



**Association of  
American Medical Colleges**  
655 K Street, N.W., Suite 100, Washington, D.C. 20001-2399  
T 202 828 0400 F 202 828 1125  
www.aamc.org

February 25, 2022

National Institutes of Health  
9000 Rockville Pike  
Bethesda, Maryland 20892

**Re: Request for Information (RFI): Request for Information on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing Policy (NOT-OD-22-029)**

*Submitted electronically at <https://osp.od.nih.gov/rfi-updating-the-nih-genomic-data-sharing-policy/>*

The Association of American Medical Colleges (AAMC) appreciates the opportunity to provide feedback to the National Institutes of Health (NIH) on proposed updates and long-term considerations for the NIH Genomic Data Sharing (GDS) Policy. The AAMC is a nonprofit association dedicated to transforming health through medical education, health care, medical research, and community collaborations. Its members are all 155 accredited U.S. and 17 accredited Canadian medical schools; approximately 400 teaching hospitals and health systems, including Department of Veterans Affairs medical centers; and more than 70 academic societies. Through these institutions and organizations, the AAMC leads and serves America's medical schools and teaching hospitals and the millions of individuals employed across academic medicine, including more than 186,000 full-time faculty members, 94,000 medical students, 145,000 resident physicians, and 60,000 graduate students and postdoctoral researchers in the biomedical sciences.

The AAMC appreciates the NIH's commitment to increase access to the results of research and to accelerate scientific progress through data sharing. AAMC's previous comments<sup>1</sup> to the agency as it developed the Final NIH Policy for Data Management and Sharing (DMS Policy), effective in Jan. 2023, detail our recommendations for increasing meaningful sharing of much of the data generated by NIH-funded research projects. While we support harmonization between the GDS Policy and the DMS Policy to the greatest extent possible, we agree that the genomic sciences require special

---

<sup>1</sup> AAMC comments re: NOT-OD-20-013 "Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental Draft Guidance." Jan. 10, 2020.  
<https://www.aamc.org/media/40536/download>

consideration, and that policies governing sharing of genomic data must keep pace with stakeholder expectations as well as the evolving state of the science. We are pleased to offer comments addressing the specific topics identified in the RFI.

### **Maximizing Data Sharing while Preserving Participant Privacy and Preferences**

Protecting the interests of research participants should continue to be a central principle of the NIH GDS Policy and will require the agency to carefully examine its existing criteria around de-identification and policies for informed consent. The AAMC appreciates the NIH's acknowledgement that "the concept of 'identifiability' is a matter of ongoing deliberation within the scientific and bioethics communities." Particularly when working with certain human genomes, such as those from a rare disease or underrepresented population, it can become significantly easier to link genomic data to an individual. Furthermore, as our understanding of genomics and analytic capabilities constantly improve, data that is marked as de-identified now would possibly be re-identifiable in the future.

We have concerns about the requirements in the current GDS Policy to de-identify human genomic data by applying standards from both the HHS Regulations for Protection of Human Subjects (The Common Rule) [45 CFR 46.102(e)] and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [45 CFR 164.514(b)(2)]. This unclear combination of two completely separate regulations is a suboptimal approach to de-identification. The removal of the 18 identifiers in the HIPAA privacy rule do not ensure de-identification, and as noted in the RFI may actually impede certain types of research due to the removal of data elements, such as granular location data, which may be needed for scientific inquiries.

In future updates, we recommend that the GDS Policy adhere to the standards set forth in the Common Rule to ensure that the identities of research subjects cannot be readily ascertained by the investigator or associated with the data. We note that the revised Common Rule includes a mechanism for periodically revisiting the concept of identifiability in light of currently existing technologies and urge that the GDS policy follow those standards as they develop. In 2016 comments in response to proposed rulemaking for the Common Rule<sup>2</sup>, the AAMC recommended against including specific safeguards for identifiability in regulation and instead suggested the use of "examples of reasonable safeguards, presented within a tiered, risk-based framework" as guidance.

---

<sup>2</sup> AAMC comments re: Docket Number HHS-OPHS-2015-0008, Notice of Proposed Rulemaking: Federal Policy for the Protection of Human Subjects (80 FR 53933). Jan. 4, 2016.  
<https://www.aamc.org/media/11936/download>

Both the scientific community and research participants would be served by the GDS Policy not requiring certain actions for data de-identification but rather focusing on developing optimal and study-specific methods for minimizing identifiability and communicating the expectation to research participants that users of data will not seek to identify individuals, but the potential for re-identification may exist.

Another element that will likely increase patient privacy concerns is the linkage of genomic sequences with data that do not necessarily meet the GDS Policy, like disease outcomes collected in clinical settings. With the rise of personalized medicine, being able to link genomic data to phenotypes, social data, and other information, can greatly increase the value of data and our ability to look for and develop diagnostics, prognostics, therapeutics, and risk assessment—however, these additional capabilities also increase the amount of information linked to each individual research participant.

Permitting data linkage also emphasizes the critical role of the consent and widespread educational processes to adequately inform research participants that their genomic and phenotypic data may be used for future research purposes and shared broadly, and what the associated risks might be. The continual emergence of new analytics and techniques may present research opportunities that cannot be specifically noted in a consent form. Obtaining sufficiently expansive consent is also critical given that locating a research participant months or years after the initial research has concluded is very difficult and itself requires re-identification.

AAMC recommends that the NIH develop sample language and/or consent form templates to clearly explain the rationale for genomic research and data linkage, as well as any associated privacy and confidentiality risks. Ideally, this guidance would assist individual Institutional Review Boards (IRBs) in understanding what should be required of investigators and communicated to research participants.

Importantly, consequences of these risks should not fall solely on the patient—as NIH moves into a new era of increased genomic data sharing and data linkage, with potential changes in risks to research participants, we believe it is important for the agency to formulate and implement appropriate enforcement actions for misuse or inappropriate sharing or transfer of data, including monetary penalties or suspension of current or future funding.

Finally, effective data linkage will require oversight and standards to ensure that data are robust and contain the appropriate metadata. Drawing connections between linked data may require not only processed data, but raw data and the appropriate metadata regarding software and scripts to be able to conduct or corroborate an analysis. While preparing high quality data is expensive and time-consuming, it is also necessary for data integration and meta-analysis.

## **Expectations for Alternative NIH-Supported Genomic Data Management and Sharing Resources that Store Human Genomic Data**

It is clear that the amount of large-scale genomic data being generated will only continue to increase and will accordingly require an expansion of platforms and repositories for data sharing, storage, and analysis. We agree that it is critical for all these NIH-supported resources to maintain appropriate standards and protections for data, and to adhere to existing principles for data access and security. Despite the fact that some of these databases exist outside of the NIH, it is key to maintain controlled access models where necessary, have systems for user authentication, and procedures for managing inappropriate or unauthorized use or access.

Although the need to develop and use new data platforms is clear, it could also result in or exacerbate inequities across the research enterprise. Institutions with more limited resources may not be set up to support investigators who want to share or access data. Typically, institutions require strong existing infrastructure and data scientists to support investigations in the genomic sciences. As NIH invests in awardee institutions to develop resources, it should consider the differences in institutions and their capabilities and ensure that the ability to conduct scientific investigation or the populations able to participate in research are not limited by disparate resources.

### **Policy Harmonization**

In its comments to NIH on the DMS Policy, AAMC emphasized that “It is critical to have as much as harmonization and standardization as possible across the NIH in both the policy requirements and implementation. This includes all grantees as well as consistency in evaluation of compliance and in institute-specific requirements.”

The AAMC appreciates NIH’s intent to harmonize the GDS Policy and GDS Plan elements, submission, and review with the DMS Policy and strongly encourages this harmonization. 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 additionally agree that these plans, including the budget for genomic data management and sharing, should be assessed during the grant review process. While the DMS Policy does set 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.

## Long-Term Consideration of the Scope of the GDS Policy

The ongoing expansion of studies in “omics” to include not just genomics, but also transcriptomics, proteomics, microbiomics, or metabolomics, raise new questions on the ideal scope of the GDS Policy. It is clear that, like genomic sequences, these data can be a valuable source of information especially when linked with other phenotypic data. The AAMC recommends that questions on whether and how these data types are integrated into the GDS Policy should involve further targeted outreach to experts, particularly as standards in these fields are still under development.

We also note the potential benefits of including smaller-scale studies under the GDS Policy. Advancements in artificial intelligence and machine learning make it easier to aggregate and analyze data from disparate sources. Additionally, meta-analysis across multiple smaller data sets can assist researchers in obtaining a broader population sample and including more underrepresented individuals in research studies.

The AAMC is committed to working with the NIH as it updates the genomic data sharing policy to keep pace with advancements and maximize the enormous potential in this field of science, while maintaining protections for human research data. Please feel free to contact me or my colleagues Anurupa Dev, PhD, Lead Specialist for Science Policy ([adev@aamc.org](mailto:adev@aamc.org)) or Heather Pierce, JD, MPH, Senior Director for Science Policy and Regulatory Counsel ([hpierce@aamc.org](mailto:hpierce@aamc.org)), with any questions about these comments.

Sincerely,

A handwritten signature in blue ink that reads "Ross McKinney, Jr., MD". The signature is stylized and includes a small circular mark at the end.

Ross McKinney, Jr., MD  
Chief Scientific Officer