10 December 2018

Response to Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research

The International Association of Scientific, Technical and Medical Publishers (STM) is the leading global trade association for academic and professional publishers. It has more than 150 members in 21 countries who each year collectively publish more than 66% of all journal articles and tens of thousands of monographs and reference works. STM members include non-profit scientific and scholarly societies, commercial publishers, and university presses who work collectively to ensure broad access to and use of the latest scientific and scholarly information. The majority of our members are small businesses and not-for-profit organizations, who represent tens of thousands of publishing employees, editors and authors, and other professionals across the United States and world who regularly contribute to the advancement of science, learning, culture and innovation throughout the nation. They comprise the bulk of a $25 billion publishing industry that contributes significantly to the U.S. economy and enhances the U.S. balance of trade.

Publishers sit at the interface between researchers, their research and the rest of the world through our work to improve the quality and availability of information related to research. STM shares our members' commitment to supporting researchers in the sharing, discoverability, and reuse of research data. Individual publishers are developing tools and services to support researchers to make their data FAIR (Findable, Accessible, Interoperable, and Re-usable), and have actively responded to community demand for citation principles for data. STM itself has been involved in numerous projects looking at data access, citation, and preservation, the most recent example of which is support for the development of SCHOLIX, an easy and universal linking mechanism between scholarly publications and research data.

In keeping with our commitment to promote sustainable open science, STM therefore supports NIH's efforts to improve data management and sharing, and welcomes this opportunity to respond to NOT-OD-19-014, "Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research," as published on October 10, 2018. This submission builds on responses that STM has submitted to previous NIH RFIs on research data and digital repositories, as well as responses that STM has submitted to two RFCs on the Federal Data Strategy.

The following comments are in response to the specific areas upon which NIH has requested that respondents focus, and have been submitted to the appropriate boxes in the response form as well as being copied here.

I. The definition of Scientific Data (Provisions I, II, and III)

STM supports the intent of NIH in promoting data sharing in order to improve reproducibility and transparency, and we are pleased to see that the goal of validation and replication is included in the definition of data. In looking at what information is necessary for this goal, it is critical to distinguish between data itself and various types of presentation of data, and appropriately consider a researcher's rights to data generated in his or her research, as well as to respect intellectual property protection and copyright laws. The Data Publication Pyramid on p. 6 of the "Report on Integration of Data and Publications" (http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf), written by a coalition representing researchers, publishers, libraries and data centers, is a comprehensive overview of research data and sharing, and many of the issues involved, but, tellingly, the report does not itself provide a definition of data.

A definition must be precise enough to make a distinction between the data and various interpretations and presentations of that data, whilst at the same time being flexible enough to encompass the data practices of a wide variety of fields. It should also be consistent with other descriptions of data in federal policy and code. The definition proposed in the "Proposed Provisions for a Draft NIH Data Management and Sharing Policy" has a significant benefit in being similar to that in the 2013 OSTP memo on "Increasing Access to the Results of Federally Funded Scientific Research," upon which other federal agencies have built their data management policies. At the same time, it would be helpful if NIH would further clarify the meaning of data as primary information and not analyses or creative presentations of the information.

Our recommendations are:

(1) that "data used to support scholarly publications" be modified to "primary data that support the finding presented in scholarly publications"

(2) that the list of materials that are not considered data include all "analyses" and all "versions of scientific papers" rather than only "preliminary analyses" and "drafts of scientific papers."

II. The requirements for Data Management and Sharing Plans (Provision IV)

STM's members publish in a wide variety of research areas, each of which has different practices with respect to data collection, use, and sharing. The plan requirements must be flexible enough to support the diverse nature of the research that NIH funds, while also providing guidance to all researchers to encourage and enable sharing. STM welcomes the opportunity for dialogue with NIH and all stakeholders to find ways to increase the impact of research data.

With the diversity of data practices and differences in the intensity of data usage in different fields, it may not be appropriate to limit data management plans to two pages in all cases. NIH may want to consider providing the limit as a guideline, or adjusting it in the case of multi-institutional or more complex data plans.

With respect to data preservation and access, NIH may want to provide guidance to researchers on criteria for an appropriate and trusted location for data, including plans for perpetual access and commitment to the FAIR Data principles. Several initiatives offer certification for or

recommendations of trusted data repositories, including CoreTrustSeal (https://www.coretrustseal.org/) and Repository Finder (https://repositoryfinder.datacite.org/about; https://www.re3data.org/).

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. (Provision V)

Whatever the timing of the implementation, STM believes that it is critically important that as the policies are implemented a robust and regular review and evaluation takes place with extensive stakeholder input. Data sharing policies are likely to have profound effects on the research enterprise which will not be fully understood until they are implemented. Regular review will enable NIH to address any unintended consequences in a timely manner, as well as help take advantage of changes in research practices or technologies.

STM, through its involvement in RDA and other initiatives, is already contributing to the development of the standards, resources, policies, and infrastructure needed to enable robust data sharing across the research community. We welcome further discussion on how NIH, STM, and our member publishers can work together to build trust in science and promote the use of research data for the benefit of research and the public.

Very truly yours,

Michael Mabe

CEO STM