COGR

Council On Governmental Relations

An Association of Research Institutions

Dr. Andrea Jackson-Dipina
Director of the Division of Scientific Data Sharing Policy
National Institutes of Health
Office of Science Policy
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

January 10, 2020

Subject: Comments to Draft NIH Policy for Data Management and Sharing

Dear Dr. Jackson-Dipina:

The Council on Governmental Relations (COGR) is an association of 188 research universities and affiliated academic medical centers and independent research institutes. COGR concerns itself with the impact of federal regulations, policies, and practices on the performance of research conducted at its member institutions.

We thank you for the opportunity to respond to the Draft NIH Policy for Data Management and Sharing and Supplemental Draft guidance (NOT-OD-20-013). COGR recognizes the importance of data sharing and generally agrees with the NIH's draft policy. However, we believe that in order to promote a culture of data-sharing across all scientific disciplines, the NIH should also prioritize developing resources and tools to better facilitate data sharing.

COGR is pleased to see the proposal to submit data management and sharing plans (DMSP) as part of the ''Just-In-Time'' (JIT) documentation for extramural awards. Requiring submission of the DMSP at the JIT phase rather than at the proposal stage minimizes administrative burden for both the applicant and peer reviewers. We also assume that the details of the plans will not be considered as part of merit review.

While submitting DMSPs during JIT will allow researchers more time to focus on the science being proposed, one potential drawback is that it will be challenging to accurately budget data management costs for a plan at time of application when the details will be later finalized with NIH Program staff at JIT. We therefore recommend that NIH allow additional data management costs to be added to the budget at JIT based on the final negotiated DMSP. We also recommend an option that allows grantees to appeal NIH Institute, Center, and Office (ICO) mandated data sharing requirements to the NIH Policy Office should the requirements be considered unreasonable or inappropriate by the grantees involved, without fear of reprisal.

Furthermore, because many institutional offices are involved in reviewing and approving components of DMSPs, having feedback available on the status of the NIH review of the plan for those involved in the development and review process at the institution will be extremely helpful in order to manage a plan.

We appreciate that the Draft Policy allows for necessary flexibility across scientific disciplines by outlining minimal specific expectations for the NIH-wide DMSPs. We are concerned, however, that allowing each of the 27 ICOs to create separate supplemental requirements will create confusion in the awardee community. We urge NIH to harmonize among the ICOs via the use of a consistent format for collecting the minimal requested DMSP information. A common form for metadata organization with standardized data fields would also be helpful to ensure that the same relevant metadata is obtained for each study. Consistent collection of appropriate metadata (such as conditions under which studies were conducted, information about the research subject populations, and journal citations) may also enhance aspects of reproducibility.

For efficiency purposes, we further recommend that NIH establish a centralized location to host ICO-specific requirements as opposed to individual institute websites. One central location for all NIH information pertinent to data sharing would improve transparency and monitoring of practices for both public and grantee communities.

Allowing researchers to create the specific plans applicable to their data is important to ensure that data are not made public before any security or privacy restrictions or concerns are addressed. We thus strongly recommend that NIH include in the policy or in its implementation appropriate options to address the myriad legal, ethical, technical, security, and privacy considerations that may impact data sharing. We further recommend that NIH provide resource information to help researchers and the public understand the meaning and implications of these various restrictions. Leaving the coordination of restrictions across sensitive data sets to researchers alone could add significant unfunded administrative burden.

Thinking more broadly, NIH has the unique opportunity to lead the community by creating field-specific data repositories that capture data elements and metadata that are relevant for that field and have the added benefit of ensuring that relevant security and privacy concerns are addressed. NIH-led data repositories would also allow both the agency and the awardees

to leverage resources, avoid duplication and disaggregation of valuable knowledge, and curate and provide data in ways that maximize the public benefit.

We appreciate NIH's recognition of protections for scientific data generated from humans or human biospecimens and ask that NIH explicitly acknowledge the role of the Institutional Review Board (IRB) in the review and approval of DMSPs, and in ensuring that such plans are appropriately disclosed in informed consent materials. NIH may want to consider the existing NIH Genomic Data Sharing (GDS) Policy and related guidance as a model, as it provides a framework for IRB considerations such as risks associated with data sharing and evaluation of informed consent, including identification of circumstances where informed consent may not adequately address data sharing. There must be consistency between the plan and the informed consent obtained from human participants.

We also ask NIH to consider issuing guidance on standards for uncontrolled access, de-identification, application of the NIH Certificate of Confidentiality Policy, consequences of participant withdrawal and ability for a participant to decline data sharing, and how requirements such as the Health Insurance Portability and Accountability Act, the European Union General Data Protection Regulation and other data protection laws apply, especially as the data could ultimately be used for commercial purposes through uncontrolled access.

The Draft Policy indicates that "*non-compliance with the NIH ICO-approved Plan may be taken into account by the funding NIH ICO for future funding decisions for the recipient institution.*" It would be helpful to gain clarity on how non-compliance will be assessed by NIH, particularly since (i) a DMSP is by definition a *plan*, subject to change, (ii) implementation of the DMSP is dependent on the progress of the research, and (iii) the DMSP requires descriptions such as anticipated timeframes and anticipated agreements that could limit the ability to share scientific data broadly. For example, if deposited data were not yet analyzed and ready for publication, the approved DSMP is unlikely to meet the overall intent of "reproducibility".

The ability of NIH to make a finding of non-compliance any time after the end of the funding period creates, in effect, an unlimited and perpetual compliance obligation for PIs and grantees. We therefore recommend findings of non-compliance be limited to failures to follow a DSMP or other actions related to data sharing and management *during* the funding period. Alternatively, NIH could consider whether the policy applies to the data set that is available at the end of the funding period, or whether the data desired and requested must necessarily rely on more fully contemplated resources needed after the end of the award period. One potential solution would be to create a data sharing mechanism using modular budgeting that could be a

supplement and extension to every award – *de facto* adding a sixth year to each standard R01 or an appropriate equivalent for each funding mechanism.

The Draft Policy contains the following statement, "*NIH encourages shared scientific data to be made available* ***as long as it is deemed useful*** *to the research community or the public*". While we appreciate that NIH encourages rather than requires this practice, as the determination of usefulness is necessarily a subjective one best made by the investigator or other experts in the same field, we are concerned NIH's "encouragement" will be used as a factor in approving a DSMP and in determining compliance with the DSMP. We ask that this statement be removed from the Draft policy or explicitly emphasized that it is only an encouragement.

On a related point, if a repository with recurring fees is the only viable option, the grantee could find itself needing to cover those costs once the project is over, potentially for years. We ask NIH to continue to discuss the allowable costs guidance of the data sharing policy with stakeholders at future roundtable meetings or other public forums.

We note the Draft Policy applies to all scientific data generated from NIH-funded research and is written with the expectation that reasonable efforts will be made to digitize all scientific data. The February 22, 2013, memo from OSTP to departments and agencies significantly applies only to digital data. This expectation that non-digital data will be digitized creates a new, complex and potentially costly burden for NIH and grantee institutions and could serve as a disincentive to participate in research.

The Draft Policy indicates that the plans should consider the life of the scientific data, and we applaud NIH for recognizing that each scientific discipline may have different life cycles for data. However, all fields will be affected by evolution of technology, which over time will render the current hardware and software necessary for accessing data obsolete. Migrating data to be compatible with future technology will be costly. In its policy guidance, NIH should recognize that technological changes are inevitable and should not require investigators to attempt to predict such changes nor require institutions to incur such costs in the future.

The recommendation to apply this Draft Policy to ***all*** projects, instead of those above the current $500K threshold, will require significant additional resources, training, and time to implement. We ask that NIH choose a policy implementation date far enough in the future to allow the grantee community to prepare sufficiently. We recommend that implementation be effective with new proposals submitted in NIH fiscal year 2021 or later, assuming NIH releases

the final Policy by March 31, 2020. A delay of at least one year for the "effective date" will benefit all parties involved by allowing sufficient time to effectively implement the Plan against the standards to be established by the ICOs. We also ask that NIH provide a reasonable embargo period for data to provide for intellectual property protection. Finally, NIH should consider clarifying that the policy does not apply to awards (or activity codes) for which no data management plan is required to be submitted as a condition of the award.

We would suggest that NIH take into account the feedback that will be received by OSTP in response to its current Request for Information, particularly with respect to research rigor and reproducibility. Prior to the implementation of the policy, NIH should consider the creation of a Good Research Data Practices (similar to Good Clinical Practices) standard that addresses DMSP, including standards for data collection/design/purpose and archival standards.

Lastly, COGR recommends that NIH consider the issues and potential solutions related to data sharing raised in the publication "Good Practices for University Open-Access Policies" published by the Harvard Open Access Project. While this work was primarily aimed at open access for scholarly articles, its principles can also be applied to data sets.

Thank you for the opportunity to comment. If there are questions, please contact Jackie Bendall at jbendall@cogr.edu.

Sincerely,

Wendy D. Streitz
President