

March 17, 2026

Matthew Memoli
Principal Deputy Director
National Institutes of Health
Department of Health and Human Services
Attention: NOT-OD-26-023
9000 Rockville Pike
Bethesda, Maryland 20892

Re: Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy (NOT-OD-26-023)

On behalf of the American Society of Human Genetics (ASHG), thank you for the opportunity to submit comments to the National Institutes of Health (NIH) Office of Science Policy (OSP) regarding the *Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy*.

ASHG is the world's largest organization for human genetics professionals. Our nearly 8,000 members—primarily based in the United States—include researchers, academics, physicians, technicians, genetic counselors, students, and others advancing the field of human genetics and genomics. ASHG's vision is that people everywhere benefit from the progress and promise of human genetics and genomics research.

Sharing research findings with the global research community strengthens the scientific enterprise by enhancing collaboration, advancing collective knowledge, and increasing transparency with the public. Furthermore, transparency regarding how data may be shared must be an integral component of the consent process for individuals participating in research. Because scientific progress fundamentally depends on the ability to build upon previous discoveries, the ethical sharing of de-identified data and results is essential for accelerating innovation and generating new insights. Accordingly, broad, responsible, and timely sharing of genomic research data and other -omics data (e.g., transcriptomic, proteomic, epigenomic) derived from NIH research is a critical component of sustaining scientific advancement and enabling evidence-based decision-making across research, clinical, and public health domains. We appreciate NIH's efforts to modernize data-sharing policies and establish harmonized and transparent policy requirements for protecting human participant research data.

1. Feedback on Draft NIH Controlled-Access Data Policy

Expectations for Appropriate Data Use

Responsible, broad data reuse is essential to ensuring that patient-contributed data achieves its maximum scientific and clinical impact, while also maximizing the return on taxpayer investment in

federally funded research. We recommend that investigators consent participants under the General Research Use (GRU) category whenever possible. More restrictive consent categories should be used only when investigators can clearly justify why broader use is not appropriate for a given study population or research context.

Furthermore, to ensure responsible stewardship of federal genomic resources, ASHG recommends that NIH clearly establish and publicly articulate the consequences for misuse of controlled-access data at the individual, departmental, and institutional levels. Penalties should be proportional to the severity of the violation and may include loss of data-access privileges, ineligibility for future NIH funding, institutional compliance actions, and, where appropriate, referral for legal or regulatory review; these expectations should be communicated transparently to all entities accessing controlled-access datasets. ASHG also continues to advocate for the robust enforcement of domestic and international ethical standards, including the principle of informed consent for research participants.

Risk-Appropriate Sharing of Genomic Variants

ASHG recommends that the NIH Controlled-Access Data Policy state more clearly that genomic summary results may include variants with allele frequencies as low as a single observed count. Many of the variants identified in large population datasets, particularly those with the greatest functional or clinical relevance, are extremely rare. Allowing these low-frequency variants to be shared is essential to realizing the full scientific value of genomic research.

Some genomic data-generating programs restrict the sharing of allele counts (AC) below a specified threshold (e.g., ACs under 20 in the *All of Us* Research Program) unless an exception is approved. However, the ability to share summary statistics for all variants, including those found in only a single individual, is essential for scientific progress and discovery. Limiting access to information about rare variants would be justifiable only if the associated privacy risks were substantial. Yet, evidence from numerous large-scale genomics initiatives demonstrates that the actual risk of re-identification from sharing such aggregated data is very low. As such, overly conservative thresholds may impede research without providing commensurate privacy protections.

Protecting Participants Through Clear Re-Identification and Data-Linking Safeguards

ASHG also highlights the growing risk that controlled-access genomic data could be indirectly identifiable through linkage to consumer genetic databases (e.g., GEDmatch, 23andMe, and Ancestry.com). Although NIH-managed genomic datasets may have appropriate protections in place, lower-resolution genetic data combined with identifiers held by commercial entities could allow malicious actors to infer identities in reverse. ASHG recommends that all users of controlled-access data be required to formally attest that they will not attempt re-identification under any circumstances, and that NIH should articulate how it will mitigate these risks.

In addition, NIH should make clear that controlled-access genomic data must not be linked to datasets held by federal agencies such as the Social Security Administration, Treasury Department, and Office of Personnel Management. Incidents reported in the past year involving improper interagency data linkages highlight the need for protections. These measures are not intended to impede appropriate, Institutional Review Board-approved scientific research involving linked datasets, but rather to prohibit the use of controlled-access data for unauthorized investigative or non-research purposes. ASHG therefore urges NIH to specify that controlled-access data may not be re-identified or intentionally linked to federal administrative datasets containing direct personal identifiers, and that these protections cannot be overridden by a declared national emergency.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy.

Guidance Needed on Federated Data Networks

ASHG recommends that NIH provide additional guidance on the use of federated data networks to enable secure querying of distributed genomic data stores. Federated approaches have the potential to expand analytical capacity while reducing the need for large-scale data transfers; however, clear standards are necessary to ensure interoperability, reliability, and consistency across platforms. More detailed expectations regarding implementation, security requirements, and governance structures would help the research community adopt these approaches more effectively.

Future Resource and Infrastructure Needs for Secure Data Stewardship

In considering future resource needs, ASHG encourages NIH to clarify how capacity for controlled-access systems is defined. As the nature, size, and location of data storage evolves, new vulnerabilities may emerge. We suggest that NIH assess whether current safeguards remain sufficient as storage infrastructures scale and diversify. This includes evaluating whether mechanisms exist to detect and prevent the uploading of corrupted or malicious files. If current protections do not adequately address these risks, NIH may need to update or expand its security protocols.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access.

Calibrating Data Controls to Scientific Risk

ASHG recognizes that the draft policy is motivated by national security considerations. However, we are concerned that, if implemented too rigidly, the policy could unintentionally fragment the global genomics ecosystem, slow scientific progress, and place disproportionate burdens on international partners who collaborate closely with NIH-supported researchers. For example, the proposal that entire classes of -omics data must always be controlled-access absent explicit open consent is overly broad. Not all data within a category carry equal re-identification or national security risk (e.g., low-coverage sequencing, methylation arrays, proteomics without identifiers, and de-identified clinical measurements).

ASHG emphasizes that sharing non-genomic -omics data (e.g., transcriptomic, proteomic, microbiome data) does not, at present, pose the same level of re-identification risk as sharing genomic data. Based on current scientific capabilities, these data types do not require the same degree of protection. We support the continued sharing of diverse -omics data that enrich researchers' ability to characterize phenotypes, provided these data do not introduce new identifiability concerns.

We encourage NIH's data sharing policy to distinguish between different levels of data sharing and access based on the risk-benefit ratio. For example, full genomic datasets containing protected health information would require controlled access (e.g., dbGaP); however, single variants interpreted for their health relevance can be shared publicly (e.g., ClinVar), even when identified in only a single individual. Similarly, a single variant with associated phenotype could be returned through a federated query with registered access and aggregated counts of variant allele frequencies can be shared publicly (e.g., gnomAD).

As technologies evolve and analytical methods change, we recommend periodic reassessment of re-identification risks to ensure appropriate protections remain in place.

Risk-Appropriate Use of Select Identifiers

As ASHG stated in our comments on the 2022 revision of the NIH Genomic Data Sharing Policy, allowing use of select HIPAA identifiers, when carefully managed, can enhance scientific value without materially increasing privacy risks. Providing flexibility within de-identification policies (e.g., Expert Determination) can enable richer, more informative datasets while still addressing higher re-identification risks for certain populations or settings. Certain HIPAA identifiers, including ZIP codes, age, and specific dates, are particularly important for advancing research on age-related conditions, environmental exposures, and gene-environment interactions. To maintain strong privacy protections, ASHG recommends limiting the use of these identifiers in circumstances where individuals face elevated re-identification risks, such as sparsely populated areas or regions near Tribal lands. We therefore support establishing minimum population thresholds for ZIP code inclusion and excluding ZIP codes overlapping Tribal jurisdictions.

We further emphasize that policies should also safeguard against group-level harms and stigmatization. Protecting confidentiality is essential to upholding ethical research principles, supporting participant autonomy, ensuring respectful and safe participation, and preventing misuse of genetic information in ways that could disadvantage individuals or communities.

Ethical Considerations for Relational and Indigenous Genomic Data

Genomic data inherently conveys information about biological relatives. If an individual provides informed consent for open sharing of their genomic data, this may inadvertently disclose genetic risks or traits of family members who have not consented. ASHG recommends that NIH acknowledge this challenge explicitly and consider whether additional guidance or enhanced consent materials are needed to help participants understand these implications.

Furthermore, ASHG emphasizes that data collected in partnership with Tribal Nations or Indigenous communities must remain under the governance of the respective Tribal authorities, consistent with principles of Indigenous data sovereignty. Policies should clearly affirm Tribal rights to control access, management, and secondary use of data originating from Tribal citizens. Equally important is ensuring that informed consent processes are sufficiently robust to support meaningful understanding of how data may be used and shared.

Assessing Suitability for Open Data Sharing

With respect to genomic data without linked health information, ASHG supports exploring mechanisms that allow limited querying of such datasets even in the absence of explicit participant consent, provided that only aggregate or minimal-risk information is returned. Many clinically generated genomic datasets, particularly those produced during rare disease testing, are generated with either no consent or limited permission for research yet contain information of substantial public health and scientific value. Although the governance of clinical laboratory data may fall outside NIH's direct remit, the clinical community is increasingly seeking federal guidance to inform best practices. Establishing principles or recommended frameworks for enabling controlled, minimal-risk queries (e.g., the ability to determine the existence of a variant or obtain high-level phenotype, or cohort-level descriptors) could significantly improve rare disease diagnosis and variant interpretation. The current reliance on manual, ad hoc exchanges between clinical laboratories is not scalable; high-level federal guidance could help promote more secure, consistent, and interoperable approaches across the clinical genomics ecosystem.

Risk-Tiered Protections That Support Both Security and Scientific Progress

The human genetics and genomics field is advancing rapidly, and these developments create policy needs that are best addressed when the scientific and national security communities work collaboratively from the outset. ASHG believes that national security and biomedical research are not competing priorities and that both can progress simultaneously when policies are informed by scientific expertise and grounded in a shared commitment to responsible data stewardship.

Diseases do not stop at national borders, and policies should ensure that research remains able to address health challenges that affect people globally. Restrictions on sharing controlled-access data with certain countries should be calibrated in a way that protects security while avoiding unintended isolation from international scientific collaboration in NIH-funded consortia. Approaches that encourage interoperability, including the use of shared definitions of controlled access and recognition of trusted research environments, can help maintain alignment with global data governance practices and ensure both security and scientific progress. ASHG recommends that NIH explore a risk-tiered data governance framework rather than categorical “always controlled-access” rules, informed by models like the Global Alliance for Genomics & Health (GA4GH) [Data Use Ontology](#), which offers a structured, risk-tiered method for classifying data sensitivity, guiding appropriate protections, and enabling proportionate access controls.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy.

ASHG agrees that it is appropriate for the revised GDS Policy to focus specifically on human data.

Regarding human genomic data collected under the GDS Policy from biospecimens or cell lines created or collected before 2015, ASHG affirms the continued importance of legacy genomic datasets, which remain essential for population genetics, rare disease research, and long-term longitudinal studies. Policies that allow responsible controlled-access use under Institutional Review Board (IRB) (or equivalent) oversight, along with practical approaches to re-consent where feasible, can help maintain access to these critical resources without creating retroactive limitations. Although consent can be withdrawn for future use, data that have already contributed to completed analyses and published results cannot feasibly be extracted from the literature.

Significant Burden of Data Preparation and Harmonization

As NIH itself contends with unappropriated mandates, researchers face similar pressures when complying with genomic data sharing requirements. Preparing, harmonizing, and formatting datasets for deposition, particularly for large-scale genomic studies, is often a substantial undertaking. These activities require significant personnel time and technical expertise, and their associated costs must be drawn from existing grant budgets. This inevitably reduces the funds available for conducting the scientific research that the grants are intended to support. If NIH were able to provide dedicated support for data preparation, harmonization, and storage, either through supplements, dedicated budget lines, or centralized infrastructure, compliance would become far more feasible and would strengthen the overall quality and consistency of shared datasets.

Given the scope of the proposed policy changes, ASHG supports the use of impact assessments to evaluate potential effects on institutional burden, costs, timelines, and international collaboration, which can help ensure that implementation approaches remain feasible, balanced, and aligned with scientific needs.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers.

ASHG encourages NIH to clearly define the metrics and evaluation criteria that will be used to assess the security of imputation servers. Given the sensitivity of the data these systems may handle, we recommend that NIH require rigorous security testing, including red-teaming exercises or equivalent adversarial assessments, to identify vulnerabilities before servers store or process controlled-access data.

ASHG urges NIH to address the unique considerations associated with imputation panels developed for populations with limited representation in reference panels. These reference panels are essential for improving imputation accuracy and reducing disparities in genomic research. However, their use may be constrained by specific governance requirements (e.g., when data originate from Tribal Nations or other groups with data-sharing restrictions). ASHG recommends that the policy explicitly clarify how researchers may perform imputation using such panels while respecting community-specific restrictions on data redistribution or secondary use. This may include technical approaches such as “impute-but-do-not-retain” workflows, controlled-execution environments, or Tribal-governed servers.

Developing Policy Frameworks for AI Interaction with Controlled-Access Genomic Data

Finally, ASHG encourages NIH to begin developing clear policy guidance on the use of artificial intelligence (AI), including advanced and general-purpose AI systems, in relation to imputation servers and controlled-access genomic data. As AI laboratories increasingly seek access to genomic datasets for model training, and as academic and clinical researchers adopt AI tools for analysis, NIH should establish guardrails that clarify when and how AI systems may access, process, or learn from controlled-access data.

We look forward to the opportunity to be a resource to NIH on this and other genetic research related issues. If you have any questions, please contact Karina Miller, PhD, (Kmillier@ashg.org) at ASHG.

Sincerely,

A handwritten signature in black ink, reading "Susan A. Slaugenhaupt". The signature is fluid and cursive, with the first name "Susan" and last name "Slaugenhaupt" clearly legible.

Susan A. Slaugenhaupt, PhD

ASHG President