MEMORANDUM

TO: **Office of Science Policy, National Institutes of Health**

FROM: **Association of American Universities**
                 Contact: Katie Steen, katie.steen@aau.edu; (202) 789-5377

                 Association of Public and Land-grant Universities
                 Contact: Kacy Redd, kredd@aplu.org; (202) 478-6022

                 Council on Governmental Relations
                 Contact: Jackie Bendall, jbendall@cogr.edu; (202) 289-6655

DATE: December 10, 2018

Re: 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"

On behalf of the over 200 universities we represent, the Association of American Universities (AAU), the Association of Public and Land-grant Universities (APLU), and the Council on Governmental Relations (COGR) greatly appreciate the opportunity to comment on the National Institutes of Health (NIH) proposed policy provisions and proposed required data elements for data management and sharing plans.

Data access, preservation, and management are complex and emerging areas of the research pipeline. Universities agree it is beneficial to make data from federally-funded research accessible both to the public and others in the research community to accelerate scientific discovery by making data more open to scrutiny and re-analysis. We are actively working with our campuses to help them fulfill agency requirements and continue to engage with federal agencies as they develop additional data access policies.

To build on a report released by AAU and APLU in November 2017 with recommendations for universities and federal agencies, AAU and APLU hosted an NSF-funded workshop in October 2018 on *Accelerating Public Access to Research Data.* The workshop convened federal agency representatives and 30 institutional teams comprised of senior research officers, data librarians, general counsels, information technology specialists, faculty members and other university administrators. The goals of the workshop were to better understand the challenges to data sharing and identify opportunities for collaboration and alignment, all in support of developing campus action plans to advance data sharing. Our universities found it helpful to discuss data access and management with stakeholders across the federal government and campuses. Representatives from NSF, NIH, Department of Energy, National

Institute on Standards and Technology, the Department of Defense, and OSTP also expressed that it was very valuable to hear the barriers and opportunities to data sharing from this cross-campus community.

We were only able to convene 30 institutional teams from primarily research-intensive institutions at this workshop, although 52 universities had applied to participate. Participants agreed there is need for additional workshops for agencies and institutions to discuss better and new ways to collaborate and to build upon the momentum generated by this first workshop. We recommend that NIH support and provide similar venues for universities--especially from a diverse mix of institutions that conduct federally funded research--and agencies to discuss public access to research data before releasing final provisions.

The Definition of Scientific Data

We support NIH's proposed definition of scientific data and that it should not include preliminary analysis, lab notebooks, and other early outputs of the research process, and should primarily consist of "individual level and summary or aggregate data, as well as metadata."

Additionally, we request that NIH provide standards for data and metadata, to the extent possible, to ensure data is reproduceable and user-friendly. If NIH is aware of standards within disciplines that it supports, it would be helpful to share these standards with the broader research community.

Requirements for Data Management and Sharing Plans

Plan Review and Evaluation

We appreciate that NIH is *not proposing* to factor data management plans into the overall impact score for extramural grants. Standards for data sharing vary significantly by discipline and both researchers and NIH program managers and reviewers may not have the data expertise to write or evaluate data management and sharing plans. Given the complexity and changing landscape of data sharing and that research plans sometimes change as novel discoveries are made, some plans may need to undergo considerable edits that warrant significant dialogue with the program manager and reviewers over the lifetime of the grant. It is critical that NIH creates and sustains opportunities for potential grantees and program managers to discuss the content of the plans and any concerns without it being factored into the impact score.

We are concerned that while researchers are experts in their disciplines, they may not be experts in data management and sustainability. Specifically, researchers may not have agreed upon "community data standards" in their discipline, making it difficult to propose where and for how long data should be stored. If data expertise is required on the part of the researchers, universities would have to spend additional resources on the writing of data management plans before applying for NIH funding.

We support NIH's proposed annual reviews of data management plans and interpret this to mean reviews will not continue after the grant has closed. In addition, it is unclear in the proposed provisions, if any party (researcher or university) is responsible for compliance with the data management plan after the grant has ended. In our report recommendations for federal agencies, we suggest that federal agencies "consider the community of interest and duration of usefulness for the data in question and

make retention and access requirements clear."[1] Clarity on this issue would be helpful to determine the kind and amount of additional resources universities will need to allocate to fulfill the agency requirements.

Plan Elements

Given the number of requirements outlined in the proposed elements of the plan, we do not believe two pages is enough for researchers to accurately and fully outline appropriate data management and sharing plans. Additionally, we think it will be very difficult to estimate the amount of scientific data that will result from NIH-funded research in advance of conducting the research. Many research projects require unexpected but necessary experiments, making it impossible to accurately estimate the amount of data resulting from the research. We also seek clarity from NIH on the standards researchers should reference in their plans and if these standards will be defined by NIH or others.

Data preservation and management standards are still evolving across the scientific research enterprise. Best practices around long-term data preservation and access are still being defined in some disciplines. If a university is required to ensure long-term access to and preservation of data, it is important that there be a mechanism to update and change how existing data is stored and plans for continued preservation. For example, if a repository shuts down, a university (assuming the onus is on the university) should be able to contact NIH and make a new plan without penalty. It would be helpful for NIH to provide guidance on what types of "newly created repositories" would be accepted. We would also suggest that NIH create its own data repository to host any NIH-funded research data. This would be especially useful for disciplines that do not already have "community repositories." Additionally, universities need clarity on if the timeline for data preservation and access is expected to continue after the grant is closed.

Many researchers will find it difficult to anticipate the commercialization opportunities before a research project starts. If NIH requires potential grantees to describe limitations on the data use with respect to intellectual property and licensing agreements too early, researchers may choose to claim that all research has commercialization potential, resulting in an unintended consequence of the NIH data sharing policy as currently proposed.

The optimal timing, including phased adoption, for NIH to consider in implementing various parts of a new data management and sharing policy.

Given the requirements for developing plans to manage and make data publicly accessible under NIH's proposed policy, it is critical that universities and researchers understand what additional expenses would be allowed as "reasonable costs" under the new policy. Clarity from NIH on this issue would help researchers in proposing their data management plans. We would strongly urge that "reasonable costs" be a standard part of the grant proposal submission.

Implementation of these proposed provisions and changes to data plans will require significant time and resources on behalf of the university and the researchers. To fully prepare for these changes, we request a two-year implementation period for the new policies. These two years would give universities

---

[1] https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Intellectual-Property/Public-Open-Access/AAU-APLU-Public-Access-Working-Group-Report.pdf

time to facilitate additional discussions across institutions to determine meaningful and effective data access collaborations while engaging with NIH and other federal agencies. We would also strongly encourage the harmonization of data policies and the required elements of data management plans within the NIH and across federal agencies.

We appreciate the opportunity to comment on the proposed policy provisions and data plan elements. AAU, APLU, and COGR encourage NIH to host convenings where universities and NIH staff can engage in dialogue about these issues before releasing final provisions.