UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

OFFICE OF THE VICE PRESIDENT – RESEARCH AND INNOVATION

OFFICE OF THE PRESIDENT
1111 Franklin Street, 11th Floor
Oakland, California 94607-5200

Submitted through: https://osp.od.nih.gov/rfi-updating-the-nih-genomic-data-sharing-policy/

February 28, 2022

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

RE: UC Comments in Response to the Request for Information on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing Policy (NOT-OD-22-029)

To Whom It May Concern:

I write on behalf of the University of California (UC) system with regard to the Request for Information on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing (GDS) Policy (NOT-OD-22-029) issued on November 30, 2021.

The UC system – comprised of ten campuses, six academic health centers, and three affiliated U.S. Department of Energy national laboratories – stands at the forefront of cutting-edge research and technology development. As a system, UC receives approximately $6 billion annually of extramural awards to support research conducted throughout all UC locations. UC generally receives 5 to 6 percent of NIH's annual appropriations for research, making UC the largest single recipient of NIH funding for academic research.

The UC system supports NIH's effort to facilitate data sharing and recognizes the benefit it offers for advancing public knowledge. We applaud NIH's proactive efforts to update the 2014 GDS policy, given new technologies and capabilities in this domain. We are also pleased to see efforts to harmonize and clarify the relationship between the GDS Policy and the new NIH-wide Data Management and Sharing (DMS) Policy. UC has specific comments to offer on considerations around identifiability, utility of data linkages, and the need for guidance and sample consent language speaking to privacy risks. Most importantly, however, we strongly suggest that NIH provide further opportunities to discuss this Policy. The research community needs more than 90 days to evaluate the implications of any changes to the GDS Policy. More information is provided below.

1. General Comment: Need for Further Public Engagement

While the UC system appreciates NIH's effort to modernize the current GDS Policy, we urge NIH to engage in further dialogues with the public on this Policy. Investigators and academic institutions directly affected by this Policy deserve a reasonable opportunity to evaluate how any GDS Policy changes may impact their research and accompanying need for human subject research protections. Moreover, projected policy changes may have unintended consequences by adding new and additional layers of compliance to an already challenging area. This not only creates institutional burden, but also could introduce new barriers to valuable research in unexpected ways. Without thoughtful and careful analysis, we cannot know how this might play out, and a 90-day comment period is simply not enough time for a thorough assessment among investigators and administrators involved in research, human subject protections, compliance, and privacy. **Therefore, the UC system recommends that NIH hold town halls and listening sessions on each of the proposed areas within the RFI. We believe NIH should not proceed with a draft revised Policy until this takes place.**

2. De-identification

The UC system appreciates NIH's recognition that genomic data requires unique considerations, as the standard of de-identification for genomic data is a moving target. As understanding of genomics and analytic capabilities constantly improve, data that was once marked as de-identified now could be re-identifiable in the future.

While UC supports expanding de-identification options to include the expert determination described at 45 CFR 164.514 (b)(1) as an acceptable method under the GDS Policy, we would like to note that Department of Health and Human Services Office for Civil Rights does not specifically address de-identification of genomic data, which can often be complex. In addition, the current GDS Policy requires that investigators follow both the Common Rule "readily identifiable standard[1]" and HIPAA Safe Harbor[2] method for de-identification. However, the two approaches are not always compatible.

Therefore, we emphasize the need for guidance specific to the de-identification standard for genomics data. Such guidance, rather than a regulatory mandate that may add additional barriers to an evolving area, can offer a principle-based approach for the use and sharing of genomic data, and more importantly considerations around downstream use of such data. Such an approach can provide examples of safeguards and a tiered risk-based framework to help researchers, their institutions, and IRBs navigate issues in an environment where definitions, standards, and legal requirements are continuously changing.

Additionally, we note the need for identifications tools in this space. For example, for projects focused on sequencing microbes that may unintentionally include a subset of identifiable human data, the National Center for Biotechnology Information (NCBI) has been helpful in providing tools

[1] 45 CFR 46.102 (e)
[2] 45 CFR 46.514(b)(2)

for de-identifying such data. It would be important that such tools across NIH continue to be developed, funded, and made available to the community.

3. Use of Potentially Identifiable Information

We recommend that NIH allow submissions of Limited Data Sets as described at 45 CFR 164.514(e) to potentially include in repositories under the GDS Policy. However, we emphasize that if identifiable data are to be shared, consent should be required.

In many cases, research participants provide their data in order to facilitate biomedical discovery. The risk of undermining that goal should be considered alongside other risks. Likewise, the magnitude of risks to individual research participants should be considered in the context of other routinely experienced privacy risks (e.g., location tracking by apps and phones) rather than as isolated risk to be mitigated at all costs. We encourage NIH to be attentive to the independent submissions from UC researchers, such as from Steven Brenner on latent privacy risks in functional genomics data.

4. Data Linkage

UC supports the ability to link participant data from diverse datasets, such as electronic health records, with genomic information. For example, to meet the goals of precision medicine, data linkage can serve as a powerful tool to understand how participants'genetic and molecular characteristics may produce different outcomes throughout their lifetime. Simply securing genomic information in a repository has limited utility compared to linking genetic variants and mutations to clinical outcomes across the lifespan. In that regard, some amount of linkage will always be necessary to advance scientific understanding, but the scientific community needs guidance from NIH on the breadth of data linkages it would allow balanced with tradeoffs for protecting research participants' privacy. **Specifically, the UC system recommends that NIH clarify the exact types of data linkage it would allow and any restrictions on combining certain datasets. Along with this guidance, we ask that NIH provide examples of what is allowable.**

To more concretely address the questions posed in section of the RFI, UC favors permitting data linkage. For datasets that may not meet the Policy expectations, research participants, upon being adequately informed about what will happen with their information and how it will be used, should be able to make the choice of whether or not to provide their data. This approach is discussed further below in Section 5. Consent for Data Linkage.

Importantly, consequences related to breaches to privacy from data linkages should not fall solely on research participants nor investigators demonstrating best efforts to abide by shifting privacy standards and requirements. As NIH moves into a new era of increased genomic data sharing, with potential changes in risks over time, we believe it is important for the agency to develop ongoing training opportunities but at the same time consider how to implement appropriate enforcement actions for misuse or inappropriate sharing or transfer of data, including monetary penalties.

5. Consent for Data Linkage

UC values the importance of properly seeking consent from those participating in research, and disclosing risks associated with data sharing. We agree that data linkages and risk considerations related to data submission to NIH repositories should be disclosed when obtaining consent for protocol under the GDS Policy. In cases where datasets are collected without providing disclosure on data linkages, we recommend that NIH note this to other investigators before providing information from the repository. It is also important to acknowledge that there is uncertainty in long-term privacy protections, and in some cases, privacy protections are not guaranteed for years to come.

The benefits of utilizing consent for data linkages include recognizing the autonomy of participants by affording them the opportunity to optimize the data they are contributing to research and maximizing societal benefits of their participation. Risks related to data linkage will be dependent on the types of data used. They may include an increased risk to confidentiality, possible damage to reputation, employability, criminal suits, among others, depending on the nature of the data.

We ask that NIH develop guidance and sample consent language that address what data linkage means, how it could affect research participants, and whether subjects can still participate in a study without agreeing to data linkage. It is important that NIH be clear on its expectations on the use multiple data sources, including cases where data obtained utilized different consents or lacked consent. Such guidance and sample language would help researchers and IRBs who may not have the expertise in the intricacies of genomic data identifiability.

6. Expectations for Alternative NIH-Supported Genomic Data Management and Sharing Resources that Store Human Genomic Data

The UC system agrees that the use and sharing of large-scale genomic data will continue to grow, and as such, the appropriate tools, resources, and platforms will need to be developed to keep pace with the needs. These NIH-supported resources, whether external or internal to NIH, should maintain appropriate standards and protections for data, and to adhere to existing principles for data access and security. For example, it would be important to have alignment in controlled access models, user authentication, and procedures for managing inappropriate or unauthorized use or access across the systems.

7. Harmonization of GDS and DMS Plans

We support inclusion of genomic data sharing within the DMS Plan, as outlined in the RFI, to avoid the duplicative submission of plans by the researcher. We also agree that these plans should allow for the budgeting for long-term genomic data management and sharing. We recommend that NIH assess these plans as part of Just-in-Time (JIT) documentation for extramural awards. Requiring submission of the plan at the JIT phase rather than at the proposal stage minimizes administrative burden for both the applicant and peer reviewers.

While the DMS Policy sets a more flexible timeline for data sharing, any decision to change GDS Policy data submission timelines should involve careful consideration of potential impacts on policy compliance, data use and re-use, and what is most effective for the genomic research community.

8. Types of Research Covered by the GDS Policy and Other Long-Term Considerations

Training and development awards (e.g., F, K, and T awards) should be excluded from this Policy. Small-scale studies, which may bear added risks to privacy, should be excluded as well.

The UC system recommends that questions on whether and how to integrate data types beyond genetic information, such as transcriptomics, proteomics, microbiomics, or metabolomics, into the GDS Policy should involve further targeted outreach to experts, particularly as standards in these fields are still under development.

Thank you for the opportunity to comment. We look forward to continued engagement on this important issue.

Sincerely,

Deborah Motton, Ph.D.
Executive Director
Research Policy Analysis and Coordination
University of California, Office of the President