Lamar Soutter Library Response to the National Institutes of Health (NIH) Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research

Request Notice Number: NOT-OD-19-014

The definition of scientific data:
We agree with the definition of scientific data as proposed, with the exception that we feel for the purposes of a Policy, scientific data includes individual level and summary data, as well as metadata. As written, the definition states that it may include such. Without this additional data, including metadata, the scientific data shared will fail at meeting the guidelines of the FAIR principle (findable, accessible, interoperable and reusable) for scientific data management.

The requirements for Data Management and Sharing Plans:
The proposed policy statement lacks any mention of (1) how NIH will know if the data sharing requirements have been met and, related, (2) data citation. Will a system that generates something similar to a PMCID (PubMed Central Identifier) be created? Will an identifier, e.g. a digital object identifier, be assigned to data sets so that others can both locate the data and attribute the creators of it accordingly? We feel these are significant gaps in the Policy at this point.

Regarding the archiving of data, we agree with encouraging researchers to deposit in no-cost repositories, however there is limited detail in how the situation will be addressed if no such repository exists. This is likely a question that researchers will have and thus is important to state from the beginning.

The creation of a checklist for preparing a data set is helpful. Providing details as to the minimal requirements for data will make it easier for researchers and their staff to prepare the data throughout their research process. We suggest the NIH’s policy include a strong working connection with the Office of Sponsored Research and institution’s library, in order to assure compliance with the proposed Data Management and Sharing Plan’s policy and procedures.

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.
We view issues related to the release and implementation of the Policy through the historical lens of the NIH Public Access Mandate, meaning lessons learned from instituting the Mandate can well-inform the same for the Policy. A phased approach, while giving researchers and administrators time to understand the issues and create the necessary new steps within their research process to share their data, also lacks the incentives necessary for adoption. As experienced in the Public Access Mandate, this type of approach resulted in confusion and delays in researchers’ compliance. Expectations of researchers
should be clearly stated and enforced from the beginning, rather than incremental steps along the way.

As proposed, making NIH-funded research data available “in a timely manner” is vague and unhelpful for anyone seeking to comply with the Policy. Be specific regarding the time allowed between generation of data, publication, and the release of the data for public use. Embargos are certainly fair, but they need to be defined clearly. Similarly, acceptable places for data deposit need to be stated, in fact, we feel it would be helpful to researchers, administrators, and those assisting with the process, if the Policy offered a list of archival entities available. Some are surely already aware of discipline-specific repositories where data can be deposited, but not everyone will.