December 10, 2018

Via Electronic Submission

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

Yale Collaboration for Research Integrity and Transparency
Comments on NIH's Proposed Provisions for Draft NIH Data Management and
Sharing Policy, NOT-OD-19-014

The Yale Collaboration for Research Integrity and Transparency (CRIT) supports the

effort to update the NIH's 2003 Data Sharing Policy. The NIH has been a leader in

requiring data sharing for funded projects, for sharing data, and for creating and

implementing standards that can be used across fields.

Below are our comments on the three specific areas on which comment is sought.

Section I   The Definition of Scientific Data

We are concerned that the proposed definition of scientific data varies from the

definition in the current *NIH Data Sharing and Implementation Guidance* (October

2003)*,* which requires sharing of "final research data." Current NIH policies, including

the *Frequently Asked Questions (FAQs) on Data Sharing* (February 16, 2004), the *NIH*

Grants Policy Statement: Access to Research Data (2010), and the *National Institutes of*

Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data

Form HIH Funded Scientific Research (February 2015), all refer to the requirement to

share "final research data" rather than "scientific data." We urge the NIH to modify all current NIH policies to adopt the term scientific data in place of final research data. This will more closely mirror current NIH data sharing policy, and clarify that data sharing of final data, rather than preliminary or incomplete data, is required. We also urge the NIH to clarify that data sharing requirements for longitudinal and cohort studies begin with the completion of data collection for each wave or module, and that data sharing is required for clinical trials within 1 year of primary outcome ascertainment.

Section II   The requirements for Data Management and Sharing Plans

We are fully in support of requirements for data management and sharing plans, and in particular, of the proposed requirement that a data management and sharing plan be made as part of the funding application process. Inclusion of a detailed data management and sharing plan at the proposal phase, at the level of detail contained in the proposed definition, is necessary to ensure that data is collected, managed and preserved using commonly accepted standards to ensure its eventual availability and usability by secondary users. The requirement contained in the proposed definition for identification of a data archive at the proposal stage is also welcome, and should be included in the final definition.

Section 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.

The current NIH general policy, in effect since 2004, requires data sharing for proposals seeking $500,000 per year in direct costs. Although various NIH subsidiaries have included proposals seeking lower direct costs when imposing data sharing requirements, the current proposal marks a significant change by requiring data sharing for all proposals, regardless of annual direct cost amount. We applaud the NIH for removing the minimum of $500,000 per year in direct costs from the requirement that data be shared. However, we recommend that the NIH carefully consider the best way to phase in data sharing requirements and ensure that already funded studies are eligible to request reasonable costs associated with data management and sharing.

The *National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data Form HIH Funded Scientific Research* and the *NIH Strategic Plan for Data Science* contain information that can be used to develop detailed standards for implementation of the new data management and sharing policy.

Our experience with data archiving leads us to propose that the NIH institute specific requirements to ensure that data collected is of high enough quality to be acceptable to data archives, and to be usable by secondary analysts. Proposals should be required to include a budget for implementation of the data preservation and access requirements of this policy, including both ongoing work to create and preserve materials to be archived, as well as final preparation of the data for archiving.

We also urge the NIH to include a provision for the final payment of grant funds to be made upon the deposit and acceptance by the archive of the final research data from the project.

All too often, researchers may exhaust their grant funds prior to archiving data, resulting in either archiving of data that does not meet the standards for use by secondary researchers, or that does not meet archive standards for data deposit.

Finally, we urge the NIH create data repositories across all of its institutes to ensure that all NIH-funded research is shared by repositories employing best practices for data archiving, safeguarding and sharing, such as those employed by NHLBI's BioLINCC.

We appreciate the opportunity to comment on the draft policy.

Sincerely,

Margaret E. McCarthy
Executive Director