharing and Supplemental Guidance**<https://www.federalregister.gov/documents/2019/11/08/2019-24529/request-for-public-comments-on-a-draft-nih-policy-for-data-management-and-sharing-and-supplemental>

Submitted January 9, 2020

To: National Institutes of Health at

<https://osp.od.nih.gov/draft-data-sharing-and-management>

Document Citation: 84 FR 60298

Document Number: 2019-24529

[**DRAFT NIH Policy for Data Management and Sharing**](https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf)

**Section I: Purpose (limit: 8000 characters)**

The Medical Library Association (MLA) and the Association of Academic Health Sciences Libraries (AAHSL) support NIH’s commitment to making funded research results and outputs available to the public and advancing biomedical research by enabling the validation of results, combining datasets to strengthen analyses, facilitating reuse, and accelerating future research. We are pleased that the Policy focuses on the importance of reproducibility and reliability of research findings. We laude the purpose of this policy and appreciate that data sharing is the bedrock of the advancement of biomedical research exploration. We make specific recommendations in the following sections for success in implementing and complying with the Policy.

**Section II: Definitions (limit: 8000 characters)**

* We recommend that laboratory notebooks either be included in the definition of scientific data or defined as a separate term. Laboratory notebooks are the primary record of research where data is recorded within context. This information is vital for reproducibility, and it preserves research integrity by increasing transparency of experimental details and by showing provenance. It is important to clarify that this document may not be necessarily shared but nonetheless should be preserved.
* We recommend, based on our comments in later sections, that the definitions section be expanded to include additional relevant definitions. Examples would include what patient data should be considered for sharing.
* We also recommend that references to existing resources be provided to researchers (e.g., the National Network of Libraries of Medicine Data Thesaurus, https://nnlm.gov/data/thesaurus).

**Section III: Scope (limit: 8000 characters)**

We applaud NIH for covering all supported research that results in the generation of data in this Policy. However we urge NIH to consider how to better share patient data in a manner that is consented for by each patient, and when patients do not consent, that patient privacy is upheld. We also ask NIH to carefully consider the costs of sharing patient data, including the effort to de-identify it before sharing so that it cannot be later triangulated with other data sets and reidentified, and provide supplemental direct funds for this activity. We recommend clarifying that all scientific data is covered under this Policy, not just data associated with a publication.

**Section IV: Effective Date(s) (limit: 8000 characters)**

We believe that any formal policy adopted by the NIH and the Office of Science Policy should allow institutions at least one year to make needed internal changes to policies and procedures in order to ensure compliance. We also implore the NIH and the Office of Science Policy to continue to take the deep care it has taken thus far as it advances the adoption and implementation of this policy. We also suggest setting effective dates that would allow researchers and evaluators to have as much educational and infrastructure support in place as possible. We expect education and training would take at least one year, and developing adequate infrastructure may take longer. For example, data repositories do not yet exist for many areas of biomedical research. Librarians and information professionals also need training to support the data management and sharing needs of their researchers. Given these parameters, we recommend effective dates no earlier than January 2022, if NIH plans to implement the Policy within the next year.

**Section V: Requirements (limit: 8000 characters)**

We are pleased that NIH is taking steps to implement data sharing requirements for all NIH-funded research. To ensure the Policy meets its goal of increasing the volume and quality of shareable research data, researchers, research administrators, program officers, ICOs, and the public will be well-served by having

* clear guidance, including clear definitions of what constitutes compliance (within both the spirit and letter of the law); and
* clear indications of both incentives for exemplary data sharing practices, and consequences for researchers who remain out of compliance (excluding those whose patients have not consented to having their data publicly shared).

While we appreciate the flexibility given to various ICOs, we are concerned that this lack of formal guidance will lead to inconsistencies in requirements and compliance, and ultimately result in uneven practices of data sharing. We recommend including assessment guidelines for ICOs (e.g. a rubric) to enable transparency in evaluations of Plans and to define what constitutes compliance. We suggest that the Plan elements be used to form the basis for a standard rubric for evaluation.

**Section VI: Data Management and Sharing Plans (limit: 8000 characters)**

* We disagree with Plans being submitted “Just-in-Time.” Such a practice would send a clear message to researchers that planning for data management is not important and is secondary to the research proposal. We recommend that Plans be considered equally important and evaluated during the research proposal review, which would be in line with other federal agencies including the National Science Foundation (NSF).
* We further recommend that Plans be reviewed by the organization’s Institutional Review Board (IRB) to ensure that there are appropriate human subjects protections in place, patient data sharing is clearly consented, mechanisms exist for identifying patient data where sharing is not consented, and data management plans for the sharing of patient data is supported at the organizational level, as ultimately institutions will be held responsible and liable. This additional step will create more administrative burden, and thus should be compensated by NIH in the direct funding of each grant.
* We strongly recommend that the Supplemental Guidance on Plan elements be incorporated into this section because it would send a stronger message that all these elements are important to consider.
* The research community also values materials that are produced alongside the scientific data as they are essential to interpret the data within its initial context, to its reuse, and to its replication. We recommend expanding the requirements to include not only scientific data but also any other essential materials (e.g. code, custom software).
* We support making Plans publicly available which would:

(1) hold the researcher accountable;

(2) allow others to learn and reuse as examples; and

(3) provide a set of textual data for mining and study.

* We suggest developing a database similar to PubMedCentral, that also supports updates to the Plans if and when necessary. Such updates should be version controlled so that changes are tracked and viewed.

**Section VII: Compliance and Enforcement (limit: 8000 characters)**

* While compliance is listed as a requirement, the Policy does not provide specific guidance for ICOs concerning how to develop robust and consistent models for defining compliance, or identifying and addressing noncompliance. As mentioned above in Section V, we are concerned that this could lead to an inconsistent approach among ICOs.
* As with the NIH Public Access Policy, we believe NIH must establish clear expectations and pathways to compliance. For example, adding a component within the Annual Research Performance Progress Reports (RPPR) that requires a description of compliance with the Policy would apply to projects across ICOs.
* We also recommend including a section in the Final RPPR for reporting how well the Plan had been completed or addressed (i.e., this is what was proposed and this is how the data was managed and then shared).
* As with publications, we expect embargoes on data sharing along with information on when and where exactly the data will become available should be included in the Final RPPR.

[**Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing**](https://osp.od.nih.gov/wp-content/uploads/DRAFT_Supplemental_Guidance_Allowable_Costs.pdf) **(limit: 8000 characters)**

* It is unclear if allowable costs are in addition to the award or if these costs should be budgeted at the time of proposal submission.
  + If the former, we recommend specifying caps on costs.
  + If the latter, we strongly reiterate that the Plan should be part of the proposal so that researchers can account for costs at the outset.
* We recommend that guidance for allowable costs include the following specifics:

(1) salary support for personnel dedicated to data management such as a data librarian or information professional;

(2) acceptable repositories and associated costs; and

(3) maximum costs for recurring fees associated with long-term preservation and what amount of time is reasonable.

We strongly request that NIH consider that smaller institutions may not have existing research support in place and enacting the Policy may come with hidden costs and unforeseeable institutional burdens.

[**Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan**](https://osp.od.nih.gov/wp-content/uploads/Supplemental_DRAFT_Guidance_Elements_NIH_Data_Management_and_Sharing_Plan.pdf) **(limit: 8000 characters)**

* As stated earlier, we strongly recommend this section be incorporated into section VI of the Policy because it would send a stronger message that this is important. It would also provide more clarity for specific elements that should be considered as part of writing a robust, sensible, and meaningful Plan.
* We agree with the elements listed here but also reiterate our recommendations submitted in December 2018 in response to NOT-OD-19-014 (https://www.mlanet.org/d/do/13787). More specifically, we strongly recommend:

(1) removing the two page limit as it would restrict much of the required elements from being described in depth;

(2) specifying what version and packages will be used for tools and software;

(3) specifying that metadata be made available even if data will not or cannot be; and

(4) prompting researchers to seek oversight assistance from specialist librarians and information professionals who have expertise in managing data.

**Other Considerations Relevant to this DRAFT Policy Proposal (limit: 8000 characters)**

These comments were developed by Medical Library Association and Association of Academic Health Sciences Libraries members who are or have been actively engaged in supporting researchers with data management and sharing.

* We reinforce our recommendation that enacting this Policy will require more support for education and infrastructure around data management and sharing. We further recommend funding support for training librarians and information professionals.
* We recommend that NIH create either a central data repository or a set of specialized central data repositories for various disciplines. In this scenario, NIH would provide expert data curation and shoulder the cost of long-term data storage, thus meeting several required elements of the Plan. We recognize that this would be a challenge but it would also be very welcome resources, especially for researchers whose data does not fit into any existing repositories. At the very least, NIH should provide a data catalog that aggregates the metadata and locations of the data and its related tools, software, and/or code made available through this Policy.
* Similar to the process for evaluating journals for inclusion in PubMed, we also recommend that NIH provide a list of acceptable vetted data repositories or point researchers to an existing resource such as the Registry of Research Data Repositories (https://www.re3data.org/). In the interest of open access and open science, we believe data should not be owned by a publisher, especially if that data is associated with a publication, and commodified.

### MLA thanks the following members who drafted these comments:

* Marisa Conte, Member, MLA Governmental Relations Committee, Data Caucus, Translational Sciences Collaboration Caucus, and Medical Informatics Caucus
* Rolando Garcia-Milian, Member, MLA Scholarly Communications Committee
* Jessica Koos, Member, MLA Scholarly Communications Committee, Data Caucus, Research Caucus, Clinical Librarians and Evidence-based Healthcare Caucus
* Terrie Wheeler, Interim chair, MLA Information Management Domain Hub and AAHSL member, Research Caucus, Translational Sciences Collaboration Caucus
* Shirley Zhao, Chair, MLA Scholarly Communications Committee, Medical Informatics Caucus, Research Caucus
* Mary Langman, Staff Liaison, MLA Scholarly Communications Committee and Information Management Domain Hub