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

DATE: January 10, 2020

Re: NOT-OD-20-013: "Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance"

On behalf of the over 200 universities we represent, the Association of American Universities (AAU) and the Association of Public and Land-grant Universities (APLU) greatly appreciate the National Institutes of Health's (NIH) efforts to seek public comment on the proposed Draft NIH Policy for Data Management and Sharing. Consultation and engagement with university stakeholders is critical to developing and implementing successful data sharing policies.

Our associations agree it is beneficial to make data from federally funded research available to the public to accelerate scientific discovery and ensure research integrity through robust replication and re-analysis. At the same time, it is imperative we achieve the appropriate balance between public access and privacy to support and enable scientific inquiry. Appropriate data sharing and access requires significant consultation, collaboration, and investment by federal agencies, universities, scholars, and the research community more broadly. Policies should not only support access to data but enable reuse through adherence to FAIR data principles. AAU and APLU are actively working with our member campuses to develop appropriate campus policies, practices, and guidance to enable public access to research data.

To support public access to data on our campuses, we hosted an NSF-funded workshop in October 2018 on *Accelerating Public Access to Research Data* that brought together 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 workshop identified challenges and opportunities for collaboration in data sharing through the development of campus action plans. In 2020 we will host a follow-up convening and two National Summits, funded by NSF with additional support from NIH, to continue this work and create a *Guide* to assist institutions in implementing appropriate data policies and practices.

General Policy Definitions and Requirements

Definition of Scientific Data

The proposed definition of scientific data is too broad. This is a departure from the definition of scientific data in NIH's 2018 Request for Information on Proposed Data Policy [Provisions](#) in that it now includes data "regardless of whether the data are used to support scholarly publications." This expanded definition will be difficult for universities and researchers to interpret and comply with because it requires extensive time and technical data expertise to assess the endless amounts of data that may be generated over the life of a grant.

The amount and type of data necessary to validate and replicate research findings is, in many cases, subjective and varies widely across disciplines. Without appropriate guard rails in place, this definition may result in the sharing of large swaths of data that are unnecessary, costly, and burdensome to manage and share. Additionally, maintaining quality and ensuring FAIR principles will be difficult, if not impossible, if researchers are required to share any and all data used in research findings throughout the project. **To ensure the data researchers share is useable by the research community and broader public, we suggest the definition of scientific data only include data underlying scholarly publications.**

Requirements for Data Management and Sharing Plans

*We support the statement that NIH may request additional information be included with the Data Management Plan***.** We interpret this to mean NIH will make it clear to the grantee if other information is expected, particularly for compliance with the Policy. This approach supports a collaborative relationship between the program officer and the grantee which is essential throughout the life of the grant. An expectation that NIH will indicate the need for more information ensures researchers will be selective and thoughtful about their Plans versus submitting an array of unnecessary information.

Data Management and Sharing Plans

It is helpful for NIH to allow for updates to be made to Plans at regular reporting intervals. This is critical as research projects often change throughout the grant. Researchers should not be expected to adhere to the initial Data Management Plan elements if changes occur during the research project that require new or different approaches to the scientific data produced. Furthermore, we are concerned the proposed Policy does not clearly state researchers may submit costs estimates after initial submission of the Plan. **To account for requirements instituted by the program officer, we suggest NIH allow for additional direct costs to be submitted after the negotiated Data Management Plan is final.**

More clarity is needed around the data NIH encourages grantees to make available. Asking researchers to make data public "as long as it is deemed useful" is not specific enough and will be confusing to researchers. The research community varies on what it deems "useful" data and researchers cannot be expected to know or understand NIH's view of what is useful. **NIH should provide additional guidance as to what this language means or remove it completely.**

Allowable Costs for Data Management and Sharing

The Guidance is confusing in that it allows "local data management considerations" as a direct cost, but then states "costs associated with collecting or otherwise gaining access to research data (e.g., data access fees)" are part of the costs of doing research and therefore not allowed. In addition to

determining what data should be shared and the metadata needed to adhere to FAIR principles, appropriate local data management, curation, and access on university campuses is essential. Establishing appropriate data storage and sharing infrastructure is one of the most difficult challenges facing research universities in their efforts to share data. Given the volume of platforms, repositories, referatories, persistent identifier (PID) generators, etc. used in the research data community, NIH should clarify what this language means.

We would suggest that the use of services and tools like DataCite, ORCID, CrossRef, figshare, and others be allowed as a direct cost in the grant proposal. Many of these tools require membership fees to participate or charge fees for additional services. These entities are critical to local data management on campus and may require significant campus investment through direct fees or human capital.

Elements of a Data Management Plan

We appreciate NIH's proposal to limit the length of Data Management Plans. However, it may be difficult to address all of NIH's suggested elements in the proposed Plan. We welcome the opportunity to work with NIH, researchers, and other campus stakeholders to develop more guidance for Data Management Plans which may include allowing supplemental materials that augment the Plan.

We agree the rationale for decisions about which data should be made publicly available is important in both the research process and as a compliance mechanism. However, this requires a broad understanding of available data services and community standards. The specific rationale may be very difficult for an individual researcher or university to determine and describe accurately. To improve understanding within disciplines and across universities, NIH should consider providing its own guidance and rationale with specific reference to costs, security, and privacy. NIH is better positioned to indicate the appropriate balance between public access and associated costs, security, and privacy. **To facilitate compliance and enhance research quality, we recommend NIH provide specific rationale on the balance between these priorities.**

Conclusion

In summary, it is helpful NIH has acknowledged the variation across disciplines in data standards and the lack of standards in some cases. In addition, the examples of metadata standards and reference to NIH's Common Data Element Resource Portal in the guidance is informative for our researchers and university staff. However, more guidance is needed around allowable costs and "useful" data to ensure full compliance. We also hope NIH will consider changing the definition of "scientific data" to only include data underlying scholarly publications. Finally, we encourage NIH to harmonize Data Management Plan formats and submission processes across the institutes and centers to streamline compliance and accelerate scientific discovery.

We appreciate NIH's dedicated work with the research community to solicit feedback on potential data sharing and management policies. A collaborative approach with stakeholders is imperative to ensure public access to federally funded research outputs and compliance with associated agency policies.