

Association of College & Research Libraries
50 E. Huron St. Chicago, IL 60611
800-545-2433, ext. 2523
acrl@ala.org, <http://www.acrl.org>



TO: National Institutes of Health
DATE: Wednesday, December 5, 2018
RE: Response to Proposed Provisions for a draft NIH Data Management and Sharing Policy

Submitted online at <https://osp.od.nih.gov/provisions-data-managment-sharing/>

Name: Mary Ellen K. Davis, Executive Director ACRL
Name of Organization: Association of College and Research Libraries
Type of Organization: Professional Org/Association
Role: Institutional Official

Research Area Most Important to You or Your Organization (e.g., clinical, genomics, neuroscience, infectious disease, epidemiology)

The Association of College and Research Libraries (ACRL) is the division in the American Library Association that serves more than 10,000 academic and research librarians and interested individuals working in institutions of higher education. ACRL develops programs, products, and services to help academic and research librarians learn, innovate, and lead within the academic community. We enhance the ability of academic library and information professionals to serve the information needs of students and researchers. For example, through a one-day workshop, ACRL presenters travel to campuses across the U.S. and train liaison librarians to enhance their skills with research data management. As reflected in our previous support for governmental policies and legislation that facilitate open access and open education -- including the NIH Open Access Policy, the Office of Science and Technology Policy mandate, and the Fair Access to Science & Technology Research Act and Federal Research Public Access Act bills -- ACRL is fundamentally committed to the open exchange of information to empower individuals and facilitate scientific discovery.

NIH welcomes comments on any aspect of the issues presented, it is particularly interested in comments on

I. The definition of Scientific Data.

NIH's definition is generally consistent with the 2013 OSTP Memo "Increasing Access to the Results of Federally Funded Scientific Research" (https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf), with OMB circular A-110 (<https://www.whitehouse.gov/wp-content/uploads/2017/11/Circular-110.pdf>) and 2 CFR § 200.315 - Intangible property (<https://www.gpo.gov/fdsys/granule/CFR-2014-title2-vol1/CFR-2014-title2-vol1-sec200-315>), which is helpful to bring alignment across the federal landscape.

The definition in the Proposed Provisions specifically excludes laboratory notebooks and case reports, which would be in agreement with the previous definitions. ACRL believes that case report

forms should not be excluded even though they may contain personnel and medical information of which a disclosure would be an unwarranted invasion of personal privacy. Instead, ACRL encourages NIH to include case reports, other medical records, or data containing PII in the definition of scientific data and clearly note that researchers should share them in accordance with federal policy and other best practices (e.g., HIPAA, restricted sharing, aggregation to a level that will reduce the possibility of disclosure).

ACRL also requests that NIH reconsider the exclusion of laboratory notebooks, as their exclusion is in tension with Section V, Part 1.2 of the Proposed Provisions, which states that the DMP must:

Describe any other information that is anticipated to be shared along with the scientific data, such as relevant associated data, and any other information necessary to interpret the data (e.g., study protocols and data collection instruments).

Laboratory notebooks include recorded information that is “necessary to interpret the data.” NIH should consider requiring that the Data Management Plan address how laboratory notebooks will be managed and how the information contained within them will be shared.

II. The requirements for Data Management and Sharing Plans.

An NIH requirement for a Data Management and Sharing Plans at all funding levels would be a new requirement, presumably overriding what is set out in NIH’s Data Sharing Policy and Implementation Guidance

(https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm). The expansion to include all funding levels, wholly or partially funded by NIH, helps bring NIH in closer alignment with other federal agencies and creates a more comprehensive treatment of data in the funding landscape. A new NIH Data Management and Sharing Policy based on the proposed revisions has the potential to clarify the importance of the data management and sharing by creating mechanisms to ensure that researchers follow it.

Part V provides the potential for stronger compliance and enforcement mechanisms, although it may be worthwhile to consider how the data management and sharing plan compliance could be integrated into eRA Commons and MyNCBI, to create a similar workflow as exists for publication public access compliance via PubMed Central. Moreover, ACRL encourages NIH to provide the guidance for making data shareable via the NIH Data Catalog, or available via PMC and linked to any published articles. Providing guidance and low-burden interfaces to researchers will help adoption of NIH-supported public access methods, which should reinforce the parameters laid out in this proposal.

In Section II, NIH proposes that scientific data should be “made accessible in a timely manner for appropriate use by the research community and broader public.” It goes on to state that any new NIH policy would establish requirements for responsible management and sharing. We suggest that any policy NIH creates should have a clear definition for what “timely” and “appropriate” mean. Given the diversity of domain engagement with NIH, “timely” may have very different interpretations by the community. Looking to other federal agencies for precedent, directorates across NSF have dictated the embargo periods in the data management plan guidance.

Within the Proposed Provisions (Section IV), NIH suggests that Data Management Plans (DMPs) remain an Additional Review Consideration. Although this is one method for considering DMPs, because Additional Review Considerations are not individually scored and do not influence the overall score, ACRL encourages NIH to consider designating the DMP as Additional Review Criteria and incorporating review of the DMP in the overall impact score. Failing that designation, ACRL encourages NIH to expand upon when and to what degree this integration would be appropriate.

A well-conceived and well-described DMP requires significant investment of time for grant applicants and conveying such may well require more than the proposed limit of two pages. Although this could be required at the time of submission, it would be more reasonable to require the detailed DMP as a condition and term of the award. A detailed DMP required at the time of award would outline specifics that would be incorporated into the terms and conditions, and NIH could provide support to ensure that investigators' plans are appropriate and actionable.

Relatedly, it is impossible to predict changes in technology standards over the life of a research grant. ACRL suggest that NIH explicitly allow the DMP to be revised as part of the annual report process. This would ensure that researchers are following the most up-to-date standards and increase the appropriate and successful preservation of data.

Section IV part 2 adds that, "the inclusion of scripts may be helpful." ACRL encourages NIH to include a stronger statement requiring the inclusion of scripts and require a justification from the researchers as a decision to use non-open source software and code. Access to scripts (which would include having access to open source software used to create and run them) is necessary for research reproducibility.

Section IV part 4 states that, "If an existing repository will not be used, indicate why not and how scientific data preservation will be assured (e.g., in a newly created repository or by the investigator's organization)." ACRL encourages NIH to offer more explicit guidance to the researcher explaining what minimally adequate preservation (e.g., exhibit a sustainable funding model, provide a succession plan) is acceptable, as long-term preservation requirements are not common knowledge across all researchers. Section V should include a requirement for long-term planning that "meets community-based standards at the time of deposition."

III. The optimal timing, including possible phased adoption, for NIH to consider in implementing various parts of a new data management and sharing policy and how possible phasing could relate to needed improvements in data infrastructure, resources, and standards.

With robust guidance and infrastructure in place from NIH, a year of community preparation could be sufficient to bring about adoption of this proposal. Researchers seeking NIH funding may have some experience in planning for the sharing of data from funded research, but a new policy as proposed in the Proposed Provisions would represent a significant change for grant applicants. We recognize the need for additional clarification and support for investigators seeking funding due to the inherent difficulties in writing a thorough and actionable DMP. NIH should provide clear guidelines and recommendations for researchers, including working with their research support partners, such as the library, on campus.

Scientific data standards are an area in which researchers may need additional information. In support of Section IV part 3, NIH could provide more assistance to proposal authors to help them better understand existing data standards, which common data elements would be appropriate, and how they should be applied. Providing tutorials or other learning objects in the call for proposals could help disseminate information to researchers. Providing embeddable learning objects also allows for librarians and other research supporting offices to reinforce these standards through other delivery avenues.

Section III of the proposed provisions states that, “[r]easonable costs associated with data management and sharing could be requested under the budget for the proposed project.” There may be significant costs associated with implementing a quality data management and sharing plan, and ACRL applauds the NIH acknowledging this in the budget allowance.