Compiled Public Comments on the Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

Guide Notice Number: NOT-OD-26-023

December 17, 2025 – March 18, 2026

Table of Contents

1. Lucina Uddin

Submit date: 12/17/2025

I am responding to this RFI: On behalf of myself

Name: Lucina Uddin

Name of Organization: University of California Los Angeles

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
In my opinion access to all de-identified human subjects data (including brain imaging) should not be subject to NIST requirements, which place a high burden on institutions. NIST security requirements cannot be readily fulfilled by scientists in lower-resourced settings; the requirement effectively prevents data sharing and impedes open science practices.

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

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

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

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

2. Susan Tapert

Submit date: 12/17/2025

I am responding to this RFI: On behalf of myself

Name: Susan Tapert

Name of Organization: University of California San Diego

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

I write as a professor of psychiatry at UC San Diego and PI on several large NIH-funded longitudinal studies (e.g., ABCD, NCANDA) that thousands of credentialed investigators access each year. I fully support NIH's mandate to safeguard participant privacy and data integrity. However, the blanket requirement that every institution handling controlled-access data achieve full NIST SP 800-171 compliance is proving disproportionately burdensome and, in practice, counter-productive to NIH's Data Management and Sharing (DMS) and open-science objectives.

Why SP 800-171 is over-scoped for de-identified research data

- Design-intent mismatch. SP 800-171 was written for Controlled Unclassified Information (CUI) with potential national-security impact. Most NIH controlled-access datasets are de-identified under HIPAA, protected by Certificates of Confidentiality, and governed by DUAs that already prohibit re-identification.

- Marginal risk reduction at very high cost. At UCSD, the quoted price for a FISMA-moderate enclave that meets all 110 controls was ≈ $100k in up-front capital and $80k per year in operating costs, with no existing funds earmarked for this purpose. Smaller institutions face even steeper proportional costs, creating a de-facto barrier to entry and tilting the playing field toward a handful of well-funded hubs.

- Scientific delays. Provisioning a compliant enclave typically takes 8–12 months, delaying hypothesis-driven research, hindering replication, and discouraging junior investigators.

- Equity concerns. Institutions with limited cyberinfrastructure budgets cannot absorb these costs, limited who can access and use the data.

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

A more proportionate, risk-based approach would:

1. Map security tiers to data sensitivity.

2. Provide an NIH-hosted shared enclave. Expand cloud workspaces so investigators can analyze data without each campus building its own environment.

3. Offer centralized attestation, in which a single site maintains the enclave and extends secure access to collaborators via a standard memorandum of understanding.

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

I respectfully ask the Office of Science Policy to re-examine the one-size-fits-all SP 800-171 mandate and consider adopting a tiered, risk-proportionate model. Even modest adjustments would free substantial resources for science, broaden participation, and still maintain strong protections against re-identification or misuse.

I would be happy to share cost analyses from my lab. Thank you for considering these concerns and for your continued stewardship of policies that protect research participants and advance open, rigorous biomedical science.

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
n/a

3. N/A

Submit date: 12/22/2025

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Not Applicable

Role: Other

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I can't believe that the NIH has suck to this level.

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

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

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

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

4. Dinesh Barupal

Submit date: 12/24/2025

I am responding to this RFI: On behalf of myself

Name: Dinesh Barupal

Name of Organization: Icahn School of Medicine at Mount Sinai

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Metabolomics Data and Exposomics Data should be included under "Human Data Types Required to be Protected through Controlled-Access (see Appendix for definitions)". These datasets are often generated using high-resolution mass spectrometry, same technique used by Proteomics, which can capture chemical and metabolic signatures for exposures, drugs, diet, health conditions and pollutants.

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

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

5. Krista Perreira

Submit date: 12/30/2025

I am responding to this RFI: On behalf of myself

Name: Krista Perreira

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The term "birthplace" is utilized and is not well defined. It could be interpreted to mean country of birth, US state or territory of birth, county of birth, or city of birth. It is essential for many demographic analyses to evaluate differences by nativity (i.e. US-born or not). Thus, the term "birthplace" should be defined more narrowly and precisely to indicate county or city of birth.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The term "birthplace" is utilized and is not well defined. It could be interpreted to mean country of birth, US state or territory of birth, county of birth, or city of birth. It is essential for many demographic analyses to evaluate differences by nativity (i.e. US-born or not) and there is no need for controlled access at such a high-level of generalization. Thus, the term "birthplace" should be defined more narrowly and precisely to indicate county or city of birth.

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

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

6. N/A

Submit date: 1/6/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Other

Role – Other: compliance

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Mention, outline, or provide a link to which repositories can be accessed via the BDC and how that complies with the NIH controlled access data policy, the pros/cons of accessing the BDC versus the academic institutions cloud solution, contact(s) at NIH or NIH's 3rd party provider to assist with cost estimation needed for proposals, etc.

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

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

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

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

7. Dan Elton

Submit date: 1/16/2026

I am responding to this RFI: On behalf of myself

Name: Dan Elton

Name of Organization: The Metascience Observatory

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I believe the government should allow agencies to get full consent from people to have their genetic data shared publicly just as any other data (no added security overhead, no access management systems, no limitations on sharing or use).

As an example, the Harvard Personal Genome Project has allowed people to donate their genetic data and make it freely available (see https://my.pgp-hms.org/public_genetic_data). I'm not sure if the above policy would allow federal agencies to host something similar, but this sort of genetic data sharing system should be allowed within the federal government as long as consent is obtained properly via adequate informing and counseling on risks.

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

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

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

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

8. Brian MacWhinney

Submit date: 1/22/2026

I am responding to this RFI: On behalf of myself

Name: Brian MacWhinney

Name of Organization: Carnegie Mellon University - TalkBank

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Applying these guidelines to new data makes sense, but applying to older data would end up wiping large amounts of accumulated knowledge off the record.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Participants in clinical studies of language function typically give their consent to have audio recordings available to researchers. Without a national database of voice prints or the types of voice prints used by commerical phone answering systems, there are currently no methods for using these recordings to identify a participant. So, the concept of "voice print" needs to be clarified. It is only functional if there is a generally accessible database of such prints, as there is in the case of fingerprints as maintained by law enforcement.

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

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

9. N/A

Submit date: 1/25/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I understand that there have been recent instances of data being improperly shared and used to inappropriately promote particular points of view (e.g., race differences). But for folks conducting human research, the combined burdens of the new scope of a clinical trial, open data sharing and immediate publication access requirements, and now very detailed requirements about what can be shared are becoming challenging to manage.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
I use neuroimaging as a tool and have placed some data in OpenNeuro. One concern is whether there will be a need to remove these data and place them into another repository. This will have new burdens, as the OpenNeuro framework is tailored for imaging studies with a variety of other co-acquired data (such as behavioral measures).

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
I am very concerned about two things. First, the limitations on individual-level clinical trial data are inconsistent with open data practices, for instance providing individual-level scores on simple behavioral tasks, and will have broad scope given current NIH definitions of clinical trials. If these restrictions are going to be implemented, then some consideration should be given to excluding work conducted under the BESH subcategory of a clinical trial. Second, the concerns about head imaging make sense, but it would be important to be clear about whether the requirements are different if techniques are used to anonymize the data (e.g., by de-facing) and/or if only derivative forms of the data (e.g., images normalized into a standard atlas space) are being shared.

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

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

10. Mitchell Berger

Submit date: 1/31/2026

I am responding to this RFI: On behalf of myself

Name: Mitchell Berger

Name of Organization:

Type of Organization: Not Applicable

Role: Other

Role – Other: Public Health

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Both policies: Reference and encourage researchers to obtain certificates of confidentiality, understand their application and share information with research participants as part of the informed consent process: NIH programs such as All of Us have obtained certificates of confidentiality. Certificates of confidentiality potentially add a layer of protection to research data and protect participants from some misuses. NIH-funded, other federal and non-NIH funded researchers may be able to obtain certificates of confidentiality. While NIH-funded research generally would automatically receive a certificate this may not always be the case and even when a certificate is issued researchers may not always understand its application. NIH should speak to certificates of confidentiality in both its genome data sharing and controlled access policies and encourage researchers dealing with sensitive information to ensure a certificate covers their research regardless of funding source. NIH also should ensure research sponsors, investigators, staff and participants fully understand its controlled access and genomic data sharing policies and certificates of confidentiality and other data protections.

While NIH may understandably prioritize its own or other federally-funded research, even non-federally funded research can discredit the research enterprise if there are misuses and abuses of research data and processes. NIH therefore has a vested interest in ensuring broad application of certificates of confidentiality as well as relevant state and federal statutes and regulations when and to the extent applicable.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Genomic data sharing policy: Cite State-level privacy requirements and 42 CFR Part 2: First, with respect to both the controlled access data policy and genomic data sharing policy, I urge NIH to cite the potential applicability of state-level policy requirements and the federal Confidentiality of Substance Use Disorder regulations (42 USC 290dd-2; 42 CFR Part 2). Specifically under #3 “NIH proposes modernizing

the following data submission and sharing practices." NIH can add 1-2 sentences stating that "Applicability of other federal and state-level requirements should be considered in NIH-supported researcher data sharing approaches."

With respect to state-level requirements, one author notes that "[a]lthough genomic data stripped of other identifiers is not considered identifiable under the HIPAA Privacy Rule, genomic data can itself be a tool for reidentification, particularly when matched with public genetic databases." Some state statutes and regulations potentially may impose higher standards for privacy, informed consent or other requirements than does the HIPAA federal baseline, including with respect to behavioral health information. Similarly, 42 CFR Part 2 (Part 2) potentially could apply and unlike HIPAA Part 2 applicability follows the information downstream and includes specific redisclosure provisions.

It is important to note that HIPAA applies only to certain covered entities that generate protected health information and Part 2 to programs that receive federal assistance and hold themselves out as providing and provide substance use disorder treatment, diagnosis or referral. In many cases, research will not fall under the HIPAA Privacy Rule or Part 2 or even state requirements. But in some cases, such as if a HIPAA covered entity or Part 2 program is involved or a state requirement applies, these requirements should be considered. At a time when more research is being performed on substance use disorders and potential genomic associations, applicability of state requirements and Part 2 may be important. It is worth noting in this regard that NIH All of US program informed consent and privacy authorization forms reference Part 2 as well as HIPAA.

Genomic data sharing policy: Clarify right of access to genomic data: There also is some potential tension between the HIPAA right of access to their own genomic information and Clinical Laboratory Improvement Act regulations that potentially provide an exception to this right for certain research. As one author explains it "The CLIA research exception lets research laboratories avoid being regulated by CLIA as long as they do not report individual-specific results for clinical purposes. Some research laboratories hesitate to provide HIPAA access because they fear they might fall under the CLIA regulations if they do so. The CMS [Centers for Medicare & Medicaid Services, which oversees CLIA along with FDA and CDC] has suggested that research laboratories that provide HIPAA access need to comply with CLIA because the data "will or could be used" for clinical purposes." NIH should in its data sharing policy clarify the applicability of the right of access to data developed or shared pursuant to NIH-funded research.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/nihcontrolledaccessgenomic-1.pdf

Description: Comment re Controlled Access and Genomic data-sharing policies, PDF

11. N/A

Submit date: 2/2/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Given that the genomics community depends on rapid sharing of variant interpretations to drive consensus, I strongly recommend that NIH explicitly clarify in implementing guidance (proposed FAQs or supplementary information) that ClinVar is excluded, e.g., "Genetic variant databases that provide aggregated clinical significance interpretations, summary phenotype data, and population-level evidence without individual-level linkages, such as ClinVar, are not subject to controlled-access requirements under this Policy, provided they do not contain individual-level clinical trial data or linkable personal health information."

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

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

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

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

12. Laura Raffield

Submit date: 2/6/2026

I am responding to this RFI: On behalf of myself

Name: Laura Raffield

Name of Organization: UNC Chapel Hill

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I appreciate the need for standardization of data security requirements, but am concerned about application of this policy to widely used, low risk genomic repositories such as https://www.ncbi.nlm.nih.gov/geo/. Similar concerns apply to public proteomics and metabolomics repositories. Such molecular phenotypes can be used to infer a small number of common genotypes, but given limited phenotypes available (and no DNA sequencing) should not pose significant data security or participant identifiability concerns. If this policy is applied to such datasets and resources, I worry that reasearchers would lose access to this data entirely.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Long queues and challenging submission requirements for public repositories like dbGaP are currently an impediment to data sharing, with data frequently taking years to be broadly available even if researchers initiate submission promptly. With new requirements, it would also be important to expedite and streamline these processes, including provision of appropriate staffing and resources at NIH, to avoid further slowing of research progress in the genomics field.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
I believe that low-risk data types like genomic summary statistics will not be subject to these controlled-access requirements, but this would be good to explicitly clarify.

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

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

13. National Society of Genetic Counselors

Submit date: 2/11/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: National Society of Genetic Counselors

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
February 11, 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: NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy (NOT-OD-26-023)

On behalf of the National Society of Genetic Counselors (NSGC), thank you for the opportunity to submit comments to the National Institutes of Health (NIH) Office of Science Policy (OSP) regarding its Request for Information (RFI) on the Draft NIH Controlled-Access Data Policy and Proposed Revisions to the NIH Genomic Data Sharing Policy.

NSGC is the leading voice of genetic counselors, advocating for their various roles, advancing the practice of genetic counseling, and fostering collaboration, education, and research to ensure equitable access to genomic healthcare. Genetic counselors are the driving force towards universal, innovative, and equitable genomic healthcare, leading to better health outcomes for all.

NSGC supports NIH's efforts to modernize data-sharing policies and urges NIH to ensure that revised requirements continue to facilitate timely, transparent, and equitable sharing of clinically relevant genetic information, balancing privacy protections with the urgent need to improve variant interpretation and patient care across diverse populations.

NSGC encourages the sharing of de-identified clinical and genetic information acquired by clinicians and laboratories from individuals undergoing clinical genetic or genomic testing. This includes evidence supporting gene-disease associations, relevant features of the individual's phenotype, and a clear delineation of the amount and strength of evidence underlying individual variant classifications. Data

should be shared in publicly accessible, non-proprietary databases that protect patient privacy in accordance with applicable state and federal regulations. Frameworks should be in place to ensure that information in such databases is reviewed systematically and kept as up to date as possible. Transparency about data-sharing practices should exist during the consent process for individuals undergoing testing.

Broadening access to genetic data is necessary for more consistent, accurate variant classification across laboratories and improves the collective understanding of the genotypic and phenotypic spectrums of genetic conditions. Responsible data sharing is essential to optimize the care of patients whose diagnosis, management, or treatment decision-making is based on genetic information.

We look forward to the opportunity to be a resource to NIH on these and other issues. If you have any questions, please do not hesitate to contact Meghan Carey at mcarey@nsgc.org.

Sincerely,

Carrie Haverty, MS, CGC

NSGC President

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NSGC_NIH-Genomic-Data-Sharing-RFI-Comment_Final_2.11.2026.pdf

14. Theodore D Satterthwaite

Submit date: 2/12/2026

I am responding to this RFI: On behalf of myself

Name: Theodore D Satterthwaite

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Human brain images (e.g., brain MRI) should *NOT* be considered controlled-access data as part of this standard. After removing facial features from structural images, these images are in no way identifiable by any reasonable standard. Other common image types -- e.g., functional and diffusion images -- are not identifiable even in their raw form. Adoption of such a standard will harm American competitiveness in the critical area of translational human neuroscience, reduce return on federal investments, increase the cost of research, and slow the pace of scientific progress in mental health research. If adopted, this standard would effectively end most open science efforts in the field of human brain imaging and be a major setback for rigorous and reproducible research practices.

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

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

15. N/A

Submit date: 2/12/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached letter for response.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached letter of response.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH_Comment_NIST_Compliance_final.pdf

Description: Comments on 1) Feedback on any aspect of the Draft NIH Controlled-Access Data Policy and 4) Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy Attached as a Letter

16. Daniel Wolf

Submit date: 2/12/2026

I am responding to this RFI: On behalf of myself

Name: Daniel Wolf

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Including face/head imaging in this controlled access policy will dramatically slow progress in the neuroimaging field, and this progress is essential to developing better neurological and psychiatric approaches to diagnosis and biomarker-based treatment. What we need are robust legal sanctions against various forms of misuse of genetic and other data (including insurance-based discrimination), not increasingly stringent control of data which impedes research and really only give the illusion of protection (until that hack comes).

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

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

17. Robin Aupperle

Submit date: 2/13/2026

I am responding to this RFI: On behalf of myself

Name: Robin Aupperle

Name of Organization: Laureate Institute for Brain Research

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
While the NIST requirement for accessing data creates an obstacle for many institutions, I think that it is a necessary step that will help constrain access to institutions that are dedicated to data security. Given prior instances where data was accessed by individuals who should not have accessed it and was misused, I think these measures are important and critical. Other input we have gotten from our Community Advisory Board members include requiring those applying for re-access of data have their publications/research products reviewed so that anyone who has misused the data is not given access to the data again.

However, it must also be recognized that there will be costs associated with institutions supporting these requirements. Consideration of this will either need to be incorporated onto the direct costs for grants that are making the data available or perhaps through indirect rates for institutions who are supporting access/storage of data.

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

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

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

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

18. N/A

Submit date: 2/15/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The current enforcement of NIST SP 800-171, which historically stems from protecting data in the defense industry, is wholly antagonistic to the reality of biomedical research computing labs. By not only degrading security by imposing compliance regimes that have no context in the computational reality of the mid-2020s, this also functions as a regressive research tax. This unfunded mandate actively diverts federal NIH dollars away from scientific discovery and into performative administrative overhead. The current enforcement of NIST SP 800-171, which historically stems from protecting data in the defense industry, is wholly antagonistic to the reality of biomedical research computing labs. By not only degrading security by imposing compliance regimes that have no context in the computational reality of the mid-2020s, this also functions as a regressive research tax. This unfunded mandate actively diverts federal NIH dollars away from scientific discovery and into performative administrative overhead.

We are being asked to trade high-performance, modern research ecosystems for a "blueprint protection" model that treats scientific data like weapons schematics.

1. The MFA Paradox (Control 3.5.3)

- The Research Tax: We must now divert engineering hours to install and maintain MFA for every internal SSH hop within our badge-protected labs.
- Security Gain: Near Zero. In a physically secure environment, adding a secondary token for internal data movement provides no defense against realistic threats but adds constant operational friction.
- The Irony: To "pay" this tax quickly, sites often route internal traffic through an external-facing VPN, artificially expanding the attack surface just to trigger a 2FA prompt for a checkbox.

2. The "Forensic Theatre" of Manual Auditing (Family 3.3)

- The Research Tax: This requirement imposes a "Labor Tax" on Principal Investigators. We are expected to manually review audit logs, system auth.log files, and even .bash_history.
- Security Gain: Zero. .bash_history is a convenience feature, not a security tool; it is trivial to manipulate. Expecting a PI to act as a forensic investigator is a profound waste of the NIH's investment in their scientific expertise.

3. The "Nix" on NFS and the 1990s Compute Model (Control 3.1.3)

- The Research Tax: Because shared clusters lack "NIST-certified" granular auditing, we are forced to abandon institution-wide super-computers.

- Security Gain: Negative. We are forced to "Nix" $20M in compute infrastructure and silo large datasets onto isolated local workstations.

- Operational Collapse: Modern neuroimaging is a high-throughput endeavor. When we are forced onto an isolated Linux machine, we lose the ability to parallelize. For instance, a single FreeSurfer run can easily take 8+ hours per subject. Batch processing a small cohort could require days or weeks of continuous compute. On a siloed workstation, one researcher running a standard data prep pipeline effectively "locks" the hardware, meaning no one else in the lab can perform any meaningful work until that job completes. We are choosing "auditability" over the ability to actually conduct science.

4. The Pipeline Kill Switch (Control 3.1.11)

- The Research Tax: NIST mandates session termination after inactivity. This ignores the technical reality of neuroimaging.

- Operational Collapse: As noted, pipelines like FreeSurfer or fMRIPrep run for massive stretches of wall-clock time. If a researcher's interactive session—used for monitoring real-time quality control or debugging a complex multi-day batch job—is killed by an aggressive security timeout, the state of the IDE and the debugger is lost.

- The Result: Researchers are forced to write "keep-alive" scripts or utilize auto-clickers to subvert the policy. We are effectively taxing the researcher's time to find ways to bypass the very security we are forcing them to buy.

Achieving this level of compliance is a massive capital expenditure that creates a "pay-to-play" barrier for research. Based on CMMC Level 2 benchmarks, the "Security Tax" for a complying site is:

Category Est. Cost (Year 1) The Compliance "Tax" Description

Bespoke Hardware $35k – $75k Purchasing dedicated, air-gapped hardware because the shared cluster is "non-compliant."

Audit Storage $5k – $15k Storing terabytes of system-call logs that offer no actual defensive value.

Administrative Labor $15k – $25k 0.25 FTE of a Systems Admin just to manage the 100+ page System Security Plan (SSP).

Total "Tax" ~$55k – $115k Estimated Per Participating Research Lab.

Source: Derived from CMMC Level 2 (NIST 800-171) cost assessments. For an average R01-funded lab, this represents a 25-40% tax on direct research costs.

The NIH is inadvertently creating a two-tiered research system where only the most well-funded labs can afford the "tax" of NIST compliance. If the environment is so hostile that a researcher cannot use a cluster, parallelize their subjects, or keep a terminal open, they will simply stop using the data.

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

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

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

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

19. N/A

Submit date: 2/16/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Simple - we will have an extremely difficult time recruiting patients and subjects for any type of human subjects study if their confidential information will be shared with the government.

From an investigator's perspective I see no scientific benefit in this policy.

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

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

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

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

20. cBioPortal for Cancer Genomics

Submit date: 2/17/2026

I am responding to this RFI: On behalf of an organization

Name: Ethan Cerami

Name of Organization: cBioPortal for Cancer Genomics

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Data sharing programs funded by the NIH must be designed to both protect patient privacy and ensure broad research access. For cancer-specific data, we believe both aims can be achieved via a two-tiered system:

1) an open-access tier containing deidentified data, including somatic mutation calls, processed molecular data, de-identified tissue images, and de-identified clinical and biospecimen meta-data; and

2) a controlled access tier containing identifiable data, including germline data, raw genomic data, e.g. DNA sequencing reads, binary alignment map (BAM) files and identifiable medical images

The NIH has been following this model for almost two decades with data generated by The Cancer Genome Atlas (TCGA) and other data sets hosted by the Genomic Data Commons (GDC); somatic variant data is freely available, as is de-identified clinical, molecular and tissue imaging data, but raw sequencing data are available only through protected access. This open access policy has broadened access of critical cancer data to the entire scientific community and significantly improved our understanding of the molecular basis of cancer.

We are concerned that the new draft fails to distinguish somatic versus germline data, and also fails to support open access to de-identified data. This will unnecessarily place extremely low risk data behind a controlled access data policy, which will limit broad research access.

As developers of the open source cBioPortal for Cancer Genomics, we have maintained an open access portal to cancer data for >15 years, and we believe that the broad access to this data is fundamental to scientific progress. The cBioPortal currently contains:

* somatic mutation data, plus very limited, non-identifying germline variants pertaining to cancer predisposition.

* processed / gene level calls for epigenomic, proteomic and transcriptomic data

* de-identified tissue imaging data, e.g. H&E images and multiplex tissue images

* de-identified clinical and biospecimen meta-data

Our open-access model currently serves > 38K unique users per month with >30K citations, demonstrating the high scientific demand for open access cancer data.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
As currently worded, it is not clear if the NIH will distinguish somatic versus germline data or support distribution of de-identified data. If the new policy requires that all genomic data (somatic and germline) plus any form of de-identified data must be restricted to a controlled-access policy, the cBioPortal for Cancer Genomics as it currently exists would need to be ended, and a new controlled-access version of cBioPortal would need to be created. We believe this would fundamentally restrict broad access to cancer data.

Furthermore, it is important to note that many peer-reviewed manuscripts already include deidentified genomics data, alongside deidentified clinical data, in open-access supplementary files. Restricting this same data to controlled-access repositories would be inconsistent with established scientific publishing practices.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
We ask for clarification on the following:

* Under the new guidance, will there be a distinction between somatic and germline data? And, will sharing of somatic data still be possible via open access mechanisms?

* Under the new guidance, will the sharing of de-identified gene level molecular data, e.g. epigenomic, proteomic and transcriptomic data, still be possible via open-access mechanisms?

* Under the new guidance, will the sharing of de-identified clinical and biospecimen meta data still be possible via open-access mechanisms?

* Under the new guidance, will the sharing of de-identified tissue images, e.g. H&E and multiplex images still be possible via open-access mechanisms?

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
None

21. N/A

Submit date: 2/19/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
If participants give explicit consent for access to voice data for research or teaching purposes, then the policy should make it clear that this is allowed. As currently written, it seems that the policy is over-riding participants informed consent by requiring various additional forms of review and control.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Strict implementation of these policies could end up making large amounts of data collected with NIH funding no longer available for research.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
There is no scientific evidence that people's identity can be determined on the basis of voice samples.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
It is not clear how voice samples contributed through informed consent for scientific research should be covered by a policy for genomic data.

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

22. TalkBank

Submit date: 2/23/2026

I am responding to this RFI: On behalf of an organization

Name: Brian MacWhinney

Name of Organization: TalkBank

Type of Organization: Research Participant/Patient Advocacy Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
There are a number of problems with this draft policy in relation to language data. Together they would serve to effectively put an end to many of the recent advances in the use of technology to understand, diagnose, and treat language disorders and delay. NIH has been at the forefront in supporting creative use of language data gathered and shared (among reseachers) with full participant informed consent. This is not genomic data and nearly all of it is not covered by HIPAA. However, the new regulations treat voice data as if it were genomic data.

It is difficult to see how anonymous recordings for young children could represent a national security threat to the United States.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Our TalkBank system is the most useful controlled access system for language data and the one that is being used now by tens of thousands of researchers. If TalkBank is forced to apply these new standards, it is difficult to see how its work can continue. Full processing of DUA or DTUA agreements for each user is way beyond the resources of TalkBank and not necessary given the fact that these data represent no security threat to our country and were contributed with full informed consent.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Treating language samples such as picture descriptions or sentence repetition as genomic data makes no sense.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
I sincerely hope that NIH recognizes the extent to which implementation of this policy would destroy its many years of great contribution to the study of language disorders, language learning, bilingualism, anbd language delay.,

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
I'm not really sure what this means.

23. GenoBank.io

Submit date: 2/25/2026

I am responding to this RFI: On behalf of an organization

Name: Daniel Uribe

Name of Organization: GenoBank.io

Type of Organization: Industry (Biotech/Device/ Pharmaceutical Company)

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
GenoBank.io strongly supports NIH's initiative to establish a unified Controlled-Access Data Policy. The proposed framework correctly identifies the breadth of human data types requiring protection and appropriately raises the bar for repository security standards. We offer the following observations:

1.1 Consent should be dynamic, not static. The Draft Policy requires informed consent for open sharing but treats consent as a binary, point-in-time authorization. In practice, participant preferences evolve. We urge NIH to recognize and encourage "Metamorphic Consent" models, where consent is not a one-time checkbox but a continuous, revocable, and auditable relationship between the data contributor and downstream users. Blockchain-anchored consent tokens (such as BioNFTs) can provide cryptographic proof of consent status that is verifiable by any repository or data access committee in real time, without exposing participant identity.

1.2 Controlled-access must include a mechanism for revocation and erasure. The Policy references protections "throughout the data lifecycle" but does not explicitly address the right to withdraw consent or request data deletion. With international collaborations subject to GDPR Article 17 (right to erasure) and similar frameworks, NIH should require that controlled-access repositories implement verifiable consent revocation. Technologies such as revocable NFT-based consent tokens, paired with mutable encrypted storage (e.g., AES-256 encrypted cloud storage on AWS S3), ensure data can be made inaccessible or deleted when consent is withdrawn, unlike immutable storage systems (e.g., IPFS) that cannot guarantee erasure.

1.3 Authentication should extend beyond human researchers to AI agents via revocable NFTs. The Policy requires "authentication of the identity of data requesters," which is essential. However, the current framing assumes data requesters are human researchers at institutions. In practice, AI agents are increasingly acting as autonomous data consumers, querying repositories, running analyses, and integrating datasets on behalf of researchers. Traditional identity frameworks such as W3C Decentralized Identifiers (DIDs) and Verifiable Credentials were designed for human-to-system interactions and do not natively handle machine-to-machine authorization, revocation, or micropayment for data access. Revocable NFTs, such as BioNFTs used by GenoBank.io for biological-origin datasets, provide a superior authentication and authorization layer for this emerging paradigm. Each BioNFT encodes the consent terms, permitted use, and revocation status of a specific dataset. An AI agent presenting a valid BioNFT can be authenticated and authorized programmatically via smart contract verification, with no human intermediary required. Critically, when consent is revoked, the NFT

state change immediately propagates to all access points, blocking both human and AI agent access in real time. This model is compatible with emerging machine-to-machine payment protocols (such as HTTP 402-based micropayment frameworks) where AI agents pay per-query or per-dataset access fees, creating an auditable economic trail that further strengthens compliance monitoring. We urge NIH to recognize that controlled-access policies must be designed not only for human researchers but for the AI agents that will increasingly act on their behalf, and that revocable NFT-based authentication is the most robust mechanism for this purpose.

1.4 Privacy-preserving access control should be explicitly encouraged. The Policy should encourage repositories to adopt privacy-preserving technologies for access gating. For example, probabilistic data structures such as Bloom Filters can verify whether a requester holds valid access credentials without exposing the credential set itself, providing efficient, privacy-preserving access control that complements NIST-SP-800-171 security standards.

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

2.1 Current repository infrastructure is centralized and may be insufficient for scale. The anticipated expansion of controlled-access data types (now including epigenomic, proteomic, transcriptomic, and imaging data) will dramatically increase demand on existing NIH-approved repositories such as dbGaP. NIH should proactively invest in and certify a broader ecosystem of repositories, including hybrid architectures that combine cloud storage with blockchain-based access control.

2.2 Blockchain-gated repositories represent a new class of compliant infrastructure. GenoBank.io operates a controlled-access architecture where: (a) data is stored in AES-256 encrypted AWS S3 buckets, meeting NIST security standards; (b) access is gated by BioNFT ownership, where each NFT encodes consent terms and permissions; (c) access requests are verified through smart contract logic and Bloom Filter-based permission checks; and (d) consent revocation immediately blocks data access without requiring repository staff intervention. This architecture satisfies all four requirements of the Draft Policy (prospective review, identity authentication, country-of-concern restrictions, and NIST-equivalent security) while adding cryptographic auditability and participant-controlled consent management. We encourage NIH to develop a certification pathway for such blockchain-gated repositories.

2.3 Interoperability between repositories is critical. As the number of compliant repositories grows, NIH should require standardized APIs and metadata schemas (building on GA4GH standards such as DRS, Passport, and Data Connect) to ensure data discoverability and federated access across repositories without requiring data movement.

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

3.1 The listed data types are appropriate and comprehensive. The inclusion of genomic, epigenomic, proteomic, and transcriptomic data alongside biometric identifiers and imaging data reflects the current understanding of re-identification risk. We support this comprehensive list.

3.2 NIH should consider adding microbiome data. Human microbiome data, particularly when paired with metadata such as geographic location, health status, or medication history, has been shown to enable re-identification. As microbiome research expands under NIH funding, this data type warrants

inclusion or at minimum, explicit guidance under the "additional policy considerations" criteria.

3.3 Thresholds for omics data types need clarification. The Policy correctly notes that "routine clinical measurements" of proteins, RNA transcripts, or epigenetic modifications are excluded. However, the boundary between "routine clinical measurement" and "systems-level analysis" is ambiguous. For example, a targeted panel of 500 gene expression markers could be argued either way. NIH should provide quantitative thresholds or exemplar scenarios to guide institutions.

3.4 Derived and aggregate data require nuanced treatment. The Policy should explicitly address the status of data derived from controlled-access sources, such as polygenic risk scores, variant frequency tables, or machine learning model weights trained on genomic data. While genomic summary results are noted as typically low-risk, model weights and embeddings from AI/ML training on genomic data may encode sufficient information to reconstruct individual-level data through model inversion attacks. This is an emerging risk that NIH should proactively address.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
4.1 The strengthened consent requirements are a significant positive step. Requiring that data collected after 2015 must have consent for use and sharing, with no sharing permitted absent consent, aligns with international norms and reinforces participant autonomy. We strongly support this revision.

4.2 Consent should go beyond permission and become an economic relationship. The GDS Policy revision appropriately strengthens consent but stops at the traditional model of "permission to use." We urge NIH to consider frameworks where data contributors maintain an ongoing economic relationship with the value generated from their data. Under GenoBank.io's "Metamorphic Consent" model, consent transforms from a static permission into an ongoing economic relationship: data contributors receive attribution and compensation (via Biodata Dividends calculated through Shapley value attribution) as their data generates downstream value in research. This model incentivizes participation, particularly from underrepresented populations who have historically been excluded from the benefits of genomic research.

4.3 The 100-individual threshold for "large scale" is reasonable but should be revisited periodically. As sequencing costs decline and single-cell multi-omics generate data from fewer participants but with greater depth, the threshold should be evaluated against re-identification risk rather than a fixed participant count.

4.4 The 6-month data sharing timeline is practical. Allowing 6 months for data cleaning, quality control, and repository processing is a reasonable balance between rapid sharing and data quality. However, NIH should clarify whether this timeline applies per-sample or per-study, particularly for longitudinal studies with rolling enrollment.

4.5 Expanding institutional review capacity beyond IRBs is welcome. Permitting HRPPs and other qualified bodies to review Institutional Certifications reduces bottlenecks without sacrificing oversight quality. This is a sensible modernization.

4.6 The Legally Authorized Representative provisions are important for equity. Accepting consent from legally authorized representatives, next-of-kin, and proxy decision-makers ensures that data from deceased individuals, minors, and those lacking capacity can be ethically shared. This is essential for

rare disease research and pediatric genomics.

4.7 AI agents acting on behalf of families and individuals must be recognized in consent and data sharing frameworks. The GDS Policy revisions address human decision-makers (researchers, LARs, next-of-kin) but do not account for the emerging reality that AI agents will increasingly act on behalf of individuals and families in managing their genomic data. Families may authorize an AI agent to monitor new research findings relevant to a child's rare disease variant, negotiate data access terms with researchers, or manage consent preferences across multiple repositories. These AI agents need a standardized mechanism to prove they are authorized to act on behalf of a specific individual or family unit. Revocable BioNFTs provide this mechanism: a family can delegate authority to an AI agent by granting it a BioNFT with scoped permissions (e.g., read-only access, consent management, or data licensing on behalf of a minor). The NFT can be revoked at any time by the family, immediately terminating the AI agent's authority. NIH should anticipate this paradigm and ensure that the GDS Policy framework accommodates authorized AI agents as legitimate representatives alongside human LARs, with equivalent requirements for verifiable authorization and revocability.

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

5.1 Allowing Approved Users to develop imputation panels and servers is a positive step, with caveats. The current restriction to NIH-operated servers (e.g., TOPMed) creates bottlenecks and limits innovation. Expanding this to Approved Users, with appropriate safeguards, will accelerate precision medicine research.

5.2 Access control should rely on blockchain signatures and Bloom Filters, not differential privacy or secure enclaves. The Policy notes interest in "privacy enhancing technologies," but not all PETs are equally suited to genomic data. Differential privacy introduces deliberate noise into query results, which is fundamentally incompatible with clinical-grade imputation where accuracy is essential for patient care decisions. Errors introduced by differential privacy mechanisms can lead to misinterpretation of imputation results, with real consequences for diagnosis and treatment. Secure enclaves (e.g., Intel SGX) have a well-documented history of side-channel vulnerabilities (Spectre, Foreshadow, Plundervolt) and impose significant computational cost without delivering the security guarantees they promise. A more robust and practical approach combines two proven technologies:

- Blockchain wallet signatures for identity authentication and authorization. Each data requester (human or AI agent) signs access requests with their cryptographic wallet, providing verifiable, non-repudiable proof of identity without relying on centralized credential stores.

- Bloom Filters for privacy-preserving permission verification. Rather than transmitting or storing full credential sets, Bloom Filters allow imputation servers to verify whether a requester holds valid access permissions with high efficiency and zero credential exposure. This combination delivers what patients actually need: privacy as a choice, not secrecy imposed by opaque systems. Patients should be able to choose who accesses their data and verify that their choices are enforced, through transparent, auditable mechanisms. Blockchain signatures provide that auditability; Bloom Filters provide that

privacy. Together they satisfy the Policy's security requirements without introducing the interpretation errors of differential privacy or the false security of enclaves.

5.3 The restriction to NIH or federal agency-funded servers may be too narrow. Requiring that imputation servers be "funded or operated by NIH or another federal agency" could exclude valuable contributions from international collaborators and private-sector entities operating under equivalent security standards. NIH should consider a certification-based approach where any entity meeting defined security and operational standards can operate an approved imputation server, regardless of funding source.

5.4 Data deletion timelines should be standardized. The current TOPMed model deletes uploaded data after 7 days. NIH should establish minimum and maximum retention periods for all approved imputation servers, with cryptographic verification of deletion.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH_RFI_NOT-OD-26-023_GenoBank_Response.pdf

Description: GenoBank.io Response to NIH RFI NOT-OD-26-023

24. The Abigail Wexner Research Institute at Nationwide Children's Hospital

Submit date: 3/3/2026

I am responding to this RFI: On behalf of an organization

Name: Peter White

Name of Organization: The Abigail Wexner Research Institute at Nationwide Children's Hospital

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NCH-Response-to-Genomic-Data-Sharing-Policy.pdf

Description: Comments and Egress Policy we Developed

25. Frontier Science

Submit date: 3/4/2026

I am responding to this RFI: On behalf of an organization

Name: Sue Siminski

Name of Organization: Frontier Science

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
On behalf of the Frontier Science organization, we submit these comments from the perspective of a data management center supporting multi-site, NIH funded HIV and cancer clinical trials, longitudinal cohort studies and global research collaborations. As a data management center, we play a central role in data collection, data aggregation, data harmonization and curation, data storage and securely stewarding sensitive participant-level data across domestic and international sites. We strongly support NIH's commitment to responsible data sharing and recognize controlled access mechanisms are essential for ensuring participant confidentiality and privacy. Our comments, questions and recommendations on the proposed policy follow below.

1. We suggest clearer, more detailed definitions on controlled access vs. open access. For example, in HIV research nearly all participant-level datasets require controlled access due to the potential re-identification concerns. Clearer criteria would ensure consistent expectations across institutes and repositories and avoid inconsistent interpretations across HIV networks and global collaborators.

2. It would be beneficial if the NIH developed a standardized NIH-wide DUA template that includes mandatory clauses and perhaps some optional clauses depending on the sensitivity of the data. Right now, different NIH repositories have different DUA requirements and review processes.

3. In the “Requirements for Controlled Access Data Sharing” section, more clarity on how authentication of a data requestor’s identity is expected to be documented and more detailed specificity on what constitutes prospective review would be helpful when operationalizing this policy.

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

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

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

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

26. dbTwin, Inc.

Submit date: 3/4/2026

I am responding to this RFI: On behalf of an organization

Name: Aditya Nanda

Name of Organization: dbTwin, Inc.

Type of Organization: Industry (Biotech/Device/ Pharmaceutical Company)

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Central Problem: The Missing Re-identifiability Standard

The draft policy lists transcriptomic data (and omics data in general) as a protected type requiring controlled access throughout the data lifecycle. This is appropriate as a default. However, the rule contains a critical internal inconsistency:

- For data types not explicitly enumerated, the policy's "Additional Policy Considerations" section states that re-identifiability is a criterion for assessing whether controls are needed — implying that adequately de-identified data can exit controlled-access.

- For enumerated data types including transcriptomic data, no such off-ramp exists. The rule says these types "must be protected throughout the data lifecycle" with no de-identification exception.

This creates an unresolved governance question: who decides whether a transcriptomic dataset is re-identifiable, and by what standard? The rule is silent. In practice, this means institutions and IRBs will default to treating all transcriptomic data as controlled-access regardless of its actual re-identification risk — including data that has been rigorously Expert-Determined to pose very small re-identification risk.

II. Differential Re-identification Risks for Counts and cohorts

The rule also draws no distinction between the two transcriptomic data types that matter most in practice: 1) Raw counts matrices (bulk or scRNA-seq) without accompanying clinical metadata and 2) Real cohort data containing transcriptomic counts paired with clinical metadata. Let us look at each in more detail.

Counts matrices alone: re-identifiable in principle, manageable in practice

Erlich & Narayanan (Nature Reviews Genetics, 2014) documented multiple routes by which transcriptomic data can be re-identified after removal of standard identifiers — including inference of individual genotype from eQTL associations in RNA expression data. Linking attacks on bulk RNA-seq have demonstrated up to 100% accuracy on controlled datasets. Walker et al. (Cell, November 2024) extended this to scRNA-seq count matrices, showing individual donors can be identified via eQTL-based linking attacks even given the additional noise in single-cell data. This study postdates the 2024 GDS Policy modernization notices and is not yet reflected in the proposed rule's framing. Importantly: these attacks require access to external reference datasets. Expert Determination — the HIPAA §164.514(b)(1)

process in which a qualified biostatistician certifies that re-identification risk is very small given the specific data environment and sharing context — is the appropriate and sufficient standard for standalone counts data. A counts matrix that has passed Expert Determination is, by regulatory definition, no longer re-identifiable. The policy should recognize this with an explicit controlled-access off-ramp for Expert-Determined transcriptomic data.

Real cohort data: presumptively re-identifiable, requiring a higher bar

When counts are linked to clinical metadata — disease subtype, treatment status, age, sex, ethnicity — re-identification risk escalates substantially, and Expert Determination becomes significantly harder to certify. The combination of a high-dimensional transcriptomic fingerprint with even modest clinical covariates narrows the population dramatically. Standard HIPAA Safe Harbor de-identification was not designed for this data type, and under §164.514(b)(2)(ii), a covered entity with actual knowledge of re-identification risk cannot rely on Safe Harbor.

We recommend that the proposed rule should explicitly designate real cohort data as presumptively re-identifiable, requiring one of two compliance pathways before open sharing: (1) Expert Determination certification with documented statistical justification, or (2) replacement with validated synthetic cohort data accompanied by a privacy QA report. Industry practice already reflects this reality: Novartis' patient-level data anonymization standards explicitly exclude genetic data from sharing entirely, citing unresolvable re-identification risk.

III. Synthetic Cohort Data: A Compliance Pathway the Rule Should Recognize

The current version of the proposed rule does not mention synthetic cohort data. This is an omission worth correcting. Synthetic cohort data generated by a privacy-preserving model trained on de-identified or consented real data is completely free from real patient transcriptomic signatures. There is no record to re-identify, no eQTL to exploit, no individual whose privacy can be breached. The re-identification risk that justifies controlled access for real cohort data is structurally absent. Yet because the rule is silent, institutions will default to treating synthetic cohort data as controlled-access - because they have no policy basis for any other determination. The result is that privacy-preserving technology designed to reduce data sharing friction ends up subject to the same DUA requirements and access review timelines as the identifiable data it was designed to replace.

We propose that synthetic cohort data, accompanied by a documented privacy QA report validating that: (a) the generating model was trained on properly de-identified or consented data, and (b) synthetic outputs are not reverse-mappable to source individuals, should be explicitly recognized as an open-sharing-eligible data type under the final policy.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
No comment on repository availability at this time

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Here are our main recommendations. Please see attached comment for details.

1. Add an Expert Determination off-ramp — and define who decides.

The rule enumerates transcriptomic data as controlled-access but provides no exit for de-identified data, inconsistent with its own Additional Considerations section. Standalone counts matrices certified by a qualified biostatistician as posing very small re-identification risk (§164.514(b)(1)) should be eligible for open sharing. The rule should specify that HIPAA Expert Determination is the governing standard for re-identifiability decisions, who is qualified to make them, what evidence is required, and how determinations are documented. The current draft is silent on all of these.

2. Designate real cohort data as presumptively re-identifiable. Counts + clinical metadata carry materially higher re-identification risk than counts alone and should be treated differently.

Real cohort data should be explicitly flagged in the policy as requiring one of two compliance pathways before open sharing: (a) Expert Determination certification with documented statistical justification, or (b) replacement with validated synthetic cohort data accompanied by a privacy QA report.

3. Recognize synthetic cohort data as open-sharing-eligible. Synthetic cohort data has no source patient — the re-identification risk justifying controlled access is absent by construction.

Synthetic cohort data generated from de-identified or consented source data, validated via a privacy QA report confirming outputs are not reverse-mappable to source individuals, should not require controlled-access treatment under the final policy.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We support revision of the GDS Policy but recommend adding an Expert Determination off-ramp for transcriptomic data and explicit governance standards for re-identifiability determinations. Please see our detailed comments in the attached PDF.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
n/a

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/dbTwin_NIH_Comment_NOT_OD_26_023.pdf

27. Lauren

Submit date: 3/9/2026

I am responding to this RFI: On behalf of myself

Name: Lauren

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Imaging data can be de identified and thus should not be subject to the same restrictions as data that has proper PII. Restrictions for imaging data here are inappropriate as many safe repositories exist. Science is international and it is not appropriate to ban non-identifiable data from countries of concern.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
There are public neuroimaging repositories that host neuroimaging data and do not have controlled access. You will cripple the field if you suddenly impose this on a field that has led the way for self correcting science and already done the yeomans work on de identifying data. Do not include human brain imaging in this policy.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
As above, human brain imaging should not be subject to this policy when data can be stripped of identifiers as is standard.

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

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

28. Jonathan D. Santoro, MD

Submit date: 3/10/2026

I am responding to this RFI: On behalf of myself

Name: Jonathan D. Santoro, MD

Name of Organization: Children's Hospital Los Angeles

Type of Organization: Academic Institution

Role: Clinician

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Existing NIH-supported repositories such as dbGaP, AnVIL, and other controlled-access data infrastructures have been instrumental in enabling responsible sharing of human genomic and clinical datasets. However, if the proposed policy expands the scope of data types requiring controlled access, particularly including proteomic, transcriptomic, epigenomic, imaging, and individual-level clinical trial data, the volume and complexity of datasets requiring secure storage and governance will increase substantially.

To meet this anticipated demand, additional infrastructure investments will likely be necessary. These include:

Expanded repository storage and compute capacity for multi-omics datasets

Scalable authentication and access management systems

Faster and standardized data access request review mechanisms

Interoperability between repositories to facilitate cross-dataset analysis

Importantly, many modern biomedical discoveries rely on integrated multi-omic analyses combining genomic, transcriptomic, proteomic, and clinical phenotypic datasets. Ensuring that repositories support federated analysis environments and standardized metadata structures will be critical for enabling efficient reuse of controlled-access data.

Finally, consideration should be given to providing institutional support mechanisms or shared infrastructure for investigators at smaller institutions who may lack the administrative or technical resources required to navigate complex controlled-access repository requirements.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
The proposed list of protected data types—including genomic, epigenomic, proteomic, transcriptomic, imaging, and individual-level clinical trial data—appropriately recognizes the increasing re-identification risks associated with modern biomedical datasets.

However, several considerations may improve implementation:

First, the policy should clearly distinguish between individual-level datasets and aggregated or summary-level outputs. Summary statistics, meta-analysis outputs, and derived biomarkers typically carry substantially lower re-identification risk and should generally remain eligible for open sharing.

Second, thresholds may be helpful for certain omics data types. For example, large-scale proteomic or metabolomic datasets containing thousands of analytes may warrant different treatment than smaller targeted panels used in clinical research.

Third, imaging datasets require careful differentiation. Imaging data that include facial structures or reconstructable anatomical features may warrant controlled access, whereas processed or feature-extracted imaging outputs may pose minimal re-identification risk.

Clear definitions and examples will be essential to avoid inconsistent interpretation across institutions and NIH Institutes and Centers.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
In some cases, open sharing of human-derived datasets may remain appropriate if the risk of participant re-identification is sufficiently low and the consent framework explicitly supports such sharing. Factors that should be considered include:

Degree of de-identification and potential for re-identification through linkage with external datasets

Population size and uniqueness of the cohort (particularly relevant in rare disease research)

Presence of sensitive or stigmatizing traits

The scope of participant consent and expectations communicated during enrollment

Whether the dataset consists of summary-level or aggregated results rather than individual-level records

In fields such as rare disease or neurodevelopmental disorder research, additional caution may be warranted because small cohort sizes can increase re-identification risk even when identifiers are removed.

At the same time, open sharing of low-risk datasets remains extremely valuable for accelerating discovery, enabling replication studies, and supporting secondary analyses.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Allowing approved users to develop or operate imputation servers using controlled-access data may provide substantial scientific benefit by enabling broader use of genomic reference panels and improving analytic reproducibility. However, strong safeguards will be necessary to ensure privacy protections are maintained.

Appropriate safeguards could include:

Operating imputation servers within secure cloud environments that meet NIH controlled-access security standards

Ensuring controlled datasets used to generate reference panels cannot be downloaded or reconstructed from server outputs

Implementing query rate limits and monitoring to prevent inference attacks

Applying privacy-enhancing technologies such as differential privacy or secure multiparty computation where feasible

Establishing clear audit and compliance monitoring mechanisms

In addition, guidance from NIH on approved technical architectures and operational standards for imputation servers would help ensure consistent implementation across institutions.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
The proposed policies represent an important step toward harmonizing NIH data sharing frameworks and strengthening protections for human participant data. However, several broader considerations should be addressed to maximize the scientific value of NIH-funded datasets:

Administrative burden and access timelines:

Access procedures should remain efficient and predictable. Excessively long review timelines for controlled-access requests could slow scientific progress and discourage secondary analyses.

Support for multi-omics integration:

Modern biomedical research increasingly relies on integrating multiple data modalities. Policies should ensure that genomic, clinical, imaging, and other omics datasets can be accessed and analyzed together in interoperable environments.

Implications for rare disease and special population research:

Many research fields—including Down syndrome, rare neurologic disorders, and pediatric conditions, depend on cross-institutional data aggregation to achieve adequate statistical power. Policies should avoid inadvertently creating barriers to these collaborative analyses.

Standardization across NIH Institutes and Centers:

Consistent implementation across NIH programs will reduce confusion and administrative burden for investigators submitting data and requesting access.

Overall, maintaining the balance between protecting participant privacy and enabling responsible scientific data sharing will be essential for ensuring that NIH-funded research continues to generate maximal benefit for patients and families.

29. University of Colorado Boulder

Submit date: 3/10/2026

I am responding to this RFI: On behalf of an organization

Name: JON REUTER

Name of Organization: University of Colorado Boulder

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Re: NOT-OD-26-023 – Draft NIH Controlled-Access Data Policy & Proposed Revisions to NIH Genomic Data Sharing Policy

The University of Colorado Boulder (CU Boulder) appreciates the opportunity to comment on NIH's Draft Controlled-Access Data Policy and proposed revisions to the Genomic Data Sharing (GDS) Policy. As a public R1 research institution with substantial NIH-funded activity, CU Boulder supports NIH's efforts to modernize data sharing in ways that strengthen participant protections while reinforcing research security and national research integrity. Many of our researchers think these improvements will help standardize data control and streamline the processes once implemented.

From the perspective of our Institutional Review Board (IRB), and institutional research security infrastructure, we offer the following considerations:

1. Controlled-Access Data as Both Human Subjects and Research Security Assets

Genomic and high-dimensional human datasets represent sensitive participant information and strategically valuable research assets. Controlled-access mechanisms therefore serve dual purposes: protecting against re-identification and mitigating risks of unauthorized access, inappropriate foreign transfer, and non-compliant secondary use.

Recommendation: NIH should explicitly acknowledge genomic and linked human data as research security–relevant assets and clarify expectations for access vetting, authentication controls, and institutional oversight.

2. Re-Identification and Data Aggregation Risks

HIPAA Expert Determination standards may not sufficiently mitigate re-identification risks in genomic or richly linked datasets. Aggregation with external datasets, domestic or international, can increase vulnerability.

Recommendation: NIH should provide guidance on enhanced risk assessment standards or tiered access controls for datasets with elevated aggregation or re-identification risk.

3. Institutional Certification and Governance

The proposal to broaden review of Institutional Certifications beyond traditional IRBs may increase operational flexibility. However, determinations affecting consent scope and participant protections require clear accountability structures.

Recommendation: NIH should define minimum governance expectations for certification review processes and clarify when IRB involvement is required to ensure consistent national standards.

4. Secondary Use Monitoring and Enforcement

Controlled-access frameworks rely heavily on Data Use Agreements and user attestations. Effective research security requires clear monitoring, reporting, and enforcement mechanisms.

Recommendation: NIH should articulate audit expectations, violation reporting requirements, institutional notification procedures, and consequences for non-compliance, including for international users.

5. Consent Alignment and Implementation

Clear consent language that distinguishes open versus controlled-access sharing, including international access, is foundational to ethical and secure data stewardship.

Recommendation: NIH should provide model consent language, implementation FAQs, and a phased transition timeline to allow institutions to align IRB review, research security workflows, and data governance systems.

CU Boulder supports NIH's commitment to advancing responsible data sharing while safeguarding research participants and reinforcing national research security objectives. Clear guidance on governance, monitoring, and consent alignment will be critical to successful implementation across the research enterprise. We appreciate the opportunity to provide input and remain committed to partnering with NIH to uphold ethical stewardship and research integrity.

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

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

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

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

30. N/A

Submit date: 3/10/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Regarding: "Imaging data of the human face or head regions. Visual representations (including functional imaging, ultrasound imaging, photographic images, 3D models, radiological scans, X-rays, and others) that depict anatomical or functional details of the human face or head regions."

The term "head regions" is too broad and ambiguous and should be removed or clarified.

Brain fMRI and other neuroimaging data should be excluded from this requirement provided that all facial features have been removed.

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

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

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

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

31. American Society of Retina Specialists

Submit date: 3/10/2026

I am responding to this RFI: On behalf of an organization

Name: Allison Madson

Name of Organization: American Society of Retina Specialists

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ASRS-Comment-NIH-Controlled-Data-Access-Policy.pdf

32. N/A

Submit date: 3/11/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
There is no reason that de-identified human neuroimaging data should be controlled access rather than open access. Controlled access creates inefficiencies, and can act as a barrier to excellent research. Resources like open neuro and neurosynth are amazing for the research capacity of American and international researchers, and would be extremely difficult to maintain and to include future NIH supported work under the revised policy. Please reconsider.

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

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

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

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

33. Cure MAPT FTD

Submit date: 3/11/2026

I am responding to this RFI: On behalf of an organization

Name: Tanya Steel

Name of Organization: Cure MAPT FTD

Type of Organization: Research Participant/Patient Advocacy Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Cure MAPT FTD, a patient advocacy organization, focused on finding a cure for FTD caused by MAPT gene mutations, supports controlling access of the NIH Data. We believe the key to curing or modifying this terminal dementia that affects people in the prime of their life, is by promoting patient and caregiving participation in clinical trials, but as such, the process needs to be secure and made efficient for sharing and accessing.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Likewise, Cure MAPT FTD supports the streamlining of access to the data so more researchers, around the country and around the world, can access invaluable data to help drive cures, especially for rare diseases.

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

34. Arizona State University

Submit date: 3/11/2026

I am responding to this RFI: On behalf of an organization

Name: Heather Clark

Name of Organization: Arizona State University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Our feedback on the DRAFT NIH Controlled-Access Data Policy is the following:

Administrative burden: Additional requirements (institutional attestations, security oversight) could significantly increase compliance costs, especially across multiple projects and agreement types.

Countries of concern:

- How should restrictions apply if a researcher relocates to a country of concern after access is granted?

- How will periodically updated country designations affect ongoing data access?

Scope and clarity:

- Confirm whether “other transactions” includes Other Transaction Agreements.

- Clarify if the policy applies broadly across all NIH funding and agreement mechanisms.

Data lifecycle: Provide a clear definition of the controlled-access data lifecycle.

- AI-generated data: Clarify whether synthetic or AI-generated data fall under the policy and any associated risk considerations.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Our feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy is the following:

Infrastructure needs: Implementation may require new or expanded IT security, data governance, and compliance staffing. Resource challenges exist under a highly distributed research environment.

Existing repository limitations:

- Few repositories allow direct deposit of restricted data.

- Not all repositories provide the administrative oversight envisioned in the proposed policy.

Guidance needed:

- Clear criteria for acceptable repositories.

- Instructions for assessing and documenting repository compliance.

- Faculty-facing guidance to explain differences from the current NIH DMS Policy.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Our feedback on the appropriateness of the protected data types designated to be controlled-access is the following:

1. Criteria clarity: Current thresholds for controlled-access designation are unclear.

2. Should thresholds be numeric (e.g., >100 participants) or qualitative?

3. Qualitative and mixed-methods data: Guidance needed on applicability to deidentified qualitative or mixed datasets.

4. Consistency: Explicit definitions, thresholds, and examples would support uniform implementation across ASU and other institutions.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Our feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy is the following:

1. Legacy data: Guidance needed for datasets collected under older consent frameworks, especially when re-consent is not feasible.

2. HRPP review capacity: Expanding review responsibilities may require additional infrastructure, clarified roles, and updated consent language.

3. Post-award verification: ASU has some processes that exist to verify data at deposit do not include identifiers. What are other best practices institutions should adopt to ensure datasets remain free of identifiers?

4. Researcher departures: Clarify responsibilities for data stewardship and compliance when a researcher leaves the institution.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Our feedback on the proposed updates to the GDS Policy for Imputation Servers is the following:

Scope of “Approved Users”:

Who qualifies as an Approved User? Individual, project, or organization?

What is the approval process and governance for granting access?

Resourcing:

Who bears the technical, financial, and operational costs of running an imputation server?

Will there be central support or is responsibility entirely with the user?

Security compliance:

How will users certify adherence to security requirements?

Will compliance be self-attested or independently audited?

How will ongoing compliance be monitored?

Description: ASU Response to the Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

35. Brian Keane

Submit date: 3/11/2026

I am responding to this RFI: On behalf of myself

Name: Brian Keane

Name of Organization: University of Rochester

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
We ask that—for now—neuroimaging data be removed from the list of protected data types so long as the neuroimaging data are defaced and deidentified. See below for an explanation.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Re-identification of individuals from defaced neuroimaging data is technically possible but the real-world risk is much lower than what some studies suggest. For example, Jwa, Koyejo, and Poldrack (2024, Imaging Neuroscience) found that when scaled to realistic population sizes, re-identification rates for defaced data drop to less than 1 percent, and that the potential harms are largely informational rather than placing participants in direct danger. Mitigation strategies beyond defacing are maturing rapidly and do not impose heavy administrative overhead. The proposed NIH Policy appropriately recognizes head and face imaging as requiring controlled access, but policy should be calibrated to avoid creating burdensome new compliance requirements that slow research without meaningfully reducing risk — particularly given this administration's emphasis on reducing regulatory burden. Overly restrictive protections carry their own ethical cost: participants consented to advance science, and locking data behind data use agreements that take months to negotiate undermines both that purpose and the efficient use of taxpayer investments. Also, there are arguably easier ways to gather personal information than through brain imaging, namely, internet search history, GPS tracking, and social media.

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

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

36. Russell Bowler

Submit date: 3/11/2026

I am responding to this RFI: On behalf of myself

Name: Russell Bowler

Name of Organization: Cleveland Clinic

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Ranking of risk by omics category is specious. 1. While DNA contains more features and therefore lower P values to identify will be statistically smaller (say 10E-500 versus 10E-200), this is only a practical difference if there are > 10E100 humans to identify); there are well under 10E11 humans in history of man, so genotypes, epigenomes, transcriptome, proteome are all equivalent at identifying a single person (using modern methods we are all using); what is the practical significance of P 10E-500 versus 10E-100 (no practical significance). 2. You don't need to start with a DNA database; people can be identified without DNA databases; this is well established, and big data scientists are aware of this (PMC7819582; PMC10247808; PMID: 22484626). 3. Your DNA (except for epigenome) doesn't change though your environment (except V(D)J recombination). However, your epigenome, transcriptome, proteome, etc all can change when you are exposed to smoke, take certain medications, use certain drugs, etc. Thus, I would consider the non-DNA datasets higher risk because they can not only identify you, but provide information about your behavior and environment in ways that germline DNA cannot.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We need to break this practical misconception that DNA is more risky than other omics because it is not correct in the current environment. Treat all large omics data equally because they have the same or greater risk than DNA. This is the way it is done in Europe (genetic equivalents).

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
The problem of regulatory environment stifling innovation is a framing problem. We know that many omics and other large datasets have identifiable data (and practically equal), we just haven't figured out an acceptable way to work with identifying data while still preserving privacy (which is why we are having this conversation). Modern problematic examples include permitted data access with research misuse (e.g. Havasupai and the University of Arizona).

37. David Rocha

Submit date: 3/11/2026

I am responding to this RFI: On behalf of myself

Name: David Rocha

Name of Organization:

Type of Organization: Not Applicable

Role: Member of the Public

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The NIH needs to coordinate with the NLM and ASTP/ONC for health information technology to elevate alignment of the genomics and clinical data ecosystems, including GA4GH, FHIR, and OMOP architectures, in addition to rare disease and Mendelian disease resources, such as OMIM, HPO, and Orphanet/ORDO. The NIH research infrastructure needs to coordinate with the interoperability frameworks by ASTP/ONC in addition to the terminologies stewarded by the NLM.

Coordinated governance among NIH, NLM, and ASTP/ONC, which would be similar to CSIRO's approach to national science and data infrastructure, could help towards an interoperable ecosystem.

The strategic alignment across these standards, ontologies, and federal interoperability frameworks need to coordinate and be interoperable in terms of infrastructure for genomic, phenotypic, and clinical data.

Such alignment will support rare disease research and precision medicine to allow the US to be a global benchmark for biomedical research and innovation.

As a result, the US could provide unified platforms for the research and academic communities to evaluate and assess the momentum towards interoperable precision infrastructure and beyond.

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

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

38. N/A

Submit date: 3/12/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Our team has had some experience working with one of the Repositories listed as a Controlled-Access Repository. The process of sharing de-identified data from our study was not seamless and took several months, with multiple emails and phone calls. If NIH is implementing using only Controlled Access Repositories for Public Use Data, it would be very important to provide support for these Repositories, including financial to ensure data is available long term, technical to assist with uploading data packages, and management to guarantee responsive communication.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Please see comments above.

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

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

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

39. Jennifer K. Wagner, Laura Y. Cabrera, and Satrajit Ghosh

Submit date: 3/12/2026

I am responding to this RFI: On behalf of myself

Name: Jennifer K. Wagner, Laura Y. Cabrera, and Satrajit Ghosh

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Wagner-Cabrera-Ghosh-Public-Comments-Submitted-2026.03.12.pdf

40. N/A

Submit date: 3/12/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The simple fact is that any data deemed shareable, should be shareable everywhere. It seems highly impractical to attempt to keep some data from a list of "Countries of Concern", while still making that data reasonably open and available elsewhere. In practice, trying to do so can only harm the openness of the data. This will greatly harm reproducibility efforts and hinder progress in health-related areas. Both of these features inherently run against the mandates and goals of the NIH. These rule changes should not be adopted.

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

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

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

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

41. Vincent Mor

Submit date: 3/12/2026

I am responding to this RFI: On behalf of myself

Name: Vincent Mor

Name of Organization: Brown University, School of Public Health, Center for Gerontology

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The policy proposes that certain data types "may only be shared without access controls if (1) there is informed consent explicitly stating data are to be shared openly without controls. In these instances, institutions must still review to determine that openly sharing these data pose very low risk when shared and used;..."

It is unclear what the policy means by 'institutions' - who will be allowed (or disallowed) from making that determination - the IRB, the PI, some other entity?

What will be the informed consent requirements for controlled access sharing? Will data sharing be required even for studies working under a waiver of informed consent? Most Institutional Review Boards will not necessarily know that controlled data access and sharing is in their purview and may be confused in the absence of some direction from the Office of Human Subjects Protection offering exemplar language that can be incorporated into Informed Consent documents in which the person signing consent acknowledges that the information that they are providing may be shared with other investigators who might use the data for alternative purposes. The identity of the individual study participants will be protected to the extent possible.

Controlled access data repositories require prospective review of requests for access to data. Who will be required to assume this responsibility and for how long? The researcher, the repository owner, the original data owner if it is secondary use? It is not reasonable to expect that a researcher or their institution will be able to support this process after their own study's funding runs out, and during the study period needed funds may have to be re-routed to these efforts. Will there be funding mechanisms to support these activities during and after the grant period ends?

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
One implication of this policy requiring controlled access to health data that has been "de-identified" but still allows connection to patients' providers and/or geography is that a mechanism has to be established and monitored to insure that the data are being used in a manner consistent with the purpose of the original study but being shared withother researchers without explicit identifiers but, to make the data useful, links to provider data and geo-location. In order to review applications to reuse the data and to monitor data use to insure that new users don't violate identifiability precepts, data repositories will have to devote considerable resources to these review and oversight functions well

after grant support is no longer available. Since few Universities have this kind of infrastructure and would be willing to sustain this effort in the absence of ongoing funding and support, unless governmental entities are willing to support this kind of infrastructure, it is unlikely that this policy will achieve the intended goal of increasing data sharing.

Each of the existing NIH CADRs are limited in scope by institute and/or focus. Are there plans to expand? E.g., will investigators funded by NIA be able to use the NIA LINKAGE program CADR for data that are not going to be linked to CMS data?

Would NIH consider keeping a registry of non-NIH CADRs that meet the requirements so that researchers know where they can safely deposit data?

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Each of the types of data indicated below are appropriate to be covered by the policy:

Covered personal identifiers

Precise geolocation data

Epigenomic data

Personal health data

Personal financial data

Individual level clinical trial data

Imaging data of the human face or head regions

Additionally, information about the identity of the providers (physicians, hospitals or other health care organizations serving the study subject should be included as protected data.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
no comment

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
no comment

42. Michael House

Submit date: 3/13/2026

I am responding to this RFI: On behalf of myself

Name: Michael House

Name of Organization: Child Mind Institute

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Implementing more controlled access policies, rather than promoting open science, will worsen access for emerging researchers and students data essential to development of novel research. The greatest threat to private and sensitive data currently comes from the corporate sector. Federal funding cuts have left researchers unemployed in some cases, and instituting control policies for data requiring significant expertise will burden new or emerging institutions and researchers by limiting access to data sources. As we advance with new technologies, we must focus on expanding research data access, not imposing controls and restrictions that historically harm marginalized populations while failing to adequately address corporate misconduct. When establishing access controls in response to research misconduct, consider the extent of personal, life-altering information that has been released intentionally, accidentally, or maliciously with little to no consequences. Let's first hold IRBs personally accountable for policy failures if we want to reduce misuse and leakage, as currently, they lack the authority to enforce regulations effectively.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
I'm not going to fill out the rest of your bureaucratic bullshit, you can read my message about it, if you want my opinion I'll give it but in my way and that will require a employee sorting it because I'm not burdening myself giving my opinion in some structured format. You're asking for a favor, don't heap more ways to reduce opinion through compliance

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Human brain and imaging data is no more identifiable than any other data source and takes a tremendous amount of skill and ability to even begin manipulating or making sense of it. It takes tremendous effort and often in times of government turmoil leaves researchers without access to any relevant data while waiting for responses from NIH to already access it and again, someone seeing a patients ears may identify them, but much more likely to happen is Equifax leaking your social security number. You'll hurt access because of a knee jerk reaction to fear and the utter denial that our corporate sector has already exposed every American's private information many times over to bad actors. NIH systems rely on trust, bad actors are more than happy to lie and only the responsible ones will face delays or hurdles.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Fire who ever proposed this while there is such a serious field wide danger to research even continuing inside the United states because of the reduction of funding to pre 2000s level. Use their salary to fund a

grant for graduate students, since guess what, the pipeline is currently stuck for 2 cycles so far and each year its stuck sends ripples for years to come.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Social engineering is still the easiest way to get access to data, it will continue to be and denying that by listening to a industry that only is employed because of a kabuki theater style gaslighting effort over access makes our data less safe.

Description: We all contribute to society and policy

43. Federation of American Societies for Experimental Biology (FASEB)

Submit date: 3/13/2026

I am responding to this RFI: On behalf of an organization

Name: Eric E. Kelley, PhD

Name of Organization: Federation of American Societies for Experimental Biology (FASEB)

Type of Organization: Professional Organization/Association

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
FASEB commends NIH's efforts to standardize Controlled-Access Data Policy across all ICs. In addition to points raised throughout our response, we also seek clarification on several aspects of the draft Controlled-Access Data Policy.

A core concern is feasibility of implementation. Several of the proposed requirements for repository eligibility and protected data types will increase demand for repository space, and it is unclear whether NIH-supported controlled-access data repositories have the capacity to accommodate this new demand. The expanded scope of protected data types will also increase administrative and institutional burden, particularly for institutions with less experience and/or fewer resources to support projects covered by this policy. Combined, these factors risk slowing adoption and implementation of the policy due to lack of resources and/or lack of understanding.

The role of the institution in protecting data collected while conducting NIH-supported research is also unclear. The Policy states that institutions must ensure that the listed data types are protected, including data not shared in a controlled-access repository. Thus, FASEB recommends that the final policy clarifies responsibilities imposed on institutions. Additional clarification is also needed for the standards for institutional repositories and criteria for determining a dataset's lifecycle.

FASEB also recommends that the final policy includes clearer parameters regarding the need for "assessment" for both determining risk associated with data shared prior to implementation of this Policy and the need for controls. As currently written, it is unclear who is responsible for assessment, what the assessment criteria include, and how often data not managed by this Policy must be assessed.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
FASEB encourages NIH to allocate funding to expand the capacity of existing repositories to accommodate the anticipated influx of controlled-access datasets. While NIH specifies that the use of other controlled-access data repositories is acceptable, many independent repositories (hyperlink: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/scientific) previously utilized for these data are unlikely to meet the new restrictive requirements, particularly those pertaining to data sharing with countries of concern (as defined by 28 CFR Part 202 [hyperlink: https://www.federalregister.gov/documents/2025/01/08/2024-31486/preventing-access-to-us-sensitive-personal-data-and-government-related-data-by-countries-of-concern]).Thus, implementation

of controlled-access data repository (CADR) restrictions may prove difficult as written if NIH controlled repositories are unable to accommodate greater demand.

Once the policy is finalized and the focus turns to implementation, FASEB urges NIH to ensure appropriate financial support for established NIH CADRs to meet the expected growth in demand. Increased funding will allow expansion of available CADR storage and increase access request review capacity. Oversight of NIH CADRs must also be expanded to ensure that repositories are operating at an acceptable level of risk, as established by the NIH Security Best Practices for CADR (hyperlink: https://grants.nih.gov/sites/default/files/flmngr/NIH-Security-BPs-for-Controlled-Access-Repositories.pdf) in accordance with NIST SP 800-53 (hyperlink: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final).

At present, established NIH CADRs (hyperlink: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/requirements) are spread across multiple NIH Institutes and Centers (ICs) and utilize several different access systems. The current effort to standardize expectations for data privacy across all ICs presents the unique opportunity to consider a new and potentially more efficient NIH repository structure. For example, NIH might consider consolidating existing repositories. Unifying repositories has the potential to reduce both administrative burden and the risk of error in implementing the proposed Controlled-Access Data Policy and future policy measures.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
While FASEB appreciates NIH's commitment to protecting the identity and personal data of human research subjects, there are questions and concerns regarding the scope of several stated protected data types:

- Personal health data. It is unclear whether NIH intends to distinguish between clinical personal health information and research health information. As personal health information is defined in this Policy, nearly all NIH-funded human subjects research would be subject to the Controlled-Access Data Policy. As this Policy is written, it will increase administrative burden and stymie open science initiatives.

- 'omics data derived from human cell lines. Language defining the scope of this Policy indicates that research generating data derived from human cell lines is subject to the Controlled-Access Data Policy. However, the Policy also states that the collection and sharing of human cell lines are not covered, creating ambiguity. FASEB recommends revisions to provide clearer and more specific guidance regarding when cell lines become subject to the Policy and whether immortalized cell lines are covered.

- Imaging data of the human head or face. Additional specification of imaging techniques should be provided. Similarly, we encourage inclusion of language regarding whether the use of "de-facing" (hyperlink: https://pmc.ncbi.nlm.nih.gov/articles/PMC10389782/) techniques to de-identify participants in images of the head and brain, are sufficient to negate the need for coverage by the Controlled-Data Access Policy.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
FASEB appreciates the inclusion of updates to the NIH Genomic Data Sharing (GDS) Policy as part of this

effort. The proposal to limit GDS coverage to only human data is a welcome update that will reduce administrative burden for researchers working with non-human subjects. However, there are several areas of concern for which additional clarification or supplementary guidance would be helpful for implementation.

First and foremost, we are concerned that defining a "large scale" genomic dataset as any dataset including data from 100 or more individuals is an arbitrary approach to an important policy matter. While risk for re-identification is often cited as a concern with small scale datasets, the Toolkit for Assessing and Mitigating Risk of Re-identification when Sharing Data Derived from Health Records (hyperlink: https://www.sentinelinitiative.org/sites/default/files/Methods/Sentinel_Report_Toolkit-Assessing-Mitigating-Risk-Re-Identification-Sharing-Data-Derived-from-Health-Records.pdf) suggests that small dataset size does not raise the risk of re-identification. Rather, high congruence between protected data and publicly available data is what drives re-identification risk. Thus, NIH might consider an alternative approach to determining inclusion or exceptions to the GDS Policy.

The application of GDS Policy to datasets including fewer than 100 subjects is particularly important for rare disease research, for which recruiting 100 participants is difficult or impossible. Provisions to include small scale datasets would minimize IC-specific interpretation (hyperlink: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/gds/developing-genomic-data-sharing-plans#:~:text=The%20National%20Institutes%20of%20Health,the%20funding%20IC's%20program%20officer.) of the GDS Policy, ultimately streamlining implementation and reporting for both researchers and program administrators.

Similarly, FASEB recommends inclusion of a clearer definition for the timing of genomic data release as part of the final policy. Specifically, our community seeks clarity regarding what is meant by "immediate" release of data to a NIH-controlled repository and whether it accommodates a reasonable delay for data cleaning and quality control, as indicated in the current policy. Current policy (hyperlink: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/gds/data-submission-and-release-expectations#:%7E:text=Level%203%20%2D%20Data%20Submission%20&%20Release%20Expectations&text=Human%20data:%20Data%20submission%20is%20expected%20after%20cleaning%20and%20quality,their%20PO%20with%20further%20questions.&text=Data%20Release%20Expectations-,Human%20data:%20Data%20release%20is%20expected%20up%20to%206%20months,data%20types%20or%20NIH%20projects.) specifies that submission of human genomic data is expected within approximately three months of collection, only after data cleaning and quality control have been completed. FASEB supports this timeline to ensure the highest quality data are available in NIH controlled-access repositories, fostering reproducibility across studies and reducing administrative burden.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
No Comments.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/FINAL_FASEB-Comments-on-NIH-Controlled-Access-Data-Policy-and-GDS-Revisions_20260313.pdf

Description: Comments from the Federation of American Societies for Experimental Biology (FASEB) formatted on organizational letterhead.

44. Peter Turkeltaub

Submit date: 3/13/2026

I am responding to this RFI: On behalf of myself

Name: Peter Turkeltaub

Name of Organization: Georgetown University Medical Center

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
This policy would be a disaster for medical research in the US. My institution and many others are not currently able to adhere to the NIST-SP-800-171 requirements. We are actively working to develop systems that meet the requirements, but doing so at scale will take years and likely cost millions of dollars. Many institutions are in the same boat and are scrambling right now to find solutions that will allow them to use public data currently covered under the policy. I understand and support the importance of data security and protection, but the NIST-SP-800-171 requirements have already slowed progress on important research.

Applying the NIST-SP-800-171 criteria so broadly would put a sudden halt to human subjects research in the US and would badly damage our competitiveness with other countries.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Established repositories cannot meet the needs of every researcher for every project. The range of research data proposed under this policy are too broad.

Researchers also need to be have ready access to their own data-- the requirement that even unshared data collected by individual investigators adhere to the NIST-SP-800-171 requirements would place a huge burden on investigators, and cause slowdowns in data processing, slowing progress on critical medical research to benefit Americans.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
"Individual level clinical trial data" is too broadly defined.

"Imaging data of the human face or head regions" should specify imaging that might allow identification of a person--- many imaging data types are too low resolution or measure signals (e.g., cerebral perfusion) that do not permit identification of a person.

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

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

45. Northwestern University

Submit date: 3/13/2026

I am responding to this RFI: On behalf of an organization

Name: Eric Perreault

Name of Organization: Northwestern University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Included in uploaded PDF.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Included in uploaded PDF.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Included in uploaded PDF.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Included in uploaded PDF.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Included in uploaded PDF.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/2026.03.13_NIH-RFI-Response_Signed-EJP.pdf

Description: Response to Request for Information - Northwestern University

46. Elissa Newport

Submit date: 3/13/2026

I am responding to this RFI: On behalf of myself

Name: Elissa Newport

Name of Organization: Georgetown University

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
NIST is not available at my university or at any other universities i know of. This requirement will make our research on human subjects impossible for years. This level of security is not necessary to preserve privacy; we can do relevant security for de identified data by methods we have available and can afford.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
We do not have any repository with NIST security and will not have any for years

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The data types are too numerous. Many can be protected without NIST

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

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

47. Laura Scott

Submit date: 3/14/2026

I am responding to this RFI: On behalf of myself

Name: Laura Scott

Name of Organization: University of Michigan

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
"Human genomic data should be submitted to an approved NIH controlled-access data repository (see: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/best-practices) to make the data available within 6 months of generation, to allow time for data cleaning, quality control, and repository release processes. Initial sequence reads and raw data do not have to be shared, consistent with previous GDS Policy expectations. Data not shared within 6 months of data cleaning and quality control should be shared consistent with DMS Policy requirements (publication date or end of the award period, whichever comes first). "

Comment on the text above: You can not hold a researcher responsible for the repository getting the data out in six months. This can take months or even years in our experience with certain NIH databases. The data sharing should only specify when the data must be submitted to the repository by the researcher. Some submission processes require complex establishment of studies, permissions of other investigators and permissions, which would take up a very large amount of time, in our experience 6 months can be too short to do even this.

Processing large amounts of genomic data after physical generation of the data can take some months of mapping just to get it to the point of being able to QC. QC can take a (very) long time to do well. You want well QC'ed data in the repository to help promote reproducible science. You also want the data that is used in publications to be in the repository so that science is reproducible. This is often impossible in 6 months. QC can take a year to 1.5 years (particularly with new data types and the newly developed tools for cleaning data that need to be tested). If this six month rule were implemented it would cause there to need to be different versions QC'd datasets in the repository. A six month dataset that wasn't well QC'ed and would lead to issues for people that used the data, and a second 1.5 year dataset that was much better QC'd and that would correspond to data used in publications. Deposit 1.5

years after data generation would be much, much better for insuring reproducible, gold standard science.

The last line of the text above is ambiguous. Does it mean if you miss the deadline for the 6 months, that other rules apply and you can wait until you publish?

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

48. Washington University

Submit date: 3/15/2026

I am responding to this RFI: On behalf of an organization

Name: Melanie Roewe

Name of Organization: Washington University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Draft NIH Controlled-Access Data Policy

Overarching Principle: WashU supports NIH's commitment to protecting research participants while advancing scientific discovery through responsible data sharing. However, effective data governance requires risk-proportionate controls that align security measures with actual privacy risks. The draft Controlled-Access Data Policy conflates fundamentally different risk profiles -applying national security frameworks to academic research and imposing identical controls to both identified clinical records and properly de-identified datasets. This response identifies critical design flaws requiring policy revision and implementation challenges requiring operational solutions. Our recommendations aim to establish a tiered governance framework that protects participants effectively while preserving the scientific utility and accessibility of research data.

DESIGN FLAWS REQUIRING POLICY REVISION

The following issues reflect fundamental conceptual problems with the draft policy that cannot be resolved through implementation guidance alone. These require substantive policy changes:

Overly Broad Scope: Essentially All NIH-Funded Human Research Would Be Covered

The definitions of the types of protected data in the draft policy are extremely broad. The policy would encompass nearly all data collected from NIH-funded research involving human participants, without differentiating between identified, identifiable, and properly de-identified data. For example, the definition of “Personal Health Data” would include any health-related information, including "basic physical measurements and health attributes (such as bodily functions, height and weight, vital signs, symptoms, and allergies); social, psychological, behavioral, and medical diagnostic, intervention, and treatment history; test results; logs of exercise habits; immunization data, data on reproductive and sexual health; and data on the use or purchase of prescribed medications." By treating all health-related data the same regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles.

This definition, combined with the other data types, would have the cumulative effect of imposing significant data-sharing restrictions on studies such as those listed below. This would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial financial compliance burdens on research that poses minimal privacy risk:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

Security Requirements Are Not Risk-Aligned

NIST 800-171 Security Standard: The draft policy requires repositories to employ "security standards for protection of controlled data (e.g., NIST-SP-800-171 or equivalent)." This requirement reflects a fundamental misalignment between risk and security measures. NIST Special Publication 800-171 "Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations" contains 110 security requirements across 14 families of controls, including requirements for physical security, personnel screening, incident response capabilities, and extensive technical controls. This standard is used to protect highly sensitive data such as Controlled Unclassified Information (CUI) in the defense industrial base and other sensitive government operations. The NIST 800-171 security controls represent a security posture appropriate for protection of information that, if disclosed, could reveal defense technologies or threaten national security interests. Applying this standard to de-identified health research data conflates two fundamentally different risk profiles: national security risks and individual privacy risks.

Table: Comparison of Threat Models

Dimension National Security Context

(NIST 800 171 / Controlled Unclassified Information — CUI) Research Privacy Context

(De identified Data)

Threat Actors

Who is trying to access the data? - Nation state adversaries

- Foreign intelligence services

- Sophisticated cyber criminal organizations

- Well-resourced actors with strategic interests - Accidental disclosure by authorized users

- Statistical re identification attempts by researchers

- Data brokers seeking to link datasets

- Limited resources and motivation

Threat Likelihood

How probable is a successful attack? - Persistent, ongoing threat

- Well resourced and sustained efforts

- High motivation and capability - Low probability when properly de identified

- Requires motivation + skills + supplementary data

- Mitigated by existing safeguards (IRBs, DUAs)

Harm Magnitude

What are the consequences of disclosure? - National Security compromise

- Military/infrastructure damage

- Damage to critical infrastructure

- Loss of strategic advantage

- International or large scale economic harm - Individual privacy breach

- Potential discrimination or stigmatization

- No systemic or national impact

- Regulatory protections (HIPAA, Common Rule) offer recourse

Appropriate Security Response - 110+ security controls (NIST 800-171)

- Physical security and personnel screening

- Continuous monitoring and incident response

- Assume persistent, capable adversary - Risk-proportionate access controls

- Researcher vetting and data use agreements

- Standard cybersecurity hygiene (encryption, access controls)

- Focus on preventing accidental disclosure

The fundamental mismatch between threat models explains why applying NIST 800-171 to de-identified research data is inappropriate. The security controls designed to protect against nation-state espionage are neither necessary nor proportionate for protecting against the statistical re-identification risks posed by properly de-identified research data.

Reidentification Risk: Most critically, the policy fails to differentiate security requirements based on the level of identifiability of the data. Data that has been de-identified according to HIPAA standards - either through Safe Harbor removal of 18 identifiers or through Expert Determination demonstrating very small re-identification risk - poses dramatically lower privacy risk than identified or identifiable data. The scientific literature on re-identification demonstrates that properly de-identified datasets with appropriate safeguards have extremely low re-identification rates. For example, research on Safe Harbor de-identified data found re-identification risk of approximately one in 3,500-a risk that "falls somewhere between one's lifetime odds of being personally struck by lightning (about one in 10,000) and the risk of being affected because someone close to you has been struck."[1] Similarly, a recent 2025 study of the National Cancer Institute cancer registry found that re-identification risk using privacy-preserving record linkage techniques was approximately 0.0002 (2 patients out of 10,000).[2] Yet the policy would impose

identical security requirements on a properly de-identified dataset as on a dataset containing names, social security numbers, and addresses.

[1] Daniel C. Barth-Jones, The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now (2012), https://fpf.org/wp-content/uploads/2025/05/DBJ_Weld_Re-Identification.pdf.

[2] Murat Kantarcioglu et al., Privacy-Preserving Record Linkage: Methods, Applications, and Lessons Learned from the National Cancer Institute's SEER Program, J Am Med Inform Assoc (2025), https://academic.oup.com/jamia/advance-article-abstract/doi/10.1093/jamia/ocaf172/8304360.

Right-Sized Alternatives: NIH needs to evaluate security standards that are "right-sized" for the level of risk posed by research data. Several more appropriate alternatives exist: First, NIH should consider the security controls being developed for the Research Security Program that is required under NSPM-33 (National Security Presidential Memorandum 33). While these security standards are not currently published, ongoing efforts by associations, in collaboration with NIST, are developing best practices. This framework would be far more appropriate for de-identified human research data than the CUI-focused NIST 800-171 standard, as it is specifically designed for the academic research environment.

Recommendation: NIH should adopt a tiered security framework that aligns security controls with actual re-identification risk. Consider adopting security standards that align with the cybersecurity element of the Research Security Program under NSPM-33.

Misapplication of National Security Framework: The DOJ Bulk Sensitive Data Rule

The policy's reliance on the Department of Justice's "Preventing Access to Americans' Bulk Sensitive Personal Data" rule (28 CFR Part 202) as the basis for defining protected data types represents a fundamental category error. The DOJ rule is explicitly focused on preventing foreign adversaries from accessing large-scale datasets that could threaten national security. The rule's stated purpose is to prevent "countries of concern" from obtaining bulk data that could be used for intelligence operations, tracking government personnel, cyber operations, etc. This is fundamentally different from protecting individual research participants' privacy in the context of scientific research.

Bulk Data Thresholds: Critically, the DOJ rule establishes specific threshold quantities that define "bulk" sensitive personal data - recognizing that small-scale data poses fundamentally different risks than large-scale datasets. The NIH policy adopts the DOJ rule's data type definitions but does not include any bulk data thresholds. This means a researcher collecting basic health information from even a handful of participants would be subject to security requirements designed to protect massive databases from nation-state adversaries.

The threat model underlying the DOJ rule - preventing foreign governments from purchasing or accessing large datasets of Americans' information - does not apply to academic research. In academic research a number of safeguards are already in place, including (1) data is typically already subject to IRB oversight and informed consent; (2) data sharing typically occurs through controlled mechanisms with approved researchers; (3) studies often involve limited numbers of participants; and (4) data is often de-identified to minimize privacy risk.

Covered Persons Vetting: Further, the draft policy requires repositories to implement “restrictions for sharing data with countries of concern as identified in Part 202.” However, the DOJ regulation does not simply limit access to individuals or entities located in countries of concern. Rather, it restricts access to “Covered Persons”, which is defined broadly to include any entity that is 50% or more owned by an entity or person in a country of concern, regardless of where that entity is located. This would require every repository housing NIH-funded data to implement vetting procedures that determine the ownership structure of requesting entities. Most academic institutions do not have the capability to conduct this type of ownership analysis, especially for activities that occur as frequently as scientific data sharing. The compliance burden would be substantial, requiring legal expertise and resources that academic institutions do not possess.

Recommendation: NIH should:

1. Adopt threshold quantities similar to those in the DOJ rule, recognizing that small-scale research data does not present national security risks;

2. Develop data type definitions specifically tailored to research contexts rather than importing wholesale definitions from a national security regulation. For example, for genomic data, distinguish between whole genome sequencing, exome sequencing, targeted gene panels, and single-gene tests, which pose vastly different re-identification risks. For personal health data, distinguish between identified clinical records, aggregate survey responses, and HIPAA deidentified research datasets. Consider whether data collection involving basic measurements from small numbers of participants warrants the same controls as large biobanks with extensive clinical phenotyping.

3. For controlled -access repositories, limit the operational standard to a prohibition on sharing with entities in a country of concern, rather than adopting the restrictions imposed by the DOJ regulation (which apply more broadly to a “Covered Person”). Additionally, NIH should develop a mechanism to approve exceptions to this prohibition for some scientific collaborations. For example, an NIH funded clinical trial where an investigational drug is coming from a pharmaceutical company located in China and is being provided free to trial participants. In exchange for providing free drug, the Chinese-based pharma company will require access to the clinical trial data and results. This arrangement would be beneficial for US study participants providing access to the investigational product at no cost. There needs to be a mechanism for NIH to approve data sharing in this scenario.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the design flaws identified above were addressed, the policy as drafted would create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: NIH requests input on "availability of established repositories for implementing the proposed Controlled-Access Data Policy." Given the vast expansion of data types requiring controlled access, current repository infrastructure is inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous

personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types "may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls." While requiring explicit consent for true open access sharing may be appropriate, the policy creates a significant practical problem by defining "protected data types" so broadly that this requirement would apply to nearly all NIH-funded human subjects research. In practice, it is uncommon for institutions to use this type of language in their informed consent documents. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of "with" or "without controls." This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed "sharing without controls." Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved-not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures rather than clarifies data sharing practices.

Data Use Agreement Administrative Burden: The policy's requirements for security and operational standards would eliminate the common practice of many repositories that use click-through data use agreements where individual researchers can accept terms directly. Instead, all access to any protected data type would require formal and binding data use agreements with the individual requester or their employer. This would dramatically increase institutional workload for processing data use agreements. Many repositories currently use streamlined click-through agreements for low-risk deidentified data, reserving formal institutional agreements for more sensitive data. The policy would eliminate this flexibility.

Data Stewardship and Retention Obligations: The policy should clarify that institutional obligations to protect controlled-access data end once the institution no longer possesses the data. Specifically, when researchers deposit data into an approved repository, the depositing institution should not retain an ongoing obligation to maintain, secure, or monitor the data. The repository accepting the data assumes responsibility for implementing the security controls and access management required by the policy. Without this clarification, institutions may face indefinite compliance obligations for data they no longer control or possess.

Risk Assessment for All Other Data: The policy requires that "data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls." This creates an

impossible compliance burden when combined with the policy's extremely broad definitions of protected data types. The policy provides no operational guidance on how this assessment should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. DMS plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources. Recommendation: NIH should explicitly allow existing oversight determinations to satisfy controlled-access policy requirements. Specifically, if an IRB has determined that data sharing with specified protections is consistent with participant consent and poses minimal risk, and if the approved Data Management and Sharing Plan describes appropriate safeguards, these determinations should be sufficient to satisfy the policy's risk assessment requirements without requiring a separate, redundant review. This consolidation would preserve participant protection while eliminating unnecessary administrative burden and conflicting review processes.

Impact on Scientific Progress and Research Equity

Beyond the operational and compliance burdens described above, the draft policy will result in harm to scientific progress and research equity that warrant serious consideration:

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet the security and administrative requirements, creating a two-tiered research ecosystem where only well-resourced institutions can effectively access and utilize shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments. Moreover, this approach is fundamentally at odds with the current administration's stated commitment to accelerating artificial intelligence in healthcare. AI development in healthcare depends critically on access to large, diverse datasets for training and validation. Machine learning models require substantial amounts of real-world health data to achieve clinical utility and to ensure algorithms work equitably across diverse populations. By imposing

excessive barriers to accessing de-identified research data that poses minimal privacy risk, the policy would significantly impede AI innovation in healthcare. The policy creates a fundamental tension between participant protection and scientific innovation that could be resolved through risk-proportionate controls rather than uniform restrictions.

These effects will be particularly pronounced for de-identified, minimal-risk data where the privacy protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Some proposed revisions to the GDS Policy are positive developments that will reduce complexity and improve clarity, while others are more complex and require further review:

- Scope Limited to Human Genomic Data: Limiting the GDS Policy scope to human genomic data is appropriate. Non-human genomic data poses no individual privacy concerns and is appropriately governed by the general DMS Policy. This revision eliminates unnecessary compliance burden without compromising participant protection. However, the policy lacks clarity on the boundaries of what

constitutes "human genomic data." While genomic sequencing data clearly falls within scope, many related molecular data types can reveal genomic information. For example, proteomics data can identify mutations in protein structure from which genetic sequences can be inferred. Similarly, transcriptomics data reflects gene expression and can reveal genetic variation. Would the GDS Policy apply to transcriptomics data? Proteomics data? Exosomes data? Metabolomics data that may correlate with genetic variants? The policy should clearly define which molecular data types are considered "human genomic data" subject to the GDS Policy versus which are governed by the general DMS Policy and Controlled-Access Data Policy.

- Large-Scale Threshold of 100 Individuals: The proposed revision lowers the threshold for "large scale" genomic data from the current standard to 100 individuals. While establishing a clear numeric threshold provides helpful clarity, the policy lacks critical implementation guidance on how this threshold will be applied in practice. Several scenarios require clarification: (1) Program vs. Study Application: Will the threshold apply to an entire NIH-funded program or to individual substudies or IRB protocols within that program? For example, could a researcher avoid the large-scale designation by administratively dividing 1,000 participants across 10 separate "studies" under the same funded program? (2) Multi-Year Studies and Evolving Enrollment: How will the threshold apply to studies that span multiple NIH funding cycles or that increase planned enrollment over time? If a study begins with 75 participants (below threshold) but later expands to 150 participants, would the original 75 participants' data be managed under different requirements than the newly enrolled participants' data? Would investigators need to create entirely new studies to avoid having legacy data managed inconsistently with newly collected data? Additionally, establishing a lower threshold will significantly affect research avenues that rely on fundamental resources that should remain open access. An example is pangenome reference work that seeks to replace the linear reference with a group of references that will better represent diversity. Genomic datasets should continue to be managed under the DMS Policy and Controlled-Access Data Policy.

- HIPAA Expert Determination: Allowing expert determination as an alternative to Safe Harbor deidentification is appropriate and provides valuable flexibility for datasets where removal of all 18 identifiers would compromise scientific utility.

- Establish consistent requirements across the NIH: We agree that consistency across the NIH is important to ensure compliance, but to not have a path for exceptions would greatly hamper, nor recognize the speed at which research takes place and evolves, a pathway for exceptions must be created in some format.

- Strengthening requirements for participant consent: It is ideal that all biospecimens or cell lines created or collected after 2015 have consent, but we maintain that it should be up to each institution conducting the research to determine use through its Human Research Protection Office in conjunction with its Institutional Review Boards, as opposed to creating an arbitrary date of compliance. Additionally, the proposed revisions do not address the need for re-consent of minors. The policy does not address whether or when researchers must obtain re-consent from participants who were enrolled as minors once they reach the age of 18. This is a significant operational and ethical question, particularly for longitudinal genomic studies that may span decades. If a participant provided assent as a minor with parental consent, does continued data sharing require re-consent upon reaching adulthood?

What happens to previously shared data if an individual declines to re-consent as an adult? NIH should provide clear guidance on re-consent requirements for participants enrolled as minors.

- Imputation Services: NIH should allow users to develop their own imputation panels or servers using controlled-access data from studies subject to the GDS Policy and we agree that the users should ensure that (1) the controlled-access data used to develop the imputation panels are protected from disclosure and attacks specific to imputation servers, (2) the imputation server operates in an environment consistent with security controls in the NIH Security Best Practices for Controlled-Access Data Repositories and, (3) the imputation servers are funded or operated by NIH or another federal agency.

- Data Availability Timeline: The proposed revision requiring genomic data to be made available within 6 months of generation to allow time for data cleaning and repository release processes is overly burdensome and insufficiently flexible. Sequencing data generation is not a single event but a process that requires extensive quality control checks and computational analysis, including alignment, scaffolding, variant calling, and validation steps. The time required for these essential processes varies substantially depending on the sequencing platform, the complexity of the organism or genomic region, the scale of the study, and available computational resources. A rigid 6-month timeline fails to account for these variables and may force release of inadequately processed data or incentivize investigators to delay formal "generation" of data until analysis is complete, undermining the policy's intent. Additionally, this requirement is too broad and risks becoming outdated as sequencing technologies continue to evolve rapidly. NIH should instead establish principles for timely data sharing while allowing flexibility based on data type, complexity, and processing requirements, with expectations described in Data Management and Sharing Plans rather than imposed as a uniform deadline.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/WashU-response-to-NIH-RFI-re-Controlled-Access-Data-Policy-and-GDS-Policy-revisions_clean_final.docx

49. Adolescent Brain and Cognitive Development Study

Submit date: 3/15/2026

I am responding to this RFI: On behalf of an organization

Name: Deanna Barch

Name of Organization: Adolescent Brain and Cognitive Development Study

Type of Organization: Other

Type of Organization - Other: Research Consortium

Role: Other

Role – Other: Data Science Strategist

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
We write as investigators involved in the Adolescent Brain and Cognitive Development (ABCD) Study, a large-scale longitudinal investigation of brain development that is widely used as an open-access scientific resource. The ABCD Study has enabled researchers worldwide to address critical questions aligned with the mission of the U.S. Department of Health and Human Services. Based on our experience managing and sharing large-scale human research data, we have several serious concerns regarding the Draft NIH Controlled-Access Data Policy.

Draft NIH Controlled-Access Data Policy

The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts involving national security risks, including defense-related research and sensitive government operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on “Preventing Access to Americans’ Bulk Sensitive Personal Data” (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule’s data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations

- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types “may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls.” While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining “protected data types” so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of “with” or “without controls.” This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed “sharing without controls.” Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that “data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls.” This creates an

impossible compliance burden when combined with the policy's extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.
- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.

- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

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

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types “may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls.” While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining “protected data types” so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of “with” or “without controls.” This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed “sharing without controls.” Importantly,

requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that “data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls.” This creates an impossible compliance burden when combined with the policy’s extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes “potential sensitivities” for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher’s computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions (“equivalent”).

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or

its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.
- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.
- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy’s effective date should be afforded for user-

side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts involving national security risks, including defense-related research and sensitive government operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on “Preventing Access to Americans’ Bulk Sensitive Personal Data” (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule’s data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations
- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types “may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls.” While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining “protected data types” so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of “with” or “without controls.” This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed “sharing without controls.” Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that “data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls.” This creates an impossible compliance burden when combined with the policy’s extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be

conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.

- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.

- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

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

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

50. Deanna Barch

Submit date: 3/15/2026

I am responding to this RFI: On behalf of myself

Name: Deanna Barch

Name of Organization: Washington University

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I write as a research who is both helping to generate data that is impacted by this policy and as someone who uses this data.

The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts

involving national security risks, including defense-related research and sensitive government operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on “Preventing Access to Americans’ Bulk Sensitive Personal Data” (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule’s data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations
- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types "may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls." While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining "protected data types" so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of "with" or "without controls." This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed "sharing without controls." Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that "data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls." This creates an impossible compliance burden when combined with the policy's extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.

- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.

- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do

not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

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

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types "may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls." While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining "protected data types" so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of "with" or "without controls." This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed "sharing without controls." Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that "data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls." This creates an impossible compliance burden when combined with the policy's extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.

- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.

- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do

not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts involving national security risks, including defense-related research and sensitive government operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific

evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on "Preventing Access to Americans' Bulk Sensitive Personal Data" (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule's data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations
- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert

substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types "may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls." While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining "protected data types" so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of "with" or "without controls." This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed "sharing without controls." Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that "data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls." This creates an impossible compliance burden when combined with the policy's extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active

Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.
- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.
- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding

individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

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

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

51. Stephen Rosenfeld

Submit date: 3/15/2026

I am responding to this RFI: On behalf of myself

Name: Stephen Rosenfeld

Name of Organization: North Star Review Board

Type of Organization: Institutional Research Oversight Committees (e.g. IRB/ IBC/ IACUC)

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Regarding the acceptability of the 18 HIPAA identifiers as sufficient to constitute "de-identification" - I believe that the concept of "de-identification" has become less meaningful with the ubiquitous collection of data on individuals in both the medical and commercial realms. While it would represent a major change to this policy, I suggest that NIH take the lead and formally acknowledge this reality. An alternative approach would be to explicitly recommend or require data obfuscation techniques like differential privacy. Such approaches have the ability to restore meaning to "de-identification," but are inconvenient and potentially costly to implement. They will not be implemented unless required, but I don't know that we can fulfill the promises we make in our informed consent forms to protect confidentiality without a new approach, given evolving technologies.

Regarding the need for assent from minors - please consider explicitly recognizing that individuals who were assented as minors should be asked for consent to continued use of their data when they become adults. Once shared, data is available for future use, and data from individuals able to give consent for themselves should not be used without such consent.

Regarding expanding institutional review capacity - the proposal invites institutions to establish additional oversight review by individuals or committees other than IRBs, Privacy Boards, etc. Given the current reality that such existing entities already struggle with structural conflicts of interest arising from institutional/commercial interests versus the interests of research participants, I'm concerned that any new entity would face similar conflicts, potentially without the statutory requirement to represent participant interests. In much research, confidentiality risk is the major risk facing participants, and must be appropriately considered by the IRB. I would suggest explicitly recognizing this role, and perhaps developing accessible training for existing groups. Such training is particularly important with the increasing use of machine learning and generative AI in research. While we increasingly see studies that involve these techniques, there has been little effort, and no resources provided, to ensure that the groups relied on to protect participant rights and welfare know enough to make informed decisions in these areas.

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

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

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

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

52. Eileen Crimmins

Submit date: 3/15/2026

I am responding to this RFI: On behalf of myself

Name: Eileen Crimmins

Name of Organization: Davis School of Gerontology, University of Southern California

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Recent changes to the IT security standard as laid out in NIST 800.171 have made it almost impossible to research on populations with genetic and genomic data held in government sponsored repositories. I have had agreements to use NIAGADS data for a number of years and now am unable to renew these agreements. We have been trying to find a computing environment which allows us the power to do what we need to do as well as the security now demanded by my institution in order to sign-off on the NIAGADS agreement. This has stopped out research for 6 months.

My University office of compliance days NIST contains "a set of 110 security measures and nearly three times as many activities that must be undertaken to meet them. " They indicate that the problem is that within the normally used high capacity computing system there is no System Security Plan (or SSP) that documents how the controls are being met, and for the controls that are not being met, an accompanying Plan of Action and Milestones (or POAM).

The reading of the NIST regulations by my University has stopped research.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
NIAGADS has been useful and helpful.

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

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

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

53. Marie Banich

Submit date: 3/15/2026

I am responding to this RFI: On behalf of myself

Name: Marie Banich

Name of Organization: University of Colorado Boulder

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ABCDColoradoResponse-to-NIH-RFI_March_27.pdf

54. Sage Bionetworks

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Christine Suver

Name of Organization: Sage Bionetworks

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Executive Summary:

Sage Bionetworks supports NIH's effort to harmonize and modernize expectations for controlled-access human data sharing. We strongly endorse the goals of protecting research participants, clarifying consent expectations, and strengthening repository security. To ensure these policies succeed in practice, NIH should expand controlled-access requirements with clear scope boundaries, scalable governance operations, and actionable implementation guidance. The NIH policies should consider the impact on international research collaborations and the emerging privacy risks associated with evolving technologies and analysis workflows, including the integration of artificial intelligence (AI) and multimodal data.

Key Recommendations:

- Scale governance capacity, not just storage. Invest in data governance staffing for Data Access Committees (DAC)/ Access and Compliance teams (ACT), standardize workflows, and promote automation for identity verification, access submissions, renewals, and auditing to prevent review bottlenecks.

- Clarify the scope to prevent over-classification of low-risk data or derivatives. Explicitly distinguish individual-level human data from aggregate, summary, and sufficiently de-identified derivatives. Publish typically open data (e.g., Summary statistics, aggregated phenotypes).

- Adopt a risk-tier framework rather than a binary open/controlled model. Provide guidance and mapping of common data types to access tier levels, taking into account the processing level, consent, identifiability risks, and technical safeguards

- Protect participant intent. Provide model consent language that standardizes pathways to open sharing through explicit opt-in consent. Include guidance on identifying and mitigating community/group risks, including IRB consultation.

- Enable certified non-NIH controlled-access repositories and services. Create a clear equivalency pathway using recognized security attestations (e.g., NIST-800-53, FedRAMP, ISO 27001) and operational standards.

- Increase operational transparency. Require annual reporting of key access metrics across designated repositories (requests, approvals/denials, median review times, categorized denial rationales (categorized at a sufficient level of generality to protect confidentiality).

- Modernize the imputation server policy. Permit vetted non-federal servers that meet defined security, audit, retention, and certification requirements. Require clear retention timeline and automated deletion.

- Facilitate compliance through automation. Provide machine-readable guidance that enables automated compliance checks to facilitate consistent policy implementation.

Sage Bionetworks' Perspective:

Sage Bionetworks (Sage) is a nonprofit biomedical research organization that enables open science and responsible use of protected data. Sage designs and operates data governance, harmonization, compliant access processes, and secure analysis workflows that support various research communities, including several AMP programs. Our experience includes translating policies into practice through tiered-access models, DAC workflows, secure analysis environments, and more.

Sage supports the NIH's move toward harmonization across the Controlled-Access Data Policy, the DMS Policy, and the GDS Policy. Our primary emphasis is on ensuring NIH policies are implementable at scale, avoid unintentionally restricting low-risk scientific outputs, and provide clear, consistent rules that support both participant autonomy and impactful research collaborations, as well as broad scientific utility.

Conclusion:

Sage Bionetworks supports NIH's efforts to strengthen protections for human participant data while harmonizing policy expectations. To ensure successful implementation, NIH should:

- Clarify scope boundaries,

- Provide practical, standardized guidance,

- Invest in scalable governance operations, and

- Enable certified, interoperable repository and analysis infrastructures.

Policies that unnecessarily suppress scientifically salient variables or create access bottlenecks reduce the return on NIH's investment in data generation. With these refinements, the proposed policies can strengthen participant protections while preserving the openness and scientific impact that NIH-funded research is intended to achieve.

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

As a provider of controlled-access infrastructure (e.g., Synapse for several AMP programs), Sage Bionetworks recognizes that limited NIH CADR capacity and access-review staffing will be significant bottlenecks to research. We recommend that NIH invest in direct funding mechanisms for certified non-NIH repositories to diversify data storage options and administer responsible data sharing. To reduce

the burden on researchers, NIH should also support a true federation of identity services (such as NIH RAS) to streamline researcher authentication across different platforms.

Recommendations:

1. Fund governance operations as core infrastructure

NIH should support:

- Dedicated DAC/ACT staffing capacity

- Workflow modernization and automation for review and renewals of data access requests

- Standardized documentation templates for data use limitations

- Auditable authentication/identity verification and verification of mandatory ethics training

Without investment in operational capacity, review timelines will become a major barrier to research reuse.

2. Enable certified non-NIH controlled-access repositories

To diversify capacity and reduce bottlenecks, NIH should create a formal equivalency pathway allowing non-NIH Controlled-Access repositories (non-NIH CADRs) to meet policy requirements, provided they demonstrate:

- Prospective review of requests to access controlled data

- Identity authentication. Federated identity systems (e.g., NIH Researcher Auth Service) should be broadly supported to streamline authentication across repositories and reduce duplicative identity checks.

- Enforceable data use agreements.

- Audit logging and monitoring

- Security controls aligned with NIH standards: NIH should allow third-party certifications (e.g., SOC 2, ISO 27001, NIST 800-53, FedRAMP Moderate) as evidence toward security compliance, paired with required operational controls specific to human data governance (e.g., restrictions on countries of concern), including institutional vetting, and cloud-region controls.

NIH should clarify how federated data access approaches, where the data remains sequestered, and analysis models go to the data, would show compliance with these policies.

3. Increase transparency of repository performance

NIH should require annual reporting of key operational metrics from designated Controlled-Access Data Repositories (CADRs), including:

- Number of access requests received

- Approval and denial counts

- Median time-to-decision

- Denial rationales (categorized at a sufficient level of generality to protect the confidentiality of individual applications)

This reporting requirement serves two purposes: it creates accountability for repositories to operate efficiently and consistently, and it provides researchers and policymakers with the data needed to identify systemic barriers to legitimate data access, compare performance across repositories, and refine policy implementation over time. NIH should establish standardized reporting templates to enable cross-repository comparability.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: Sage Bionetworks generally supports the NIH's designated list of protected data types but recommends refinements to the definitions and a shift toward a more nuanced classification system to avoid unnecessary barriers to open science. NIH should also clarify how the Controlled-Access Data Policy interfaces with the DMS Policy for mixed human/non-human or multi-omic studies.

1. Clarify scope: individual-level vs aggregate/summary outputs

NIH should explicitly clarify that the Controlled-Access Data Policy applies primarily to:

- Individual-level human data

- Individual-level derivatives that retain material re-identification risk

It should not automatically apply to:

- Aggregate or summary outputs: GWAS summary statistics (e.g., allele frequencies, effect sizes, p-values), aggregated phenotypes across cohorts

- Meta-analysis results

- De-identified summary of clinical trial outcomes

- Quality Control metrics

- Population-level statistics

- Synthetic datasets that are not reasonably re-identifiable

- Pathogenic or somatic genomic data

NIH should publish a "typically open" guidance document listing such examples to prevent unnecessary barriers to sharing valuable scientific data.

2. Replace binary access logic with a risk-tier framework

Rather than the proposed binary approach treating data as either "controlled" or "open," NIH should issue guidance aligned with a tiered model that considers:

- Data type and modality

- Processing level and dimensionality

- Identifiability context (e.g., rare cohorts)

- Consent provisions

- De-identification methods

- Privacy-enhancing safeguards employed

- Emerging risks due to technologies and data linkage

Sage recommends that NIH adopt the following five-tier risk stratification matrix:

- Tier 0 (Open): Anonymous users, any use. Includes summary results and aggregate data, project descriptions, and minimal non-identifiable attributes.

- Tier 1 (Registered Use): Registered users (e.g., individuals with known email addresses who agree to data use conditions) are bound by Terms of Service. Includes individual metadata, individual-level summarized data, and aligned/processed data.

- Tier 2 (Limited Use): Registered users with specific use limitations (e.g., non-profit only)

- Tier 3 (Controlled Use): Approved, registered users (e.g., authenticated individuals who have completed ethics training, agree to specific terms, and demonstrate institutional support); includes individual-level raw or minimally processed data.

- Tier 4 (Enclave Use): Secure data enclave only; no data download. For sensitive data with high re-identification risk.

By using this matrix, NIH can provide clearer pathways for "Tier 0" open sharing of low-risk data while reserving "Tier 3" (Controlled Use) or "Tier 4" (Enclave Use) for the most sensitive individual-level data.

An NIH-provided mapping of common data products to these access tiers, along with published decision trees to classify data appropriately, would reduce inconsistent interpretations across institutions.

NIH should also clarify expectations for derived data such as embeddings, feature matrices, and trained model weights. Guidance should distinguish between derived data that retain individual-level signal and those that function as aggregate summaries.

3. Define boundaries to prevent scope creep

Several protected data types require clearer operational definitions:

- Imaging Data: We suggest clarifying imaging boundaries, such as explicitly excluding de-faced MRI scans from mandatory controlled access.

- Omics and Analyte Thresholds: NIH should consider establishing specific thresholds for analyte counts or data resolution that trigger controlled-access requirements.

- Low-Risk FAQ: We strongly recommend that NIH publish an FAQ or guidance document listing content typically suitable for open sharing, such as genomic summary results, aggregated phenotypes, and de-identified summary clinical trial results.

The policy should more clearly define the conditions under which the designated protected data types can be shared openly

- Participant Consent: We encourage NIH to propose standardized consent language that explicitly permits open sharing, allowing participants to provide clear, informed, and "opt-in" consent for unrestricted release, provided there is no documented risk to third parties or identifiable groups (e.g., family members). Consent guidance should also include the use of participant data for AI/ML uses.

- Privacy-Enhancing Technologies (PETs): Data that has undergone robust de-identification or has been transformed using PETs should have a documented path toward less restrictive sharing.

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

4. Identifiers vs pseudonymous linkage tools

In its risk assessment policy, we recommend that the NIH distinguish between direct identifiers and pseudonymous tools, such as Global Unique Identifiers (GUIDs). GUIDs are created specifically to enable data linkage without revealing someone's identity. GUIDs are critical for multi-site longitudinal research and data integration, and should not be treated as "protected data" that trigger controlled-access requirements unless there is a demonstrable risk of readily re-identification.

5. Operationalize "very low risk" determinations

Where the draft permits open sharing if risk is "very low," NIH should provide standardized guidance, including:

- Designated decision authority (e.g., IRB/HRPP/data governance body)

- Required documentation and justification

- Downstream communication of determinations

- Periodic reassessment

- Appeals pathway

Without standardized guidance, institutions will interpret "very low risk" inconsistently, leading either to over-restriction or inadvertent under-protection. Guidelines for standard communication procedures (e.g., an updated Institutional Certification process) will ensure the streamlined, accurate, and end-to-end application of controls, including open access.

NIH should ensure that Institutes and Centers do not impose divergent interpretations of protected data categories that undermine the harmonization goals of this policy.

6. Ensure scientifically salient variables are shared

NIH should formally stipulate that scientifically important variables should not be omitted or that datasets should not be artificially modified solely to avoid triggering controlled-access classification when appropriate safeguards are available. When a variable poses an elevated re-identification risk, the

default approach should be to share it under appropriate access controls rather than suppress it entirely.

For example, exact ages over 89 are frequently omitted under HIPAA Safe Harbor standards. However, in aging and longevity research, exact age is a scientifically critical variable. Replacing exact ages with broad categories (e.g., “90+”) can materially undermine research validity. In such cases, NIH should require that exact ages be shared under controlled access. NIH’s allowance of HIPAA Expert Determination provides an appropriate mechanism for sharing such variables under controlled access rather than suppressing them.

More broadly, NIH should state that:

- Scientifically salient variables necessary for reproducibility and secondary analysis must be shared under either open or controlled access.

- Suppression of key variables should be permitted only when risk cannot be mitigated through controlled-access safeguards.

- Policies should discourage over-application of Safe Harbor de-identification when Expert Determination or controlled-access protections provide a viable alternative.

This approach preserves participant protection while ensuring NIH-funded research remains scientifically meaningful and reusable.

7. Protect participant intent with procedural safeguards

When participants have provided explicit opt-in consent for unrestricted sharing, NIH should permit open sharing unless there is a documented, concrete risk to third parties or identifiable groups.

As with any ethics determination, the decision to override a participant's intent should not rest solely with the primary researcher. Any override of explicit consent should require independent review and time-limited determination by an IRB or ethics board, a transparent risk-assessment framework, and an appeals mechanism.

In addition to risks to individual participants, certain datasets may pose potential harms to identifiable communities or populations represented within the data. NIH should encourage researchers and oversight bodies to consider such group-level risks when evaluating data-sharing decisions.

This approach preserves participant autonomy while ensuring careful ethical oversight.

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

Sage supports the use of imputation servers beyond a single NIH-operated environment, provided they adhere to stringent security and privacy standards. NIH should encourage secure enclave/TRE-based implementations as a baseline and clarify which PET approaches are sufficiently mature for production use in imputation workflows, to avoid inconsistent or impractical expectations across institutions.

Our recommendations focus on balancing technical flexibility with robust data protection:

Recommendations

1. Permit vetted non-federal imputation servers

NIH should allow non-NIH-operated servers that:

- Operate within secure enclave or Trusted Research Environment architectures

- Prohibit uncontrolled data egress, make “bringing code to the data” the default for Tier 4 and recommend it for Tier 3

- Log and monitor access

- Undergo periodic independent audits and publish conformance criteria for certified secure enclaves

- Align with NIH security standards for controlled-access repositories

2. Establish a certification registry and checklist

NIH should maintain a registry of approved imputation servers that have passed a standardized certification checklist addressing:

- Infrastructure security controls

- Access management

- Provenance and versioning

- Protections against imputation-specific attacks

- Logging and auditability

- Incident response procedures

3. Require strict retention and deletion controls

To preserve non-transferability expectations:

- Temporary uploads and outputs should be automatically deleted after a defined time window.

- Clear policies should prevent the reuse of uploaded controlled-access data beyond the approved purpose.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Sage-Bionetworks-Response-to-RFI-NOT-OD-26-023-Controlled-Access-Data.pdf

55. Luke Hyde

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name: Luke Hyde

Name of Organization: University of Michigan

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
We write as investigators involved in the Adolescent Brain and Cognitive Development (ABCD) Study, a large-scale longitudinal investigation of brain development that is widely used as an open-access scientific resource. The ABCD Study has enabled researchers worldwide to address critical questions aligned with the mission of the U.S. Department of Health and Human Services. Based on our experience managing and sharing large-scale human research data, we have several serious concerns regarding the Draft NIH Controlled-Access Data Policy.

Draft NIH Controlled-Access Data Policy

The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts involving national security risks, including defense-related research and sensitive government operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on “Preventing Access to Americans’ Bulk Sensitive Personal Data” (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule’s data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations
- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

IMPLEMENTATION CONSEQUENCES AND OPERATIONAL CHALLENGES

Even if the policy flaws identified above were addressed, the policy, as drafted, would still create substantial operational challenges. Some of these may be addressable through implementation guidance, resource allocation, and phased rollout, but they warrant NIH's careful consideration:

Implementation Challenges and Resource Implications

Repository Infrastructure: Given the vast expansion of data types requiring controlled access, current repository infrastructure is very much inadequate. Most institutional repositories and many specialized repositories do not have and cannot rapidly develop NIST 800-171-compliant environments. The cost to build and maintain such infrastructure is substantial - involving not just technology investments but ongoing compliance monitoring, regular security audits, continuous personnel training, and documented security plans. For many academic institutions and smaller research organizations, these costs are prohibitive. Compliance with NIST 800-171 requires 110 distinct security controls, formal system security plans, and regular assessment – this is not trivial infrastructure that can be implemented quickly or inexpensively. The controls place a heavy burden on users of the system, which substantially slows down research progress, while at the same time substantially driving up costs. The policy would divert substantial resources from research activities to compliance activities, with minimal corresponding benefit to participant protection for de-identified data.

Informed Consent Requirements: The policy states that protected data types “may only be shared without access controls if there is informed consent explicitly stating data are to be shared openly without controls.” While requiring explicit consent for true open access sharing is appropriate, the policy creates a significant practical problem by defining “protected data types” so broadly that this requirement would apply to nearly all NIH-funded human subjects research. Standard research consent language typically describes data sharing in terms of whether the data will be identifiable, or deidentified, whether it will be publicly available or restricted to approved researchers, and what protections will be in place – but rarely uses the specific framing of “with” or “without controls.” This requirement essentially makes open sharing of any protected data types impossible for most existing studies and difficult for future studies. Even studies collecting minimal-risk data would be prohibited from open sharing unless the consent explicitly addressed “sharing without controls.” Importantly, requiring technical regulatory terminology ("with" or "without controls") in consent documents may actually reduce participant understanding rather than enhance informed decision-making. Participants are better served by plain-language explanations of who will have access to their data, how it will be protected, and what risks are involved- not by legal classifications that have little meaning outside regulatory compliance contexts. Effective informed consent should promote participant trust through transparency and comprehensibility, not through regulatory jargon that obscures, rather than clarifies, data sharing practices.

Risk Assessment for All Other Data: The policy requires that “data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls.” This creates an impossible compliance burden when combined with the policy’s extremely broad definitions of protected data types. The policy provides no operational guidance on how these assessments should be

conducted and what documentation is required. The criteria are subjective – different institutions and reviewers will reach different conclusions about what constitutes "potential sensitivities" for example. Further, this requirement duplicates existing oversight mechanisms. IRBs already assess privacy and confidentiality risks to participants. Data Management plans already require researchers to describe data sharing approaches and safeguards. Adding another layer of individualized risk assessments provides no clear value but would require significant personnel resources.

Clarification of End-User Requirements: The policy does not specify the security level required for the researcher's computing environment after controlled-access data has been accessed, downloaded, or analyzed. Several new Data Use Certification (DUC) agreements, which are the primary mechanism outlining users security obligations, require broadly that data be stored securely, and access be limited to approved personnel, and in the case of ABCD Study data, as of release 6.0, now requires NIST SP 800-171 or an equivalent standard. Thus, repositories face well-defined, standards-based requirements, while user-side computing environments operate under comparatively vague obligations that can be interpreted differently across institutions ("equivalent").

Further, NIST SP 800-171 compliance at the institutional level is a major undertaking. It requires institutions to develop and maintain a System Security Plan (SSP), implement 110 security controls across 14 control families, conduct regular self-assessments, and produce a Plan of Action and Milestones (POA&M) for any gaps. These requirements have become familiar to institutions with active Department of Defense contracts through the Cybersecurity Maturity Model Certification (CMMC) program. But the vast majority of NIH-funded research institutions — including many large academic medical centers and R1 universities, let alone R2 or R3 institutions — do not yet have certified NIST 800-171 environments covering the computing systems used for biomedical research analysis. Building, documenting, and maintaining such environments requires institutional investment measured not in months but in years, and in costs that can easily reach hundreds of thousands to millions of dollars per institution.

The policy would implement unfunded and unanticipated compliance burdens: A consequence that the proposed policy does not appear to account for is the effect on investigators who are currently funded under active NIH awards that include access to controlled-access data. These awards were reviewed, scored, and funded under existing policy expectations. The budgets were constructed accordingly. No provision was made for the cost of achieving NIST SP 800-171 compliance at the institutional level, because no such requirement existed or was signaled at the time of submission. If the new policy — or its implementation through updated DUC agreements or repository access terms — effectively requires NIST-compliant computing environments as a condition of data access, then investigators with multi-year awards already in progress will face one of three challenging situations:

- They must divert existing award funds to cover institutional compliance infrastructure, reducing resources available for the scientific aims of the award.

- They must seek supplemental funding for costs that were not part of the original budget, adding administrative burden and uncertainty.

- They lose the ability to access the controlled-access data central to their approved specific aims, potentially rendering the award non-executable.

None of these outcomes serves the interests of NIH, the research community, or the participants whose data the policy is designed to protect. This is not a hypothetical concern: the ABCD end-user community has experienced significant disruption due to the introduction of these enhanced security requirements, with limited lead time, in the ABCD Study Data Use Agreement for release 6.0. A policy change of the magnitude proposed here requires proportionate planning time and resources. The scope of this concern is broad. An estimated tens of thousands of NIH-funded investigators currently hold or have applied for access to controlled-access repositories. The majority are affiliated with institutions that do not have pre-existing NIST 800-171 compliant research computing infrastructure. Achieving compliance across this community represents an enormous, largely unfunded collective burden that, if not explicitly addressed in this policy, will fall inequitably on smaller programs, institutions with fewer resources, and early-career investigators and trainees who do not have or cannot easily access the administrative infrastructure that larger research enterprises can mobilize.

Deleterious Effects on Scientific Progress

Because of the operational and compliance burdens described above, the draft policy will result in substantial harm to scientific progress diverting financial resources and time to compliance instead. Furthermore, it will undermine the very rationale behind publicly available data sets by precluding individuals at all career levels at a broad range of institutions and universities from gaining access to the data.

(1) Slowed secondary analyses: Increased barriers to data access will delay meta-analyses, replication studies, and hypothesis generation from existing datasets. Secondary data analysis is a critical driver of scientific discovery, allowing researchers to answer new questions from existing data without additional participant burden or research costs. Imposing controlled-access requirements on low-risk de-identified data will create months-long delays in data access approvals, substantially slowing the pace of discovery.

(2) Reduced reproducibility: If data that could be safely shared openly requires controlled access, fewer researchers will be able to verify published findings. The reproducibility crisis in science has led to widespread recognition that data transparency is essential for scientific integrity. Requiring controlled access for minimal-risk data will reduce the number of researchers who can access data to verify results, undermining confidence in research findings.

(3) Inequitable access to data: Researchers at under-resourced institutions and early-career investigators as well as trainees will face disproportionate barriers to accessing data, exacerbating existing disparities in research capacity. Institutions with limited legal, compliance, and IT infrastructure will struggle to meet security and administrative requirements, creating a two-tiered research ecosystem in which only extremely well-resourced institutions can effectively access and use shared data.

(4) Chilling effect on data sharing: Faced with overwhelming compliance requirements, investigators may choose to share less data or design studies with smaller sample sizes to avoid triggering controlled-access requirements. This perverse incentive undermines NIH's fundamental goal of maximizing the scientific value of research investments.

These effects will be particularly pronounced for de-identified, minimal-risk data, where the privacy-protection benefits do not justify the scientific costs. A risk-proportionate approach would preserve robust protections for truly sensitive data while maintaining accessibility for properly de-identified datasets that pose minimal privacy risk.

Recommendations

1. Adopt a risk-tiered security framework that explicitly distinguishes between identified, identifiable, and properly de-identified data, with security and access requirements proportionate to demonstrated re-identification risk rather than applying uniform controls across all human research data.

2. Replace national-security-oriented standards with research-appropriate security controls for de-identified and low-risk datasets, drawing on the Research Security Program under NSPM-33 and established academic best practices, rather than defaulting to NIST SP 800-171.

3. Incorporate data scale and context thresholds (similar to those in the DOJ Bulk Sensitive Data rule) so that small-scale, minimal-risk research datasets are not subject to controls intended for large population-level or biobank-scale data holdings.

4. Align controlled-access requirements with existing oversight mechanisms by allowing IRB determinations, Data Management and Sharing Plans, and repository governance processes to satisfy risk assessment and documentation requirements, avoiding redundant review layers.

5. Conduct a formal implementation and impact assessment prior to finalization, evaluating repository readiness, institutional compliance costs, effects on data sharing and reuse, and the net benefit to participant protection, particularly for HIPAA de-identified data.

6. Extend the User-Level Compliance Timeline Substantially. We very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, at the very least the NIH should establish a phased implementation timeline for any user-side NIST security requirements that is distinct from and longer than the repository-level timeline. Achieving NIST SP 800-171 compliance at the institutional level for biomedical research computing environments is a multi-year effort involving information security staff, legal review, procurement, system hardening, documentation, and self-assessment. A minimum of 24 to 36 months from the policy's effective date should be afforded for user-side compliance, with active NIH-funded projects grandfathered for the duration of their current award periods.

7. Provide NIH-Funded Resources and Implementation Support. Again, we very much hope that the NIH will revise this policy along the lines outlined above. However, should it not, NIH should recognize that imposing NIST compliance requirements without corresponding resources constitutes an unfunded mandate with the potential to widen disparities in research access. We recommend that NIH:

1. Issue NIH-developed guidance documents, System Security Plan (SSP) templates, and self-assessment tools tailored to the biomedical research context, reducing the cost and effort required for institutions to document compliance;

2. Explore mechanisms to allow supplemental funding for compliance infrastructure on currently active awards that include controlled-access data access, recognizing that investigators could not have anticipated these costs at the time of award;

3. Establish or support shared institutional compliance resources — analogous to shared research computing cores — that smaller institutions can draw on, reducing duplicative investment across the community.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The Policy Is Overly Broad in Scope

As currently written, the policy would apply to nearly all NIH-funded research involving human participants. The definitions of protected data types are extremely broad and fail to distinguish meaningfully between identifiable and properly de-identified data. For example, the definition of "Personal Health Data" includes virtually all health-related information, including basic physical measurements and health attributes (e.g., height, weight, vital signs), behavioral and psychological measures, medical histories, test results, and exercise habits. As written, this definition would encompass nearly all human subjects research funded by NIH, regardless of whether the data are identifiable or pose any meaningful risk of re-identification. By treating all health-related data equivalently, regardless of identifiability or re-identification risk, the policy conflates fundamentally different privacy risk profiles. Properly de-identified datasets with minimal re-identification risk are treated identically to datasets containing direct identifiers such as names, Social Security numbers, or addresses. In practice, the breadth of the definitions would impose controlled-access requirements on studies such as:

- Community health surveys collecting weight and blood pressure
- Exercise intervention studies tracking vital signs
- Nutritional studies recording dietary habits and basic measurements
- Educational interventions measuring health knowledge

This policy would dramatically expand the number of studies requiring controlled-access infrastructure, overwhelming existing repository capacity and imposing substantial compliance burdens on research that poses minimal privacy risk.

Security Requirements Are Not Aligned With Risk

The draft policy requires institutional repositories to implement security standards such as NIST SP 800-171 or equivalent controls. This requirement reflects a fundamental misalignment between the level of risk posed by most research datasets and the level of security required. NIST SP 800-171 was designed to protect Controlled Unclassified Information in nonfederal systems and is typically used in contexts involving national security risks, including defense-related research and sensitive government

operations. The standard includes 110 security requirements across 14 control families, including physical security, personnel screening, incident response capabilities, and extensive technical safeguards. These controls are appropriate for protecting information that could threaten national security if disclosed. However, applying these standards to properly de-identified health research data conflates national security risks with individual privacy risks. The threat model underlying NIST SP 800-171 — including protection against nation-state espionage — is not proportionate to the statistical re-identification risks associated with properly de-identified research datasets. Applying identical security standards to both identifiable and de-identified datasets fails to account for these differences in risk. Most critically, the policy does not differentiate security requirements based on the level of identifiability. Data that have been de-identified according to HIPAA standards — either through Safe Harbor removal of identifiers or Expert Determination — pose dramatically lower privacy risks than identifiable datasets. Yet the policy would impose identical security requirements on both. Scientific evidence indicates that properly de-identified datasets with appropriate safeguards have very low re-identification risk. A risk-proportionate framework would recognize these differences and align security requirements accordingly and would avoid opportunity costs (restricted research innovation) due to a mismatch between risk and security level.

Misapplication of a National Security Framework

The draft policy draws on the Department of Justice rule on “Preventing Access to Americans’ Bulk Sensitive Personal Data” (28 CFR Part 202) as the basis for defining protected data types. This represents a fundamental mismatch between the purpose of the DOJ rule and the needs of biomedical research. The DOJ rule is designed to prevent foreign adversaries from obtaining large-scale datasets that could be used for intelligence operations or national security threats. In contrast, NIH data-sharing policies are intended to protect research participants while maximizing scientific benefit. The NIH draft policy adopts the DOJ rule’s data type definitions but does not include bulk data thresholds. As a result, even small-scale research studies would be subject to security requirements designed for massive commercial or governmental datasets. This approach ignores the existing safeguards present in academic research, including:

- Institutional Review Board oversight
- Informed consent requirements
- Controlled-access sharing mechanisms
- Limited study populations
- De-identification procedures

The threat model underlying the DOJ rule does not align with the realities of academic research data sharing.

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

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

56. Heather Griffis

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name: Heather Griffis

Name of Organization: Children's Hospital of Philadelphia

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
see attached comments

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

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
see attached comments

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Arcus-Omics-Comments-on-Draft-NIH-Controlled-Access-Policy-and-GDS-Policy-Revisions-Submitted.pdf

Description: CHOP Omics Comments on Draft NIH Controlled Access Policy and GDS Policy Revisions - Submitted

57. Nicholas Breitnauer

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name: Nicholas Breitnauer

Name of Organization: University of Colorado

Type of Organization: Not Applicable

Role: Clinician

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
In my overall review of this draft data policy, the spirit appears to be good. I think a large motivator of this policy can be traced to historic injustices, including the most notable case of Henrietta Lacks.

With all new policy, even well intended, comes the risk of unintended consequences. I am a medical expert in the care for adults with Down syndrome and my patients and this community have started to see benefits of large data sets. Some of this data is in the form of genetic information. Our research colleagues are starting to crack into the mystery disease states associated with Down syndrome with the goal of improving quality and quantity of life for this deserving community. Notably, strides have been made in furthering science for the understanding and ultimate treatment of Alzheimer's disease.

I would ask the committee review the policy as it stacks against the goals of INCLUD Data Coordinating Center. From my non-legal read, there would be restrictions on the data sharing of this vital information source. This in turn would hinder and slow the progress we have seen improving the quality of life of adults with Down syndrome.

I wonder if restrictions on the back end like this are more prudent compared to more transparency and discussion at time of consent for research? Additionally, sharing data allows for collaboration and reanalyzing data and limiting data sharing could be costly and could impact reproducibility. Lastly, many high impact journals require open-access data policies, meaning the proposed policy changes would contradict and limit publication opportunities.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Overall, I think data sharing policy could be applied to data sets that are deemed insufficient to protect participants/extend from the original intention of the study. Participants with Down syndrome who consent to genomic datasets intended for their information to be used to improve the quality and quantity of life for those with and without Down syndrome. I am afraid this new policy would restrict the intention of INCLUDE.

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

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

58. University of Illinois at Urbana Champaign

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Heidi Imker

Name of Organization: University of Illinois at Urbana Champaign

Type of Organization: Academic Institution

Role: Other

Role – Other: Professor and Service Provider Leadership

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The University of Illinois Urbana-Champaign appreciates NIH's efforts to thoughtfully move its proposed Controlled-Access Data Policy forward. Clear expectations will help investigators themselves as well as institutional support (e.g., research administration, grants and contracts, core facility, IT, library, technology transfer, and general council staff) comply effectively and efficiently. Our responses to each of the RFI prompts are provided below.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
The current data sharing policy tends to focus more on genomic data, as shown by the data types in the definitions appendix. The reality is that many modern techniques or studies can (and do) include multiple data types (multiple sequence data types, images, LC/MS spectra, etc) that can interrelate and can vary in type from raw to intermediate to 'final' data (summarized and potentially reused for ML/AI applications). Additionally, many NIH-funded data types that are subject to sharing are not molecular in nature. Some explicitly fall into a very large bucket of "biometric identifiers" but others are less codified. Examples include:

* Neurophysiological recordings (e.g., EEG, MEG, intracranial recordings)

* Structural and functional neuroimaging (e.g., MRI, fMRI, DTI, PET)

* Digital health data from wearables and mobile devices (e.g., actigraphy, heart rate, geolocation, passive sensing streams)

* Audio recordings (e.g., speech samples, voice biomarkers)

* Video recordings (e.g., behavioral assessments, motor function tasks, facial expression analysis)

* Survey responses, patient-reported outcomes, and qualitative interview transcripts

* Educational, cognitive, and behavioral assessment data

* Social, environmental, and geospatial exposure data

It is not clear if NIH-supported CADRs support all the data types within the definitions appendix, let alone the broader list above. It is also not clear how community repositories–just a few examples being

Databrary (NYU), Physionet (MIT), or ICPSR (Michigan)–fit into this policy when they are not operated by NIH but have been supported via NIH funding. How is a PI to make sense of the list in the Required Security and Operational Standards for NIH Controlled-Access Data and Repositories being different from the controlled access options in the National Library of Medicine's Repository Finder tool (https://www.nnlm.gov/resources/finder)? Identifying the best repository becomes even more complex then when considering that while some data types may fall under controlled access restrictions, others from the same study may not need this constraint. Therefore, another challenge then is to ensure related data can be either explicitly linked or are co-located and accessible, basically in a manner that makes identification of relevant studies straightforward and data reuse reliable and consistent.

Reducing ambiguity throughout is essential, especially for multidisciplinary studies and hybrid datasets. For example, resting-state fMRI combined with cognitive task performance or EEG recordings may result in data suitable for multiple repositories, but repositories that seem most appropriate depending on the context of the grant project. In this scenario, data may have the best reuse potential within NIA Data Archive for aging studies or OpenfMRI (among many others!), but resources may be housed at different NIH Institutes than the funding announcement's issuing Institute or not within the NIH at all. This creates uncertainty for investigators and institutional support staff. There is some precedent for omic data that could be used as a start. For example, the GEO database has clear instructions that allow deposition of multiple data types, some which can be under controlled access restriction. However the database's implied focus is gene expression which essentially limits its use outside of that scope, even though other experimental types exist within it (ChIP-Seq, ATAC-Seq, etc). Why limit it in this manner? This model could be used in a more general manner to encourage deposition of and linking to other genomic feature data: microbiome/metagenome feature data, genotype data (non-human), epigenomic, genome assemblies, etc.

Great strides have also been made to make the deposition process simpler by many listed CADRs, but datasets are often deposited at the last minute as a direct requirement for publication or to observe GDS policies. To alleviate this, additional effort to make deposition more straightforward and accessible to researchers and core facilities is needed. We also encourage NIH to consider requiring earlier deposition when possible, such as done now for some NDA-bound data for NIMH grants, and to allow for the combined deposition of different data types as well as processed results. Researchers here at Illinois report staged 6 month deposition into NDA to be good pacing and further report that being able to talk with experts at NDA was extremely helpful--essential, in fact. Therefore, appropriate staffing at the CADR's is also critical to achieving NIH's data sharing goals.

To best promote efficient implementation of the proposed Controlled-Access Data Policy, we argue a key resource needed is also the most familiar: Funding Opportunity Announcements (FOAs). We recommend that NIH clearly specify expectations for specific repository use, deposition timelines, and submission requirements within individual FOAs. This approach ensures that NIH's data strategy is carried out as intended, because investigators will clearly understand which data the agency seeks to collect. Ultimately, this approach helps NIH grow its intended data collections strategically and demonstrates progress and accountability in meeting its research and data-sharing objectives. Importantly, FOAs should explicitly state that repository deposition costs are allowable and expected to be included in grant budgets, with direct links to repository-specific cost guidance (e.g., the NIH NDA Data Contribution cost structure). Transparent cost guidance will enable applicants to prepare realistic

budgets and avoid underestimating the personnel, curation, and submission expenses associated with controlled-access data deposition.

Providing standardized language across FOAs, along with precise links to repository cost estimation resources, will improve compliance, reduce administrative burden, and support timely, high-quality data submission.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
To ensure scalable, secure, and compliant data sharing, additional investments are needed in clear policy definitions, de-identification standards, and access protocols. Without these complementary resources, increases in storage capacity alone may not be sufficient to support a growth in controlled-access datasets. As listed above, there are many additional data types that warrant definitions and thresholds.

We support a policy framework in which controlled-access requirements are proportional to actual re-identification risk, but reiterate that clarity is critical at the FOA level. With NIH's new streamlined DMSPs that do not list specific data types, this is an opportunity to flip the direction and have individual FOAs explicitly identify the data types NIH expects to be generated and deposited under that award, along with the anticipated level of access (open vs. controlled). This will: 1) reduce uncertainty about expectations for submission, consent, and budgeting, 2) enable realistic planning for controlled-access deposition workflows and de-identification needs, and 3) align NIH expectations for expert determinations and informed consent scope—all things that are critical at the beginning of a grant project.

As mentioned above, to ensure appropriate categorization of data as controlled or open, FOAs should also include clear standards for de-identification, e.g., expecting use of established regulatory and technical frameworks such as the HIPAA Privacy Rule (Safe Harbor or Expert Determination), 45 CFR 46 (Common Rule) definitions, and other published privacy guidance (e.g., NIST) that has been deemed appropriate by NIH for that specific FOA.

Finally, FOAs should articulate whether summary-level data are expected to be open and outline relevant exceptions. Clear, FOA-level definitions and expectations—combined with budget language that explicitly allows and directs applicants to include deposition and data management costs—will promote compliance, reduce administrative burden, and ensure that controlled-access oversight is appropriately tailored to risk.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We offer no additional comments.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
We agree that strong safeguards for imputation server handling controlled-access data is important, and clear technical requirements will be critical for ensuring consistent protection. While critical, this also requires that NIH recognize that the costs of meeting these requirements may limit the number of organizations able to operate such servers in the future.

59. Non-NIH Members of the PRIMED Consortium Data Sharing Working Group

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Johanna Smith

Name of Organization: Non-NIH Members of the PRIMED Consortium Data Sharing Working Group

Type of Organization: Other

Type of Organization - Other: Consortium Working Group

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
The PRIMED Consortium Data Sharing Working Group suggests that the NIH may be able to fund and monitor access to repositories for data derivatives, such as AnVIL workspaces, under data sharing guidelines determined by data sources. This could make derivative dbGaP cohorts available within AnVIL for pre-cleaned data source options, decreasing noise or error in genomic analyses resulting from differences in data cleaning. PRIMED has developed multiple data derivative workspaces in AnVIL already that could be used for such a task.

We also suggest defining the term “data derivative” and including it in discussion. The PRIMED Data Sharing Working Group has developed a drafted definition for this term based on previous NIH documentation and current genomic field relevance. This paves a path for derived data release by secondary data users as well. Defining data derivatives can help clarify whether data should be controlled-access, if it includes genomic summary results or individual-level data, and help further refining other definitions that are reliant upon such designations for use in data sharing policies.

We recommend the data derivative definition as follows:

Data derivatives are data stemming from one or more source datasets that are further processed or cleaned into individual-level data or aggregate data. Data derivatives may be controlled access or open access. Controlled-access data derivatives inherit access requirements from the source datasets. Examples of data derivatives include imputed genotype datasets, harmonized phenotypes, annotated data, Genomic Summary Results (GSR), or any data further defined by NIH as derivative.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
We suggest revisiting whether social and structural determinants of health datatypes are sufficiently represented in the list of protected data types, as these data types are increasingly used in genetic and other types of biomedical research (e.g. religion, housing, transportation, education). Ensuring that these SDoH data types can be robustly and consistently ingested and released by controlled access repositories is critical for their downstream use.

We also suggest revisiting whether other types of 'omics data beyond those listed in the draft policy (e.g. metabolomic, microbiome data, etc.) should be included in the list of protected data types.

Based on the PRIMED Consortium's experience navigating release and sharing of summary-level genomic data, the Data Sharing Working Group recommends clarifying whether the listed data types (and associated definitions) refer only to individual-level (i.e. participant or "row level") data, or if and where summary or aggregate level information is also included in the expectation for controlled access. PRIMED's experience navigating the NIH Genomic Summary Results (GSR) policy is that it is often unclear what types of summary-level data are within scope of a given policy and why or why not. Downstream use of summary-level data is not without re-identification or other potential risks to participants and their associated communities; however, overly restrictive access policies for summary-level data that are not proportionate to these potential harms also pose undue burdens on data submitters and users. For Genomic Summary Results, we also suggest NIH policymakers revisit the risk-benefit analysis for multi-study GSR sharing and clarify in NIH policy when and why multi-study GSR require controlled-access sharing.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The Draft Revisions to the GDS Policy mentions a "large scale data" threshold of 100 or more individuals and specifies that protections differ between those designated as large scale data and not large scale data. The PRIMED Consortium Data Sharing Working Group suggests inclusion of the reasoning stated for the "large scale data" threshold choice and further clarification of the treatment of those data choices.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
The proposed policy states "NIH currently does not allow users to develop their own imputation panels or servers." If the proposed policy update is not adopted, we suggest clarifying that (based on our understanding) it is currently allowed for a user to develop an imputation panel for their own/local use (as distinct from then making the panel more widely available via an imputation server).

60. Lauren Spor

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name: Lauren Spor

Name of Organization: University of Pittsburgh

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
The proposed update incorrectly states that the "NIH currently does not allow users to develop their own imputation panels or servers." Specifically, the NIH has funded grants to create population-specific imputation servers for groups that are underrepresented in biomedical research. Using an imputation panel that does not represent the target population will result in poor imputation due to missing population-specific haplotypes., which is testament to why researchers should be allowed to continue creating their own population-specific panels. This should be clarified in this section, as imputation panels and imputation servers are distinct, and developing imputation panels without an imputation server are currently allowed.

61. N/A

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The policy is overly broad and does not distinguish between identifiable and deidentified data. Moreover, requiring this level of NIST compliance will slow if not stop universities and smaller colleges from being able to utilize data. There are ways to increase security but also allow access to data for researchers.

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

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

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

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

62. Matthew Galbraith

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name: Matthew Galbraith

Name of Organization:

Type of Organization: Not Applicable

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

• Overall Impact: There is significant concern that these proposed policies will impose heavy administrative and logistical burdens that hamper research progress and medical breakthroughs, further exacerbating existing barriers to the efficient utilization of federally funded research data.

• Administrative and security burdens: Mandating NIST SP 800-171 standards for all data download locations imposes significant logistical and financial burdens that may exacerbate resource disparities between institutions. These requirements may also limit an investigator's ability to perform low-risk analyses on local computer systems that are not part of a certified institutional framework.

• Controls on de-identified data: Applying extensive new security requirements even to data de-identified via HIPAA standards (Safe Harbor or Expert Determination) may be an overreach that complicates research without significantly increasing actual participant privacy.

• Barriers to utilization of taxpayer-funded Data: The additional administrative, logistical, and financial burdens imposed by the proposed new policies could act as barriers to accessing publicly funded data, potentially hindering broad scientific utility and return on taxpayer investments.

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

• Resourcing for access and certification: There is a critical concern regarding whether Data Access Committees (DACs) and institutional bodies like Human Research Protection Programs (HRPPs) will be provided with the necessary resources to handle the anticipated "increased demand" for reviews and certifications in a robust and timely manner.

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

• Inclusion of additional "Omics" data types: Restricting access to transcriptomics, proteomics, and epigenomics (and potentially metabolomics) based on perceived re-identification risk or "systems-level" sensitivity could create unnecessary barriers and slow the progress of broad molecular research.

• Uncertainty of analyte thresholds for ‘systems-level’ datasets or analyses: The lack of established thresholds for "systems-level" datasets and the risk of "data mosaics"—where combined datasets might meet a threshold that individual ones do not—create significant technical and administrative ambiguity. The current definitions should be modified to provide more certainty for investigators regarding which datasets require controlled access. Under the current draft proposal, there are no established numerical

thresholds (such as a specific count of proteins or transcripts) for defining a "systems-level analysis". Instead, the primary distinction appears to be whether the measurements are routine clinical measurements used for individualized patient care. If the data is collected for targeted clinical purposes, it is generally excluded from the "systems-level" definition for epigenomic, proteomic, and transcriptomic data.

• Ambiguity regarding processed data: It remains unclear whether specific processed outputs, such as variant calls or gene-level counts, will be treated as individual-level protected data or as open-access summary results.

• Restriction of row-level clinical trial data: Mandating controlled-access for row-level clinical trial data, even with research consent, limits accessibility. Because institutions must still determine that open sharing poses a "very low risk," common standards are needed to ensure consistency across the research enterprise.

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

• 6-month submission timeline: This requirement may not allow primary investigators enough time to clean, analyze, and publish their findings, potentially disincentivizing data generation and unfairly favoring secondary users. It remains unclear how this timeline interacts with the standard DMS Policy timeline, which typically ties sharing to publication or the end of the award.

• Conflicts with journal data Sharing requirements: Researchers may face difficulty publishing if journals require open data sharing, as the proposed policy mandates controlled-access for protected data types unless explicit "open-sharing" consent was obtained.

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

63. Association of Public and Land-grant Universities

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Kacy Redd

Name of Organization: Association of Public and Land-grant Universities

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Association of Public and Land-grant Universities (APLU) appreciates the opportunity to provide feedback on the Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy (NOT-OD-26-023). APLU shares NIH's commitment to safeguarding national security, protecting research participants, advancing science through responsible data sharing, maximizing the return on federal investments in research, and reducing researcher burden.

APLU is a membership organization that fosters a community of university leaders collectively working to advance the mission of public research universities. The association's U.S membership consists of more than 240 public research universities, land-grant institutions, state university systems, and affiliated organizations spanning across all 50 states, the District of Columbia, and six U.S. territories. The association and its members collectively focus on increasing student success and workforce readiness; promoting pathbreaking scientific research; and bolstering economic and community engagement. Annually, its U.S. member campuses enroll 4.5 million undergraduates and 1.4 million graduate students, award 1.3 million degrees, employ 1.3 million faculty and staff, and conduct $70 billion in university-based research.

APLU has four top-line recommendations for NIH to consider for the Controlled Access Data (CAD) policy:

1. Retain the exemption for federally funded research from the DOJ rule (28 CFR Part 202) to preserve critical health research.

2. Utilize existing mechanisms, including Institutional Review Board (IRB) review under 45 CFR 46 and Data Management and Sharing Plans, to provide risk-calibrated protections for research participants.

3. Build on the Health Insurance Portability and Accountability Act (HIPAA) data protection standards to address national security concerns. HIPAA-equivalent standards combined with controlled-access mechanisms that prohibit access by Countries of Concern may be more appropriate for biomedical research than NIST SP 800-171 controls used for Controlled Unclassified Information (CUI)

4. Establish a community working group as recommended by the explanatory statement in the Consolidated Appropriations Act, 2023 (H.R.2617) to address overlapping and intersecting NIH data policies to alleviate confusion and eliminate administrative burden.

APLU has responded to NIH's questions 1 through 4 below.

Question 1: Feedback on any aspect of the Draft NIH Controlled-Access Data Policy.

Retain the DOJ Exemption for Federally Funded Research

NIH's proposed CAD policy (NOT-OD-26-023) bases the 11 data types that must be protected for national security concerns on the Department of Justice final rule (28 CFR Part 202). The DOJ rule was developed through the rulemaking process with public comment and with input from NIH, FDA, CDC, NSF, and dozens of other agencies. DOJ's final regulation exempts federally funded research, which is explained in the accompanying DOJ Factsheet on the rule:

"In addition, the rule contains exemptions meant to preserve critical health research, including exemptions for federally funded research, for sharing data pursuant to international agreements (including certain pandemic-related and global-health-surveillance agreements), for submissions of regulatory approval data for medical drugs, devices, and biological products, and for certain clinical-investigation data and post-marketing surveillance data." (DOJ Fact Sheet, 2024)

NIH's proposed policy adopts 28 CFR Part 202's data type definitions but does not adopt its exemption for federally funded research.

APLU Recommendations

- APLU urges NIH to adopt the 28 CFR Part 202 exemption for federally funded research.
- APLU encourages NIH to coordinate with DOJ, NSF, DOE, DOD, other federal agencies, and the regulated research community to establish consistent controlled-access requirements for human research data. Consistent federal standards enable institutions to build compliance infrastructure that serves all funders efficiently and reduces researcher confusion and likelihood of errors.

IRBs and Data Management and Sharing Plans are Risk-Calibrated

Existing oversight mechanisms provide risk-calibrated protections for human subjects data. For each research study, Institutional Review Boards (IRBs) are required under 45 CFR 46 to determine adequate protections for research participants and for sharing their data. IRBs evaluate individual re-identification risk, risks to populations including potential for group harm or stigmatization, and whether obtained consent permits the proposed data sharing. IRBs are best positioned to help researchers and institutions make a risk-calibrated response for data sharing.

Additionally, the required NIH Data Management and Sharing Plans (DMSPs) enable a shared understanding between the researcher, institution, and NIH about how data will be protected, managed, and shared. The DMSP helps institutions develop budgets for these management and sharing costs and provides NIH and its peer reviewers with project-specific information to assess whether proposed plans are appropriate. The elements in a DMSP have gone through robust community conversations and tool development (i.e. DMPTool), and there has been some success in harmonizing DMSP sharing requirements across agencies. Harmonization helps reduce researcher burden. Research security is enhanced if researchers and institutions articulate project-specific data protections.

APLU Recommendations

- Continue to rely on the DMSP and IRB review as the most appropriate mechanisms for addressing project-specific protections in a risk-calibrated manner.

- APLU requests that NIH work with the research community to provide additional training for IRB committees and DMSP guidance to address project-specific national security risks.

Security Standards for Fundamental Biomedical Research

By requiring NIST SP 800-171 or equivalent standards, this policy would subject fundamental research to the same security requirements as Controlled Unclassified Information (CUI). NIST SP 800-171 requires 110 security controls and was designed for defense contractors handling CUI. APLU shares NIHs concern about protecting sensitive data from national security risks and ensuring participant privacy. However, there has not been a thorough examination of whether NIST SP 800-171 is risk-proportionate for fundamental biomedical research or what impact applying these controls at scale would have on the research enterprise.

APLU Recommendation

- Before finalizing this policy, engage the research community in assessing whether NIST SP 800-171 is necessary and appropriate for fundamental biomedical research, or whether HIPAA-equivalent standards potentially combined with specific controlled-access and Countries of Concern restrictions, would adequately address national security concerns.

Establish a Working Group to Address the Varied Goals of this Policy

APLU is concerned that the CAD policy attempts to address too many issues simultaneously, which include controlled-access requirements across data types, revising the Genomic Data Sharing Policy (NOT-OD-14-124), aligning with DOJ national security rules (28 CFR Part 202), requiring new consent standards for an expanded set of data types, and establishing NIST SP 800-171 security standards for fundamental biomedical research. The interactions between this policy and other recent NIH actions, including the proposed changes to the DMSP format (NOT-OD-26-046) and NIH Controlled-Access Data Repository (CADR) standards (NOT-OD-25-159), create unintended consequences that the RFI process cannot adequately address.

As referenced in the policy, the 2023 Consolidated Appropriations Act directed NIH to develop a framework for managing national security risks in biomedical research. That Act does not require NIST SP 800-171, the 11 data types and volume thresholds, or extend research security requirements to institutions below the $50 million threshold established in the National Security Presidential Memorandum – 33 (NSPM-33) and reinforced by the CHIPS and Science Act, all of which this policy (NOT-OD-26-023) would require.

APLU Recommendations

- Before finalizing this CAD policy, APLU urges NIH to convene a research community working group to assess how these interlocking policies and regulations interact, identify and mitigate unintended consequences, and develop implementation guidance and training. The explanatory statement accompanying the Consolidated Appropriations Act urged NIH to "convene a working group to develop and disseminate best practices on genomic data sharing for use by entities engaged in biomedical research and international collaboration." That approach remains appropriate.

Clarify Scope of NIST SP 800-171 Security to Controlled-Access Data Repositories Only

The policy states data must be protected throughout the data lifecycle but does not clarify whether NIST SP 800-171 equivalent standards apply to only repositories sharing data externally for public access, sharing between research teams on a campus, or sharing between campuses during data collection and analysis.

NIH has already recognized distinctions between NIH CADRs and research environments in a previous policy. NOT-OD-25-159 states that "repositories that only facilitate direct sharing between investigator teams, cloud workspaces that only temporarily store data, data coordinating centers, and similar activities" are not NIH CADRs and are not subject to repository requirements.

Limiting NIST SP 800-171 to CADRs concentrates security resources where data is most accessible to external parties, which is where national security risk is highest.

APLU Recommendation

- Confirm that NIST SP 800-171 equivalent requirements apply only to CADRs and not all research environments, specifically adopting the language in NOT-OD-25-159, which states that "repositories that only facilitate direct sharing between investigator teams, cloud workspaces that only temporarily store data, data coordinating centers, and similar activities" are not CADRs and are not subject to repository requirements.

Assess Impact on Emerging Research Institutions

The proposed requirements would disproportionately burden emerging research institutions that are not already subject to NSPM-33 and CHIPS and Sciences Act research security requirements. Preliminary estimates from research institutions preparing to meet existing requirements illustrate the scale of this burden. One R1 institution estimates $2 - $5 million initial investment and $500,000 - $1 million annually to bring a single repository into NIST SP 800-171 compliance. If requirements extend to the full data lifecycle, including collection and analysis environments, initial compliance costs could exceed $5 - $10 million, with ongoing costs requiring 2-5 additional full time equivalent (FTE) staff and 10-20% increases in project startup time. Another R1 university estimates that ongoing maintenance for NIST SP 800-171 controls would be $30,000 per user per year, meaning a single ten-person research lab would face annual costs of approximately $300,000. These costs would add additional cost burdens to project budgets.

APLU Recommendation

- Before finalizing the policy, publish an estimate of how many institutions below the $50 million NSPM-33 threshold conduct NIH-funded research with the controlled-access data types and might be impacted.

Apply the Policy Only to New Awards

The policy creates new costs for active grants that may not have been budgeted. These include costs to download and use controlled-access data, which will require computing environments that meet NIST SP 800-171. Many standard institutional secure servers do not meet this standard, which may require researchers to purchase access to compliant secure research computing enclaves that were not originally budgeted. It also may include costs to prepare the data for stricter standards for deposition in

a CADR for an expanded list of controlled-access data. Grants in their final year face the most acute challenge, as there may be insufficient time to request supplements or renegotiate scope.

APLU Recommendations

- Apply the policy prospectively to new awards only to allow compliance costs to be built into grant budgets.
- Alternatively, ensure administrative supplements remain available to cover newly required compliance costs for grants awarded before the policy effective date.
- Allow the cost of storing data in a NIST SP 800-171 compliant environment and preparing the data for deposition into a CADR to be treated as an allowable grant expense.
- Consider providing a federally managed cloud environment that meets NIST SP 800-171 standards, which institutions could use for controlled-access data storage and analysis.

Continue Support for Institutional Stewardship and Community Commitments

Institutions conducting NIH-funded research have made commitments to participants and communities to protect their data. These commitments reflect years of work to build trust, particularly with communities that have historical reasons to be cautious about research participation. Institutions address participant and community concerns through the following mechanisms:

- IRB authority to require additional protections. IRBs may determine that specific data requires protections beyond federal minimums as a condition of approving data collection and sharing.
- Community governance arrangements. Some institutions have established Community Advisory Boards or similar review processes that reflect community concerns about research uses. Institutions have committed to honor these governance requirements as a condition of community participation in research. Data sharing from these communities will decline if institutions cannot fulfill these commitments.
- Data Use Limitations (DUL). Institutions specify restrictions on secondary data use based on consent, IRB determinations, and community requirements. For example, an IRB may limit data to specific disease research based on community concerns, or consent may prohibit commercial use. These DULs are one guardrail for ensuring that approved users respect the conditions under which participants agreed to share their data.
- Certificates of Confidentiality (CoC). The 2014 GDS Policy (NIH NOT-OD-14-124) acknowledged that genomic data can be re-identified even when de-identified under HIPAA and Common Rule standards. NIH obtained a Certificate of Confidentiality for the database of Genotypes and Phenotypes (dbGaP) and encouraged investigators and institutions to seek CoCs as safeguards against compelled disclosure. This is another guardrail that helps ensure that participant data is used only in ways participants understood and accepted.

APLU Recommendations

- Confirm that Data Use Limitations, Certificates of Confidentiality, community governance agreements, and IRB designated limitations will be maintained and applied as data is deposited to NIH

CADRs. If any of these stewardship mechanisms will change under the proposed policy, APLU requests that NIH identify which mechanisms are affected and explain the rationale for the change.

- Clarify any exceptions that may permit disclosure of data held in NIH CADRs to other federal agencies so that the community can update consent documents.

- Confirm that if an IRB determines that depositing data in a NIH CADR would not adequately protect research participants, the institution may decline to share consistent with existing IRB authority under 45 CFR 46.

- Ensure that data deposited in NIH CADRs remains immutable and that any modifications to datasets require confirmation from the depositing researcher and, where applicable, the IRB of record.

Address Unintended Impacts on Science

The proposed policy may create barriers to multi-institutional collaboration and stymie replication studies and independent verification of research results.

Multi-institutional collaborations face new administrative barriers. Before transferring protected data to a collaborator, institutions may need to verify the recipient has implemented NIST SP 800-171 controls. Researchers may decide not to conduct critical studies like the Adolescent Brain Cognitive Development (ABCD) study, limit collaborations to partners with existing NIST SP 800-171 infrastructure, or avoid multi-institutional projects involving controlled-access data types.

Research integrity depends on access to underlying data. Peer review, replication studies, and independent verification require reviewers and other researchers to access the data. Controlled-access processes may delay these reviews that help ensure science is accurate and trustworthy.

APLU Recommendation

- Engage a working group as recommended in the explanatory statement in the Consolidated Appropriations Act, 2023 (H.R.2617) to address the potential impacts on collaborations, replication studies, and research integrity.

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

Capacity constraints are a concern. If CADRs do not have the capacity to absorb data, researchers cannot deposit data, cannot comply with sharing requirements, and cannot publish findings that depend on that data. Additionally, institutions may currently lack the capacity to certify that data can be deposited, in a timely manner. Both of these constraints could interact with the proposed changes to the DMSP template (NOT-OD-26-046) that require sharing data underlying preprints and conference presentations. If underlying data involves controlled-access types, researchers cannot post a preprint until the data is deposited in a CADR, which will take time. Researchers may prioritize sharing findings via journal publications rather than sharing more rapidly via pre-prints and conference proposals.

The potential need to move legacy data held in legacy repositories into a CADR could also put strain on the system. The policy states that “Human data or data derived from human cell lines or biospecimens already shared prior to the effective date of this Policy should be assessed for risk but are not required

to be controlled to comport with this Policy." However, the policy is unclear on whether the legacy repositories holding that legacy data must meet NIST SP 800-171 equivalent standards. If legacy repositories must upgrade to continue sharing, some repositories may be unable to comply and be forced to take data offline, making federally funded research inaccessible. If the legacy data must be migrated to a CADR because the repository holding it is not able to achieve NIST SP 800-171 equivalent controls, this may be a significant amount of data to migrate, further straining CADR capacity.

APLU Recommendations

- Clarify whether NIH CADRs have sufficient repository capacity to accept the amount of data expected. If capacity is insufficient, what is the gap and timeline to address it?

- Assess compliance costs, including the number of active grants that would face new obligations under this policy and whether current supplement funding is adequate to meet anticipated demand.

- Clarify that legacy repositories with legacy data can continue to share legacy data but not accept new controlled-access data after the effective date of the policy.

- If legacy repositories must upgrade to continue sharing, assess the cost to migrate a typical dataset from a repository to a CADR, including possible reconsenting, data preparation, metadata harmonization, de-identification review, and coordinating transfer, and clarify what funding mechanism NIH anticipates will cover these costs.

- To reduce the burden on CADRs, provide funding for infrastructure upgrades for institutional repositories to implement NIST SP 800-171 standards.

- Consider allowing a three-year implementation period before the CAD Policy takes effect to give institutions time to upgrade institutional repositories to be CADRs or migrate data to CADRs.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The proposed 11 protected data types are based on 28 CFR Part 202, which was developed to address national security risks from data transactions and not federally funded fundamental research. Several data types (e.g., precise geolocation data, personal financial data) are uncommon in NIH-funded research, while the inclusion of all genomic, epigenomic, proteomic, and transcriptomic data may be overly broad for research contexts where IRBs have determined risk is low.

APLU Recommendation

- Follow the recommendation in the explanatory statement in the Consolidated Appropriations Act, 2023 (H.R.2617) to form a working group. The working group could assess whether the proposed 11 data type are appropriately designated for biomedical research and whether the proposed security controls are risk proportionate.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
There are several proposed changes to the Genomic Data Sharing Policy. One is around consent standards. The policy states that human genomic data collected from biospecimens or cell lines created

or collected after 2015 must have consent for use and sharing. If consent has not been obtained that align with this policy, the data cannot be shared. For data collected under protocols where participants are deceased, lost to follow-up, or where protocols prohibit re-contact, it may not be feasible to reconsent.

The 100-individual threshold for large-scale genomic data is not risk-calibrated and may create confusion. Data below 100 individuals is still subject to the CAD Policy requirements. The threshold only determines whether the full Genomic Data Sharing Policy also applies. This means even small pilot studies face controlled-access requirements, which may inhibit small studies, pilot studies, and research involving special populations that have difficulty finding funding, such as orphan disease research.

Additionally, the proposed revision to the Genomic Data Sharing policy in the CAD policy proposes data submission within six months of data cleaning and quality control. This assumes data collection is a discrete event. Longitudinal studies, clinical trials, and multi-site cohorts collect data continuously over months or years. Often, researchers cannot meaningfully clean data until a scientifically coherent dataset is assembled, making the six-month trigger ambiguous for a significant portion of NIH-funded research.

APLU Recommendations

- Clarify what options exist for continued genomic data sharing when reconsent is not feasible.
- Assess whether the 100-individual threshold for large-scale genomic data should allow for risk-calibrated exceptions or safe harbors for small studies, pilot studies, and research involving special populations.
- Clarify that the 100-individual threshold is study-specific and not institution-wide and cumulative across all research studies.
- Align data submission timelines with the NIH Data Management and Sharing Policy, which requires data sharing at time of publication or by the end of the award period.
- If NIH retains a fixed submission timeline, extend it to 12 months, or align submission with annual Research Performance Progress Reports (RPPR), to accommodate researcher practice, data cleaning, quality checks, and repository processing.
- Clarify that the clock begins when data collection for the relevant study aim or cohort is complete, not when individual data points are generated.

Thank you for your consideration of APLU's perspectives. Addressing our shared goals of protecting research participants, safeguarding national security, and maximizing the impact of the investments in science requires a community-wide approach. APLU welcomes the opportunity to continue discussions with NIH and the research community on the implementation of controlled-access data policies.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/APLU-RFI-NOT-OD-26-023.pdf

Description: APLU letter on NOT-OD-26-023.

64. N/A

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
NIST SP 800-171 was developed by the Department of Defense to protect Controlled Unclassified Information in the defense industrial base. Its 110 controls assume a threat model dominated by nation-state adversaries conducting persistent, targeted intrusions against high-value proprietary assets. Genomic research data does not fit this threat model. The data is not proprietary. It was generated with public funds, collected from volunteers who consented to its use, and deposited in repositories built for sharing. The empirical record confirms the mismatch. Through July 2018 (the most recent publicly available statistics), 38 policy violations were documented among 42,292 approved dbGaP data access requests: a rate of 0.09%. None involved nation-state actors or cyberattacks. The most common category was premature publication before an embargo date; the second was accidental disclosure due to administrative error. These violations are structurally incapable of producing re-identification harm. They involve timing and process failures, not exfiltration of individual-level data. We urge NIH to publish updated violation data so that policy decisions are grounded in current evidence. Encryption-at-rest does not prevent an embargo violation. Intrusion detection systems do not catch a researcher who presents findings at a conference one month early. MFA does not address the human judgment failures that account for nearly all documented incidents. The proposed framework applies a technical solution to what is fundamentally an administrative problem. The attestation framework compounds this problem. Under the False Claims Act's "implied certification" doctrine (Universal Health Services v. Escobar, 579 U.S. 176, 2016), institutions drawing down grant funds while operating under POAMs acknowledging they meet a fraction of required controls could face material legal exposure. The DOJ has already settled with Georgia Tech ($875,000, Sept 2025) and Penn State ($1.25M, Oct 2024) for cybersecurity misrepresentations -- neither involving actual breaches. Wake Forest temporarily suspended dbGaP access in October 2025 rather than face this exposure. A system that punishes good faith disclosure and rewards inflated self-assessment produces the opposite of security.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
In 2013, the Supreme Court held in Association for Molecular Pathology v. Myriad Genetics (569 U.S. 576) that naturally occurring DNA sequences cannot be patented, because locking up the genome behind proprietary barriers would prevent others from studying it and stifle the scientific progress that benefits everyone. The NIST SP 800-171 compliance requirement is not a patent, but it produces the same result: access to the genomic commons determined by institutional wealth rather than scientific merit. The cost of NIST SP 800-171 compliance creates a financial barrier that determines access to the genomic commons by institutional wealth rather than scientific merit. Industry estimates for bringing a

single research computing environment into compliance range from $50,000 to over $200,000 in initial costs, with comparable recurring annual costs. For a major research university with centralized IT, these costs are absorbable. For HBCUs, Minority Serving Institutions, and smaller regional universities, they are prohibitive. NIH has identified health equity as a strategic priority. The institutions most likely to conduct research with underrepresented populations and recruit diverse cohorts are precisely those least able to absorb unfunded compliance costs. A compliance mandate costing $200,000 represents a meaningful fraction of many of these institutions' entire research budgets. If a $500,000 NIH grant requires $100,000 in compliance infrastructure before a single analysis can run, the effective research budget drops by 20%. For a PI who cannot spread compliance costs across dozens of grants, the arithmetic is disqualifying. The PI does not apply. The research does not happen. The proposed policy extends NIST-equivalent requirements to 11 categories of human research data. The compliance burden scales from a specialized subset of genomics researchers to the majority of NIH-funded investigators working with human data. NIH cannot simultaneously mandate defense-grade security and decline to fund it. Repository availability is meaningless if the institutions and researchers who need access are priced out of compliance.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The extension from genomic data to 11 protected categories -- including personal health data, proteomic data, transcriptomic data, epigenomic data, and head/facial imaging -- will subject virtually all human subjects research conducted with NIH funding to NIST-level security requirements. This expansion is disproportionate to demonstrated risk. The 20-year empirical record at dbGaP shows a 0.09% violation rate with zero confirmed participant harms, even for the narrower category of genomic data. Extending the same defense-industrial security framework to all 11 categories multiplies the compliance burden by an order of magnitude without evidence that the previous framework caused harm. The categories themselves vary enormously in re-identification risk. Aggregate proteomic or transcriptomic data from cell lines poses fundamentally different privacy risks than whole-genome sequences linked to clinical records. A uniform NIST 800-171 requirement across all 11 categories fails to distinguish between these risk profiles.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The proposed GDS revisions should be evaluated against the mechanism that already governs participant risk: informed consent. Every human subjects protocol requires IRB approval and informed consent covering data types, storage, sharing, and re-identification risks. Participants who find those risks unacceptable decline. Those who consent do so knowingly. NIST SP 800-171 overrides that decision. It substitutes a cybersecurity standard written for defense contractors for the judgment of the individuals whose data is at stake. Patient advocacy communities -- rare disease organizations, disability groups, research participant networks -- have consistently supported broader, faster data sharing. Surveys of genomic research participants (including myself) confirm willingness to accept meaningful privacy risks in exchange for scientific advancement. The current policy does not reflect their preferences. Current consent processes are not perfect, and comprehension of genomic privacy risks is uneven. But the solution to imperfect consent is better consent and not a cybersecurity standard that removes the participant from the decision entirely. NIH should invest in consent innovation (e.g., dynamic consent models, participant-facing dashboards) rather than compliance infrastructure. Specific recommendations: (1) Withdraw the blanket NIST SP 800-171 requirement for research data. The pre-NOT-OD-24-157 framework produced a 0.09% violation rate over 20 years with zero confirmed harms.

(2) Reform or abandon the attestation framework, which generates widespread known non-compliance and subjects institutions to FCA liability for disclosing it. If NIH wants to assess security postures, it should conduct direct assessments, not self-attestation backed by punitive enforcement. (3) Publish updated violation data so policy decisions are grounded in current evidence. (4) Acknowledge that NIH operates under FISMA constraints but use the "or equivalent" language already in the draft policy to define proportionate alternatives.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Imputation servers built from controlled-access reference panels are critical infrastructure for genomic research, enabling genotype inference at a fraction of the cost of whole-genome sequencing. The restriction to servers "funded and operated by NIH or a federal agency" is unnecessarily narrow. Researcher-operated imputation servers, operating under the same access controls that governed dbGaP successfully for 20 years (access authentication, data use agreements, institutional review), should be permitted. The requirement that controlled-access data used to build these tools be restricted to a single federal operator bottlenecks a capability that benefits the entire field. If the goal is to ensure security of the reference panels, the policy should specify security requirements for imputation servers rather than restricting who may operate them.

65. COGR

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Kristin West

Name of Organization: COGR

Type of Organization: Other

Type of Organization - Other: Non-Profit Association of Academic Research Institutions

Role: Other

Role – Other: Director, Research Ethics & Compliance

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached letter.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Please see attached letter.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Please see attached letter.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached letter.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see attached letter.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/PDF-Final-Submission-COGR-Response-to-NIH-RFI-on-CADR-and-GDS-Policy-March-2026.pdf

Description: Letter detailing response to NIH's Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

66. University of Pittsburgh

Submit date: 3/16/2026

I am responding to this RFI: On behalf of an organization

Name: Bill Yates

Name of Organization: University of Pittsburgh

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
INSTITUTIONS CANNOT COMPLY WITH A POLICY THEY CANNOT INTERPRET

Our research compliance and IT security teams have reviewed the CAD Policy, but they cannot confidently determine what it requires. The policy instructs that covered data "must be protected throughout the data lifecycle" and directs institutions to employ NIST SP 800-171 or equivalent security standards. But it does not define "controlled-access repository," does not identify what constitutes "the data lifecycle" in an operational sense, and does not clarify which systems within an institution's research infrastructure fall within its scope.

This ambiguity has direct, high-stakes consequences. The University of Pittsburgh operates clinical research activities that involve pharmaceutical company-provided electronic case report forms, commercial laboratory information systems, imaging platforms built by device vendors, and electronic medical record modules used simultaneously for clinical care and research documentation. None of these systems were designed to meet NIST SP 800-171 — a standard developed for federal contractors handling controlled unclassified information — and their vendors have no regulatory obligation or commercial incentive to modify them to do so. If the CAD Policy is interpreted to reach these systems, compliance may be technically impossible, not merely expensive.

THE CAD POLICY DUPLICATES HIPAA WITHOUT ACKNOWLEDGING IT

A substantial proportion of the human participant data covered by the CAD Policy at our institution already exists within HIPAA-regulated systems. UPMC is a covered entity under HIPAA, and data that flows from clinical care into research — including electronic health record extracts, claims data, laboratory results, and imaging files — is subject to the full HIPAA Privacy and Security Rule framework. These rules are not weak: they require administrative, physical, and technical safeguards; they mandate risk analysis and risk management; and they are actively enforced by the Office for Civil Rights.

The CAD Policy does not acknowledge HIPAA or the considerable protections it already provides. Instead, it would layer NIST SP 800-171 on top of HIPAA compliance for the same data, in the same systems, managed by the same personnel. The practical consequence is not a more secure research environment; it is duplicative documentation, redundant auditing, and conflicting standards that compliance staff must somehow reconcile — all without meaningfully improving participant protection. A patient whose blood pressure measurements are used in an NIH-funded trial should not be subject to a different security standard than a patient whose identical measurements are used in an industry-funded trial, simply because one funding source requires NIST and the other does not.

We strongly recommend that NIH exempt from CAD Policy controls any data that is already subject to the HIPAA Security Rule and maintained in systems that have completed a compliant HIPAA security risk analysis. This exemption is consistent with the DOJ Rule's approach of avoiding duplicative federal regulation and would significantly reduce compliance burden without creating any gap in participant protection.

RETROACTIVE APPLICATION WILL HARM ONGOING AND FUTURE RESEARCH

The University of Pittsburgh has researchers currently conducting NIH-funded studies that rely on institutional data repositories assembled over years or decades — repositories containing deidentified clinical data, biospecimen-derived measurements, and imaging archives for which participant consent was obtained under the standards applicable at the time of collection. The CAD Policy's requirement that openly shared data be accompanied by informed consent, "explicitly stating data are to be shared openly without controls," cannot be satisfied for this data. In most cases, the individuals who contributed these samples and records cannot be recontacted, and even if they could, the datasets are typically de-identified in ways that preclude linkage back to individual consent records.

We are also concerned about studies that are currently funded and underway. Investigators who designed studies around institutional analytical environments — environments that provide capabilities unavailable in any NIH-managed repository, such as integration with clinical data warehouses or specialized neuroimaging pipelines — have no path to compliance under the CAD Policy as proposed without abandoning their current infrastructure. For some studies, this would mean early termination. NIH should not inadvertently defund active research through a retroactive policy change.

We urge NIH to apply the CAD Policy prospectively, covering only awards made and data collected on or after the effective date. For existing datasets, informed consent should not be required as a condition of continued use when the data has already been de-identified per the HIPAA Safe Harbor (removal of the 18 enumerated identifiers) or the Expert Determination standard. We also urge NIH to provide a formal transition mechanism for currently funded studies that need additional time or resources to assess and address their compliance status.

THE FINANCIAL BURDEN WILL STRATIFY ACCESS TO NIH FUNDING

Cost is a legitimate policy consideration — one that is often underweighted in regulatory comment processes, and one we want to address directly. The University of Pittsburgh has explored the development of research data enclaves compliant with NIST SP 800-171, including an enclave specifically built and recently third-party certified for Department of Defense Cybersecurity Maturity Model Certification (CMMC) Level 2.

Standing up a single enclave — encompassing start-up costs, fixed infrastructure, configuration, validation, staffing, and ongoing maintenance — is estimated to cost between $1 million and $3 million over an initial five-year period. For a project requiring access for three individuals, fixed costs alone translate to an annualized per-project expense exceeding $16,000. Variable costs, which scale with actual usage of the environment, add another $16,000 or more annually under the same assumptions — and can rise substantially depending on utilization.

Combined, a single project with three users can expect to pay more than $32,000 per year to operate within one NIST 800-171-compliant, CMMC Level 2-certified enclave — before accounting for any surge in usage-driven expenses.

Given the breadth of data the CAD Policy would cover, institutions would not need one such enclave — they would need several, configured for different data types, research populations, and analytical environments. For a large research university like Pitt, this is an extraordinary but manageable undertaking if approached over a multi-year horizon. For smaller institutions, regional universities, historically Black colleges and universities, and community-based medical centers — all of which play irreplaceable roles in ensuring that clinical research reflects the full scope of the American population — these costs may be simply prohibitive. The predictable result would be a contraction of the NIH-funded research enterprise into a smaller number of well-resourced institutions, which we believe is contrary to NIH's stated goals for a "broad range of smaller and emerging research institutions.

To mitigate these effects, we recommend that NIH: (a) establish an effective date no sooner than three years after final policy issuance to allow institutions adequate time to plan, budget, and build; (b) make compliance infrastructure costs allowable as direct costs on NIH awards; (c) create a dedicated funding mechanism — similar in spirit to the S10 instrumentation program — to support institutional development of compliant research data repositories; and (d) survey institutions prior to finalizing the policy to develop realistic estimates of the data volume that would need to be housed in NIH-managed repositories, and assess whether current NIH infrastructure can absorb it.

THE POLICY SHOULD BE DEVELOPED IN COORDINATION WITH OTHER FEDERAL AGENCIES

The University of Pittsburgh receives research funding from NIH, NSF, DOD, VA, and a number of other federal agencies, many of which are independently developing or implementing research data security requirements. Managing compliance across these parallel frameworks is already a significant administrative challenge. The CAD Policy as proposed would add another layer of requirements that may not align with — and could conflict with — those of other agencies.

Executive Order 14117 specifically directed HHS, DOD, VA, and NSF to develop data protection guidance in consultation with each other. That consultation does not appear to have informed the CAD Policy. We urge NIH to engage in that interagency process before finalizing the policy. A coordinated approach — modeled on the Common Rule framework that harmonized human subjects protections across federal agencies, or on the interagency implementation of NSPM-33 — would significantly reduce the compliance burden on institutions while ensuring consistent, enforceable standards. At a minimum, we ask that NIH commit to publishing revised draft language for stakeholder comment before finalizing any policy that differs materially from what has been proposed in this RFI.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
NIH describes the CAD Policy as a response to "emergent privacy and security risks," and we agree that those risks are real. The DOJ's Bulk Data Transfer Rule — which our institution has worked to comply

with — represents a carefully calibrated federal response to the threat posed by foreign adversaries' access to sensitive personal data. The CAD Policy, however, does not replicate that calibration. Where the DOJ Rule establishes meaningful thresholds (e.g., genomic data from at least 100 U.S. persons, precise geolocation from at least 1,000) and limits its reach to transactions with countries of concern, the CAD Policy would apply to a sizeable fraction of human participant data generated in NIH-funded studies, regardless of quantity, format, sensitivity, or whether the data is ever shared with anyone outside the institution.

Consider a practical example: a clinical coordinator records a single study participant's ZIP code to confirm eligibility for a local trial. Under the CAD Policy as written, that single data point — geolocation tied to one individual — would require controlled-access protections. There is no credible national security or privacy argument for treating that record the same as a whole-genome sequencing dataset from 10,000 participants. A policy that cannot distinguish between these cases is not risk-based; it is simply broad.

We urge NIH to restructure the CAD Policy around a tiered, risk-proportionate framework. Data categories should be stratified by sensitivity and quantity, with security requirements scaled accordingly. At a minimum, the CAD Policy should incorporate bulk data thresholds comparable to those in the DOJ Rule before imposing its most stringent controls. NIH should also clearly distinguish between data types that are inherently high-risk (e.g., whole-genome sequences, imaging data enabling facial recognition) and data types that become sensitive only in combination or at scale.

THE CAD POLICY DUPLICATES HIPAA WITHOUT ACKNOWLEDGING IT

A substantial proportion of the human participant data covered by the CAD Policy at our institution already exists within HIPAA-regulated systems. UPMC is a covered entity under HIPAA, and data that flows from clinical care into research — including electronic health record extracts, claims data, laboratory results, and imaging files — is subject to the full HIPAA Privacy and Security Rule framework. These rules are not weak: they require administrative, physical, and technical safeguards; they mandate risk analysis and risk management; and they are actively enforced by the Office for Civil Rights.

The CAD Policy does not acknowledge HIPAA or the considerable protections it already provides. Instead, it would layer NIST SP 800-171 on top of HIPAA compliance for the same data, in the same systems, managed by the same personnel. The practical consequence is not a more secure research environment; it is duplicative documentation, redundant auditing, and conflicting standards that compliance staff must somehow reconcile — all without meaningfully improving participant protection. A patient whose blood pressure measurements are used in an NIH-funded trial should not be subject to a different security standard than a patient whose identical measurements are used in an industry-funded trial, simply because one funding source requires NIST and the other does not.

We strongly recommend that NIH exempt from CAD Policy controls any data that is already subject to the HIPAA Security Rule and maintained in systems that have completed a compliant HIPAA security risk analysis. This exemption is consistent with the DOJ Rule’s approach of avoiding duplicative federal regulation and would significantly reduce compliance burden without creating any gap in participant protection.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We generally support the proposed GDS Policy revisions and believe they represent a sensible modernization of a policy that has served the research community well since 2014. Limiting the GDS Policy's scope to human genomic data, simplifying the "large scale" threshold to studies involving 100 or more individuals, and permitting HIPAA Expert Determination as an alternative to Safe Harbor de-identification are all practical improvements. We also support allowing Human Research Protection Programs to certify data submissions beyond the traditional IRB or Privacy Board — this reflects the reality that many institutions have developed robust human research protection infrastructure that extends well beyond the IRB.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/University-of-Pittsburgh-Response-CAD-Policy.pdf

Description: Response Letter

67. N/A

Submit date: 3/16/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Other

Role: Member of the Public

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I am a former biomedical researcher and someone who has participated in research studies. The NIH Controlled Access Data policy leads with a national security framing, but NIH has not provided sufficient evidence that the national security threat justifies the scale of what is being proposed in this policy. We already have Institutional Review Boards specifically designed to make risk-calibrated judgments about data protection. NIH should demonstrate why that system is inadequate before replacing it.

My biggest concern is centralization. I have revoked my consent for my medical records to be used for research purposes because of this proposed policy change. This policy would move personal health data into federally managed repositories. Once it is there, what prevents other federal agencies, including the Department of Justice, from accessing this, genomic, or other potentially identifiable information? The current system, where data is held across many institutions, provides stronger protection for individuals and communities. That federated structure is a feature worth preserving. Additionally, research participants may not have consented to sharing their data in a federally managed repository.

Congress directed NIH to convene a working group to make recommendations on appropriate data sharing, which NIH should do. NIH should also conduct a thorough analysis of whether less burdensome alternatives would achieve the same security and privacy goals before implementing this policy. Finally, NIH should make clear what protections are in place to ensure participant's data will not be accessed by other federal agencies.

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

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

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

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

68. Foundation for Defense of Democracies

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Foundation for Defense of Democracies

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached file.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Please see attached file.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Please see attached file.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached file.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see attached file.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/FDD-NIH-Public-Comment.docx

Description: Please see the attached document.

69. Yale University Cushing/Whitney Medical Library

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Levi Dolan

Name of Organization: Yale University Cushing/Whitney Medical Library

Type of Organization: Academic Institution

Role: Other

Role – Other: Librarian

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/2026_NIH_CADR_RFI_CWML.docx

Description: This document was reviewed and commented on by the Yale Cushing/Whitney Medical Library Technology & Innovation Team and library leadership.

70. Barbara J. Evans, Ph.D., J.D., LL.M.

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: Barbara J. Evans, Ph.D., J.D., LL.M.

Name of Organization:

Type of Organization: Not Applicable

Role: Member of the Public

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please refer to the attached PDF file, which provides detailed comments of Barbara J. Evans, Ph.D., J.D., LL.M., on the draft NIH Controlled-Access Data Policy and proposed revisions to the NIH Genomic Data Sharing (GDS) Policy (NOT-OD-26-023).

To summarize, the proposed policies described in NOT-OD-26-023 do not appear to ensure that NIH Controlled-Access Data Repositories (CADRs) will operate in full compliance with the U.S. Department of Justice Data Security Program (DSP) regulations at 28 C.F.R. Part 202. U.S.-based academic medical centers and research institutions could face significant legal risks (large monetary fines and potential criminal liability) for contributing data to a repository, knowing that the repository reshares data under policies that do not comply with the DSP regulations. It does not appear that NIH CADRs and those who contribute data to them will be exempt from the DSP regulations. However, if NIH has determined that its CADRs are exempt from DSP compliance, it would be helpful for its Controlled-Access Data Policy to set out the legal basis for that determination. Otherwise, U.S.-based institutions that receive NIH funding will need stronger reassurances that NIH CADRs plan to operate in full compliance with the DSP regulations. The attached comments offer specific suggestions for the NIH to consider as it works to finalize the draft policies described in NOT-OD-26-023. More generally, I applaud the NIH's efforts to sustain scientific data sharing in the face of recent regulatory changes that profoundly altered the ground rules for sharing genomic, health, and other types of data essential to biomedical research. Creating a new, financially sustainable national infrastructure for secure, privacy-preserving scientific data sharing is a crucial challenge of our time, and there are no easy solutions. This challenge can be met through NIH leadership and through close collaboration between the NIH and its funded research institutions, many of which have long experience navigating the ever-more-complex U.S. legal and regulatory environment.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The attached comments of Barbara J. Evans, Ph.D., J.D., LL.M. are relevant to the proposed revisions to

the NIH Genomic Data Sharing (GDS) Policy as well as to the draft NIH Controlled-Access Data Policy (NOT-OD-26-023).

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Barbara-J-Evans-Comments-on-RFI-NOT-OD-26-023.pdf

71. International Society for Biological and Environmental Repositories (ISBER)

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Rita T. Lawlor

Name of Organization: International Society for Biological and Environmental Repositories (ISBER)

Type of Organization: Other

Type of Organization - Other: global organization that addresses the harmonization of scientific, technical, legal, and ethical issues relevant to repositories of biological and environmental specimens

Role: Other

Role – Other: chair of science policy community of practice of ISBER

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attachment

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

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see attachment

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ISBER-Comments-on-NIH-Datasharing-RFI-3-10-26_signed.pdf

Description: ISBER comments on NIH data sharing as it pertains to biobanked samples and data in an international context.

72. Kayte Spector-Bagdady

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: Kayte Spector-Bagdady

Name of Organization: University of Michigan

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Statements of Support

Benefits of controlled-access databases: The proposed NIH Controlled-Access Data Policy's emphasis on centralized, repository-based data access represents an important opportunity to advance equity in biomedical research. Researchers have reported that moving away from a model requiring them to download, store, and analyze large volumes of data on institutional servers can substantially lower barriers for access by less well-resourced researchers and institutions. Secure data-analysis enclaves and restricted cloud environments, in which researchers use NIH-provided or approved tools to analyze data without removing it from the repository environment, simultaneously expand access and strengthen the privacy protections central to this proposed policy.

Benefits of streamlining and harmonizing administrative processes: Researchers have reported that current controlled-access processes can be slow and inconsistently applied, and that the Data Use Agreement process between federal agencies and institutions can create significant administrative burden, sometimes inequitably applied. These tensions can delay or deter research, particularly for investigators at smaller institutions with limited administrative infrastructure.1 The proposed harmonization of requirements across NIH Institutes, Centers, and Offices – including the proposal to prohibit individual ICOs from expanding GDS Policy scope through program-specific expectations – is a promising way to reduce complexity, improve consistency, and make controlled-access data more accessible to the broader research community.

Additional Recommendations

NIH should continue to prioritize data harmonization across databases: A persistent barrier to the use of controlled-access data is the burden of harmonizing datasets within or across repositories. This is often described by researchers as time-consuming and technically demanding, particularly when aligning samples across different reference panels or data collection protocols. Centralizing harmonization at the repository level, rather than requiring each research team to independently resolve these incompatibilities, would substantially reduce duplicative effort and improve the interoperability and reusability of shared data. NIH should invest in repository-level harmonization standards, and, where possible, require that data submissions conform to common data elements and formats to facilitate cross-study analysis.

NIH should integrate demographic and ancestral diversity within high-use controlled-access repositories: A critical limitation of many existing NIH controlled-access databases is the underrepresentation of

participants from non-European ancestral backgrounds. Expanding the diversity of data within widely used repositories is an operational investment that directly increases the scientific value and utility of controlled-access infrastructure. Researchers report that, even when they actively seek to conduct ancestry-diverse analyses, they are constrained by the composition of available cohorts. Researchers also report prioritizing ease of access and phenotype availability over diversity when diverse samples are not readily available. NIH should prioritize integrating data from underrepresented populations directly into existing, widely used databases. Evidence suggests that researchers are more likely to incorporate diverse data into their analyses when it is accessible within the platforms and resources they already use routinely. Recruitment and data collection efforts targeting underrepresented communities should be designed with integration into established controlled-access repositories as the goal, and NIH should consider requiring standardized collection and reporting of ancestry and demographic metadata.

The NIH should increase resources for monitoring and enforcement: Meeting increased demand for controlled-access data sharing requires not only expanded repository capacity, but also the operational resources necessary to ensure that data submitted to those repositories meets consistent quality and usage standards. The proposed policy establishes important requirements for data sharing and protection, but its success depends on NIH's capacity to enforce those requirements consistently and transparently. Researchers report that inconsistent enforcement can create a “free-rider” dynamic in which some researchers benefit from others’ shared data without contributing their own, placing researchers who comply in good faith at a competitive disadvantage. Repository-level quality control is equally important, without enforced standards, secondary researchers encounter datasets that lack the clinical or methodological detail necessary for robust reanalysis, wasting significant time and resources. NIH should establish clear consequences for non-compliance with data sharing and quality requirements and commit resources to auditing and enforcing them. Identifying dedicated resources for compliance monitoring is an operational prerequisite for realizing the capacity investments NIH is making in controlled-access repositories.

NIH should incentivize the labor of data sharing: The NIH will not realize repository capacity investments if submissions remain low-quality or are delayed. Long-term success depends on researchers perceiving data sharing not merely as a compliance obligation but as a professionally rewarding activity. Researchers frequently report that the work of cleaning, documenting, and preparing data for submission to controlled-access repositories is a significant burden that yields little academic advancement. NIH should create mechanisms, such as data contribution metrics in grant review, that provide meaningful professional credit for the production of high-quality, reusable datasets. Reducing the perceived cost of compliance while increasing its professional reward would strengthen both the quantity and quality of data submitted.

NIH should invest in repository administrative support: The proposed policy’s success will depend not only on technical infrastructure, but also on sustained administrative support – particularly for junior researchers or those at institutions with fewer mentorship opportunities. Researchers value consistency and efficiency in their interactions with data steward representatives. Dedicated, long-term personnel serving as technical guides would meaningfully lower barriers to data use by helping researchers navigate repository-specific data structures and address study-specific questions. NIH should fund stable administrative and technical support positions within controlled-access repositories, rather than automated systems or short-term contract staff. This investment would be particularly beneficial for researchers at institutions with limited informatics infrastructure or mentorship.

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

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

Benefits of including "data on reproductive and sexual health" in the definition of personal health data: The designation of "personal health data" as a controlled-access data type and explicitly including reproductive and sexual health data are consistent with NIH's stated commitment to participant protection and reflects a responsive approach to emergent privacy risks. These data are among the most sensitive that NIH-supported research collects, and their unauthorized disclosure could expose research participants to discrimination, stigmatization, or legal harm.

Recommendation

NIH should use repositories as infrastructure for dynamic risk-management of "covered personal identifiers": Although the proposal includes "covered personal identifiers" as a controlled-access data type, its definition treats identifiability as a fixed property rather than a function of context, capability, and time. Even data stripped of traditional identifiers can be re-identified through aggregation of disparate datasets, and advances in artificial intelligence have dramatically lowered the technical barrier to doing so. Because the availability of data that can be used to re-identify previously de-identified datasets changes continuously, identifiability determinations have an implicit expiration date. NIH should consider time-limited de-identification certifications, after which risk must be reassessed against the current technological landscape. Tiered access repository models, such as those in use with All of Us, can provide such risk-proportionate architecture. Paired with secure data-analysis enclaves, this approach can enforce data use restrictions and limit linkage to external datasets.

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

Clarify the de-identification standard that applies under the GDS Policy: The proposed revision to the GDS Policy incorporates de-identification standards from both HIPAA and the Common Rule, but does not resolve meaningful inconsistencies between them. Under HIPAA, de-identification can be achieved through either Expert Determination (requiring that re-identification risk be "very small") or removal of 18 identifiers (also requiring "no actual knowledge" that the data could be used to identify an individual). The overarching HIPAA de-identification standard – that there be "no reasonable basis to believe" that data can be re-identified – arguably applies to both methods, but it is unclear how to reconcile these seemingly different scienter, or intentionality, requirements. The proposed policy also references the Common Rule's "cannot be readily ascertained" standard, without clarifying how they all interact.

Institutions need additional clarification regarding at least three questions:

1. When an institution removes the 18 HIPAA identifiers, must it also ensure that identities "cannot be readily ascertained" under the Common Rule (presumably a more protective standard than HIPAA's Safe Harbor "no actual knowledge" requirement)?

2. When HIPAA Expert Determination is used, do “cannot be readily ascertained” and “very small risk” impose the same standard, or are both independently required? If so, what is the difference?

3. Does the overarching HIPAA de-identification standard, that there is “no reasonable basis to believe” data can be re-identified, apply to GDS-governed data? If so, how are we to apply it within the context of the other standards?

The above ambiguities will create compliance uncertainty for institutions and researchers. Researchers might respond by over-restricting data sharing and increasing administrative burden, undermining NIH goals of simplifying policies and promoting maximal data sharing. Or they might apply the least protective interpretation, leaving some participants vulnerable to re-identification. This could create a situation where some participants will receive stronger privacy protection than others, based solely on where their data was collected or processed. Clearer guidance on de-identification, or preferably a unified standard across DHHS, is essential for consistent implementation. More broadly, the complexity of reconciling these overlapping frameworks reinforces the case for moving away from a binary identified/de-identified model toward a risk-proportionate, tiered, controlled-access architecture.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Spector-Bagdady-RFI-on-Draft-NIH-Controlled-Access-Data-Policy-and-Proposed-Revisions-to-NIH-Genomic-Data-Sharing-Policy-03.17.26.pdf

Description: Full comments

73. Eric S. Rosenthal, M.D., and Barbara J. Evans, Ph.D., J.D., LL.M.

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: Eric S. Rosenthal, M.D., and Barbara J. Evans, Ph.D., J.D., LL.M.

Name of Organization:

Type of Organization: Not Applicable

Role: Member of the Public

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The attached PDF file provides detailed comments of Eric S. Rosenthal, M.D., and Barbara J. Evans, Ph.D., J.D., LL.M., on the draft NIH Controlled-Access Data Policy and proposed revisions to the NIH Genomic Data Sharing (GDS) Policy (NOT-OD-26-023). To summarize, real-world clinical data (RWD) are an increasingly important resource for biomedical discovery and for developing and validating AI tools for clinical care. Scientific data sharing of RWD requires heightened privacy protections and careful regulatory compliance to protect the rights of patients whose RWD fuel biomedical discovery. Following amendments to the Common Rule that took effect in 2019, HIPAA-covered academic medical centers that use RWD in NIH-funded research often are "Common Rule-exempt" under 45 C.F.R. § 46.104(d)(4)(iii). The Common Rule itself calls for their research to be regulated by the HIPAA Privacy Rule (rather than by the Common Rule), and the HIPAA Privacy Rule also governs future sharing of RWD from Common Rule-exempt research. Ironically, in order for NIH CADRs to comply with the amended Common Rule, they now need to be able to comply with the HIPAA Privacy Rule – the Common Rule requires them to do so if they host and share any RWD from Common Rule-exempt research that the NIH funds. Unfortunately, many of the NIH's legacy data repositories predate the Common Rule amendments and lack capacity to administer scientific data sharing of RWD from Common Rule-exempt research. This situation raises serious concerns for NIH-funded academic medical centers that use RWD in Common Rule-exempt research. They have non-delegable duties to comply with the HIPAA Privacy Rule and more-stringent state laws that survive HIPAA preemption. They cannot lawfully entrust HIPAA-protected RWD to NIH CADRs without reassurance that NIH CADRs are capable of managing future data sharing in compliance with those laws. The draft NIH Controlled-Access Data Policy and proposed revisions to the NIH GDS Policy seem to assume that data entering NIH CADRs comes from Common Rule-regulated research. The proposed policies do not offer satisfactory assurances that NIH CADRs will be able to administer the sharing of HIPAA-regulated RWD, which represent a growing part of the modern research data ecosystem. Our comments offer specific suggestions for NIH to consider as it finalizes the proposed policies to address this gap. We look forward to ongoing dialogue with the NIH as it works to resolve privacy protection concerns with scientific sharing of real-world clinical data.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The attached comments of Eric S. Rosenthal, M.D., and Barbara J. Evans, Ph.D., J.D., LL.M. are relevant to the proposed revisions to the NIH Genomic Data Sharing (GDS) Policy as well as to the draft NIH Controlled-Access Data Policy (NOT-OD-26-023).

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Eric-Rosenthal-and-Barbara-Evans-Comments-RFI-NOT-OD-26-023.pdf

74. Massachusetts Institute of Technology - MIT Libraries

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Chris Bourg

Name of Organization: Massachusetts Institute of Technology - MIT Libraries

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Thank you for this opportunity to comment on the Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy. Given our institutional commitments "…to advance knowledge and educate students in science, technology, and other areas of scholarship that will best serve the nation and the world in the 21st century," and "...to generating, disseminating, and preserving knowledge, and to working with others to bring this knowledge to bear on the world's great challenges," along with our organizational principle to "...democratize access to data, and provide expertise and advocacy to advance socially responsible computational research and learning," these comments focus on how these proposals support the ability of our and the nation's collective research community to continue conducting the bold and barrier-breaking research for which they have been historically known.

A key to successful implementation of securing genomic and related research is that our researchers, their teams, and their institutions be able to manage security responsibilities efficiently. In this proposal, the requirement that the listed data types be protected at a level equivalent to NIST 800-171 throughout the data lifecycle results in significant burden, which could be addressed more effectively. Recognizing the duty to protect these data types from being shared with "countries of concern," there are many ways to approach this responsibility and to design systems that meet the complexity of the work and its demands for protection. Systems should focus on identifying and effectively vetting responsible people and organizations in reliable, sensible, and trustworthy ways. Processes for determining that collaborators are responsible people who will handle data with the care, concern, and protectiveness it demands are more effective, nuanced, and appropriate to the applicable systems.

As represented in these proposals, the protective requirements present significant challenges and substantial burdens that institutions will struggle to meet. The requirements as drafted imply simple control systems across the range of institutions, which in fact vary widely in their size, complexity, and resources. In even relatively small institutions, research environments are often decentralized given the varying structures within which they operate. The reality is that this monolithic approach could significantly harm the nation's research productivity, by introducing significant new burdens.

Beyond these challenges in the service of protection, other aspects of the proposals may lead to stifling our national scientific advancements. These include imprecise wording in the "Additional Policy Considerations" section that is subject to varied interpretations. Concerns that these considerations may be employed inconsistently and punitively will cause researchers to avoid potentially advantageous research paths. Additionally, an apparent gap in guidance around the disposition of and timelines

related to non-human genomic data sharing may inhibit the timely sharing of this information and associated advancements. Providing direction on when and where these data types should be shared, or to an applicable framework would help mitigate these impacts.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Reviews of DMSPs submitted to NIH may provide actionable information on expected volume of data for the listed types, and support prediction growth models to assess the adequacy of current availability. Further funding should be provided to each of the repository organizations currently serving the needs of storage and distribution of controlled-access data to fully adhere to the NSTC Desirable Characteristics for Repositories. Standard increases in funding should be guaranteed to support their continued development and future sustainability. Future development should align with global standards and practices such as those currently in progress by the RDA Health Data Commons GORC Profile WG and Trusted Research Environments for Sensitive or Confidential Data: FAIRness for Controlled Data and Processes WG.

- https://www.rd-alliance.org/groups/health-data-commons-gorc-profile-wg/activity/

- https://www.rd-alliance.org/groups/trusted-research-environments-for-sensitive-or-confidential-data-fairness-for-controlled-data-and-processes-wg/activity/

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

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

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

75. San Diego State University

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: John Crockett

Name of Organization: San Diego State University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
"Protected throughout the data lifecycle" needs operational guardrails.

The draft policy requires these data types be protected "throughout the data lifecycle," including when not shared through a controlled-access repository. NIH should define, at a minimum, what institutional safeguards are expected in routine operations versus repository deposit versus broad downstream reuse. NIH should explicitly confirm that institutions may meet this expectation via a defined, project/system-scoped secure research environment (i.e., an "enclave" model), rather than implying campus-wide compliance across all institutional systems. In the absence of minimum operational definitions, institutions will implement inconsistent practices and accept inconsistent risk. NIH could publish an FAQ and sample SOP language describing acceptable protections (e.g., role-based access control, encryption at rest/in transit, access logging, incident response) and how these differ by context.

NIH should also clarify what constitutes "authentication of the identity of data requesters" (e.g., whether federated SSO + MFA and standard institutional identity proofing is sufficient) to avoid institutions adopting unnecessary and costly identity proofing processes.

Security standards, attestations, and institutional burden.

NIH's current trajectory requires users of controlled-access data to align institutional and third-party systems to NIST SP 800-171 (or comparable equivalents in limited contexts). For higher education, NIH should clarify whether attestation is expected at the institution-wide level or at the system boundary used for NIH controlled/protected data; we strongly recommend system-/environment-scoped attestation (SSP/POA&M at the enclave boundary) as the only scalable model for decentralized campus research. If the new policy expands controlled-access requirements beyond genomics, NIH should provide standardized attestation language, a consistent approach to Plans of Action & Milestones (POA&Ms), and clear guidance on responsibility allocation across prime/subawardee institutions. NIH should also publish practical implementation guidance on restrictions for "countries of concern" in common academic scenarios (e.g., travel, visiting scholars, dual affiliations, cloud tenancy/residency) to reduce inadvertent noncompliance. NIH should also invest in NIH-supported secure analysis workspaces to reduce duplicative, inequitable campus-by-campus buildout.

Budgeting and potential unfunded mandates.

NIH's DMS budgeting guidance recognizes data stewardship costs but can treat shared secure infrastructure as unallowable when accounted as institutional overhead. NIH should clarify which

security/compliance costs are allowable direct costs when required by policy (e.g., secure environment costs, security assessments, repository deposit fees) and consider mechanisms to avoid shifting these costs to institutional funds. NIH should also provide reference architectures and/or pre-approved security patterns (including cloud options) to reduce duplicative spending and speed compliance. A phased implementation with adequate lead time is essential.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Repository availability, certification, and NIH-supported platforms.

We encourage NIH to expand the capacity of established controlled-access repositories and associated access management systems before the policy is effective. Many disciplines (imaging, behavioral, mixed clinical/real-world data) rely today on domain repositories that may not meet NIH's minimum controlled-access definition. NIH should provide a transparent certification pathway for non-NIH repositories that can meet the minimum requirements, and should publish repository onboarding timelines and capacity planning assumptions. NIH should also provide a practical "minimum viable" control set and implementation examples for domain repositories (beyond stating what is not controlled access), so repositories and institutions can converge on consistent, auditable practices without excessive reinvention.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Scope, definitions, and risk-tiering.

We support NIH specifying which data types warrant controlled access; however, several proposed "protected data types" are defined so broadly that they risk making controlled access the default for most human participant research data (particularly "personal health data" and "imaging data of the human face or head regions"). We recommend NIH adopt a tiered, risk-based framework:

• Tier A (always controlled): direct identifiers, precise geolocation, biometric templates, and imaging containing facial features/surfaces reasonably usable for identification.

• Tier B (controlled by default, may be open with documented low-risk assessment): individual-level clinical trial data; -omics data; and other high-dimensional participant-level datasets.

• Tier C (typically open): summary statistics and other low-risk outputs, with clear safe harbors.

NIH should publish explicit examples and safe-harbor categories that "typically should not be controlled," including genomic summary results and clinical trial summary results, to preserve appropriate open science while focusing controls on higher-risk data.

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

We support reducing duplicative requirements and clarifying that "large-scale" genomic data is defined by a clear threshold. We request clarification on how the "6 months from generation" expectation will be applied across diverse study designs and repositories. NIH should explicitly provide an exception/extension pathway for academic realities that drive delay (consent limitations, multi-site agreement/DUA timing, de-identification and QC workload, repository release queues), with clear documentation expectations and without punitive assumptions of noncompliance. NIH should also

reconsider whether “100 individuals” as the definition of “large scale” is too low for many routine academic studies, or provide a risk-based modifier to avoid unnecessary administrative burden where re-identification risk is demonstrably low.

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

For imputation servers, NIH should prioritize privacy-enhancing approaches and provide a clear operational pathway for academic or consortium-operated servers when they meet defined security and governance requirements, rather than restricting innovation only to federally operated servers. Specifically, NIH should allow non-federal (university/consortium) imputation services when they meet defined controls (e.g., strong access controls, isolation of controlled-access inputs, encryption, logging, secure execution environment, defined retention/deletion) and can demonstrate compliance via independent assessment/attestation, rather than conditioning eligibility solely on being funded/operated by NIH or another federal agency. NIH should also clarify how institutions can validate and document imputation-server-specific risks and mitigations.

76. Memorial Sloan Kettering Cancer Center

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Memorial Sloan Kettering Cancer Center

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
A key aspect of MSK's work to spur advances in the identification of somatic cancer mutations to drive treatment advances is our cBioPortal for Cancer Genomics, an open-source, open-access portal to cancer data. The cBioPortal has been available for more than 15 years and currently contains:

- somatic mutation data;
- processed /gene level calls for epigenomic, proteomic and transcriptomic data;
- de-identified tissue imaging data, e.g. H&E images and multiplex tissue images; and
- de-identified clinical and biospecimen meta-data.

MSK typically makes somatic variant data as well as de-identified clinical data available to the research community through supplemental tables in journal articles, as well as through the cBioPortal. Our open-access model currently serves more than 38,000 unique users per month with more than 30,000 citations, demonstrating the high scientific demand for open access cancer data. As such, we are particularly concerned about clarifying the application of the proposed changes to these types of data. Our comments on each of the questions posed in the NIH proposal follow:

Broad access to the type of data available in the cBioPortal for Cancer Genomics is fundamental to scientific progress. Data sharing programs funded by the NIH can and should be designed to both protect patient privacy and ensure broad research access. For cancer-specific data, we believe both aims can be achieved via a two-tiered system:

1. An open-access tier containing de-identified data, including somatic mutation calls, processed molecular data, de-identified tissue images, and de-identified clinical and biospecimen meta-data

2. A controlled access tier containing identifiable data, including germline data, raw genomic data, e.g., BAM files, and identifiable medical images

The NIH has been following this model for almost two decades with data generated by The Cancer Genome Atlas (TCGA) and other data sets hosted by the Genomic Data Commons (GDC); somatic variant data is freely available, as is de-identified clinical, molecular, and tissue imaging data, but raw sequencing data are available only through protected access. This open access policy has broadened access to critical cancer data to the entire scientific community and significantly improved our understanding of the molecular basis of cancer.

We urge the NIH to clarify the application of the proposed Genomic Data Sharing Policy to clarify that de-identified data, e.g., somatic mutation calls, may continue to be made available on an open-access basis.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
As currently worded, it is not clear if the NIH will distinguish somatic versus germline data or support distribution of de-identified data. If the new policy requires that all genomic data (somatic and germline) plus any form of de-identified data must be restricted to a controlled-access policy, the cBioPortal for Cancer Genomics as it currently exists would need to be ended, and a new controlled-access version of cBioPortal would need to be created. We believe this would fundamentally restrict broad access to cancer data.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
We ask for clarification on the following:

- Under the new guidance, will there be a distinction between somatic and germline data? And, will sharing of somatic data still be possible via open access mechanisms?

- Under the new guidance, will the sharing of de-identified gene level molecular data, e.g., epigenomic, proteomic and transcriptomic data, still be possible via open-access mechanisms?

- Under the new guidance, will the sharing of de-identified clinical and biospecimen meta data still be possible via open-access mechanisms?

- Under the new guidance, will the sharing of de-identified tissue images, e.g. H&E and multiplex images still be possible via open-access mechanisms?

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
In addition to our commentary above relating to the importance of maintaining open access to somatic cancer mutation and other de-identified clinical data, we note the following:

MSK commends the proposal to expand institutional review capacity by specifying that "institutions are permitted to appoint an individual or institutional body technically and legally capable of reviewing Institutional Certifications for data submission beyond an Institutional Review Board, Privacy Board, or equivalent." This will improve workflows and efficiency, while continuing to safeguard data submissions.

In addition, we seek clarification on the proposed informed consent changes. The current policy notes: "For studies proposing to use genomic data from cell lines or clinical specimens that were created or collected after the effective date of the Policy, NIH expects that informed consent for future research use and broad data sharing will have been obtained even if the cell lines or clinical specimens are de-identified. If there are compelling scientific reasons that necessitate the use of genomic data from cell lines or clinical specimens that were created or collected after the effective date of this Policy and that lack consent for research use and data sharing, investigators should provide a justification in the funding

request for their use. The funding IC will review the justification and decide whether to make an exception to the consent expectation."

Please clarify whether the ability to provide compelling scientific reasons for use that lack consent is being removed and replaced with the requirement to seek consent from next of kin. We would ask the NIH to consider maintaining its current policy of enabling researchers to request a waiver of informed consent due to compelling circumstances. Specifically, we ask the NIH to consider maintaining the option to request such a waiver in order access sample data from a deceased individual, rather than mandating the informed consent of the next of kin or legally authorized representative. The proposed approach may impose new burdens on families of deceased patients, while also limiting research.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
n/a

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/2026-3-17-MSK-Comments-on-NIH-RFI-on-Genomic-Data-Sharing-Policy-Revisions_final-signed.pdf

Description: Full comment letter

77. Data Management and Sharing Policy WG including ODS/DERT/DIR representatives, NIEHS/NIH

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Maria Shatz

Name of Organization: Data Management and Sharing Policy WG including ODS/DERT/DIR representatives, NIEHS/NIH

Type of Organization: Other

Type of Organization - Other: Government

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

- It is important to stress that the policy serves to support sharing sensitive information under appropriate controls and data use agreements rather than providing justification not to share. Human subject research funded by NIH should use common standardized consent language that promotes broad research data reuse under the policy.

- The statement "the following data types, listed below, must be protected throughout the data lifecycle" requires clarification – our understanding is that during collection, cleaning and analysis stages and after data download from CADR for secondary use that data will be stored internally with the investigator's institution with necessary strict protections but not requiring all the CADR controls, tracking, and auditing features. Perhaps, a table listing lifecycle stages and corresponding data protection requirements for each stage would be helpful.

- Exposing sufficient metadata for searching and for building virtual cohorts before requesting data access is essential. CADRs should maximize use of ontologies and standardized vocabularies to improve search.

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

- There are gaps in the use of NIH-supported repositories for epidemiology cohorts. Centralized support to enable existing CADRs to accommodate additional types of epidemiology cohorts (e.g., environmental epidemiology) would help maximize sharing rather than storing at grantee's institutional repository and making data available on request.

- To improve efficiency and unified approach and policy implementation it will be beneficial if CADRs are maintained and managed by the NIH rather than individual ICs or research institutions. If there are IC-managed domain appropriate CADRs we suggest opening them for the entire community and rethinking the sustainability and financial responsibility model. We want to avoid medium/small IC having to re-invent the wheel, duplicative investments, and uncontrolled proliferation of the list of CADRs.

- We need clarification on how different controlled-access data types generated in the same study will be linked (e.g. genomics and metabolomics). Will it be shared in the same repository and in this case, we need to expand repository capacity to multiple data types or if there will be linkage of multiple co-generated data types across different repositories.

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

- Please clarify that human pathogen data type is excluded from the list of protected data types to align with Department of Justice's final rule "Preventing Access to Americans' Bulk Sensitive Personal Data and United States Government-Related Data by Countries of Concern or Covered Persons" (28 CFR Part 202, § 202.224

- Please clarify whether human microbiome data type is included in the human 'omics data.

- Datasets shared via CADRs are required to be de-identified. Therefore, all Listed identifiers should be removed before submitting to CADRs and inclusion of Covered personal identifiers in the Protected data types in context of CADR is not necessary and confusing.

In context of data protection during the collection, cleaning, and analysis stages all Listed identifiers must be protected whether standalone or in combination including demographic or contact data that is linked only to other demographic or contact data (such as first and last name, birthplace, ZIP code, residential street or postal address, phone number, and email address and similar public account identifiers). Please revise and simplify this statement.

- New capabilities are required in NIH-managed CADR infrastructure systems to accommodate sensitive data like precision geospatial or temporal exposure data.

The policy designates precision geolocation data within 1,000 meters resolution (28 CFR 202.242) as a protected datatype. While this definition provides clarity, it may unintentionally incentivize researchers to degrade spatial resolution to avoid controlled-access requirements, thereby limiting downstream scientific utility.

High-resolution geolocation data are often essential for evaluating environmental exposures linked to health outcomes (e.g., wildfire smoke, flooding, urban heat, traffic-related pollution, and proximity to emission sources such as coal plants or CAFOs). While some exposures can be assessed at ~1,000-meter resolution, others—such as flooding, heat islands, and traffic pollution—require sub-kilometer precision (5–250 meters). Restricting shared data to coarse resolution may therefore significantly reduce reusability.

NIH should support technical solutions that preserve both privacy and analytic utility, such as:

- Depositing data at the highest resolution with tiered access to lower-resolution derivatives based on approved use

- Repository-based tools for geocoding, spatial aggregation, and resolution adjustment

Similar challenges apply to temporal data (e.g., date of birth, clinical encounter dates), which are often masked to reduce identifiability. However, precise dates are critical for studying time-sensitive exposures (e. g. symptom worsening, hospital visits due to an air or water pollution event, wildfire

smoke, release of toxic chemicals or an environmental disaster during vulnerable developmental windows). Additionally, date of birth is necessary for linking to external databases such as the National Death Index, cancer registries, Medicare, etc. Therefore, we need a differential approach to sharing these dates with secondary users depending on their research study needs.

NIH should enable a differential, use-based approach to sharing temporal data, allowing access to appropriately granular information under controlled conditions aligned with specific research needs.

A general comment: Data types proposed for controlled access under this policy (e.g., de-identified electronic health records (EHR), medical and pharmacy claims, genomic, clinical, imaging, and geolocation-linked data) are, in some cases, also available through commercial channels such as health data systems data brokers under less stringent access controls and oversight mechanisms. NIH and HHS should consider steps to reduce discrepancies between controlled-access research data and commercially available health-related datasets that contain comparable information types and re-identification risks to ensure that governance of sensitive data is risk-based and source-agnostic.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The simplification to 100 individuals is helpful, however it does not address longitudinal data or data from multiple experimental conditions (e.g., 25 participants/cell lines at 4 time points with and without an exposure, is that still 25 subjects or is it 200 subjects [25x4x2]). Please clarify distinction between individuals and data points.

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

78. Henry Chang, MD

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: Henry Chang, MD

Name of Organization:

Type of Organization: Not Applicable

Role: Other

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I am a physician-scientist who retired from NIH in 2019, and just happened to see this request. I may not be updated on the latest policies, but then we HSAs were supposed to ensure that data was usable and shareable, yet we did not have free access to do so ourselves. This was frustrating because we could not check publications or proposals against NIH databases for validation or refutation. Please be aware that excessive privacy can limit data curation that saves money and effort.

An example of this issue was the Systolic Blood Pressure Intervention Trial (SPRINT), a clinical trial of 9361 patients from 2010-2015 to find the best range to maintain BP. After it was published in the New England Journal, the data were shared with 200 qualified applicants to find new results (Learning What We Didn't Know - The SPRINT Data Analysis Challenge | New England Journal of Medicine (nejm.org). Without being asked, two teams found errors in the supposedly curated data, and one of them published a report (Accurate estimation of cardiovascular risk in a non-diabetic adult: detecting and correcting the error in the reported Framingham Risk Score for the Systolic Blood Pressure Intervention Trial population | BMJ Open. While there was no significant effect on the SPRINT results, one wonders what might have happened if all researchers had been asked to look for and report errors. Such requirements should be imposed when NIH databases are released, to assess their error rates and quality, and to inform prior users or publishers of potential flaws.

It is not clear from the draft policy if the shared data can be used by artificial intelligence (AI) because it is uncertain if processing will be private in cyberspace. If discussed elsewhere, a link should be provided, since this may determine if a grantee should bother to apply. While privacy may not be assured, AI could be useful within the NIH to find data errors and biases. Please let me know the current guidance on this. Thanks.

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

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

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

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

Description: Excessive privacy may limit data curation and underestimate errors

79. Human Pangenome Reference Consortium

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Lucinda Antonacci-Fulton

Name of Organization: Human Pangenome Reference Consortium

Type of Organization: Other

Type of Organization - Other: Research Consortium

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The HPRC mission is to create a pangenome reference that represents global human genomic variation and our work relies on the high-quality, telomere-to-telomere assembly of diverse genomes.

Informed Consent: HPRC already recognized the need for informed consent explicitly stating data are to be shared openly without controls and we continue to work with our institutions to determine that openly sharing these data pose very low risk when shared and used. The number of samples utilized in the pangenome resources is relatively low when compared with the whole of the population available for such research.

Simplify Thresholds: The 100 individual threshold is arbitrary. Instead the applied threshold should be interpreted by the researchers and their institution because they will have the project specifics and be in a position to best determine when the policy should be applied.

Data Access: For the pangenome, it is vital that the data remain in an unrestricted access model to maximize scientific utility while individual-level assemblies remain protected. These assemblies serve as a reference set for many research and clinic laboratories across the world and have been consented for open access data sharing and institutions have found that data pose very low risk when shared and used.

Submission Timelines: The HPRC notes the proposal to remove specific data processing levels and require data availability within 6 months of generation. While we support timely sharing, we emphasize that pangenome reference creation involves complex, iterative graph-based assemblies. We advocate for flexibility in the "6-month" window to ensure data undergoes rigorous Quality Control (QC) before broad release to maintain the reference’s high standards.

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

80. University of Illinois Chicago

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Joanna L. Groden, PhD

Name of Organization: University of Illinois Chicago

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/RFI-Controlled-Data-Access-Response-FINAL.pdf

81. University of Virginia

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: University of Virginia

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIHControlledAccessRFI_UVAResponse.pdf

Description: University of Virginia response

82. Association for the Accreditation of Human Research Protection Programs

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Nichelle Cobb

Name of Organization: Association for the Accreditation of Human Research Protection Programs

Type of Organization: Professional Organization/Association

Role: Other

Role – Other: Organization that accredits human research protection programs, including institutional review boards (IRBs)

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see the attached document for feedback.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Please see the attached document for feedback.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Please see the attached document for feedback.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see the attached document for feedback.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see the attached document for feedback.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIHcommentsGDS_final_aahrpp.pdf

Description: AAHRPP joint feedback with PRIM&R in response to Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

83. Endocrine Society

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Sophia Kaska

Name of Organization: Endocrine Society

Type of Organization: Professional Organization/Association

Role: Other

Role – Other: Government and Public Affairs staff

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The protection of human participant data is important for maintaining patient privacy while simultaneously allowing researchers to accelerate our understanding of human health and disease. While we appreciate that NIH is aiming to balance these needs, the goal and rationale for the specific policy changes are unclear to us and should be laid out in greater detail so that the research community can more accurately comment on the costs and benefits of the proposed changes. We note two significant issues that should be considered in the establishment of the policy:

1. Cost: Setting up and maintaining highly secure environments for human data is expensive; unless funding for these environments is provided, only highly resourced institutions will be able to adhere to this proposed policy.

2. Administrative Burden: These policies will create significant work for research teams and support staff related to ensure compliance with strict access standards and controls.

Because the policy may conflict with some aspects of the existing NIH data management and sharing policy, NIH should provide detailed guidance for how research teams can ensure compliance with both of these policies.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
In principle, established repositories provide opportunities for researchers to view and analyze data sets from different teams and even combine datasets to increase statistical power or make new discoveries. We appreciate that controlled access to certain datasets may be necessary; however, this would create barriers to access and collaboration that may discourage innovation and discovery. For example, many research teams develop and use custom-made, novel analytical tools to answer their research questions. It may not be possible to use these resources within a controlled environment, necessitating new approaches that may not be suitable to quickly address the specific biological question under investigation.

Regional and international collaborations may be disrupted if researchers are unable to access and combine datasets from different environments, especially for projects that are already underway. We are concerned about the ability of NIH-funded research teams to share data with other research teams that are not funded by NIH and may be generating, storing, or analyzing data governed by different

regulations and policies. Different access rules may create barriers to collaboration or integrated analysis which is necessary for areas such as rare disease where existing data sets are already small and challenging to collect.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
None.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
None.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Similar to the concerns expressed in question two, without policies that allow for copies of data to be transferred to other servers, any meaningful analysis of large data sets will ultimately be impossible for data that exists on different servers. This level of restriction, to only use servers that "are funded or operated by NIH or another federal agency" will hinder research efforts, particularly international collaborations, rather than promote access and collaboration.

Description: Endocrine Society's comments on the Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy.

84. Vivli

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Rebecca Li

Name of Organization: Vivli

Type of Organization: Other

Type of Organization - Other: Non-profit Data Respository

Role: Other

Role – Other: CEO

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Vivli agrees that NIH's investments in controlled-access data repositories (CADRs) are essential and timely particularly given the increasing privacy, security, and national security risks associated with controlled-access data. We recognize that the proposed policy is intended to promote responsible data sharing while strengthening protections. However, as currently drafted, the policy presents significant implementation challenges for non–NIH-funded repositories, including substantial start-up costs, complex security and compliance requirements, and the need for NIH-specific specialized operational expertise. Without a clear and attainable pathway for externally funded repositories to align with these expectations, there is a risk that data sharing will become increasingly concentrated within NIH-funded CADRs, reducing diversity, innovation, and flexibility in the broader data-sharing ecosystem. We encourage NIH to consider these potential downstream effects as part of its implementation planning, particularly given the important role that non-NIH-funded CADRs play in supporting NIH-funded investigators and long-term data stewardship.

Non-NIH-funded CADRs play a critical role in ensuring the long-term stewardship, accessibility, and reuse of NIH-funded research data. NIH budgets and funding priorities necessarily evolve over time, and reliance on time-limited grants alone is not a sustainable model for the ongoing preservation, governance, and secure access of clinical research data across decades. Externally funded repositories, including nonprofit and mixed-funding models, often provide continuity, infrastructure stability, and dedicated expertise that extend beyond individual funding cycles, while still supporting NIH-funded investigators in meeting data-sharing expectations. We encourage NIH to consider these potential downstream effects as part of its implementation planning, particularly given the essential role that non-NIH-funded CADRs play in maintaining a resilient, diverse, and sustainable data-sharing ecosystem that supports both current research needs and long-term public benefit.

For non-NIH-funded CADRs, NIH-funded data typically represents only a subset of a broader and more diverse data catalogue, which may include data supported by other U.S. agencies, international funders, foundations, industry sponsors, or public–private partnerships. As a result, policy changes specific to NIH-funded data are not uniformly applicable across the entire repository, unlike in NIH-funded repositories where the vast majority of hosted data is generated under NIH awards and governed by NIH policy frameworks. Requiring non-NIH-funded CADRs to substantially redesign their technical

infrastructure, governance models, and operational workflows to meet a specific set of NIH standards for a minority of datasets creates a disproportionate burden on the repository. Without flexibility or phased approaches, this expectation risks discouraging repositories from accepting NIH-funded data altogether, despite their long-standing role in supporting NIH-funded investigators and facilitating responsible data sharing at scale.

Vivli requests additional clarity regarding NIH's expectations and support mechanisms for repositories operating outside of NIH-funded CADRs. Specifically, it would be helpful for NIH to clarify whether and how it will formally recognize or designate repositories that meet controlled-access requirements but are not NIH-funded, and whether additional funding opportunities, technical assistance, or other support mechanisms will be made available to help repositories implement these policy requirements and scale to meet anticipated increases in user demand. In addition, Vivli encourages NIH to establish a realistic and consultative implementation timeline, informed by direct engagement with non–NIH-funded repositories, to ensure that compliance expectations are achievable without disrupting existing data-sharing activities that serve the community. Finally, more robust resources to facilitate connection and coordination between non-NIH-funded CADRs and the NIH-funded data ecosystem such as guidance, forums, or designated points of contact would support consistent implementation and help preserve a diverse, interoperable, and sustainable controlled-access landscape.

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

Vivli is a U.S. nonprofit, generalist clinical research repository that also operates a secure analysis environment through which approved users analyze individual participant data (IPD) from Vivli, non-NIH sponsors, and NIH-funded investigators. Because we work with multiple funders and data contributors, NIH-funded studies comprise only a subset of our catalogue, which makes repository-wide alignment to NIH-specific requirements operationally complex and, without support, cost-prohibitive for a self-sustaining model.

We support NIH's aims to harmonize expectations and strengthen protections for human participant data; however, implementing the proposed requirements at an established, externally funded repository will require substantial effort across: (1) software development and system hardening; (2) information-security governance, risk, and compliance; and (3) data governance, access workflows, and agreement updates. Several elements in the draft policy and its referenced materials are written specifically for NIH-funded CADRs (e.g., reliance on NIH Data Access Committees and NIH Developer Access Committees), which do not map cleanly to externally funded repositories. We therefore request that NIH revise and re-issue the referenced security/operational standards (currently framed for NIH CADRs) in a form that is not NIH-program specific, so non-NIH repositories can implement them without ambiguity (e.g., updating guidance included in NOT-OD-24-157 and related notices cited in the RFI).

Clarity and support needed for practical implementation:

1. Recognition/designation pathway. Please clarify how NIH will recognize or designate repositories that meet controlled-access expectations outside of NIH-funded CADRs, including expectations for identity proofing, access review, restrictions related to countries of concern, and security control baselines. This is especially important because the draft policy lists minimum controls (e.g., prospective access review,

requester authentication, restrictions tied to 28 CFR Part 202, and security standards such as NIST SP 800-171 or equivalent) without specifying how non-NIH repositories demonstrate equivalency in practice.

2. Equivalency and attestation guidance. Please provide explicit guidance on what documentation NIH will accept to evidence compliance or equivalency (e.g., third-party audits/certifications), and how "attestations" by Approved Users and institutions should be handled, including recognition of ISO/IEC 27001/27002 as an accepted international equivalent where appropriate. Clear, uniform procedures will reduce risk for repositories asked to rely on depositor and user attestations

Funding and technical assistance. We request additional funding or support mechanisms (e.g., targeted grants, cooperative agreements, technical assistance) to help externally funded repositories implement required controls and scale for increased demand anticipated under the new policy. The RFI acknowledges expanded controlled-access practices and references harmonized oversight; extending comparable resources beyond NIH-funded CADRs would enable sustainable adoption by the broader ecosystem.

3. Realistic, consultative timeline and on-ramp. Establish a phased implementation timeline, developed in consultation with non-NIH repositories, with clear milestones, transition periods, and practical "on-ramp" supports (implementation guides, FAQs, workshops/listening sessions, pilot programs, and designated NIH points of contact). NIH provided structure and resources when NIH-funded CADRs implemented prior changes. If NIH provides similar structure to external repositories it will minimize disruption and reduce the risk of data withdrawal or sharing delays during transition.

Taken together, these actions would allow established, externally funded CADRs to implement the proposed policy credibly and sustainably, preserving investigator choice and ecosystem diversity while advancing NIH's goals for responsible, secure, and high-value data sharing.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: It would be valuable for NIH to clarify whether all levels of clinical research individual participant data (IPD) (e.g. directly identifying data, to de-identified data, to fully anonymized datasets) are uniformly considered controlled-access data and therefore required to be managed within a CADR subject to the NIH CAD policy. Clear guidance on this point would help investigators, institutions, and repositories make consistent determinations, align informed consent language appropriately, and avoid unnecessary over-restriction of low-risk data.

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

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

85. VALERIE A ARBOLEDA

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: VALERIE A ARBOLEDA

Name of Organization: UCLA

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I strongly recommends that NIH establish clear transition provisions for legacy datasets collected prior to the policy's effective date. Many previously collected datasets do not contain explicit language authorizing open sharing without controls. Retroactive consent is infeasible under the Common Rule. NIH should provide a safe harbor for pre-policy datasets and clearly articulate how such data will be treated under the new framework.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
This is confusing as to which repository fulfills this requirement.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
While I understand the need for harmonizing and clarifying protected data types, but as a physician scientist I am very concerned with expansion and burden that this places on smaller labs and research groups that focus on rare diseases. Our patient community advocates are often clamoring for their data, often from cell lines and for RNA-sequencing and other multi-omic modalities) or derivitive lines, to be shared with the broader community. They in fact dislike all the barriers that are in place for "protection" that in fact hinder the easy sharing of their data sets. Increasing the scope of protected data that is controlled acess will limit the fundamental access and discovery that is the basis of science. While protected data types should include ones that include obvious identifiers, items like MRI images or RNA-sequencing and proteomics data from cell lines (both established and patient derived) should not be included. These cannot be linked to individuals based on the limited data in the samples alone and would create mass confusion and increased costs to data analysis. Training the next generation of scientists and bioinformaticians will be much more difficult if all of this happens in a controlled access environment that is more costly and wont allow for training of student scientists.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
I. support the proposal to limit the GDS Policy to human data and to harmonize it with the broader NIH Data Management and Sharing Policy. This revision will be enabling researchers to focus resources on responsible data stewardship rather than navigating potentially redundant requirements. There is significant redunancy in these documents.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
I support NIH's interest in modernizing governance of imputation servers. We recommend that NIH permit Approved Users to operate institutional imputation servers, provided that:

- Adequate technical safeguards are demonstrated.
- Security controls align with NIH Best Practices.
- Robust access review and monitoring mechanisms are implemented.

86. Population Association of America/Association of Population Centers

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Mary Jo Mitchell

Name of Organization: Population Association of America/Association of Population Centers

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/PAA-APC-comments-re-NIH-data-policy-3-26.docx

87. Adolescent Brain Cognitive Development Study

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Rebekah Huber

Name of Organization: Adolescent Brain Cognitive Development Study

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ABCDesponse-to-NIH-RFI_022126_v2.docx

88. American Society of Human Genetics

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Karina Miller

Name of Organization: American Society of Human Genetics

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the 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.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ASHG_NIH-Genomic-Data-Sharing-Policy-RFI_03.17.2026-FINAL.pdf

Description: Please see the attached letter from ASHG for our full response to this RFI.

89. The Global Alliance for Genomics and Health

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Gemma Brown

Name of Organization: The Global Alliance for Genomics and Health

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The draft NIH Controlled-Access Data (CAD) Policy represents an important step toward balancing participant privacy, national security considerations, and scientific utility. The Global Alliance for Genomics and Health (GA4GH) strongly supports the goal of harmonizing controlled-access requirements across NIH and implementing risk-based frameworks that protect participants while enabling responsible data sharing.

However, several clarifications and refinements would strengthen the policy and its implementation.

Scope and international collaboration:

The proposed CAD Policy applies broadly to designated protected data types wherever the repository is located globally. As a result, the policy may apply to NIH-funded projects across the international research community regardless of whether collaborators are located in countries identified as "countries of concern" (CoC). In contrast, the Department of Justice Final Rule implementing Executive Order 14117 focuses restrictions specifically on access by covered persons associated with designated CoCs. Because the CAD Policy appears to operate more broadly than this rule, it may introduce constraints that exceed existing U.S. legal requirements and could unintentionally slow international scientific collaboration. Clarifying how the policy aligns with the DOJ rule would help avoid unnecessary barriers to global research partnerships. It may also be helpful for the NIH to reference lifecycle protections (e.g., controlled access, stronger repository security, enhanced auditing).

Equivalence of consent and oversight frameworks:

Given the global nature of genomic research, NIH should clarify whether consent frameworks developed under other regulatory systems may be recognized where participant protections are substantively equivalent. This recognition would benefit from a clear and transparent process for assessing equivalence to ensure participant protection and governance are meaningfully comparable. Similarly, confirmation that ethics oversight bodies operating under non-U.S. regulatory frameworks may fulfill institutional certification requirements—when competency and accountability standards are met—would support international collaboration while maintaining appropriate protections. This could be achieved through the existing Federal Wide Assurance (FWA) program.

Timely and efficient access:

Controlled access should maintain strong protections while minimizing unnecessary administrative burden. To support efficient data sharing, the policy should encourage:

- Standardized and predictable review timelines for data access requests

- Minimization of duplicative Data Use Agreements across NIH repositories

- Support for federated and cloud-based analysis models that reduce data movement while preserving oversight

Clear guidance for international collaborators, particularly institutions in low- and middle-income countries (LMICs), would also help ensure that controlled-access requirements do not inadvertently disadvantage researchers operating in resource-limited environments.

AI-related privacy risks:

Advances in artificial intelligence introduce new re-identification risks that are directly relevant to controlled-access data, including model inversion, membership inference, multimodal linkage, and leakage from trained models. NIH may wish to explicitly acknowledge these risks within the policy and consider safeguards such as secure compute environments that allow analysis without data export. It may also be useful to address AI-derived artifacts, and clarify how these outputs should be handled, shared, or retained. Guidance on whether controlled-access datasets may be used for AI model training—and under what conditions trained models may be shared—would also improve clarity.

Machine-readable compliance guidance:

Researchers frequently face challenges interpreting complex policy language when implementing data-sharing requirements. Publishing machine-readable versions of policy provisions—such as structured definitions and policy rules—would enable the development of automated compliance tools that assist researchers in identifying applicable requirements and permissible uses. This approach could reduce compliance burden and improve consistent interpretation of policy requirements.

Sharing of genomic summary statistics:

The policy should also clarify that genomic summary statistics—including allele counts for extremely rare variants—can be shared. Many variants discovered in large population datasets occur only once or a small number of times, yet they may have important functional or clinical implications. The privacy risk of sharing low-frequency or single-count variants is not uniform, given other factors such as cohort size or ancestry can influence de-identification risk. Applying a context specific risk assessment could help maintain participant protection. Restrictions that obscure allele counts below certain thresholds can significantly limit the scientific utility of genomic datasets. Evidence from large genomics initiatives suggests that the privacy risks associated with sharing such summary statistics are low, while the benefits for variant interpretation and discovery are substantial.

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

NIH-designated repositories provide important infrastructure for implementing controlled-access data sharing. However, additional guidance on the use of federated data networks would be beneficial. Federated approaches allow researchers to query distributed genomic datasets without requiring centralized data movement. This model supports both privacy protection and international

collaboration, particularly when data cannot legally or ethically be transferred across jurisdictions. Clear guidance on how federated repositories or distributed query systems may satisfy CAD policy requirements would help enable modern data-sharing architectures while maintaining appropriate oversight.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: The protected data types identified in the draft policy—including genomic and multi-omics data, detailed phenotypic data, clinical datasets, and facial imaging—are appropriate candidates for controlled access because they may pose re-identification risks.

Several refinements could further strengthen the framework.

Clarification of definitions and thresholds:

Some definitions would benefit from greater specificity. For example:

- For omics datasets, clearer criteria for what constitutes “systems-level analyses” would support consistent implementation.

- Facial or head imaging definitions should explicitly include high-resolution 2D photographic, 3D camera surface models, as well as high-resolution 3D anatomical, diffusion, and functional imaging using MRI, which are increasingly vulnerable to AI-based re-identification.

The policy should also clarify how thresholds such as the proposed 100-participant “large-scale” definition interact with data sensitivity, as fixed thresholds may not consistently reflect risk across different data types. A complementary risk-based assessment approach may also be useful.

Risk-tiered governance

A risk-stratified model may help distinguish between:

- Highly identifiable individual-level datasets (e.g., whole genome sequencing, facial imaging, MRI, 2D photographic, and/or 3D camera surface models)

- Context-dependent datasets

- De-identified aggregate or summary datasets

This approach would enable governance proportional to the re-identification risk while preserving scientific utility.

Multimodal linkage risks:

The policy should acknowledge that datasets that appear low-risk individually may become identifiable when combined. AI-enabled multimodal linkage across genomic, phenotypic, biometric, or imaging datasets can significantly increase re-identification risk. The policy should also address whether the HIPAA Expert Determination remains sufficient, given the difficulty for experts to determine that the risk

of re-identification is small when faced with AI-enabled linkage. Explicit guidance on such combined modalities would improve risk management.

Community-level considerations:

Genomic datasets may pose risks not only to individuals but also to communities or populations represented in the data, including families, ancestry groups or geographically limited populations. Such impact can consist of discrimination, stigmatization, and targeting of communities. Explicit recognition of potential group-level harms would strengthen the policy's ethical framework.

Public health and pathogen genomics:

The policy should clearly distinguish human genomic data from pathogen genomic datasets used for public health surveillance. Microbial whole-genome sequencing datasets without patient identifiers typically carry substantially lower re-identification risk. Applying overly restrictive controlled-access requirements to such datasets could slow outbreak response and antimicrobial resistance surveillance. Having a flexible risk-based approach that permits urgent access to public health use cases can better promote safety, privacy, and rapid scientific progress.

Query access for clinical genomic data:

Limited query-based access to genomic data without direct identifiers may also support important clinical applications, such as rare disease variant interpretation. For example, allowing researchers to query the presence of specific variants and receive minimal contextual information (e.g., disease cohort or study context) could facilitate diagnosis while preserving privacy protections. To avoid misuse, such systems should be designed with clear boundaries and measures, including strong logging and authorization procedures, to ensure these interfaces do not allow reconstruction of the underlying data. Guidance from NIH on such models would be valuable, particularly as clinical laboratories increasingly seek scalable approaches to responsible genomic data sharing.

Automated data classification:

Finally, automated systems to assist with the classification of protected data types at the time of repository submission could reduce human error and improve consistent designation of controlled-access status across datasets.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The proposed revisions to the NIH Genomic Data Sharing (GDS) Policy appropriately modernize the framework in light of expanding genomic technologies and increasingly integrated multi-omics datasets. Alignment with the NIH Data Management and Sharing Policy also reduces redundancy and improves policy coherence. Several clarifications may further strengthen the revisions.

Consistency in thresholds and data types:

The revised policy identifies datasets containing genomic data from more than 100 individuals as "large-scale." While this threshold provides clarity for human genomic datasets, the policy does not specify equivalent thresholds for other omics data types such as epigenomic, proteomic, or transcriptomic

datasets. This lack of consistency could create ambiguity in implementation. NIH may wish to clarify how these data types should be treated under the policy.

Flexibility in data-submission timelines:

The proposed six-month timeline for submitting data to repositories is generally reasonable. However, flexibility may be necessary for complex multi-omics datasets or projects with significant infrastructure constraints. Allowing justified extensions where appropriate would support data quality and responsible submission practices.

AI-related considerations:

Given the increasing use of genomic and multi-omics datasets in machine learning applications, consent language and policy guidance should explicitly address the use of data for AI model training and the potential downstream sharing of trained models. This could also clarify how outputs from these analyses should be handled, since they can carry residual information about the data.

Recognition of international governance structures:

To support global collaboration, the policy should clarify whether substantively equivalent de-identification standards outside the HIPAA framework may be recognized and whether non-U.S. oversight bodies may satisfy certification requirements when appropriate standards are met.

Structured and consolidated policy definitions:

Policy requirements related to genomic data sharing are currently distributed across multiple NIH notices and documents with overlapping terminology. Publishing a consolidated and versioned reference for key definitions and requirements would significantly improve clarity for researchers and institutions implementing compliance workflows.

Equitable global collaboration:

Genomic research increasingly relies on global datasets and partnerships. The revised policy should encourage equitable data governance principles in international collaborations, including recognition of joint data stewardship models and support for infrastructure development in partner countries. Ensuring that contributing institutions—particularly in LMICs—can participate meaningfully in downstream analyses and publications would strengthen the fairness and sustainability of global genomic data sharing.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Imputation servers provide important analytical capabilities for genomic research, but they can introduce privacy risks if individual-level genotype data are uploaded without appropriate safeguards.

The proposed updates represent a positive step toward clarifying operational expectations. Additional guidance could further strengthen implementation.

Security and privacy protections:

NIH should require strong security standards for imputation servers, including encryption of uploaded data, robust authentication mechanisms, and detailed audit logs. Privacy-enhancing technologies—such as secure enclaves, differential privacy, or secure multiparty computation—may further reduce risks associated with data processing. It could also be of benefit to indicate that these servers should operate in governed and monitored environments that reduce exposure of the uploaded data and the results.

Server-side computation and data minimization:

Encouraging server-side computation models that minimize exposure of raw genotype data would help reduce privacy risks. Secure cloud-based environments where imputation occurs within controlled infrastructure may be preferable to models requiring users to upload sensitive datasets to external servers.

Governance and compliance clarity:

Clarification on security standards, compliance assessment criteria, and expectations for internationally governed infrastructures would assist researchers operating imputation platforms.

Additionally, clear policies should address whether imputation panels or models derived from controlled-access data may be exported or reused, as such models may encode sensitive information. These policy considerations should also aim to integrate stewardship, scientific openness, and participant protection.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/GA4GHs-Response-to-the-NIHs-Request-for-Information-.pdf

Description: The above responses with acknowledgements and individual technical responses.

90. UCLA HVP VCC

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Yvonne Kapila

Name of Organization: UCLA HVP VCC

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

• Proposed Data Management Sharing (DMS) Plan updates points towards controlled access databases for virome and microbiome data. Filtering human (host) and virome/microbiome sequencing data is critical.

• A complicating factor for virome characterization is the handling of Human endogenous retroviruses (HERVs) – these viral sequences are integrated into the human genome. Metagenomic data informatics are designed to remove human sequences prior to analyses, however, guidelines as to how to handle metagenomic sequencing data are critical as they generally include human host sequences. Clear distinctions between human and non-human sequences are required so that data sets can fit into a DMS plan.

• The requirement that data are to be deposited immediately into NCBI may be problematic due to time needed for analyses, publications, etc. Setting a release date is helpful, but with large scale data production and characterization studies, this may also be challenging due to the need to integrate large volumes of data generated over longer periods of time.

• Non-NIH/NCBI databases for sequencing data deposition and storage should require considerable vetting and scrutiny to ensure that human data are secure and not susceptible to data leaks.

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

• Non-NIH/NCBI databases for sequencing data deposition and storage should require considerable vetting and scrutiny to ensure that human data are secure and not susceptible to data leaks.

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

• Proposed Data Management Sharing (DMS) Plan updates points towards controlled access databases for virome and microbiome data. Filtering human (host) and virome/microbiome sequencing data is critical.

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

• A complicating factor for virome characterization is the handling of Human endogenous retroviruses (HERVs) – these viral sequences are integrated into the human genome. Metagenomic data informatics are designed to remove human sequences prior to analyses, however, guidelines as to how to handle metagenomic sequencing data are critical as they generally include human host sequences. Clear

distinctions between human and non-human sequences are required so that data sets can fit into a DMS plan.

• The requirement that data are to be deposited immediately into NCBI may be problematic due to time needed for analyses, publications, etc. Setting a release date is helpful, but with large scale data production and characterization studies, this may also be challenging due to the need to integrate large volumes of data generated over longer periods of time.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/DMS-Update-comments-UCLA-VCC.docx

91. University of Utah

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Lisa Young

Name of Organization: University of Utah

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
(See our uploaded PDF submission)

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
(See our uploaded PDF submission)

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
(See our uploaded PDF submission)

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
(See our uploaded PDF submission)

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
(See our uploaded PDF submission)

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/U-Utah-Response-to-NIH-RFI-NOT-OD-26-023-2026-03-17.pdf

Description: Our Campus-wide response (uploaded) was authored by 9 University officials and faculty indicated in the cover letter

92. Yale University

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Yale University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Yale-response-to-NIH-NOT-OD-26-023.pdf

93. Ellen Wright Clayton, Camille Nebeker, Joseph Yracheta

Submit date: 3/17/2026

I am responding to this RFI: On behalf of myself

Name: Ellen Wright Clayton, Camille Nebeker, Joseph Yracheta

Name of Organization: Vanderbilt University Medical Center, University of California San Diego, Native BioData Consortium

Type of Organization: Other

Type of Organization - Other: 2 Academic Institutions and Nonprofit organization

Role: Other

Role – Other: Investigators and advocates

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Clayton-Nebeker-Yracheta-Response-to-RFI-final.docx

Description: summary of our response on several issues

94. Coalition for Academic Scientific Computation (CASC)

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Kathryn Kelley

Name of Organization: Coalition for Academic Scientific Computation (CASC)

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Coalition for Academic Scientific Computing (CASC) appreciates the opportunity to provide input to the National Institutes of Health on changes to the Controlled-Access Data Policy and Genomic Data Sharing Policy.

CASC strongly supports the goals of the RFI, specifically harmonizing and clarifying expectations for protecting human subject data in light of increasing concerns about participant privacy. However, in the spirit of constructive partnership we hope to identify areas of the policy that could benefit from further refinement.

Broadly, CASC has concerns about the scoping within the Draft NIH Controlled-Access Data Policy. Both the "Scope and Applicability" and "Requirements" sections indicate an extension of Controlled-Access Data protections to "all NIH-supported research generating human data … throughout the data lifecycle." With NIST SP 800-171 now being the standard for protecting data from the Controlled Access Data Repositories, the implication is that all listed data types will now be subject to the NIST SP 800-171 controls from the point that they are first produced. While CASC applauds NIH's efforts to ensure the protection of identities and sensitive medical information of individuals that take part in NIH funded research projects, the proposed policy raises serious practical concerns. The listed data types are widely distributed within the information systems operated by CASC member institutions as part of clinical, translational and basic research workflows.

Specific controls within the NIST SP 800-171 control set are poorly aligned with some environments. For example, NIST SP 800-171r2 control 3.10.3 requires that visitors be escorted within the controlled environment. In a clinical setting where research data collection is intermingled with patient care, this control cannot be feasibly met without disrupting clinical operations. This impact seems contrary to NIH's aim of accelerating translational successes. We ask that NIH explicitly define the protection measures they would like to see at each stage of the data lifecycle, so that institutions can accurately access and comment on the operational impacts.

While NIST SP 800-171 is a recognized baseline for controlled access data, the standard is focused on the types of information systems found in a classic office setting, or their virtualized equivalent in the cloud. This focus has made it challenging to develop large scale computational systems, commonly called High Performance Computing (HPC) or High Throughput Computing (HTC) systems, that meet the letter of these controls. CASC members are reporting increasing demands from researchers for access to HPC/HTC resources as they look to work with Controlled-Access Datasets at the systems level.

We respectfully urge NIH to explore NIST SP 800-234 as a security overlay for Controlled Access Datasets to provide guidance on how to implement NIST 800-171 controls in HPC/HTC environments. NIST SP 800-234, as an overlay to the NIST SP 800-53 moderate baseline, provides guidance for implementing compensating controls that impact the high-performance mission of HPC systems, Since NIST 800-171 is a tailoring of NIST 800-53 moderate, the guidance from NIST 800-234 applies.. The meaning over tailoring and overlay is defined in the NIST Risk Management Framework (RMF), see https://csrc.nist.gov/projects/risk-management/about-rmf.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Secure workspaces designed for CAD users intentionally restrict external connections to safeguard CAD. However, the CAD Repositories (CADR) documentation for accessing data does not account for these secure system's architecture, or clearly provide the technical details required to establish connections from isolated environments (e.g. S3 bucket information, documented endpoints, HTTP URL, etc). Further, the quality of documentation varies significantly from CADR to CADR and is often outdated, particularly with the January 2025 shift in guidance. Setting common expectations/policies across NIH Institutes, as well as modernizing and standardizing documentation, will help researchers access the CAD they are authorized for and prevent researchers from trying to access the data outside of a secure environment.

Additionally, CADR should explore using industry standard services for transferring data between secure systems to ensure CAD are protected during the project's lifecycle. 75 Regulated Research Community of Practice (RRCoP) members have signed a letter to the NIH requesting they explore Globus for CADR due to its high adoption by research institutions.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
CASC Members have expressed frustration with the current level of guidance on what level of summarization is needed to allow for sharing of data (and publication). Guidance concerning the line at which data transitions from controlled to not controlled would help open up more and more efficient computational capacity to researchers who can move more intensive calculations in a lower security zone. Practical examples or templates of summary tables would be helpful.

While the definitions used for data types are defined in the CFR, the use of these definitions verbatim raises challenges. The definition for "Genomic Data" does not include the statement requiring 'a systems level analysis' that is found in proteomic, transcriptomic and epigenomic definitions. The genomic data definition will pull data into scope beyond what is appropriate for coverage under the CAD policy, the transcriptomics data definition excludes data that is likely to be sensitive, both because transcripts can present highly similar information to genomic data, and because the description appears to carve out data collected without a clear experimental design (under specific conditions or specific cell lines).

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
CASC has no comments with regards to this element of the request for information.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
CASC has no comments with regards to this element of the request for information.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH-Controlled-Access-Data-Policy-and-Proposed-Revisions-to-NIH-Genomic-Data-Sharing-Policy-031826.pdf

95. Emory University

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Kimberly Eck

Name of Organization: Emory University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attachment.

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/RFI-NIH-Data-Emory-Response-March-2026.pdf

96. ICPSR

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: John Kubale

Name of Organization: ICPSR

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH_RFI_response_final.pdf

97. Cleveland Clinic

Submit date: 3/17/2026

I am responding to this RFI: On behalf of an organization

Name: Lara Jehi, MD

Name of Organization: Cleveland Clinic

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please refer to Cleveland Clinic's attached letter which contains comments on the proposed revisions to the Genomic Data Sharing Policy.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/3_17_2026-NIH-Data-Policy-RFI_Cleveland-Clinic-final.pdf

Description: Cleveland Clinic comment letter in response to the NIH's data policy RFI and proposed revisions to the Genomic Data Sharing Policy.

98. See statement for full list of names

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: See statement for full list of names

Name of Organization: Department of Biological Chemistry, University of California, Los Angeles

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The faculty of the Department of Biological Chemistry in the David Geffen School of Medicine at the University of California, Los Angeles discussed the "Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy" (NOT-OD-26-023) at a faculty meeting. We, the undersigned faculty, have also reviewed a draft of the official University of California response and agree with the points raised in that letter. While we appreciate the importance of enhancing NIH policies around controlled-access data and data sharing, we also ask that NIH carefully evaluate for each policy change whether it could lead to unintended or disproportionate additional financial or administrative burdens on researchers relative to its expected benefits, as such burdens could risk slowing down discoveries that advance human health.

Jason Ernst, Professor of Biological Chemistry, Computer Science, and Computational Medicine

Siavash Kurdistani, MD Professor and Chair

Kathrin Plath, Professor

Avi Samelson, Assistant Professor

Aparna Bhaduri, Assistant Professor

Thomas Vallim, Associate Professor

Peter Tontonoz, Professor

David Eisenberg, Professor

Gabriel H. Travis, M.D. Professor of Ophthalmology and Biological Chemistry

Feng Guo, Professor

Gregory S. Payne, Professor Emeritus

Debora Sobreira, Assistant Professor

Keriann Backus, Associate Professor

Heather Christofk, Professor

Rosalie Lawrence, Assistant Professor, Biological Chemistry

Alexander van der Bliek, Professor

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

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

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

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

99. The Ohio State University

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Stephen Gavazzi; John Horack; Damon Jaggars

Name of Organization: The Ohio State University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/20260316_NIH_RFI_response.docx.pdf

100. Jessica Turner

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Jessica Turner

Name of Organization: The Ohio State University

Type of Organization: Academic Institution

Role: Institutional Review Oversight Committee Member

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
This proposed change to the data access policy is overkill. Human subjects research data is not a national security issue. Clearly, participant confidentiality and privacy need to be protected--and they are. This policy change is not addressing any known problem with that.

What this policy change as written would do is irreparable harm to the large scale, global research consortia which are critical for moving our understanding of human disease and treatment development forward. Re-analysis of existing data is a key part of human research--it allows us to develop new algorithms, test AI approaches, and train the next generation efficiently. Putting additional barriers in the way of that data sharing, is unnecessary and wasteful of resources. Open access data repositories are already carefully curated to protect individual participant privacy and confidentiality. Data which cannot be anonymized is already not being shared, but being stored in approved repositories under specific policies to protect all parties. This policy change would not improve any aspect of security, confidentiality, or benefit to the common good.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Human neuroimaging using Magnetic Resonance Imaging (MRI) is one of the richest resources we have for understanding mental health and resilience. Scientists and privacy experts have spent the past twentyfive years figuring out ways to make it possible, easy, and efficient to share de-identified neuroimaging data, with no damage to the participants or the US government and national interests. MRI falls under the images of the head or face, as the policy is now written. The implementation of this policy as written would set us back decades in our speed of evaluating and reproducing reliable findings, with no benefit to anyone from that reduction in productivity.

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

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

101. Public Responsibility in Medicine and Research (PRIM&R)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Daniel McLean

Name of Organization: Public Responsibility in Medicine and Research (PRIM&R)

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached.

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

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see attached.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/031726NIHcommentsGDS_AAHRPP_PRIMR_FINAL_LH_IRT-Signed_03.17.26.pdf

Description: Comments from Public Responsibility in Medicine and Research (PRIM&R).

102. N/A

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization:

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Retina-scans-as-biometric-data-signed.pdf

103. Steven Joffe

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Steven Joffe

Name of Organization:

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
We write in our individual capacities to comment on NOT-OD-26-023, which proposes the creation of an NIH Controlled-Access Data Policy. The proposed policy applies to "all NIH-supported research generating human data or deriving data from human data, cell lines, or biospecimens," and describes 11 categories of data that fall within its scope. Notably, although microbiome or virome data derived from individual human subjects would seem to be in the proposed policy's scope because they are derived from human biospecimens, the draft does not mention them, leaving their status under the policy ambiguous. We believe these data categories should be added to the list of data types that explicitly fall under the Controlled-Access Data Policy.

The 11 categories include several types of data, such as geolocation data, personal health data, and personal financial data, that are not measures of human biology but rather are data derived from individual human subjects that may be sensitive and pose informational risks if reidentified or misused. Microbiome and virome data derived from individual humans should be handled in the same way. These data may reveal that a subject has (or had) an infectious disease, such as a sexually transmitted infection, that may be stigmatizing. Alternately, they may reveal that a subject has a condition that could have implications for insurability, employment, or other entitlements or services. In this way, microbiome and virome data are no different than personal health data, which explicitly fall within the proposed policy.

Beyond their potential consequences for individuals, microbiome and virome data could cause group harms if, for example, analyses suggest that certain stigmatizing infections are more common among some groups than among others. We note that NOT-OD-26-023, in introducing the proposed Controlled-Access Data Policy, is concerned with the possibility of group as well as individual harms. Microbiome and virome data pose a high risk of group harm and therefore belong within the policy's scope.

For the reasons above, we urge addition of microbiome and virome data derived from human subjects to the list of categories explicitly covered by the proposed Controlled-Access Data Policy.

Steven Joffe, MD, MPH

Art and Ilene Penn Professor and Chair of Medical Ethics and Health Policy

Professor of Pediatrics

University of Pennsylvania Perelman School of Medicine

joffes@upenn.edu

Holly Fernandez Lynch, JD, MBE

Associate Professor of Medical Ethics and Health Policy

University of Pennsylvania Perelman School of Medicine

Associate Professor of Law

Penn Carey Law School

lynchhf@pennmedicine.upenn.edu

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

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

104. American Academy of Ophthalmology

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Rachael George

Name of Organization: American Academy of Ophthalmology

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The American Academy of Ophthalmology (the Academy) appreciates the opportunity to submit comments in response to the National Institutes of Health (NIH) Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to the NIH Genomic Data Sharing Policy. The Academy is the largest association of eye physicians and surgeons in the United States. A nationwide community of over 20,000 medical doctors, we protect sight and empower lives by setting the standards for ophthalmic education, supporting research, and advocating for our patients and the public. We innovate to advance our profession and to ensure the delivery of the highest-quality eye care.

Comments on NIH Controlled-Access Data Policy

The Academy commends NIH for seeking stakeholder input on these policies, as they have the potential for wide-reaching impact on the research community. We have significant concerns about the implications of the implementation of the new NIH Controlled-Access Data policy on data access. As the policy is currently written, there could be no sharing without the requirement of individual patient consent for each human data element specified in the policy. This would impose an impossible requirement upon researchers and institutions for past and current NIH-supported research involving the potential re-identification of previously de-identified data elements. This would require the identification of contacts and the development of a system to secure consent from tens of thousands of patients who had contributed data. This would severely constrain collaborative research activities across institutions without the ability to pool and share information that has been considered very low risk when shared or used. This would result in stalling advances in artificial intelligence across medicine, where such collaboration and sharing of information is vital.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Retinal Scan Inclusion in the Definition of Biometric Identifiers

The Academy wishes to outline concerns with the inclusion of retinal scans in the definition of biometric identifiers. We believe that there is a general misunderstanding and confusion which has lead to retinal scans being considered similar to iris scans. While there are databanks that link iris scans with individuals, there is no existing databank that links retinal scans to the individual, and thus, it cannot be a biometric identifier. As the preeminent global research organization, it would be incumbent upon the

NIH to be clear on the exclusion of retinal scans from biometric identifiers because there are no practical means for re-identification, and thus, these data should be considered very low risk when shared or used.

Because of the misunderstanding and confusion leading to the categorization of retinal scans in the same category as iris scans, we have been actively engaging stakeholders and policymakers to educate on this issue and promote clarity. In June 2024, the Academy issued a Policy Statement – Balancing Benefits and Risks: The Case for Retinal Images to be Considered as Nonprotected Health Information for Research Purposes. We are submitting the full statement as an attachment to our written comments to the RFI but wish to underscore the statement's conclusion:

The Academy concludes that the risk of re-identification with de-identified retinal images is low, based on principles guiding expert determinations of de-identified information under the HIPAA privacy rule. The real benefits to advancing scientific knowledge and innovation, addressing eye disease and treatment for individual patients, and accelerating approaches to address public health issues to improve population vision health warrant the collection and collation of retinal images for research and quality improvement purposes.

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/2026-03-18_PHA_NIHRFI_DataModernization_AAO.vf_.docx

105. LungMAP Phase 3 Data Coordination Center

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Dr. Nathan Salomonis, MPI of the LungMAP Phase 3 Data Coordination Center

Name of Organization: LungMAP Phase 3 Data Coordination Center

Type of Organization: Other

Type of Organization - Other: NHBLI Data Coordination Center

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The NHLBI LungMAP consortium is composed of pulmonologists, pathologists, basic scientists, clinical experts, bioinformaticians, AI/machine learning scientists, software developers, and outreach experts. We produce an ecosystem of platforms to explore multi-omics single-cell, bulk, and spatial profiling datasets. We perform extensive outreach using these platforms, including AI-based navigation tools, to educate and enable the research community to develop mechanistic disease hypotheses and predict novel therapeutic targets and drugs for experimental validation.

The proposed policy changes represent an incalculable risk to the biomedical sciences and could lead to increased irreproducible science, a severe loss of scientific transparency and integrity, reduced scientific education and communication, and a substantial limitation on secondary analyses and large-scale reporting of existing studies. The products of large NIH-supported initiatives could become largely inaccessible to the majority of researchers, with no clear corresponding benefit to research participants. Manuscripts reporting research findings may also be limited in their ability to accurately display results (e.g., heatmaps and similar visualizations that contain underlying quantitative information), and reviewers may be restricted in their ability to critically evaluate the underlying data. The development of intelligent “work around” systems (i.e., pseudo-random synthetic or summarized data and metadata), to prevent individuals from human cellular analyzing molecular profiles will ultimately cost the research community tens of millions of dollars and hundreds of thousands of research hours, that otherwise could be spent translating research discoveries to cures.

While we understand that the proposed guidelines are intended to affect datasets moving forward, within the LungMAP consortium the proposed recommendations could still impact >90% of the data products and analysis tools currently provided to the research community, including:

- Sample and single-cell gene expression profiles (processed data)

- Proteomics quantification results

- Metabolomics quantification results

- Lipidomics quantification results

- Chromatin accessibility results

- Spatial transcriptomics results

- Heatmap-based quantitative visualizations

- CellxGene and ShinyCell data visualization portals

- Vitesse and R-Shiny spatial omics visualization

- LungChat AI-based single-cell and pseudobulk quantification results

Limiting access to such resources would likely result in: 1) decreased scientific literacy and reduced training opportunities for the broader research community, 2) an inability to rigorously replicate results across studies through quantitative re-analysis, and 3) a reduced ability for advanced AI platforms to effectively synthesize knowledge from large datasets. Collectively, these limitations would slow research discovery and significantly hinder therapeutic development.

There is little to no clear benefit for research participants or the American public in proceeding with these recommendations in their current form. Rather than democratizing science, the proposed guidelines risk creating a gatekeeping effect in which scientific knowledge becomes accessible only to a restricted group of individuals with specialized expertise and controlled access infrastructure.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Unlike the Gene Expression Omnibus (GEO), which provides more standardized file formats for processed data files, object types (RDS, h5ad, validated matrix files) and allows for rapid release and dissemination of research products, resources such as BioDataCatalyst and dbGAP/SRA do not provide structured results with clear requirements in terms of formatting and validation for specific end-points of genomic analyses. Current dbGAP and BioDataCatalyst outputs comprise what these investigators would refer to as a "wild-west" of formats, data types and structured requirements. Historically, GEO provides such semi-structured outputs for linked datasets with managed access sequencing files (normalized expression, h5, count files, etc.) with associated metadata that directly relates to the published study (i.e., supplemental tables). This is not typically true with data in dbGAP, SRA and BioDataCatalyst. In many cases, IRB exemption or IRB approval of non-human subjects research will be required, significantly limiting access to data. In LungMAP, while open-access data files are downloaded hundreds or thousands of times per month, only 5 LungMAP external people have gained approval to LungMAP controlled access datasets in the last 7 years. Thus, researchers are more likely not to conduct research on a subject than to redirect their efforts to controlled access authorization for gene expression counts, which are nearly impossible to re-associate with a subject, than to traverse the proposed new requirements. Thus, this policy will have a significant negative impact on research globally, which is not an underestimate.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The re-identification of human subjects based on non-sequence based molecular readouts remains a largely hypothetical and speculative area of research. This is particularly true for metabolomics, lipidomics and proteomics datasets in which spectral match analyses are subject to high false positives and negatives, limiting the ability to distinguish or hypothetically infer single nucleotide variants with reasonable confidence. Re-classifying data types such as gene expression counts (bulk or single-cell), metabolomics, lipidomics, proteomics, chromatin accessibility or other direct quantitative biological

measurements (i.e., serological) as controlled-access, is highly inappropriate, will not benefit any research participants.

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

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

106. The Pennsylvania State University

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Andrew Read

Name of Organization: The Pennsylvania State University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached letter

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Please see attached letter

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Please see attached letter

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached letter

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Please see attached letter

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Comments_NOT-OD-26-023.pdf

107. St. Jude Children's Research Hospital

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Mackenzie Bloom

Name of Organization: St. Jude Children's Research Hospital

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Increased access to cloud infrastructure will help maintain centralized secure data access, but cost remains a hurdle for many institutions. This disproportionally limits the ability of small labs and early-stage investigators to participate in genetic analysis of large cohorts. With these increases to controlled access data management, the additional administrative burden further limits these less-resourced investigators.

Additionally, while there is an understandable need for protecting data collected in the U.S. from countries considered adversaries to the United States, NIH should strive to maintain the collaborative nature of biomedical research. To better understand rare and ultrarare pediatric diseases, such as childhood cancers, it is necessary to have a process to effectively share, and often pool, data with colleagues around the world.

Some additional considerations for genomic study data housed within cloud infrastructure. The Cloud Infrastructure of hosted genomic study data should include a method for checking to find potential duplication of samples across studies. Having overlap of patients without transparent access to raw genotype data within the Cloud environment could lead to misinterpretation of what we consider replicated signals, as they could be influenced heavily by having the same patient data across multiple studies. Additionally, for some study cohort genotype call data in NIH approved cloud (e.g. CAVATICA, BioData Catalyst), no joint genotype calls are available for the whole cohort level. This creates challenges for small labs to perform quick validation of their genomic discoveries or query specific variants to subset the cohort for downstream analyses. NIH-supported tools for cross-study sample fingerprinting could help address this challenge.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Undoubtedly, additional resources will be required to maintain quick and effective access to data as larger amounts of controlled-access data are stored/managed. Consolidation and federation of data storage to improve data access will be essential, as, currently, datasets are available across different Cloud Computing platforms without the ability to access across platforms. This state of affairs creates challenges such as inhibiting joint variant calling from large cohort study data. And these consolidated cloud repositories will need to adapt to new types of ‘omics data, such as current challenges being seen in the community with storing raw sequencing data for spatial ‘omics assays. Additionally, if the permitted data access will be determined by the data depositor, there is likely to be a significant bottleneck in data-request approvals as the amount of data to be managed increases, especially if

previously open access data becomes controlled access under the new guideline. There will thus be a significant need for increased administrative support to review data access requests. Individual investigators, especially those with smaller research groups and/or at smaller institutions, may not have the resources to verify the credentials and propriety of a requesting user in a timely fashion, resulting in insufficient access to publicly funded data sets. NIH-supported centralized or shared Data Access Committee services could help mitigate these bottlenecks without placing undue burden on individual investigators.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: Ultimately, the goal needs to be to ensure data is as freely accessible as possible to allow for science to progress, especially for studies focused on rare and ultrarare diseases, while still preventing identifiability. Different types of raw genomic data present different levels of risk (e.g. bulk RNA expression matrices and summary results pose low risk of identifiability, while some single cell or raw proteomics data have the potential to identify individuals). Additionally, the biological source of data poses additional considerations. NIH should form a working group to establish data-access tiers based on identifiability risk, expanding on NIH's current policies that already somewhat address this. The NIH should also consider how its policies apply to data collected with informed consent by participants for their data to be openly shared. NIH should further consider the burden that it will take to apply new rules retroactively to already deposited data. Reconsent often cannot be obtained. In brief, there should be controlled access for raw, individual-level data while supporting open release of summary-level and low-risk derived data whenever possible. Please see the accompanying appendix, "Data Classification Guidance: Open vs. Controlled Data Types" for a proposed, more systematic framework of St. Jude's current recommendation for classifying data generated from pediatric patient specimens and model systems as either controlled or open.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy: Increased access to cloud infrastructure will help maintain centralized secure data access, but cost remains a hurdle for many institutions. This disproportionally limits the ability of small labs and early-stage investigators to participate in genetic analysis of large cohorts. With these increases to controlled access data management, the additional administrative burden further limits these less-resourced investigators.

Additionally, while there is an understandable need for protecting data collected in the U.S. from countries considered adversaries to the United States, NIH should strive to maintain the collaborative nature of biomedical research. To better understand rare and ultrarare pediatric diseases, such as childhood cancers, it is necessary to have a process to effectively share, and often pool, data with colleagues around the world.

Some additional considerations for genomic study data housed within cloud infrastructure. The Cloud Infrastructure of hosted genomic study data should include a method for checking to find potential duplication of samples across studies. Having overlap of patients without transparent access to raw genotype data within the Cloud environment could lead to misinterpretation of what we consider replicated signals, as they could be influenced heavily by having the same patient data across multiple studies. Additionally, for some study cohort genotype call data in NIH approved cloud (e.g. CAVATICA, BioData Catalyst), no joint genotype calls are available for the whole cohort level. This creates challenges for small labs to perform quick validation of their genomic discoveries or query specific

variants to subset the cohort for downstream analyses. NIH-supported tools for cross-study sample fingerprinting could help address this challenge.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
If the NIH implements similar policies as those used for the TOPMed imputation server (ie, secure environment, encryption, etc), there is not apparent additional risk. The TOPMed server is a sensible security model in that they enforce (1) uploaded inputs deleted when no longer needed for processing, (2) outputs are encrypted using a one-time password that is not retained server-side, (3) results are available for only 7 days before automatic deletion. NIH may wish to define minimum required security controls for approved imputation services to ensure consistent privacy protections.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/St.-Jude_Data-Classification-Guidance.pdf

Description: This document provides recommendations for guidance for classifying data generated from pediatric patient specimens and model systems as either controlled or open. This framework reflects pragmatic operational norms, community expectations, NIH Genomic Data Sharing principles, and institutional privacy risk tolerance.

108. Digital Twins for Health Society (DT4HS)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Jun Deng

Name of Organization: Digital Twins for Health Society (DT4HS)

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached document.

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

1. Expand existing controlled-access repositories to support digital twin workflows by (1) enabling model + data co-hosting, (2) supporting version control for simulations and outputs, and (3) capturing full provenance metadata (e.g., inputs, model parameters, and update history).

2. Promote cloud-based secure research environments to (1) encourage "compute-to-data" paradigms where researchers access digital twin data within secure enclaves rather than downloading them, and (2) align repository infrastructure with scalable AI/ML workloads required for simulation.

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

1. Designate digital twin-generated individual-level data as controlled-access when (1) the data are derived from genomic or other sensitive human data, (2) the data are capable of being linked back to individuals directly or indirectly, and (3) the data contain predictive or inferential health information.

2. Align with informed consent to (1) require explicit disclosure of simulation and predictive data generation, and (2) clarify whether future digital twin analyses are covered under existing consent or require re-consent.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached document.

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

1. Classify imputed genomic data as controlled-access under the GDS Policy: treat imputed variants as controlled-access human genomic data when linked to individuals.

2. Extend policy coverage to server-based processing environments: (1) require imputation servers to meet NIH security and privacy standards, and (2) mandate data minimization and retention limits.

3. Define governance for intermediate and derived outputs: explicitly address whether temporary files, logs, and derived datasets are retained and how they are protected.

4. Encourage federated or privacy-preserving imputation approaches: (1) support methods that minimize raw data transfer, and (2) promote secure multi-party computation or enclave-based imputation.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/DT4HS-response-to-NOT-OD-26-023.pdf

Description: DT4HS response to NOT-OD-26-023

109. Association of American Universities

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Lizbet Boroughs

Name of Organization: Association of American Universities

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Association of American Universities (AAU) appreciates the opportunity to respond to the Request for Information on Draft Controlled-Access Data (CAD) policy and Proposed Revisions to NIH Genomic Data Sharing (GDS) Policy.

Founded in 1900, AAU is composed of America's leading research universities. AAU's 69 research universities in the United States transform lives through education, research, and innovation. Research universities, including AAU's member institutions, have a long-standing partnership with the federal government to advance science and technology in the national interest. This partnership, which has roots going back to World War II, has been central to facilitating U.S. global leadership in science and technology.

AAU and its member institutions are committed to responsible data stewardship and share the National Institutes of Health's (NIH) goal of balancing data security with appropriate data access. We take seriously our obligation to safeguard research and the data that underpins it. We understand the landscape of economic and national security threats, and universities have implemented a range of staff and resource-intensive measures to strengthen the protection of data and research.

In addition to this memo, AAU supports the Request for Information (RFI) responses submitted by the University of Pittsburgh, and by colleagues at COGR, the Association of Public and Land-grant Universities (APLU), and the Association of American Medical Colleges (AAMC). AAU's feedback on specific questions is provided below.

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy

Simply put, the draft policy's scale and complexity will significantly affect institutional compliance with existing research policies. The draft policy is not consistent with NIH's goal of improving data access to enhance scientific rigor and reproducibility, because it would restrict an unprecedented amount of data and does not appear to be calibrated to actual privacy and security risks. The policy is challenging to implement without first resolving how it integrates with existing, interlocking federal data privacy standards — such as the Department of Justice (DOJ)'s "Rule Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons" — and without clearly defining ambiguous terms in the policy. AAU requests definitions of key terms in the RFI, including: "data lifecycle," "Controlled-Access Data Repository (CADR)," and "equivalent security standards." For example, is the data lifecycle meant to include early-stage institution-level Institutional

Review Board (IRB) oversight and approval, as well as collection prior to its validation? Is a Controlled-Access Data Repository considered a federal or institutional repository?

It is unclear what NIH means by "equivalent security standards" with respect to the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-171. Clinical research data systems and electronic record formats are widely used by institutions for research documentation. As the University of Pittsburgh stated, neither system type was designed to meet NIST SP 800-171 requirements, "and vendors of clinical research data and electronic health records have no regulatory obligation or commercial incentive to modify them to do so. If the CAD Policy is to be interpreted to reach these systems, compliance may be technically impossible, not merely expensive." This raises substantial concerns about feasibility and the potential for unintended impacts on essential research infrastructure.

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

In AAU's view, as more data is regarded as "controlled, unclassified," the greater the friction will be in fulfilling NIH's goal of improving data access and use. The proposed CAD requirements are a significant expansion of current institutional practice and other established federal requirements. This expansion may be inconsistent with NIH's goal to "improve overall performance" of NIH-funded research.

Please clarify whether the NIST SP 800-171 requirements apply only to CAD repositories or also to repositories that facilitate direct sharing between investigator teams, cloud spaces that temporarily store data, data coordinating centers, and similar activities. Under NOT-OD-25-159, such repositories and centers are not currently subject to CAD requirements. Expansion of NIST requirements to all sharing mechanisms, such as intra-laboratory team members working on the same project, would require considerable funding support for institutions to implement.

The CAD policy is not harmonized with existing federal regulations governing consent and re-consent for data use. The RFI states, "NIH accepts data when collected under informed consent for research use…consistent with the Common Rule, 45 CFR 46, if the "consent meets other expectations of the GDS policy." However, 45 CFR 46 is not applicable to the use of decedent data. Research with deidentified decedent information is allowable under the HIPAA Privacy Rule (45 CFR Part 160 and Subparts A and E of Part 164).

AAU requests that NIH, working with research community stakeholders, the Department of Health and Human Services the DOJ, and other federal agencies, determine how to navigate interlocking federal privacy policies that balance security with scientific access.

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

The proposed scope of the applicable protected data types represents a substantial expansion and diverges from traditional concepts of human genomic data. As AAU's colleagues at COGR have previously stated, "Various categories of 'omic data encompass a wide set of measurements related to human physiological, pathological, or genetic measurements that are used to help understand basic mechanisms or functions of human health states and that do not contain identifiable information." The policy fails to describe how these types of 'omic data pose national security risks.

The RFI states that data collected from NIH-funded research in amounts below the threshold will still be subject to the "expectations of the Data Management and Sharing (DMS) Policy and proposed NIH CAD policy." In essence, all listed participant data, regardless of amount, would be affected by the draft policy. Requiring CAD "equivalent" storage for all data volumes will greatly restrict, if not prohibit, the use of legacy datasets and disrupt research already in progress, as existing institutional repositories may lack sufficient resources to comply with the proposed CAD standards.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
The complexity of working within the constraints of the new GDS Policy may discourage researchers from engaging with NIH data. The proposed threshold of "100 individuals" to be defined as "large scale" and therefore subject to GDS Policy's consent and data sharing requirements is simply too low, given the staffing and financial demands of data stewardship. We request that NIH consider increasing the threshold for large-scale data sharing. At a minimum, NIH should harmonize its proposed threshold to match 28 CFR Part 202, from the Department of Justice's "Rule Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons," in effect since April 8, 2025.

Institutional concerns about applicable federal consent regulations may also limit data sharing and hamper research that relies on data access, diluting the impact and quality of previously collected critical datasets containing deidentified genomic information. AAU urges NIH to clarify whether the proposed policy will apply only to data collected after 2015, rather than to legacy datasets, or whether it is prospective upon final policy implementation. NIH should also clarify whether the agency will allow compliance costs to be included in grant awards.

There may be unintentional disruptions to research, scientific development, and innovation due to the compliance costs institutions must bear. Higher data protection thresholds may have the unintended consequence of slowing the pace of research and reducing institutional capacity for collaboration. These factors, outlined in COGR's May 11, 2023 blog on compliance costs, will most acutely burden smaller institutions, potentially exacerbating existing disparities in health research throughout the United States. Additionally, there is no clear policy roadmap for researchers to follow in contexts where data sharing with researchers in countries of concern may be scientifically or legally necessary. This was recognized in the DOJ rule and included in specific exemptions in 28 CFR §§ 202.510 and 202.511 for drug, biological product, and medical device authorizations, as well as other clinical investigations.

Recommendations

AAU supports APLU's recommendation for NIH to create an active advisory group comprising representatives from federal agencies, industry, and academic research institutions to ensure a full understanding of potential national security risks and the unknown implications of enhanced regulatory restrictions. AAU requests that the group:

- Clarify operational terms including "data lifecycle," "Controlled-Access Data Policy," and "equivalent security standards," and that the proposed CAD policy is prospective and specifically allows safe harbor for the use of legacy datasets.

- Examine the operational balance of risk-based data management and scientific access, including the expansion of NIH-supported data repositories.
- Resolve interlocking NIH data policies and existing regulations, including required IRB review of data management and sharing plans, and Health Insurance Portability and Accountability Act (HIPAA) requirements that impact data security.
- Develop consistent controlled access requirements for human research data across federal research agencies to avoid duplicative federal regulation and to examine the appropriateness of integrating existing mechanisms,

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/AAU-RFI-comments-NIH-Controlled-Access-Data-and-Genomic-Data-Sharing-March-18-2026-final.pdf

110. Amanda Del Giacco

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Amanda Del Giacco

Name of Organization: University of Southern California

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I write as a postdoctoral fellow and young investigator involved in the Adolescent Brain and Cognitive Development (ABCD) Study. The ABCD Study has attracted a wide variety of researchers worldwide as its missions are driven by having open-access scientific resources available. Given the new controlled-access data policy, I have several serious concerns.

1) Slowed scientific progress

2) Lack of alignment in security requirements and proposed risk

3) Unfunded mandate of NIST compliance

Proposing to mandate NIST compliance requirements without corresponding resources will widen disparities in research access and impede career development for young investigators at large.

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

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

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

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

111. Li-San Wang

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Li-San Wang

Name of Organization: NIAGADS

Type of Organization: Other

Type of Organization - Other: CADR

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Challenges with Managing Parallel Data Access Requests (DARs)

Current NIH CADR policy requires that external collaborators submit separate Data Access Requests (DARs) with identical titles and Research Use Statements, and that each DAR explicitly lists all approved collaborators. While this ensures institutional accountability and independent review, it creates significant administrative challenges for multi-institutional collaborations.

In practice, research collaborations are often dynamic, with new institutions and investigators joining over time. Under the current model, any change in collaboration membership requires updates across all existing DARs to maintain consistency in collaborator lists. This results in:

- Repeated administrative effort across multiple institutions
- Delays in onboarding new collaborators
- Increased risk of inconsistencies across parallel DARs
- Reduced scalability for large or evolving consortia

These challenges create friction in the data access process and may slow the pace of collaborative research, particularly for large, multi-site projects.

We recognize that the current structure supports important requirements for institutional certification, accountability, and compliance with data use agreements. However, the growing complexity of collaborative research highlights a need to evaluate whether additional guidance or mechanisms could reduce administrative burden while preserving these core principles.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Censoring of age data. HIPAA regulations require that any age greater than 90 be recorded as "90+" to protect privacy.

This rule was established when individuals aged 90+ were rare and identifiable. However, contemporary dementia and aging research faces a critical scientific limitation:

- Individuals living to 90+ years are now common (approximately 2–3% of the population and rising).
- Dementia research fundamentally requires understanding cognitive decline in advanced age; loss of exact age data eliminates precision in very-old populations.
- Re-identification risk from age alone has decreased dramatically.
- Many dementia cohorts specifically enroll individuals 85+ years old; censoring exact ages eliminates age as a stratification variable.

We recommend NIH establish special guidance for dementia, aging, and gerontology research permitting exact ages above 90 when data are managed under controlled-access through NIST-compliant repositories, participant consent explicitly permits age sharing, and study population is specifically recruited as an aging/dementia cohort. Alternatively, allow institutional review bodies to conduct Expert Determination on age data in aging/dementia studies, determining that exact age poses minimal re-identification risk in the context of controlled access.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy: Research Use Limitations and AI/Machine Learning Applications.

An additional topic that requires in-depth discussion is the use of controlled-access genomic data for artificial intelligence and machine learning (AI/ML) model development and how research use limitations apply to AI models trained on restricted data. Researchers increasingly use GDS-regulated genomic data combined with clinical, imaging, and other modalities to develop AI models for disease prediction and risk stratification. Yet current policies provide no guidance on whether AI/ML development constitutes an authorized research use, whether trained models inherit research use restrictions from their training data, or how to document data provenance when models combine data from multiple cohorts with heterogeneous restrictions.

We recommend that NIH develop a comprehensive AI/ML strategy for controlled-access data, grounded in three principles: (1) AI is a technology that should be governed by general GDS framework principles rather than treated as a special category—research use limitations and consent requirements that apply to data should apply equally to AI models derived from that data; (2) the strategy should anticipate and align with emerging federal policies and technical standards for AI governance (e.g., NIST AI Risk Management Framework, Executive Order guidance on AI); and (3) NIH should provide practical implementation guides and tools (e.g., model documentation templates, consent language examples, institutional review checklists) that make compliance straightforward and enable researchers to innovate responsibly.

Determination of research use limitations using data from multiple studies.

The most critical implementation challenge in determining research use limitations is ensuring consistency across Institutional Certifications (ICs) derived from multiple studies. Under the current

process, ICs are often collected during Just-In-Time (JIT) for newly funded grants, including those that will generate genomic data from existing cohorts or analyze data from multiple datasets to generate summary statistics. However, the Institutional Review Board (IRB) at the awardee institution is typically not in a position to comprehensively evaluate:

- all consent forms associated with the original cohorts contributing primary data, nor
- all consent and data use restrictions associated with secondary data requested from repositories

As a result, IRBs may be asked to certify data sharing permissions without full visibility into the underlying participant consent conditions. This creates a significant risk of inconsistent or inappropriate determination of research use limitations, particularly when downstream projects generate new or derived data from multiple existing cohorts.

These challenges are further amplified for aggregated or derived data products generated from multiple contributing studies. When cohorts with heterogeneous consent and data use limitations are combined, there is currently limited consistency in how research use limitations are applied to the resulting data. In practice, aggregation may obscure cohort-specific constraints, yet those constraints remain relevant to participant intent. At the same time, there is limited guidance on whether, or under what conditions, aggregation or derivation of data may permit broader sharing than would be allowed for the underlying individual-level data.

Unless a project is limited to data generated from a cohort under the direct oversight of the same investigative team, downstream data-generating projects should not be responsible for determining data sharing limitations for participant data.

We recommend that repositories should rely exclusively on the Institutional Certification provided by the institution responsible for the original cohort study to determine research use limitations. Downstream ICs associated with secondary analyses or data generation should not override or reinterpret these constraints.

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

112. Bridge2AI-Voice and AI-Readi Consortium

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: yael bensoussan

Name of Organization: Bridge2AI-Voice and AI-Readi Consortium

Type of Organization: Other

Type of Organization - Other: NIH funded consortium

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached form

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached form

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/RFI_response_focused_on_Biometric_Yael_Cecilia_-_for_signature_CL.docx.pdf

Description: Response focused on discussion around the label of "biometric data" for voice and retinal imaging data

113. American Association for Dental, Oral, and Craniofacial Research

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Makyba Charles-Ayinde

Name of Organization: American Association for Dental, Oral, and Craniofacial Research

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The American Association for Dental, Oral, and Craniofacial Research (AADOCR) is the leading professional community for multidisciplinary scientists who advance dental, oral, and craniofacial research. We appreciate AADOCR appreciates National Institutes of Health (NIH)'s efforts to establish harmonized, transparent, and risk-proportionate requirements for protecting human participant research data while enabling responsible data sharing. To respond to this request for comments, AADOCR engaged its Science Information Committee.

We support the NIH's goals to clarify which data types should be shared via controlled-access systems under NIH sharing policies, and to simplify and harmonize requirements by revising the NIH Genomic Data Sharing Policy. The dental, oral, and craniofacial (DOC) research ecosystem includes data modalities that carry unique privacy risks (e.g., high-dimensional imaging, facial morphology, longitudinal dental EHR data, oral microbiome paired with clinical phenotypes). Harmonized controlled-access expectations will be particularly valuable for DOC datasets where re-identification risk may be underestimated if decisions rely solely on de-identification labels rather than data type and context. NIH's existing guidance on when data may warrant controlled access under the DMS framework is an important foundation; AADOCR supports NIH's emphasis that privacy risk can remain even when data are de-identified.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
NIH has invested in controlled-access repository capacity and best practices. However, an expanded controlled-access requirement will likely strain capacity and introduce new operational burdens for disciplines that rely heavily on imaging and clinical systems data, including DOC research. AADOCR recommends the NIH (i). address repository capacity for high-volume, high-dimensional data, this is essential as DOC research increasingly relies on large imaging files and therefore controlled-access storage and compute must scale accordingly, (ii). provide standardized repository intake templates (metadata requirements, consent elements, IRB language examples) to reduce burden on smaller institutions, and (iii). fund training and implementation support for disciplines with less experience using controlled-access systems.

The National Institute of Dental and Craniofacial Research (NIDCR) Data-Driven Science (DDS) Hub serves as a centralized resource to scientific data, biospecimens and other experimental materials, and tools that support data science–driven research and training. AADOCR recommends NIH leverage the DDS Hub to: (i) help investigators identify appropriate data sources, repositories, and analytic resources; (ii)

promote high-quality data generation and FAIR-aligned practices; and (iii) expand training and support for data science–enabled DOC research. Although the DDS Hub is not a controlled-access repository, it can play a critical role by helping the NIH community navigate repository options and evolving compliance expectations as controlled-access requirements expand.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: AADOCR supports NIH's approach of designating certain sensitive data types as controlled-access and recommends the explicit mention of facial/craniofacial and dental imaging as high-risk modalities. In DOC research, imaging often contains face-identifying or uniquely identifying anatomy, therefore, NIH should ensure that the Controlled-Access Data Policy clearly encompasses (i). facial photographs and 3D facial images, (ii). CT/CBCT imaging that includes facial structures, and (iii). longitudinal imaging series that can be linked across time. These modalities warrant controlled access due to re-identification potential and sensitivity, especially when linked with clinical histories and demographics.

AADOCR also supports clear definitions and examples for common gray areas. This may include terms such as:

i. Imaging data: differentiate low-risk derived measures versus high-risk raw images,

ii. Clinical trial data: clarify what can be open versus what should be controlled,

iii. High-dimensional molecular data: clarify boundaries between controlled-access "omics" and lower-risk summaries.

Additionally, AADOCR recommends NIH avoid thresholds that may quickly become obsolete as technologies evolve. AADOCR supports a risk-based, context framework that (i). considers identifiability, uniqueness, and linkage potential including whether the dataset is combined with clinical/demographic variables, imaging, geolocation, or rare disease status and (ii). provides decision support examples illustrating when smaller analyte sets still pose meaningful risk (e.g., rare variants, distinctive biomarker panels paired with clinical phenotypes).

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy: AADOCR appreciates NIH's commitment to harmonizing controlled-access designations and simplifying genomic data sharing requirements while strengthening participant protections. AADOCR would welcome continued engagement as NIH refines these policies and develops implementation guidance.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
AADOCR supports NIH's focus on maintaining privacy protections for imputation servers and reference panels while enabling responsible scientific use. AADOCR recommends NIH explicitly encourage and/or pilot validated privacy-enhancing and secure-compute strategies that reduce risk without compromising scientific utility. Additionally, AADOCR recommends NIH recommended minimum security controls for institutions accessing controlled-access data, consistent with NIH security best practices expectations.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Draft-NIH-Controlled-Access-Data-Policy-and-Proposed-Revisions-to-NIH-Genomic-Data-Sharing-Policy_AADOCR.pdf

114. Carnegie Mellon University

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Carnegie Mellon University

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/CMU_NIH_GDS_RFI_Response.pdf

Description: Carnegie Mellon University's Response to Draft NIH Controlled-Access Data Policy and Proposed Revisions to the NIH Genomic Data Sharing Policy

115. National Alliance for Eye and Vision Research

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Dan Ignaszewski

Name of Organization: National Alliance for Eye and Vision Research

Type of Organization: Research Participant/Patient Advocacy Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Final-NAEVR-Comment-NIH-NOT-OD-26-023.pdf

Description: NAEVR Response to RFI re retinal imaging

116. Muñoz Torres et al. All contributing authors are listed in the attached PDF.

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Muñoz Torres et al. All contributing authors are listed in the attached PDF.

Name of Organization: All organizations are listed in the attached PDF.

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
N/A

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
See attached PDF, titled "Response to RFI NOT-OD-26-023 Repositories 2026-03-18."

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
N/A

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Response-to-RFI-NOT-OD-26-023-Repositories-2026-03-18.pdf

Description: We submit this response as members of (not on behalf of) the Bridge2AI consortium with direct and substantial experience in the data types and arguments at the center of this policy discussion. Thanks for giving us the opportunity to share our expertise and perspectives.

117. Association for Research in Vision and Ophthalmology (ARVO)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Iris M. Rush

Name of Organization: Association for Research in Vision and Ophthalmology (ARVO)

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/ARVO-Comment-NIH-NOT-OD-26-023-.pdf

Description: ARVO Response to Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

118. American Academy of Pediatrics

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: American Academy of Pediatrics

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/AAP-Comments-on-NIH-Research-Participant-Data-RFI-FINAL.pdf

119. Dartmouth College

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Dartmouth College

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH_Data_Sharing_RFI_March_26.pdf

120. American Physiological Society

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Mark Eichelberg

Name of Organization: American Physiological Society

Type of Organization: Professional Organization/Association

Role: Other

Role – Other: Science Policy Manager

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The American Physiological Society (APS) represents nearly 8,000 biomedical researchers and educators in academia, private industry and government. While APS acknowledges the importance of data security in human subject research, the proposed policy would bring important human research, including clinical trials, under equivalent requirements to NIST-SP-800-171, a level of security typically applied to defense contractors. This would be a dramatic increase to regulatory burden that reaches beyond what is necessary to protect the privacy of human research participants. The costs of implementing this policy far exceed the data security benefits, and would result in duplicative or conflicting requirements, undermining of collaborative research, and significant disruption of ongoing studies. APS strongly urges the National Institutes of Health (NIH) to consider the following recommendations:

1) Reduce the required security environment for covered human data types to be compatible with the standards set by the HIPAA Security Rule.

2) Evaluate the expected costs to researchers and institutions, and accommodate the time it will take for institutions to budget, plan, and build any necessary infrastructure to comply with the policy.

3) Clarify that adequately de-identified data may be used or shared without restriction by the policy.

4) Apply the policy prospectively so that ongoing studies are not disrupted and previously collected data may continue to be used and shared.

Currently, personally identifiable human subject research data is often secured under the standard of the Health Insurance Portability and Accountability Act (HIPAA) Security Rule. The difference between these two security standards is significant. The latest revision of NIST-SP-800-171 outlines more than 90 security controls across 17 families of security requirements, each of which must be met to maintain compliance. Transitioning into a compliant environment is extraordinarily expensive, requiring substantial up-front infrastructure development, and continued compliance involves significant maintenance, staff training, and monitoring costs. The estimated cost of setting up a Cybersecurity Maturity Model Certification (CMMC) Level 2 environment, which shares identical requirements as NIST-SP-800-171, extends into the hundreds of thousands of U.S. dollars for a medium-sized institution, with annual upkeep costs likely in a similar range.(1)

The standards of NIST-SP-800-171 are incongruent with the data security of the academic environment. Data and samples collected through hospitals and clinics must already comply with the HIPAA Security Rule, which under most circumstances continues to cover the data while in the hands of researchers. All human data types outlined in the proposed policy fall under the 18 HIPAA identifiers covered in the Privacy Rule. The HIPAA security environment, which is much closer to CMMC Level 1, effectively ensures research participant privacy when clinicians share data with researchers. It also provides guidance for de-identification of protected health information so that data can be more easily shared and re-used. This level of security is substantially more affordable to maintain (see Ref. 1 for estimated costs), and already in place at most research institutions that conduct human subject research.

The consequences of implementing the proposed Controlled-Access Data Policy would be severe. Because of the expenses associated with NIST-SP-800-171 compliance, many researchers, particularly at smaller or less research-intensive institutions, would be unable to continue conducting research with human subjects. Studies currently in progress would be disrupted, including important clinical studies. Furthermore, it is not clear how the policy would affect the interface of data collection – often in a HIPAA security environment – or scientific instruments used to collect or analyze data, which are not designed to meet the security standard of controlled unclassified information.

APS strongly recommends that NIH revise the proposed policy so that compliance with NIST-SP-800-171 security standards is not required for researchers to use and share data that is more appropriately covered by HIPAA-level security. NIH should also provide researchers and institutions with adequate time and resources to build any new infrastructure necessary for compliance. Finally, NIH should ensure that the policy does not disrupt data collection, sharing, or collaboration between institutions by outlining necessary steps for data de-identification and clarifying how the policy covers the interface between researchers and HIPAA-regulated systems.

1. The True Cost of CMMC Compliance: What Defense Contractors Need to Budget For [Online]. Kiteworks. https://www.kiteworks.com/cmmc-compliance/compliance-costs/ [2026 Mar 6].

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
APS has no comment on this component of the policy.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Each protected data type outlined in the proposed policy is covered under the 18 identifiers in the HIPAA Privacy Rule. Importantly, the outlined data types would cover nearly all biomedical research with human subjects, and particularly all clinical research. As previously explained, APS considers a HIPAA-level security environment to be a more appropriate data security level for these data types, and NIH should provide guidance for the de-identification of data, as well as what it considers to be the life cycle of the data.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
APS has no comment on this component of the policy.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
APS has no comment on this component of the policy.

121. Han Yi, Brock Wester, Bree Christie, Erik Johnson, Rahul Hingorani

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Han Yi, Brock Wester, Bree Christie, Erik Johnson, Rahul Hingorani

Name of Organization: Johns Hopkins University Applied Physics Laboratory

Type of Organization: Non-profit Research Organization

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Our team at the Johns Hopkins University Applied Physics Laboratory is part of the BRAIN BBQS (Brain Behavior Quantification and Synchronization) Program Consortium, which consists of neuroscientists, data scientists, device engineers, experimentalists, ethicists, and other researchers. The goals of the BBQS program are to advance our understanding of human and animal behavior, and most of the data being generated and analyzed is not genomic in nature. We oversee the development of the EMBER (Ecosystem for Multimodal Brain-behavior Experimentation and Research) Data Archive, which will support the BBQS program in multiple ways including data storage and sharing. As part of the Consortium, we wholeheartedly support the goals of the Draft NIH Controlled-Access Data Policy aiming to protect the sensitive human data collected by the Consortium.

As a part of designing, developing, and overseeing the EMBER Data Archive for BBQS, we will work with the NIH to institute tools and workflows to implement appropriate controls for data access, as needed, including the CADR-relevant Country of Concern or Covered Person designations. We understand that these designations and access controls/policies for various use cases and for specific data may be adjudicated by appropriate bodies outside of BBQS. Importantly, as will be elaborated further below, much of the data generated by BBQS are not sensitive human data as defined by the Draft NIH CADR Policy.

In this response, we focus on the relevant technical controls that are active or pending to ensure adherence to the new and emerging policy requirements as will be announced by the NIH. Specifically, this pertains to the sensitive human data that constitute a small subset of all data generated by the BBQS Consortium, which will be supported under a special-purpose resource (named EMBERvault), quarantined from the rest of our open-access repository (e.g., EMBER-DANDI).

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
EMBER is the newly established NIH BRAIN Data Archive for brain-behavior data. Within this data ecosystem, a special-purpose data storage resource called EMBERvault is being developed to host CADR-pertinent, sensitive human data. EMBERvault will be a HIPAA-compliant cloud storage system designed for the storage of PII/PHI data. This HIPAA compliance process flow for EMBERvault will help to satisfy requirements for NIH CADR certification. It is very important to note that these highly secure data systems require additional resources, and increase costs associated with storage, access, tracking, and

audits/reviews. Other data storage components of the EMBER data ecosystem (i.e., non-EMBERvault) will host non-sensitive data, such as animal data, and these should not be subject to CADR.

For context, the BBQS Consortium is generating predominantly non-genomic brain and behavior research data. All sensitive human data will be securely deposited in EMBERvault. Non-human data and any non-sensitive human data, as defined by CADR-policy, is planned for deposition in separate data storage components for the EMBER data ecosystem (e.g., EMBER-DANDI).

EMBERvault will have the requisite controlled-access safeguards as instituted through state-of-the-art technical/engineering controls for authentication and authorization. This includes crucial steps that incorporate external adjudication of requests to access the EMBERvault resource as a whole, as well as individual repositories governed by IRB oversight.

Here, we propose the pipeline for granting access to controlled data in EMBERvault, which will involve the following steps in series: (1) a request is received from a user to access EMBERvault; (2) pursuant to CADR, an external body (e.g., Data Access Committee) outside of EMBER determines whether the request shall be authorized, considering multiple factors including potential association with countries of concern or covered persons; (3) if and when the authorization from the external body is received, EMBER creates accounts for services that enable secure authentication (e.g., Keycloak) and secure file transfer (e.g., Globus); (4) using the authentication method, the user logs into EMBERvault; (5) the user requests access to a specific data repository on EMBERvault; (6) through a process that shall be established and implemented by the BBQS consortium involving the data owners, the institutional review board (IRB), and others, the determination for granting the user access to the repository is made; (7) if and when the determination is made, the user is provided with an access key to the secure, repository-specific cloud storage (e.g., AWS S3 bucket).

We emphasize that the EMBERvault development is still ongoing, and it is our highest priority to institute a system that can be compliant with current, new, and emerging NIH policies as pursuant to CADR. More generally, for NIH funded data system developments that may not have anticipated additional requirements for controlling data access (as compared to requirements in the funding announcement or as proposed), adding new requirements mid-stream could create financial risk for the projects and thus the research programs being supported by the resources.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access: The EMBER team believes that the protected data types in the current Draft Controlled-Access Data Policy are appropriate, and align with current designs for EMBERvault, the secure data store for sensitive data, to be compliant with CADR as necessary. To this end, it would be helpful to achieve further specification for protected data types of potentially sensitive human data in the next version of the policy.

For instance, the following categories, as written, could be interpreted in a fairly broad manner and need clarity and specificity to ensure that both data systems are properly designed and implemented to ensure the goals of the CADR policy and also that safe secondary science by the research community is maximized: “individual-level clinical trial data” and “imaging data of the human face or head regions.” While data generated from individuals in clinical trials may involve health information, it can also include

results from simple psychological tasks (e.g., reaction times in an attention task) that alone (with no identifying information) bear minimal or no risk for breach of privacy or confidentiality. Similarly, imaging data from the human brain can be de-identified (e.g., by cropping out image data other than the brain tissue), with little risk for downstream societal consequences. Thus, we request that a higher level of granularity be implemented for specific data types to avoid scenarios where data is unnecessarily stored in costly and science-limited higher security locations.

Additionally, we request additional clarification on the how these protected data types may be applied in IRB-controlled studies with informed consent where participants have agreed to broad sharing of data. As currently stated in the RFI, "These data types may only be shared without access controls if (1) there is informed consent explicitly stating data are to be shared openly without controls. In these instances, institutions must still review to determine that openly sharing these data pose very low risk when shared and used; or (2) open sharing is required or authorized by Federal law or international agreements to which the United States is a party". Here we suggest providing more details on the procedures necessary to deem data to "pose very low risk" and offer the possibility of maintaining the open data sharing posture for such data types, which we suggest could be maximally beneficial for scientific advances while minimizing harms from data misuse, including by those associated with countries of concern or covered persons.

Again, the EMBER team is ready to comply with all required policies, including instituting technical controls in accordance with guidance received from NIH or other U.S. Government entities. Such design elements for EMBER could include tiered access approaches for multiple data types and data sensitivities (as outlined above), including solutions for authorization, authentication, and administrative approvals.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
None; the BBQS Consortium is not anticipating the generation of significant genomic data. EMBER is not planning to support the storage of genomic or multi-omic data.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
None; the BBQS Consortium is not anticipating the generation of significant genomic data. EMBER is not planning to support the storage of genomic or multi-omic data.

122. Broad Institute

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Jonathan Lawson

Name of Organization: Broad Institute

Type of Organization: Non-profit Research Organization

Role: Other

Role – Other: Senior Director, Data & Federal Partnerships, Data Sciences Platform

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Broad Institute is an independent, non-profit research organization that aims to discover the root causes of all common and rare diseases, and to use this insight to help develop safe and effective therapeutic interventions. We tackle big scientific questions that no single lab can address alone. We empower cross-disciplinary teams to solve the most important challenges in biomedicine. We invent and openly share cutting-edge technologies and tools to accelerate research and catalyze improvements in human health throughout the United States and beyond. The Broad innovates at the intersection of scientific disciplines, convening scientists and other experts from genomics, cell biology, chemistry, engineering, neuroscience, therapeutics, artificial intelligence/machine learning, computational biology, and public health.

Broad appreciates the opportunity to provide feedback on the proposed Draft NIH Controlled-Access Data Policy and the revisions to the Genomic Data Sharing (GDS) Policy. As an institution dedicated to accelerating the pace of biomedical research through collaborative data sharing, we offer the following comments based on our experience generating data, and developing and operating data ecosystems at a large scale, from omics and other types of research and clinical data.

Harmonization of Data Policy Frameworks

We strongly commend the NIH for its efforts to synthesize and harmonize a diverse set of policies that have historically been difficult for researchers to navigate. The current landscape of overlapping and occasionally inconsistent requirements, spanning the GDS Policy, the DMS Policy, and various Institute-specific notices, creates administrative burden that inhibits scientific progress and limits the integration of important NIH datasets particularly when working across NIH ICs. By providing a transparent, unified framework for human participant research data, the NIH is taking a critical step toward accelerating scientific discovery while upholding participant protections.

Addressing Security and Ethical Risks in the Age of AI

The rapid advancement of AI-driven analytical tools has fundamentally altered the risk profile of "de-identified" data, making re-identification more technically feasible than ever before.

In this context, we believe it is both wise and necessary to protect research participants and maintain public trust by obligating data users of increasingly sensitive data types to the formal terms and conditions of data access agreements (DAAs), like the NIH Data Use Certification (DUC), and general

terms of use, like the NIH Genomic Data User Code of Conduct. These legal frameworks are essential for binding researchers to appropriate codes of conduct and ensuring that the use of sensitive biological and phenotypic data remains aligned with the original participant consent.

Furthermore, as the potential for data misuse grows, we believe there is a critical necessity to enhance and expand the specific questions asked of researchers during the application process. Standard Research Use Statements (RUS) often fail to capture the full scope of a researcher's ethical intent or the potential secondary implications of their methodologies. By requiring more granular disclosures regarding the ethical guardrails and specific intentions of the proposed research, the NIH can more effectively draw out concerns that might otherwise remain latent. This proactive transparency is vital; it provides the NIH with clear recourse against users who might otherwise misuse data in ways that harm participants or damage the reputation of the global research community. Establishing these explicit expectations at the point of access ensures that accountability is a foundational component of the data sharing lifecycle.

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

Reducing Friction in Controlled Access through Novel Governance Frameworks

While we support the expansion of controlled-access protections, we must also note the significant challenges that the research community faces regarding these changes. Historically, controlled access often leads to delays, burdensome manual reviews, and inconsistent institutional requirements.

However, we believe these obstacles are no longer a technical necessity. The Broad Institute developed and deployed a novel controlled-access data governance framework for Broad's controlled access data that expedites and sometimes automates data access, while ensuring consistent compliance procedures. This is a promising framework that reduces administrative burden, enhances compliance adherence, and helps researchers access data faster.

Standardized & Automated Review

Broad's DUOS system enables the automated processing of Data Access Requests (DARs) by matching research purposes against structured, machine-readable consent terms (e.g., Global Alliance for Genomics and Health [GA4GH] Data Use Ontology [DUO] standard). This functionality is already being utilized by the Broad Institute Data Access Committee and has been tested by multiple NIH DACs. Employing such a framework while significantly increasing the size of NIH's controlled access data corpus may alleviate researcher concerns and potential access delays, while upholding compliant controlled access procedures.

Pre-Authorizing Researchers to Request Data

We have proposed that DAAs empower Signing Officials (SOs) to pre-authorize trusted researchers to submit DARs automatically (aka Library Card Agreements), with the SO retaining the right to revoke access at any time. This model places scientific oversight of the research activities into the institutional administrative oversight, rather than placing the burden on NIH. By shifting the model to pushing relevant datasets to investigators working on relevant problems, we will speed scientific progress and reduce the time spent on duplicative data access requests for the same scope of work. The Broad

Institute and its DAC already operate under such a model, and legal agreements reflecting this approach have already been drafted and approved for use by NIH's legal office.

Data Access Tiers

Originally, open and controlled access tiers were binary concepts. Later, the idea of registered access (with varying definitions) was employed to expedite researcher access to less sensitive datasets. We encourage evaluating when tiered access models may be appropriate for a single dataset and for subsets of a dataset, such that data elements from a single dataset may be allocated across multiple tiers (i.e. controlled, registered, open). This approach adquates the rigor of access request with the sensitivity of data, making sure that researchers are not burden with a more complex than necessary request process. Ultimately, expediting researchers access while upholding necessary protections.

If NIH Data Access Committees (DACs) were to adopt similar policies and software infrastructures, researchers would see a transition from months of waiting to near-instantaneous turnaround times. We emphasize that the expansion of controlled-access data types must be accompanied by a commitment to modernizing the mechanisms of access to ensure that scientific progress is not throttled by administrative lag.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Defining Raw Readouts and Scale Parameters for Controlled Access

The Broad Cancer Program generates large scale multi-omic datasets as resources for the oncology community. For new data types included in the updated policies, especially the epigenomic and proteomic data types, it would be helpful to provide the types of "raw" readouts that would be considered as requiring controlled-access. Also, for situations where the whole genome or full proteome are not being analyzed, it would be a benefit to understand the scale of the data that requires controlled-access.

Improving Access to Low Allele Count Data to Enable Rare Disease Diagnosis and Discovery

At the Broad Institute, we support many genomic data resources that support rare disease diagnosis and discovery. This includes the gnomAD database which provides aggregate allele frequencies from previously generated datasets across ancestrally distinct populations. We also support the expert interpretation of variants for rare disease through the NIH-funded Clinical Genome Resource (ClinGen), working jointly with the NBCI-supported ClinVar database. Critical to the maintenance of this work, and its routine use by clinical diagnostic pipelines, is the ability to access variant data with allele counts down to a single individual. Approximately 80% of variants submitted to ClinVar, primarily by clinical labs, have been submitted by a single lab, likely due to observation in a single individual or family. Despite this rarity, sharing this data openly is critical for rare disease diagnosis and discovery. Yet, some genomic data generating programs limit the ability to share allele counts below a certain threshold. However, sharing summary statistics for all variants, including those found only in a single individual, is critical for scientific progress. Obscuring information about this rare variation would be reasonable if the risks to participants were large, but the experiences of many other large genomics initiatives illustrate that the actual risks to privacy are quite small. Follow this link to read our full statement supporting public release of low frequency allele count summary statistics. https://gnomad.broadinstitute.org/AC1

We would like the policy to be clearer that, for the sharing of genomic summary results, variants with allele frequencies as low as a single count can be shared publicly.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Urgent Need to Unblock AI-Specific Data Use Policies

We wish to express concern regarding the delay in the release of a clear NIH policy on the use of AI and Large Language Models (LLMs) with controlled-access data.

As the NIH expands the list of data types requiring controlled access, including high-dimensional "omics" and imaging data, it is imperative that researchers have clear guidelines on how these data can be analyzed using the most novel technologies. Blocking the use of AI on these datasets, or leaving the policy in a state of ambiguity, effectively prevents the research community from making the best use of the very data NIH is working to protect. We urge the NIH to complete and release its AI data use policy immediately to ensure that security measures do not inadvertently become barriers to innovation.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Support for Imputation Servers and Term Clarifications

We are supportive of enabling use cases such as imputation servers, where the genomic data are encrypted and ephemeral. We thank the NIH for recognizing the scientific opportunities that such capabilities enable.

The RFI "...requests input on clarifying that Approved Users may operate imputation servers…". The term Approved Users in the controlled access framework tends to imply secondary researchers approved for access to data submitted by primary researcher teams. It may be important to clarify if NIH views a future where users requesting access to data via a data access request process such as dbGaP (i.e., Approved Users), may be authorized to establish and operate Imputation Servers without those systems being under federal security oversight, such as an NIH Authorization to Operate (ATO) or demonstrable compliance with the NIH Security Best Practices for Controlled-Access Data Repositories through a third-party audit report submitted to and reviewed by NIH, or if those Imputation Servers should indeed be subject to more stringent federal security oversight. Similarly, we believe the third point stating "(3) the imputation servers are funded or operated by NIH or another federal agency" may better accommodate desirable future use cases if it read "(3) the imputation servers are funded or operated by or under the governance of NIH or another federal agency", such that an entity outside of NIH could be permitted to operate an Imputation Server without direct funding from NIH, so long as NIH maintains oversight as they deem appropriate.

123. University of Washington Genetic Analysis Center

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: University of Washington Genetic Analysis Center

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/GAC-response-to-RFI_-NOT-OD-26-023.pdf

Description: Responses to all 5 prompts

124. Association of American Medical Colleges

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name:

Name of Organization: Association of American Medical Colleges

Type of Organization: Professional Organization/Association

Role: Other

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
In broad conversations with the academic medical research community on the proposed controlled-access data policy, AAMC received significant feedback on 1) questions and concerns regarding critical details and definitions which are unclear in the policy, 2) concerns about the increased breadth of data and data types covered by the policy without consideration of whether there is an increased risk to privacy or security, and 3) the institutional resources of time, infrastructure modifications, and financial investment that it would take to implement a policy of this magnitude. In particular, the funds necessary to implement this policy as written would be extremely challenging for lower-resourced institutions and it would restrict the ability of many organizations to carry out research with human data. AAMC is extremely concerned that institutions which are unable to meet the requirements defined by the policy would be forced to cease ongoing studies and that the policy will have a widespread, dampening effect on the data ecosystem which is so crucial for scientific progress.

AAMC's overarching recommendations for a new controlled-access data policy are that NIH conduct a careful, risk-based analysis of the data which should be subject to this policy, harmonize the policy with other federal standards and regulations, define a phased, multi-year approach for policy implementation and enforcement, and limit the scope of the policy to new awards so that institutions can include compliance costs in the budget for a grant application. To facilitate these recommendations and assist in the policy revision process, we strongly encourage NIH to develop an expert working group comprised of members from the federally-funded research community to assist in the creation of a revised draft, which would draw on institutional knowledge and provide specific input to facilitate appropriate levels of data security while minimizing administrative burden and ensuring clear, consistent, and implementable requirements.

Below we address specific provisions in the current draft policy.

Scope and Applicability. AAMC appreciates the stated exemptions in the policy for NIH-funded research that only involves (1) generation and sharing of non-human data or (2) collection and sharing of human cell lines and biospecimens. To streamline the language and increase clarity, we suggest that NIH add a third exemption in this section, that: "human data or data derived

from human cell lines or biospecimens already shared prior to the effective date of this policy," which is currently listed separately from the first two exemptions. We also urge NIH to reconsider the decision to apply the policy to derivative data sets, even when these may have little to no security risk. We

recommend that these data sets only be subject to the policy in cases where a security or privacy risk has been identified.

Required Human Data Types. The AAMC has significant concerns about implementation of the draft policy requirements, which state that the listed data types "must be protected throughout the data lifecycle." This descriptor is vague and will undoubtedly be interpreted differently across the regulated community. To the extent that the "data lifecycle" is considered to include early stages of data generation, the policy would create an enormous burden on institutions and would be virtually impossible to implement. This interpretation would require additional security considerations and the use of specialized data enclaves at every step of the research process, including at the level of the individual laboratory collecting raw data, institutional core services which engage in data processing, and other cleaning and analytic steps prior to data sharing. Such a broad scope would considerably hamper researcher workflows for data analysis, which often utilize open-source tools, and increase the burden of generating data use agreements and sharing data with collaborators. AAMC strongly recommends that NIH revise the draft policy to limit its applicability to the "data sharing lifecycle," defining that term to begin at the time data from the initial study are required to be shared.

We also note that the draft policy states that "Institutions conducting NIH-supported research" must ensure that all the listed data types in this section are protected, suggesting that the policy applies to any research conducted by an institution that receives any NIH funding, regardless of the funding source for that research. We do not believe that was the intention of the policy, as the Applicability section states that the policy only applies to "NIH-supported research." We request that NIH make this language consistent throughout, clarifying that the policy applies only to data generated from NIH-funded research.

AAMC has strong concerns with the significant expansion in the proposed policy of the scope of data which institutions would be required to manage under extremely specific and rigorous standards. The eleven data types which would require additional protection under the proposed policy, "even when not shared through a controlled-access repository," encompass almost all human-related research. Given that institutions are already managing these data under a complex network of federal requirements, including the recently instituted Bulk Data Rule from the Department of Justice, we strongly urge the NIH to ensure that the policy does not conflict with or reach beyond the existing federal framework for research data security.

AAMC supports the current exemptions to sharing data with access controls in situations where the data is low-risk or where data sharing is required by federal law or international agreements. However, we are concerned by the inclusion of a broader requirement for "informed consent explicitly stating data are to be shared openly without controls," as this may not be possible for many long-standing or legacy data sets currently in use for research. We suggest that this provision be removed from the policy.

Requirements for Controlled-Access Data Sharing. In order for institutions to effectively implement new requirements for controlled access repositories, the policy must be clear as to the necessary security and operational standards which institutions must meet to be in compliance, as well as when these standards should be applied. It is imperative that the requirements allow researchers and institutions to engage in long-term planning and the type of financial investment required to bring institutional technical systems into compliance on this vast scale.

AAMC received feedback from many members of the research community that the lack of availability – and significant cost – of repositories which would meet the requirements for all the data types specified in the policy will make compliance difficult and could decrease the ability of the research community to share data. The policy currently states that controlled-access data repositories should meet "NIST-SP-800-171 or equivalent" standards. We recommend that NIH specifically list any additional security standards which would fulfill the agency's requirement or the process that would be required for an institution to demonstrate that the employed standards are equivalent to NIST-SP-800-171.

While some institutions have built infrastructure which can support the NIST SP 800-171 standard, we received input that this is an enormously time-consuming and expensive update, requiring the institution to completely re-write its institutional repository infrastructure over a period of several years. In the context of the large amount of data which would fall under the policy as proposed, we underscore the critical importance of instituting our earlier recommendation on the definition of the data sharing lifecycle to facilitate implementation of this aspect of the policy.

There is broad agreement across AAMC's biomedical research community and beyond that the currently available NIH-funded controlled-access repositories would not be sufficient for the volume of data that would need to be stored in this environment under the proposed policy and may not possess the needed analytic capability for certain research projects. We additionally note the complexity of applying a security standard not just to the data stored in an environment, but for all of the tools used during the research process. We urge NIH to revise this policy such that it is limited to research data for which there is a justified scientific and security rationale to apply the outlined standards and operational requirements.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
In advance of the AAMC's detailed recommendations for the description of the proposed revisions to the Genomic Data Sharing (GDS) Policy contained in the RFI, we are making both the assumption and the request that, as it has done in the past, after evaluating the RFI response from the impacted community, NIH will release a draft policy in its entirety so that it is clear whether the proposed language in the RFI will be replacing or adding to the existing policy. In some cases, proposed language would have a different meaning or impact, depending on how it is incorporated into the existing policy. Both the NIH and the research community would benefit from the review of and comments on the full text of a new draft GDS policy, including the incorporation of feedback from this RFI, before a final policy is issued. The proposed revisions also have elements that are dependent on the new controlled-access data policy discussed above, which has not yet been finalized, further complicating our ability to evaluate the GDS policy.

Finally, we note that the existing NIH GDS Policy was finalized in 2014, and there have been exponential advancements in the field since that time, all of which should be considered when developing an updated policy.

Scope. AAMC agrees with the exclusion of non-human genomic data from the GDS policy, although we suggest that NIH provide a specific definition for "human genomic data" to ensure clear and consistent policy implementation. Particularly, the agency should specify the circumstances under which cell lines and/or biospecimens would be included in the definition, since a broad definition would capture genomic data types with variability in identifiability and potential security risk.

We strongly support the provision that individual NIH institutes and centers (ICOs) will not be permitted to "expand the scope of the GDS policy through individual program or policy expectations." However, we also recommend the creation of additional guidance or FAQ from NIH on the additional data protections or use of controlled-access repositories that will be allowed by the ICOs, as this will allow institutions to prepare for the potential range of actions they will need to address when conducting NIH-funded research.

Finally, regarding the proposal that "any amount of human genomic data collected from 100 individuals or more will be defined as 'large scale' and required to comply with the GDS Policy's consent and data sharing requirements," we refer to previous comments from AAMC to the Department of Justice where we note that if the primary intent of this policy is to ensure appropriate privacy and security considerations, the risk profile of genomic data cannot be determined in a volume-based manner by the number of individuals represented in the data. This definition as written would apply to an enormous swath of genomic data projects, which by scientific necessity, require large sample numbers and are analyzed using high-throughput methods.

Timelines for Data Processing. The multiple timelines listed in the proposed policy for data sharing are unclear, and do not take into account the significant time needed to process genomic data. We recommend that NIH simplify the language to state that data should be shared in a manner that is consistent with DMS Policy requirements.

Modernization of Data Submission and Sharing Practices. This section introduces references to existing federal research regulations, which when taken out of context, are ambiguous and extremely challenging to interpret. We particularly cite issues within the section "Strengthening requirements for participant consent." The proposed policy states, more than once, that the policy is consistent with the Common Rule. However, this is not the case. In fact, the draft revisions set forth requirements that in some cases directly conflict with the Common Rule, a complex set of connected and integrated provisions, which already create a framework and requirements for seeking consent from human subjects. The consent requirement in the proposed revisions suggesting that "next of kin" should be consulted for the use of biospecimens from deceased individuals is inconsistent with the Common Rule, which, specifically does not apply to biospecimens from deceased individuals. Further, the draft revisions invoke the concept of requiring assent from minors, a requirement that is not in the Common Rule, which instead stipulates that an IRB should make that determination. Finally, the consent language in the revision is an excerpted definition from the Common Rule and concept known as "broad consent." We suggest, to avoid confusion and the potential for conflict with an existing regulatory scheme, that the GDS policy revision not create new consent standards that invoke but go beyond the Common Rule.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/AAMC-Comments-re-NIH-NOT-OD-26-023.pdf

Description: AAMC comments to NIH re: NOT-OD-26-023 Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

125. Federation of Associations in Behavioral & Brain Sciences (FABBS)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Molly Madzelan

Name of Organization: Federation of Associations in Behavioral & Brain Sciences (FABBS)

Type of Organization: Other

Type of Organization - Other: Non-Profit Association of Scientific Societies

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Federation of Associations in Behavioral & Brain Sciences (FABBS) represents 34 of the nation's leading scientific societies in the psychological, cognitive, and behavioral sciences. Our mission is to advance the sciences of mind, brain, and behavior; promote evidence-based policymaking; and support the integrity and impact of the federal scientific enterprise. FABBS researchers often rely on human participants to study critical public health challenges, such as maternal mortality, youth mental health, and chronic illnesses (e.g., diabetes and heart disease). Thus, we value the opportunity to comment on the National Institutes of Health's (NIH) Draft Controlled-Access Data Policy.

We appreciate that NIH is carefully considering data security during a time of increased data sharing and collaboration between scientists. FABBS agrees with the goal of optimizing open sharing while keeping appropriate data protections in place. We have several concerns about the Controlled-Access Data Policy as it is currently proposed in this notice and want to bring attention to the potential unintended consequences of these changes.

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy.

General Concerns

FABBS is concerned that the draft policy is too broad in scope and treats all data types identically, regardless of the privacy risks posed, which vary substantially. We encourage NIH to instead consider developing a tiered controlled-access framework that better aligns the risk proposed by certain data and the corresponding security required to minimize that risk and properly protect participant data.

In its current form, the draft policy would apply to almost all data collected from NIH-funded research that involves human participants. This policy is built on the Department of Justice (DOJ) rule concerning foreign adversaries' access to Americans' sensitive personal data. However, the data types covered by the policy differ widely in terms of the risks they pose and should be handled accordingly. For example, the draft policy treats "covered personal identifiers" (e.g., contact information such as address or phone number) and "personal health data" (e.g., height, weight) as identical in terms of security needed, despite the former posing far more risk to personal privacy than the latter. Similarly, the data categories themselves are too broad, overlooking distinctions between various kinds of data within these categories. For example, "genomic data" can include whole gene sequencing or single-gene tests, which differ greatly in potential for re-identification.

The current policy does not distinguish between identified and de-identified data, a difference that has important implications for risk. In certain contexts, researchers will collect no identifying information at all. Further, even de-identified data can vary significantly in terms of potential for re-identification. As a result, the application of a single security framework to all these data types creates a mismatch between the risk posed by certain data and the security needed to minimize those risks.

Importantly, in developing this policy, NIH turned to several existing rules and standards aimed at minimizing national security threats posed by access to large-scale datasets containing sensitive information. These include the aforementioned DOJ rule ("Preventing Access to Americans' Bulk Sensitive Personal Data and United States Government-Related Data by Countries of Concern or Covered Persons" (28 CFR Part 202)) and National Institute of Standards and Technology requirements (NIST-SP-800-171). The NIST security standards are entirely appropriate and necessary when dealing with highly sensitive information that could affect our national security interests – however, few datasets coming out of NIH-funded research reach this level of risk. As such, it is ill-advised to apply the same security standards to both. Moreover, by treating all datasets as national security threats, the draft policy places considerable additional burdens on scientists conducting minimal risk research without meaningfully benefiting participant privacy, as researchers and institutions already have safeguards in place to keep these kinds of data secure (e.g., data de-identification, Institutional Review Boards often require Data Use Agreements).

Instead of a one-size-fits-all approach, FABBS encourages NIH to develop a controlled-access data policy in which security standards are proportional with the risk posed by the specific type of data. Perhaps most critically, the framework should factor in whether data are de-identified, how easy it would be to re-identify them, and how identifiable certain pieces of data are (e.g., brain images are not as identifiable as many believe them to be). We also suggest that the agency consider additional factors when aligning security and risk, including, for example: actors (i.e., who could access the data and do they pose a threat?), risk probability (i.e., how likely is it that the data will be accessed by an unsanctioned actor?), and magnitude of harm (e.g., what are the consequences of unsanctioned data access?). Such an approach would allow NIH to develop "right-sized" security frameworks – or borrow others currently in development or use elsewhere (e.g., the Research Security Program required under National Security Presidential Memorandum 33; the policies used by the Inter-University Consortium for Political and Social Research (ICPSR) and OpenNeuro) – that are better suited for academic research data.

Potential Unintended Consequences

FABBS is also concerned about potential unintended consequences of the draft Controlled-Access Data Policy. Before moving forward with revisions or implementation, we urge NIH to conduct a thorough assessment of potential consequences and how to mitigate them. Especially for minimal risk research involving de-identified data, the potential drawbacks outweigh the potential (limited) benefits to participant protection and privacy.

By way of example, the recent NIH decision to eliminate Basic Experimental Studies Involving Humans (BESH) from the definition of clinical trials reveals the value of doing such a pre-implementation assessment. In 2014, the agency redefined clinical trials to include basic behavioral and social science research. As a result, scientists in these fields had to comply with registration and reporting requirements designed for clinical trials rather than tailored for basic discovery research. This proved to

be untenable leading NIH to attempt to remedy the problem with a new classification, but this was still unworkable. Twelve years later, NIH reversed course and will no longer consider BESH to be clinical trials. An initial, more thorough consideration of the possible unintended consequences of changing the clinical trials definition may have saved the agency and researchers from substantial confusion, time, and effort.

The following list of potential consequences is not exhaustive.

- As currently proposed, the draft policy would significantly increase the number of studies requiring controlled-access security. In turn, this would stretch the capacity of the current repository infrastructure as well as increase compliance and administrative burdens on scientists, with little if any benefit to participant privacy.

  - At this time, repository infrastructure is inadequate for handling the expansion of data types requiring controlled access that will occur if the draft policy is implemented. Improving this infrastructure will be difficult due to the required costs and resources, which may be prohibitive to many researchers and institutions. To develop compliant repositories, researchers and/or institutions must have the funds to cover, for example, technology investments, ongoing compliance monitoring, personnel training, and regular security audits. (The NIT-SP-800-171 standards cited in the policy require 110 distinct security controls, formal system security plans, and regular assessment.)

    - Certain institutions – in particular R2 institutions and other Research Colleges and Universities (RCUs), including community colleges – may not have the resources to make or keep repositories compliant. Even the most well-resourced R1 institutions may have trouble funding compliance activities, but they will still have an advantage over smaller, less-resourced institutions in developing and maintaining repositories. As a result, this policy may exacerbate the gap in funding and research output between R1s and other research institutions, especially those that, historically, have been underfunded.

    - The costs involved may also force researchers to recruit smaller samples than originally intended, resulting in less reliable findings due to lower statistical power.

  - In other new policies (e.g., limiting allowable publishing costs), NIH has committed to maximizing the amount of grant funds actually spent on research versus administrative and other costs. The draft Controlled-Access Data Policy, however, will likely further divert funding from research activities to data security and compliance activities as, noted above, individual researchers and institutions may not have the resources to fund these themselves.

    - Recognizing the time and resources necessary to comply with these new standards, the policy risks diverting scientists' time away from research itself.

- There is a risk that the draft policy could slow secondary data analysis by increasing the barriers to data access for many researchers. Secondary data analysis efforts include meta-analyses, reproducibility efforts, and replication studies, the latter two of which are critical to Director Dr. Jay Bhattacharya's vision for the NIH. Limiting secondary data-analyses would further undermine confidence in scientific findings.

  - As noted above, the stricter requirements on data repositories required by the draft policy would limit accessibility to data, especially for researchers from lesser-resourced institutions. The ability

to do data analysis and share data may be further centralized to a few well-resourced institutions that can afford to stay in compliance.

o Researchers-in-training and early career scientists may be uniquely impacted due to their reliance on existing datasets in their work and training.

o The new requirements may also discourage researchers from making their data available to other scientists or may lead them to design studies in ways that avoid overly burdensome compliance activities.

▪ Large-scale data sharing is critical to understanding many health problems, including those intractable disorders that have substantial human and economic impact (e.g., neurological and autoimmune conditions). Reducing or even disincentivizing data sharing will significantly harm our ability to develop cures and treatments.

FABBS is grateful for NIH's work on this issue as well as the opportunity to share our feedback. We are happy to serve as a resource to NIH as it continues to develop this policy.

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/FABBS-Comments-on-NIH-Controlled-Access-Data-Policy-Submitted-03.18.2026.pdf

Description: FABBS' comments in a PDF file.

126. University of Washington

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Carol Rhodes

Name of Organization: University of Washington

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Please see attached letter.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Please see attached letter.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
Please see attached letter.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/University-of-Washington.RFI-response.NOT-OD-26-023.pdf

Description: University of Washington.response to RFI.NOT-OD-26-023

127. N/A

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name:

Name of Organization:

Type of Organization: Academic Institution

Role: Other

Role – Other: Health sciences data librarian, genomicist

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
I submit these comments as both a genomicist and a health sciences data librarian at a research university. I work daily at the intersection of data generation, data management, and researcher support — and I engage with NIH-funded investigators across a range of biomedical disciplines who depend on timely, accessible data to advance their science. I appreciate NIH's stated goals of harmonizing data protections, clarifying requirements, and responding to legitimate national security concerns. The creation of a single, coherent Controlled-Access Data Policy has the potential to reduce confusion that currently arises from inconsistent requirements across NIH Institutes, Centers, and Offices. However, I have significant concerns that several provisions of this proposal, as drafted, would impose disproportionate compliance burdens, restrict data types whose re-identification risk is poorly characterized, strand large volumes of ethically collected legacy data, and further strain an already-limited controlled-access repository infrastructure. Taken together, these provisions risk chilling legitimate biomedical research and undermining NIH's longstanding commitment to open, responsible science. I urge NIH to revise the proposal in the areas described below.

The proposal mandates controlled access for a dramatically wider range of data types without confirming that repository infrastructure can support the demand.

The Draft Controlled-Access Data Policy extends controlled-access requirements to numerous new data categories beyond genomic data, including epigenomic, proteomic, transcriptomic, imaging, and geolocation data, as well as individual-level clinical trial data and personal health data broadly defined. From my position supporting researchers at a major research university, I can attest that the volume of data generated across these categories in NIH-funded research is enormous and growing rapidly as multi-omic study designs become the norm. NIH acknowledges this challenge in its own Request for Input, specifically asking whether additional resources will be needed to meet increased demand. That NIH is asking this question after proposing the policy, rather than before, is itself a concern. A mandate should not precede a resource assessment.

I urge the NIH to:

- Commission and publish a capacity analysis of dbGaP, AnVIL, and other approved NIH Controlled-Access Data Repositories (CADRs) before the policy takes effect, including projected data volume, processing timelines, and access request turnaround times.

- Establish minimum service-level standards for access request review. Researchers, especially trainees and early-career investigators, cannot afford to wait six to twelve months for data access approval. If controlled-access is expanded, the access pipeline must keep pace.

- Provide dedicated supplemental funding to repositories and institutions to offset compliance costs. Smaller institutions and under-resourced research programs will be disproportionately affected by the increased administrative burden.

- Clarify what non-NIH repositories must do to meet the policy's security and operational standards, and publish a list of pre-approved third-party repositories so that researchers who cannot use NIH-operated systems have workable alternatives. In the interim, NIH should consider a phased implementation schedule that prioritizes data types with established infrastructure support (e.g., genomic data in dbGaP) before extending requirements to newer or less-established data categories.

Treating all instances of a data type as equally sensitive ignores the highly variable re-identification risk within each category.

The proposal mandates controlled access for all data within eleven enumerated categories, with essentially no consideration of context, study design, aggregation level, or actual re-identification risk. As a genomicist, I find this approach scientifically untenable. Risk is not a property of a data type in the abstract- it is a function of the data's content, the population studied, the degree of aggregation, and the context of disclosure.

Several specific concerns:

Proteomic and transcriptomic data: The proposal requires controlled access for all proteomic and transcriptomic data derived from human samples. While systems-level omics data can be sensitive in specific contexts, the re-identification risk of routine RNA-seq or mass spectrometry-based proteomics from non-rare-disease, non-family-based cohorts is poorly established and almost certainly lower than that of whole-genome sequencing. Treating these data types identically to whole-genome sequence data will impose substantial costs without commensurate privacy protection. NIH should commission or cite empirical re-identification risk assessments for proteomic and transcriptomic data before including them in the mandatory controlled-access tier, and should consider a risk-stratified framework that accounts for study design, sample size, and population characteristics.

Imaging data of the human face or head regions: The inclusion of this category is understandable given facial recognition concerns, but the definition as written is extremely broad. Structural MRI, functional fMRI, EEG, and PET imaging of the brain are among the most widely shared data types in cognitive and clinical neuroscience, and are currently shared openly through platforms such as OpenNeuro. Requiring controlled access for all head imaging would effectively shut down open neuroscience data sharing as it is currently practiced. NIH should clarify that de-identified, defaced neuroimaging data, where facial features have been algorithmically removed using tools such as pydeface or mri_deface, is not subject to mandatory controlled access. Guidance should also be provided on accepted defacing standards that satisfy the policy's intent.

Personal health data: The definition of personal health data in the Appendix is extraordinarily expansive, encompassing height, weight, vital signs, exercise logs, allergy information, and data on the purchase of medications. Under this definition, nearly any clinical or health survey dataset would require controlled

access. This will capture a vast quantity of data that is currently shared openly with no meaningful privacy risk and for which open sharing is scientifically valuable. The NIH should establish meaningful thresholds or carve-outs for non-sensitive health data collected in large population studies, particularly data that has been collected under informed consent specifically permitting open sharing.

Rather than a binary controlled/open categorization by data type, the NIH should adopt a risk-tiered framework that classifies data based on assessed re-identification risk (taking into account data type, study design, population, and aggregation level), the terms of informed consent, and institutional data governance review. This approach would better reflect the scientific reality of modern biomedical data and avoid the chilling effects of overbroad categorical mandates.

Requiring affirmative consent for all post-2015 genomic data collected from biospecimens will exclude vast quantities of ethically sound data from sharing.

The proposal states that human genomic data collected under the GDS Policy from biospecimens or cell lines created or collected after 2015 must have affirmative opt-in consent for use and sharing; if consent has not been obtained, the data cannot be shared. As someone who regularly works with researchers trying to navigate data sharing requirements for legacy cohort studies, I consider this provision one of the most consequential and potentially damaging in the entire proposal. Many prospective cohort studies, biobanks, and clinical research programs that began enrolling participants after 2015 obtained IRB-approved consent that was entirely appropriate under the ethical standards of the time, but that did not use the specific opt-in language now required. These studies cannot feasibly return to participants to re-consent, particularly for deceased participants, those lost to follow-up, or rare disease cohorts where population sizes are already small.

I raise several specific concerns:

- The hard cutoff of 2015 is arbitrary. The NIH GDS Policy took effect in January 2015, but implementation across institutions was uneven, and many institutions did not update consent templates immediately. A rigid cutoff will penalize institutions that were in the process of updating their practices in good faith.

- No exception is provided for data where re-consent is demonstrably infeasible, such as data from deceased participants or participants with advanced illness. The proposal acknowledges consent from next-of-kin for deceased participants, but does not address the large volume of data for which no next-of-kin is available or identifiable.

- This provision creates a perverse incentive: researchers who collected data carefully under strong IRB oversight but with slightly different consent language are penalized relative to those who collected data under weaker but technically compliant consent forms.

I urge the NIH to:

- Establish a formal infeasibility waiver process, modeled on the Common Rule's existing provisions, by which institutions can demonstrate that re-consent is not practicable and obtain approval for sharing based on a risk-benefit assessment.

- Clarify that institutional data governance bodies (not only IRBs, who are not trained on data management and sharing) may conduct these reviews, consistent with the proposal's expanded institutional review provisions.

- Provide a specific transition period of at least 24 months during which existing studies can assess their consent landscape, pursue re-consent where feasible, and apply for waivers where it is not.

- Publish model consent language and retrospective consent amendment templates to assist institutions in updating existing consent documentation where possible.

The genomic datasets at risk from this provision represent decades of participant time, research investment, and public funding. NIH should exhaust every mechanism for preserving their utility before imposing a categorical sharing prohibition.

The requirement that imputation servers be funded or operated by NIH or a federal agency is overly restrictive and inconsistent with how modern genomic research is conducted.

I support NIH's goal of ensuring that imputation servers operating on controlled-access data meet robust security and privacy standards. Imputation is a powerful technique with real privacy implications and poorly secured imputation panels can expose information about the reference individuals. I appreciate that NIH is seeking to extend thoughtful guidance to this area. However, the current proposal would require that imputation servers be funded or operated by NIH or another federal agency. This is far more restrictive than what is necessary to achieve the stated privacy and security goals, and it ignores the substantial investment that academic institutions, non-profit consortia, and international collaborators have made in building secure, high-performance imputation infrastructure. Well-established imputation servers such as the Michigan Imputation Server and others operated by non-federal academic institutions already implement security controls comparable to those required by NIH's own Best Practices, including encrypted data transfer, user authentication, non-transferability enforcement, and data deletion after use. Excluding these platforms from permissible use without cause does not strengthen privacy protection; it simply reduces access to high-quality imputation resources.

I recommend that NIH:

- Replace the federal-operation requirement with a certification framework. Any imputation server (regardless of operator) should be eligible for use with controlled-access data if it undergoes and passes a formal certification review against NIH's Security Best Practices for Controlled-Access Data Repositories.

- Establish a public registry of certified imputation servers, with regular recertification requirements, so that researchers can identify approved options and NIH can maintain oversight without centralizing all operations in federal systems.

- Engage with privacy-enhancing technology (PET) researchers and the broader computational genomics community to define appropriate technical standards for next-generation imputation approaches, including federated learning and secure multi-party computation, before regulatory language is finalized.

- Provide clear guidance on what constitutes a 'privacy attack' specific to imputation servers (e.g., membership inference, reference panel exposure) and what countermeasures satisfy NIH's

requirements. The current proposal is vague on technical specifics in ways that will create uncertainty for server operators.

The goal of protecting reference panel privacy is legitimate and important. But the mechanism proposed — limiting operation to federal agencies — is both unnecessarily narrow and unlikely to remain technically current as the field evolves. A certification-based framework is more durable, more inclusive of existing high-quality infrastructure, and better aligned with how the research community actually works.

In conclusion, I strongly support NIH's commitment to responsible human data stewardship and appreciate the effort to harmonize data protections across a complex and evolving policy landscape. The goals of this proposal are sound. But in several important respects, the means chosen are too blunt, too broad, or insufficiently attentive to the practical realities of research data management and the infrastructure available to implement these requirements.

I urge the NIH to:

- Conduct and publish a repository capacity assessment before the policy takes effect, and phase implementation to match available infrastructure.
- Adopt a risk-tiered framework for controlled-access designations rather than categorical mandates by data type, and provide empirical evidence for the re-identification risks attributed to each category.
- Establish a robust infeasibility waiver process for legacy data, with clear criteria and a meaningful transition period.
- Replace the federal-operation requirement for imputation servers with an open certification framework based on demonstrated adherence to NIH security standards.

These revisions would preserve the policy's protective intent while ensuring that the research community (and the patients who have invested their data in that research) can continue to benefit from broad, responsible, and efficient data sharing.

Thank you for the opportunity to comment.

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

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

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

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

128. National Cancer Institute (Trans-NCI Data Management and Sharing Working Group)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Emily Boja

Name of Organization: National Cancer Institute (Trans-NCI Data Management and Sharing Working Group)

Type of Organization: Other

Type of Organization - Other: Government

Role: Other

Role – Other: Coordinator

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Concerns were raised regarding the implementation of the CADR policy at non-NIH institutions due to the security requirements. More clarifications are needed as to whether other data types alone need to be protected under controlled-access, and whether there is any threshold for those data types that are not genomics.

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Proposed data types other than genomics under controlled-access need to be evaluated at a more granular level. For example, aggregated somatic variants (MAF files) derived from genomic data, a differentially expressed gene matrix table processed from transcriptomic data, or a list of identified proteins with relative quantities from raw mass spec data can be open access due to low re-identification risks even if the original data are sensitive and more identifiable that need protected under controlled-access.

Additional clarification is needed in terms of whether these other data types that are proposed to be under controlled access are considered as the sole data types, or in combination with genomic data since multi-omic or multi-modal data have been postulated to be more identifiable.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
1. Appropriateness of the threshold for “Large-Scale” Genomic Data (Proposed: ≥100 Individuals):

• The proposed threshold of 100 individuals may be too low to apply uniformly across all research contexts.

• However, in rare cancers (e.g., childhood cancers with 20–50 cases annually), datasets representing 100% of a disease population may fall below the threshold and yet have extremely high scientific value, but with higher re-identification risks.

In addition, it was unclear whether this proposed threshold applies to other proposed data types for CADRs, such as proteomics, metabolomics, or personal identifiers.

• Concern was raised that an absolute numeric threshold may not adequately account for disease prevalence versus re-identification risk.

• Consideration: Should thresholds incorporate a percentage of the affected population rather than a fixed number?

2. Scientific Value vs. Administrative Burden (GPA, PO, and Repositories):

• For genotyping datasets (e.g., imputation-based datasets), scientific utility is typically realized at much larger scales.

• Requiring submission of every dataset involving ≥100 individuals could create significant administrative and repository capacity burdens.

• A need to balance scientific value with operational feasibility.

• The proposal appears to place responsibility on POs to assess whether repositories outside NIH adhere to policy standards. Concern was expressed that this could substantially increase PO workload and shift administrative responsibilities.

3. Data Depth vs. Number of Individuals: For example, how thresholds will account for high-dimensional datasets (e.g., single-cell sequencing) that will provide large volumes of data generated from small numbers of individuals?

4. ICO-Level Implementation Flexibility: Concern was raised that the proposed language may limit ICOs' implementation flexibility, despite programmatic differences and varying mission needs.

5. Genomic vs. Non-Genomic Data:

• Genomic data must be registered in dbGaP although data submission and storage may occur in other/external (CADRs, such as CRDC (NIH or non-NIH-supported that meet the NIST standards).

• Proteomics and other non-genomic molecular data have fewer requirements when repositories are publicly accessible currently.

• The differing treatment of genomic data compared to other molecular data types was viewed as inconsistent and potentially confusing, e.g., will the thresholds apply to other data types? What about single-cell data?

6. Timeline Changes: Proposed revision moves from the currently staggered timelines to a flat 6-month submission requirement for all data levels.

Concerns:

• May push additional workload towards the time of publication.

• PIs frequently request accession numbers urgently for manuscript submission, adding potential administrative strain on GPAs and support staff.

• Need to evaluate feasibility and downstream impacts.

Clarifications are needed regarding the proposed 100-individual threshold, potential loss of ICO flexibility, administrative burden, repository compliance expectations, and revised timelines. There is a strong interest in ensuring that policy updates:

• Reflect scientific context and disease prevalence.

• Balance data sharing goals with operational feasibility.

• Preserve appropriate ICO discretion.

• Provide clarity on non-human data and emerging technologies.

• Avoid unintended administrative burdens on POs and staff.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
Establishing criteria for a secure and controlled environment for imputation servers is a great step to protect sensitive information. Concerns were raised about whether imputation servers can meet all proposed criteria due to infrastructure compliance requirements and operational burden.

129. Global Down Syndrome Foundation

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Bryn Gelaro

Name of Organization: Global Down Syndrome Foundation

Type of Organization: Other

Type of Organization - Other: Non Profit Advocacy Organization

Role: Other

Role – Other: Employee at Advocacy NPO

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

- We value efforts to protect participant privacy.
- Open-Access data sharing has been integral to the rapid proliferation of Down syndrome research clinical trials and publications in the past 8 years.
- Controlled Access Policy reinforces disparities between institutions who do not have the team to meet the increasingly complex standards.
- Data is already de-identified and meets HIPPA Compliance standards. The risk level for re-identification differs between the human data types outlines.
- Public access to data is one way NIH can give back to the community, especially participants. Controlled Data stands to make the research participation process less transparent and disincentivizes participation. Public access should be protected, especially for publicly tax funded projects, such as NIH research.
- Sharing and reanalyzing data is important; limiting data sharing could be costly and could impact reproducibility

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

- The proposed policy requires additional layers of review and vetting. We value protecting the privacy of participants, however, it's unclear if the repository or HRPPs will be able to function in a timely manner under increased demand/volume.

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

- Some of the data types included or more high risk than others. It would seem appropriate to establish more individualized protections based on a clear risk-benefit analysis.

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

- Threshold for "large scale" being reduced to 100 may be too burdensome.

- Consistent requirements across all NIH Institutions seems administratively burdensome. Unique populations like Down syndrome could
- Open-sharing is favored by many of the top journals. Controlled-Access with conflict with publication requirements and limit publication opportunities, thus slowing dissemination.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/2026-RFI-NIH-Data-Sharing-LOS-03-17.docx

130. The Regents of the University of California, Office of the President

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Brian Russ

Name of Organization: The Regents of the University of California, Office of the President

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
see attached

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/UC-Comment-Letter-on-NIH-Controlled-Access-Data-and-Genomic-Data-Policies_finalsubmission.pdf

Description: The letter details the University of California's recommendations regarding the subject RFI on NIH's draft Controlled-Access Data Policy and proposed genomic data-sharing policy revisions, focusing on definitions, scope, consent, security, costs, and implementation.

131. Robert Carroll

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Robert Carroll

Name of Organization: Vanderbilt University Medical Center

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Regarding open access data, it is unclear what "openly sharing these data pose very low risk when shared and used". Risk of reidentification? Of harm to the consented participant? Of national security? Very low risk is stringent and likely prohibitive (even to quantify) depending on the "risk of".

The policy mentions CADRs and NIST-SP-800-171, but the standard policy is 800-53. This policy should be considered with that substantial resource investment in mind- hundreds of thousands of dollars a year just to prove compliance.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
Increasing compliance burden (eg, by requiring more stringent NIST 800 standards) and increasing what data types are covered makes it challenging to meet communities' needs around portals and data distribution. If everything that shows data must be a CADR and require data access approvals, it will be become increasingly infeasible to support research communities without resorting to "here is your secure download link" only. It's possible this will stifle innovation of platforms and researchers.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
Stringent limitations on personal health data could lead to the inability to even describe a dataset in detail, eg, public metadata about a cohort of consented participants describing them as Type 2 Diabetes cases could be disallowed.

It's unclear why functionally all individual-level data used in research is considered sensitive. It does not seem necessary, and will likely slow discovery as openly available data greatly speeds time-to-analysis, by reducing administrative barriers and enabling more tools to provide direct access to data. However, if that is the requirement, is it possible to make explicit specific exceptions to support users understanding the nature of data in the datasets or finding data?

Many of these data types, sufficiently de-identified, do not appear to present risk of a serious adverse effect to participants (or to operations/assets). That's the standard at which a Moderate Impact system is ruled, which is the default for CADRs. While researchers and repositories are committed to protecting participant privacy and consent, it doesn't follow that step counts and weights would merit this level of compliance oversight (as an example).

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
There are overlapping policies around de-identification introduced in the appendix when compared to

HIPAA de-identification, eg, around geographic identifiers. The "precise geolocation data" requirement of 1000m is also far more specific in many cases than the HIPAA rules.

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

132. Pittsburgh Supercomputing Center

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Philip Blood

Name of Organization: Pittsburgh Supercomputing Center

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/PSC_Response-NIH_RFI_NOT-OD-26-023-1.pdf

Description: Please use this copy. I sent one earlier without my name or organizational name in the submission field.

133. Franco Pestilli

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Franco Pestilli

Name of Organization: brainlife.io

Type of Organization: Other

Type of Organization - Other: NIH Data Infrastructure

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
We strongly support NIH's commitment to protecting research participants in an era of increasingly rich and computationally tractable data. At the same time, we are concerned that the current draft adopts a categorical framework that is not sufficiently aligned with the structure of modern neuroscience data. In particular, the designation of "imaging data of the human face or head regions" as presumptively requiring controlled access is overly broad and does not reflect the heterogeneity of neuroimaging data types.

We respectfully recommend that NIH adopt a risk-based framework that distinguishes between identifiable imaging data, properly de-identified datasets, and derived outputs with negligible re-identification risk. Current scientific practice already includes removal of direct identifiers, defacing of anatomical scans, and metadata de-identification under institutional oversight. Once these steps are taken, the residual risk profile changes substantially and should not be treated as equivalent to raw identifiable data.

We further recommend that NIH clarify how risk should be assessed in practice, including the role of metadata, linkage potential, and realistic re-identification pathways. The current draft leaves key operational questions unresolved, which may lead institutions to adopt overly conservative interpretations that unnecessarily restrict access. Finally, we encourage NIH to distinguish between empirically demonstrated risks and speculative risks associated with emerging computational methods, and to ensure that policy decisions are grounded in evidence.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
We emphasize that many existing NIH-supported infrastructures, including platforms such as brainlife.io, are not traditional repositories but integrated computational environments that combine data storage, analysis, and provenance tracking. These systems have been explicitly designed to enable reproducibility, data reuse, and scalable scientific workflows.

As currently written, the proposed policy would require substantial redesign of these infrastructures to comply with broad controlled-access requirements and security standards such as NIST SP 800-171. These changes would affect not only data access mechanisms but also computational workflows, user environments, and institutional compliance processes. In many cases, this would place prior federal

investments in open scientific infrastructure at risk and reduce the functionality that enables reproducible science.

We also note that the capacity to implement such requirements is uneven across institutions. While well-resourced institutions may be able to absorb the additional compliance burden, many others may not. We therefore recommend that NIH ensure that implementation pathways remain feasible for a wide range of repositories and institutions, and that compliance requirements are proportionate to the actual risk profile of the data.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
We respectfully submit that the current designation of all imaging data of the human head or face as requiring controlled access is not appropriate as a general rule. Neuroimaging data encompass a wide range of modalities and processing states, each with distinct risk characteristics.

Structural MRI data that retain reconstructable facial anatomy may warrant stricter protections, particularly when linked with sensitive metadata. However, this rationale does not extend uniformly to defaced anatomical images, functional MRI data, diffusion MRI data, or derived products such as statistical maps, connectomes, and regional summaries. These data types do not meaningfully preserve facial identity information and, in many cases, present minimal re-identification risk.

We recommend that NIH adopt a framework that evaluates data sensitivity as a function of data type, processing status, and metadata context, rather than relying on a single categorical designation. Such an approach would allow controlled access to be reserved for genuinely sensitive datasets while preserving broader access where risk is demonstrably low.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We encourage NIH to ensure that revisions to the Genomic Data Sharing Policy remain aligned with the principle of proportionality between data sensitivity and governance requirements. In particular, we caution against the direct application of governance models and security standards developed for one data domain to other domains with fundamentally different risk profiles.

The experience of the neuroscience community demonstrates that large-scale data sharing, when implemented with appropriate safeguards, enables reproducibility, secondary analysis, and methodological innovation. We recommend that policy revisions preserve these capabilities by supporting tiered access models and avoiding unnecessarily restrictive default classifications.

We also emphasize the importance of maintaining adaptability in policy design, as both data modalities and computational methods continue to evolve. Policies that are evidence-based, risk-aware, and responsive to scientific practice will be more effective in achieving both participant protection and scientific progress.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
We recommend that updates to the GDS Policy for imputation servers reflect the specific operational and risk characteristics of these systems. Imputation servers often function as controlled computational intermediaries, where data are processed within defined environments rather than broadly distributed. As such, their governance should not default to the same level of restriction applied to primary sensitive datasets without consideration of their actual risk profile.

We further encourage NIH to support the development and use of federated and distributed computational models, in which data remain under local control while enabling secure and interoperable analysis. These approaches can reduce the need for large-scale data transfer while maintaining scientific utility and protecting participant privacy.

More broadly, we recommend that governance of imputation services be integrated into a tiered, risk-based framework that aligns access controls with demonstrated risk, supports technical innovation in secure computation, and preserves the scalability and usability of these essential research tools.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/brainlife.io-Response-to-NIH-Request-for-Information.pdf

Description: brainlife.io response to NIH RFI on Data Policy

134. Erica Jonlin

Submit date: 3/18/2026

I am responding to this RFI: On behalf of myself

Name: Erica Jonlin

Name of Organization: University of Washington

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
The Policy needs to be explicit in what genomic information is considered "identifying" as we see this as the foundational motivation for the NIH data repositories to be controlled access.

The data that we upload to NIH repositories include human DNA sequence of select genes, occasionally whole genome sequencing, transcriptome analysis (including single cell), and epigenomics studies. We never disclose or upload to databases personal identifiers (either "covered personal identifiers" or "listed identifiers"), precise geolocation data, biometric identifiers, or imaging data of the human face or head regions. We do upload the above-mentioned genomics data and some linked clinical information and demographics data such as sex, stated ethnicity, and age at diagnosis (if relevant). In some cases, the risk of re-identification is extremely unlikely, although the risk increases as technology improves.

This leads us to our first question which we'd like NIH to address in the new policy - please specify which of the following the NIH considers potentially identifying:

i. Gene sequences (usually not whole genome sequence) - often cancer cell sequences, not germline; often only a few genes that have been deeply sequenced

ii. Transcriptomics (may or may not be whole transcriptomes) - frequently these are cancer cell transcriptomes; sometimes we perform spatial transcriptomics (this is not sequence)

iii. Epigenomics

iv. Proteomics

v. Mitochondrial DNA

Usually, only one or two of the above are uploaded

We would like NIH to specify which of the above, including which combination of the above, places a subject at risk of re-identification and why. References would be helpful.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
We understand that the NIH proposes to "simplify" the definition of large-scale genomic data, "setting a clear threshold of data from 100 or more individuals."

i. Please explain what the concern is. How is this large-scale genomics data? Is the NIH thinking of unique populations?

ii. What is the NIH's position on providing sequence data from fewer than 100 individuals? What if subjects have a rare disease?

iii. What if the sequence data is only from cancer cells and is not germline?

iv. Journals require sharing of all the data presented. Why does GDS need to refer only to "large scale"? Shouldn't the GDS policy apply to genomic data from any number of individuals?

We sequence only a few genes from each individual subject specimen. We seek guidance on how many genes fall UNDER the GDS threshold. And specifically, how many genes would be considered to be identifying?

The RFI states, "Human genomic data collected under the GDS Policy from biospecimens or cell lines created or collected after 2015 must have consent for use and sharing. . . If consent has not been obtained, the data cannot be shared." We note that data acquired from leftover archived clinical samples collected after 2015, in IRB-approved protocols, would thus not be shareable. Often, journals ask that data be made available via NIH databanks, including for studies not funded by the NIH. This could represent a significant loss of communication of important findings. If genomic data is allowed to be posted only in a restricted/controlled database, does that not overcome the risk of reidentification and inappropriate sharing, even if there was no informed consent, or incomplete informed consent (e.g., the consent does not include elements such as risks to relatives)?

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Jonlin-Response-to-NIH-RFI-NOT-OD-26-023.pdf

Description: Letter compiling Erica Jonlin's responses to the Request for Information on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

135. OpenNeuro Data Archive

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Russell Poldrack

Name of Organization: OpenNeuro Data Archive

Type of Organization: Other

Type of Organization - Other: Data Archive

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
As a research group that maintains and operates one of the largest platforms for openly sharing MRI and other brain imaging data, OpenNeuro.org, we are the embodiment of NIH's longstanding commitment to a simple principle: scientific data generated by NIH-funded research should be shared with the broader scientific community to catalyze novel biomedical discoveries. Open and responsible data sharing has played a central role in advancing neuroscience and biomedical discovery, and we vigorously agree that these practices must be implemented in ways that appropriately protect the privacy of research participants. We applaud the draft NIH Controlled-Access Data Sharing Policy for its effort to harmonize policy across multiple federal directives, as well as its explicit consideration of data management throughout the data lifecycle.

At the same time, aspects of the proposal raises serious concerns. In particular, the proposal's requirement that "Imaging data of the human face or head regions" be considered "protected" by default is operationally underspecified, as is the evidentiary basis for making a determination that controlled access is necessary to mitigate privacy risks associated with those data. As drafted, the policy's categorical inclusion of "Imaging data of the human face or head regions" as a protected data type - regardless of whether they are robustly defaced/anonymized and shared under security-aware open-science norms - would impose administrative and compliance costs on investigators and institutions that would be sufficient to collapse the open neuroimaging ecosystem.

In the comments below, we elaborate on these concerns in greater detail and provide alternative recommendations that we believe accomplish the same goals without incurring the costs to investigators, institutions and the reliability of scientific inference .

1) Uncertainties regarding the definition of "imaging data of the human face or head regions"

See Section 3 below for feedback regarding the designation of imaging data of the human face and head regions as a category of data required to be protected under controlled access.

2) Uncertainty Regarding the Empirical Basis for Requiring Controlled Access

The proposed policy does not provide a clear justification for including "imaging data of the human face and head regions" among the data types required to be placed under controlled access. However, the policy does outline criteria for determining whether other, non-listed data types should require controlled access, which may provide insight into the rationale for designating certain imaging data as

requiring controlled access. According to the proposed policy, such determinations may consider the following factors:

- Explicit limitations on subsequent use, such as those imposed by laws, regulations, policies, informed consent, and agreements.

- Potential sensitivities, such as information regarding potentially stigmatizing traits, illegal behaviors, or other information that could be perceived as causing group harm or used for discriminatory purposes. Sensitive data may also include data from individuals, groups, or populations with unique attributes that increase the risk of re-identification.

- Lack of adequate data de-identification or the possibility of re-identification cannot sufficiently be reduced.

Aside from cases in which explicit limitations on data sharing or secondary use apply, the inclusion of "imaging data of the human face and head regions" among the categories designated for controlled access appears to imply that such data are presumed either to contain potentially sensitive information or to lack sufficient deidentification to adequately mitigate the risk of reidentification.

Some imaging datasets may indeed contain highly sensitive information or remain highly identifiable, particularly when collected from clinical populations or very small and distinctive groups, even after conventional deidentification procedures are applied. In these circumstances, requiring controlled access may be justified to mitigate potential privacy risks. Nevertheless, this consideration alone does not justify applying controlled-access requirements to all imaging data of the human face and head regions. Although there is currently no scientific consensus regarding the privacy risks associated with sharing different types of brain imaging data, the available empirical evidence does not indicate a level of risk that would warrant such a broad restriction.

The proposed policy appropriately recognizes re-identification risks. As discussed above, facial features in structural MR images can be used to reidentify data subjects, and thus neuroimaging researchers have removed facial features before sharing to reduce reidentification risk. Some studies have demonstrated that reidentification may still be possible using advanced face-recognition algorithms even when structural scans have been defaced (e.g., Schwarz et al., 2019, N Engl J Med; 2021, Neuroimage). However, most empirical demonstrations of reidentification from neuroimaging data have been conducted under controlled experimental conditions involving relatively small populations and reference datasets containing similar types of imaging data. As such, these findings cannot be directly extrapolated to the likelihood of reidentification in real-world settings. A recent simulation analysis conducted by our group suggests that the real-world probability of reidentifying individuals from defaced neuroimaging data using face-recognition approaches is substantially lower than estimates suggested by prior experimental studies, and may be unlikely to pose a practical concern in real-world settings (Jwa, Koyejo, & Poldrack, 2024, Imaging Neuroscience).

Further, the empirical grounds for including functional imaging (i.e. imaging using a contrast that does not contain information about facial geometry, for example, T2* images used in functional magnetic resonance imaging or fMRI) among the data types required to be placed under controlled access is also not clear. The inclusion of functional imaging modalities, such as functional MRI (fMRI), may be overly broad if the underlying concern motivating the policy is the risk of reidentification through facial

reconstruction. For functional MRI or diffusion MRI, the endogenous contrasts used to acquire these data generally do not resolve facial geometry to a degree that would permit reconstruction of identifiable facial anatomy.

Instead, the primary privacy concern associated with functional neuroimaging relates to the possibility that distinctive patterns of brain activity or connectivity could be used to match datasets from the same individual across studies (often referred to as connectome fingerprinting). However, such cross-dataset matching does not itself reveal an individual's real-world identity. Unlike structural MRI—which may permit reconstruction of facial anatomy comparable to a photographic image—functional activation patterns do not inherently contain identifiable personal information.

As noted by Eke et al. (2021, NeuroImage: Reports), identification based on functional data would require access to a separate database that explicitly links brain imaging data to identifiable individuals, or voluntary public disclosure by the individual themselves. This suggests that unique functional signatures represent a secondary rather than primary reidentification risk. Similarly, Clunie et al. (2023, ArXiv) and Wachinger et al. (2015, Neuroimage) observe that, given the absence of widely accessible population-scale brain databases linked to verified identities, the practical feasibility of reidentification through such matching remains limited.

Another emerging privacy concern related to functional MRI involves the possibility that advances in artificial intelligence and machine learning could enable new forms of inferential analysis of neuroimaging data. Through reanalysis or reprocessing of existing datasets, it has been suggested that brain imaging data might conceivably reveal information about an individual's future health condition, mental state, or behavioral tendencies that were not anticipated at the time of data collection. Some have speculated that these inferences may include the sensitivities outlined in the proposed policy's criteria to determine whether data should be protected under controlled access (e.g., potentially stigmatizing traits, illegal behaviors, or other information that could be perceived as causing group harm or used for discriminatory purposes).

Speculation aside, the empirical evidence supporting such risks remains limited. The current state of neuroimaging methods suggests that the likelihood of this threat is extremely low. Recent methodological reviews indicate that much of the existing literature relies primarily on in-sample statistical associations, rather than demonstrating robust, out-of-sample and generalizable individual-level predictions (Poldrack et al., 2021, JAMA Psychiatry). In many cases, validation procedures are limited, and model performance does not reliably generalize beyond the specific experimental context or study population in which the models were developed.

It is therefore important to clearly distinguish between speculative future risk and currently demonstrated empirical risks in the proposed policy. Policies that do not draw this distinction may unduly restrict open science practice based on hypothetical capabilities rather than evidence-based assessments of privacy risk. Moreover, such inferential risks would generally constitute a secondary privacy risk, as they could only materialize if neuroimaging data were first linked to an identifiable individual through other means.

We would also like to emphasize that privacy risk assessments for neuroimaging data should account for not only the likelihood of re-identification but also the magnitude of harm that re-identification would cause (Meyer, 2018, AMPPS). In most cognitive and psychological neuroscience studies, data are

collected to investigate neural mechanisms underlying cognitive processes and typically do not include sensitive information that could expose participants to significant material or reputational harm. Thus, the overall privacy risk remains very low when both the likelihood of reidentification and the magnitude of potential harm are considered.

In assessing re-identification risks, it is important to consider not only the technical feasibility of such attacks, but also the incentives and cost–benefit considerations for potential actors. Much of the current discussion focuses on what is technically possible, yet the likelihood of re-identification in practice depends critically on whether there is meaningful motivation or gain for a bad actor relative to the effort required (Meyer, 2018, AMPPS). In the context of openly shared neuroimaging datasets (e.g., OpenNeuro), there is limited evidence to suggest that re-identification would offer substantial value to malicious actors, particularly when weighed against the technical challenges and costs involved.

As currently framed, the designation—particularly with respect to defaced structural MRI scans and functional brain imaging data—does not appear to be based on a clear assessment of the balance between potential privacy risks and the benefits of data sharing.

Recommendations: (1) Given the current lack of empirical evidence supporting the need to impose controlled-access requirements for all "imaging data of the human face or head regions," we recommend that NIH reconsider the inclusion of this broad category of data under the proposed policy. (2) We urge the NIH to articulate a clear policy distinction between imaging data that does not contain reconstructable facial geometry, imaging data that contains reconstructable facial geometry, and imaging data that has been processed to remove potentially identifying facial features. (3) We encourage the NIH to explicitly consider imaging data that does not contain facial geometry (e.g. standard T2*-weighted images) and imaging data in which facial geometry has been removed to prevent identification as "very low risk" by default. Without this clarification, institutions and IRBs will predictably take the most conservative stance ("all head imaging is controlled"), because the policy makes controlled access the default unless explicit open-consent and "very low risk" determinations are met. However, while the policy implies a risk-based review could allow open sharing if "very low risk" criteria are met, it never explicitly defines what "very low risk" means, who determines it, or what technical criteria satisfy it.

3) Implementation challenges for the institutional risk assessment requirement

The proposed policy will introduce new administrative and compliance burdens for data repositories and research institutions responsible for implementing its requirements. Under the policy, the listed data types may be shared without access controls if informed consent explicitly states that the data are to be shared openly without access restrictions, in which case institutions must still review and determine that open sharing poses very low risk when the data are shared and used.

Under the Common Rule, informed consent must specify whether the data collected as part of the research may be used or shared for future research (45 C.F.R. §46.116 (b)(9)). When investigators intend to share data openly, it is also common practice in the neuroimaging community to explicitly state in consent materials that the data will be deidentified and shared under fully open access (e.g., through initiatives such as Open Brain Consent). However, the additional requirement under the proposed policy for institutions to conduct a separate review to determine that openly shared data pose "very low risk" could impose substantial new administrative burdens on both researchers and institutions.

Notably, informed consent and institutional review were also identified as relevant considerations in the Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data (NOT-OD-22-213). During the development of that guidance, public comments raised concerns about the potential burden on institutions to certify that datasets had been appropriately deidentified and posed a very low risk of reidentification. Commentators also questioned how such reviews would be integrated into existing oversight processes and which institutional body would be responsible for conducting them. Similar concerns may arise under the current proposed policy requirement for institutional review prior to open data sharing.

More specifically, the supplemental information did not designate a specific institutional office responsible for conducting such reviews, noting instead that NIH did not intend to prescribe specific features of the institutional review process and encouraged a flexible approach for institutions. Similarly, the draft policy provides little guidance on how this review requirement should be implemented in practice. However, in the absence of clear guidance, identifying or establishing an appropriate institutional office with the expertise needed to assess reidentification risks—which may vary across data types and depend on the availability of other datasets that could potentially be linked—may prove challenging and could create significant administrative bottlenecks that delay or hinder data sharing. Moreover, without sufficient expertise to evaluate these risks, institutions may adopt a more risk-averse approach and unnecessarily restrict open sharing and beneficial secondary uses of data, even in cases such as defaced structural MR scans or functional imaging data for which current empirical evidence suggests relatively low risk of reidentification.

Recommendations: NIH should operationalize “very low risk” to reduce compliance ambiguity. The current draft allows open sharing of otherwise protected data types only with explicit consent and institutional determination that open sharing poses “very low risk,” but offers no rubric for making that determination. NIH should provide a short, auditable decision framework (even as an FAQ or guidance) that includes: clear examples for neuroimaging (defaced NIfTI vs non-defaced T1/T2 vs raw DICOM), examples of “safe” vs “high-risk” metadata bundles, and a recommended documentation standard that investigators can reference in DMS Plans and repositories can enforce consistently.

4) Administrative and compliance burdens of controlled-access data sharing requirements

The proposed policy also imposes substantial additional burdens on neuroimaging data repositories responsible for implementing controlled-access requirements, which may be disproportionate given the currently understood real-world risk of reidentification. Under the draft policy, repositories designated to host controlled-access data must meet several minimum requirements, including:

- Prospective review of requests to access controlled data;

- Authentication of the identity of data requesters;

- Restrictions for sharing data with countries of concern as identified in Part 202; and

- Employing security standards for protection of controlled data (e.g., NIST-SP-800-171 or equivalent).

We envision three major issues raised by this burden: Cost escalation and inequity, reduced sharing, and lower reliability. First, controlled-access workflows (access committees, DUAs, identity proofing, security audits, logging, user support) introduce recurring costs that are significant for individual PIs and

potentially existential for some repositories. These costs will disproportionately burden smaller labs and under-resourced institutions, and could reduce participation in NIH-funded data sharing. Second, when the parameters of compliance are both ambiguous and burdensome, investigators and/or institutional decision-makers may “minimize risk” by sharing less - creating the same closed system of scientific production that NIH’s DMS objectives were designed to open. Neuroimaging research benefits from the broad reuse of neuroimaging data (Milham et al., 2018, Nat. Commun.; Markiewicz et al., 2021, eLife), which enables reproducibility checks, advances methods development, facilitates discovery, and generally maximizes the societal benefit of human subject participation (which is an obligation of researchers under the Belmont Principles; Brakewood & Poldrack, 2013, Neuroimage). If defaced imaging is pushed into controlled access by default, access frictions will measurably reduce downstream reuse - dramatically and unnecessarily reducing return on investment (ROI) for NIH-funded neuroimaging data.

Recommendation: To balance concerns about participant privacy with the concerns raised above, we recommend that NIH require repositories to adopt a hybrid model for neuroimaging that preserves open sharing where risk is demonstrably low, while placing truly sensitive linkages behind controlled access. A first step would be to define a Defaced/Anonymized “Neuroimaging Safe Harbor” eligible for open access (see above Tier recommendations). NIH should specify that neuroimaging is eligible for open sharing when it meets objective criteria that materially reduce re-identification risk, for example: defacing/anonymization that removes facial surface geometry and other directly identifying features (with documented QC/validation steps), removal of direct identifiers from headers/sidecars and exclusion of raw DICOMs when they contain embedded identifiers, and risk-aware accompanying metadata - limiting or de-risking potentially identifiable demographic data (e.g., top-coded or binned age, broader geography bins) when feasible and consistent with scientific aims. This would preserve responsible open sharing for imaging data that is widely used for method development, reproducibility, and discovery while aligning with the policy’s approach to reducing risk. Second, we suggest that NIH designate low-risk imaging datasets as very-low risk by default, even when they may link to high-risk metadata that are protected. We encourage NIH to explicitly recognize that the highest re-identification and harm risk in many neuroimaging datasets often comes from metadata linkability, not from the neuroimaging data per se. Examples that should be strongly presumptive for controlled access could include: sensitive questionnaire content (e.g., illegal behaviors, stigmatizing traits, detailed trauma/sexual history, immigration/legal vulnerability), biospecimen quantification and multi-omic linkages (especially individual-level high-dimensional profiles), rare disease flags or unique combinations of attributes that increase singling-out risk, and granular geolocation that could enable triangulation. This approach advances participant protection while avoiding unnecessary friction for low-risk imaging reuse.

Before implementing restrictions that could significantly affect current data-sharing practices in the neurosciences, we also recommend that NIH support or conduct empirical research to assess the real-world likelihood of reidentification in imaging data of the human face or head regions, both with and without the application of commonly used deidentification methods. Such work could also help inform the development of more advanced deidentification techniques to mitigate emerging privacy risks associated with advances in reidentification technologies. It would also be important to identify and examine other factors that may increase the sensitivity of imaging data in particular contexts and thereby justify sharing data under controlled-access.

In addition, it is critical to evaluate whether—and to what extent—controlled-access mechanisms meaningfully reduce the risk of reidentification in practice. Such evidence is necessary to appropriately weigh potential privacy risks against the benefits of data sharing, particularly given that controlled-access also increases administrative burdens, delay access to data, and limit beneficial secondary uses of research data.

Open neuroimaging repositories are engines for discovery, catalyze the development of new methods, and are a lynchpin for ensuring reliable, reproducible neuroscience. They are a force-multiplier that dramatically increases the ROI on taxpayer-funded neuroimaging research. The current proposal threatens to collapse the open neuroimaging ecosystem, evaporating that enormous ROI by incentivizing investigators and institutions to return to the closed system that current DMS policy was designed to open. We believe that our recommendations balance the privacy concerns at the heart of the draft proposal with the FAIR guiding principles for scientific data management that have made the open neuroscience movement such a success. Thousands of papers have been published using data from INDI, OpenNeuro and other open repositories. We urge the NIH to consider the chilling effect the current proposal will have on scientific progress and remove imaging data of the human face and head regions as a category of data required to be protected under controlled access.

Signatories:

Russell A. Poldrack, PhD

Principal Investigator, OpenNeuro.org

Chair and Albert Ray Lang Professor, Department of Psychology, Stanford University

Anita Jwa, JD, JSD

Chief Bioethicist, OpenNeuro.org

Research Scholar, Department of Psychology, Stanford University

Joshua W. Buckholtz, PhD

Executive Director, OpenNeuro.org

Research Scholar, Department of Psychology, Stanford University

Investigator, Department of Psychiatry, Massachusetts General Brigham

Michelle N. Meyer, PhD, JD

External Advisory Board Member, OpenNeuro.org

Chief Bioethics Officer, Geisinger

Chair & Associate Professor, Department of Bioethics & Decision Sciences, Geisinger College of Health Sciences

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

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
The proposed policy lists imaging data of the human face and head regions as a category of data required to be protected under controlled access. The policy further defines this category as visual representations—including functional imaging, ultrasound imaging, photographic images, 3D models, radiological scans, X-rays, and other modalities—that depict anatomical or functional details of the human face or head regions.

With respect to imaging data of the human face, it is unclear whether brain imaging data from which facial features have been removed (i.e., defaced images) would fall within the scope of the proposed definition. The primary privacy risk associated with structural brain imaging data is the possibility of reidentifying research participants through facial features reconstructed from structural MR scans. Such reconstructions could potentially be considered comparable to the HIPAA direct identifier "full face photographic images and any comparable images (45 C.F.R. § 164.514(b)(e)(i))." For this reason, it is common practice within the neuroimaging community to remove facial features from structural MRI scans prior to data sharing—particularly when datasets are made openly accessible—in order to reduce the risk of reidentification. While the draft policy appropriately recognizes re-identification and misuse risks, it does not operationalize a distinction between (a) imaging that contains reconstructable facial geometry and (b) imaging that has been processed to remove face-identifying features.

Recommendation: A tiered system that distinguishes between neuroimaging data types based on risk

Tier 1 (default controlled): raw DICOMs; non-defaced T1/T2; images with face geometry preserved; high-resolution head scans.

Tier 2 (eligible for open if criteria met): defaced/anonymized NIfTI images where face geometry is removed to a specified standard.

Tier 3 (open by default): derived measures and non-invertible representations (MRI sequences that do not acquire facial images - e.g. standard T2*-weighted images, regional volumes, statistical maps, connectomes, parcellated maps, group-level summaries).

This tiered system distinguishes between (1) structural imaging data that retain identifiable facial anatomy and may permit reconstruction of facial features, (2) structural imaging data in which facial features have been removed (e.g., defaced images), and (3) neuroimaging data types and derivatives that do not contain information that could be used to re-identify participants. Such a system would allow repositories and investigators to continue applying widely used privacy-preserving techniques while focusing controlled-access requirements on data types that meaningfully increase the risk of participant reidentification.

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

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

136. University of Michigan Medical School

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Sachin Kheterpal, MD, MBA

Name of Organization: University of Michigan Medical School

Type of Organization: Academic Institution

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
March 18, 2026

The University of Michigan Medical School's Office of Research is pleased to submit this response to the draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy. We are deeply committed to cutting edge discoveries and thread the needle between making our research data accessible in accordance with open science advancement AND protecting the privacy and dignity of our research participants and the data, biospecimens they entrust to our research community.

We appreciate the stated intention of the policy to harmonize requirements for grant recipients in the handling and sharing of data produced from NIH funding mechanisms. However noble a pursuit, the harmonization represents a costly and unnecessary burden on the research community. More specifically, our concerns are outlined below:

1. Current protections working well: Existing protections of data, including role-based access, adherence to controls set forth in HIPAA and the Common Rule, data encryption, next-generation firewalls are highly effective at keeping data secure when adhered to. Data privacy issues are due to lack of enforcement and adherence to existing standards, not the lack of rigor within them.

2. One size does not fit all: The foundation of institutional data security is to match level of safeguards with level of risk. The level of controls proposed in the NIH policy are disproportionate to the risk. Existing physical, cultural, and process-level protections on the list of data types are sufficient.

3. "Duplicative burden" is not onerous: The proposed "holistic approach" to research data protections is neither an efficient nor strategic. While the draft policy states that it is providing "a standard set of expectations" to address increasing privacy and security risks, we would submit that there already exist a clear set of regulations that research institutions are measured against, have implemented, and are typically working well. Gaps in effectiveness are due to gaps in enforcement, not standards.

4. Financial burden: Adherence to the proposed policy would require significant infrastructure build, increased human resources for managed access processes, and reliance on vendor solutions. Research institutions are under increased financial constraints due to declining clinical reimbursements, increased costs, and an uncertain research funding landscape.

5. Stifled Innovation: The University of Michigan has achieved its success in human subjects research because of the creative ability to do more with less when it comes to advanced uses of data.

The proposed security and regulatory changes put forth in the Controlled-Access policy would be so resource-intensive that our ability to innovate for the betterment of our research participants would suffer. Many projects would simply never start because of the regulatory and financial fixed costs.

Thank you once again for the opportunity to comment on this important topic. Should there be questions or need for further explanation of our position, please contact:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/OoR-Response_NIH-controlled-access_Kheterpal.pdf

137. American Medical Informatics Association (AMIA)

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Tayler Williams

Name of Organization: American Medical Informatics Association (AMIA)

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
AMIA supports NIH's objective of clarifying when human participant data warrant controlled access rather than open dissemination. We offer the following key recommendations to strengthen the policy:

1. Adopt a proportionate, risk-based framework

AMIA urges NIH to move beyond categorical designations based solely on data type and instead adopt a flexible, risk-based model. Re-identification risk increasingly depends on context, linkage potential, and analytic capability, particularly in multi-modal environments that combine genomic, clinical, imaging, and social determinants data. A framework that accounts for contextual risk, technical safeguards, and intended scientific use would better balance participant protection with scientific utility.

2. Ensure harmonization across NIH data policies

NIH should align the Controlled-Access Data Policy with the Data Management and Sharing (DMS) Policy and the revised GDS Policy. Fragmented implementation across Institutes and Centers creates unnecessary complexity. Clear guidance should explain how controlled-access determinations intersect with DMS Plans, repository selection, and informed consent expectations, improving compliance while reducing administrative burden.

3. Invest in infrastructure, governance, and workforce capacity

Effective controlled access depends on robust repository ecosystems. Consistent with prior AMIA recommendations, policy mandates must be paired with sustained support for infrastructure and workforce development. NIH should:

- Define minimum technical, governance, and cybersecurity standards for controlled-access repositories;
- Provide sustained funding for data stewardship, auditing, and access review;
- Support workforce development for biomedical data stewards and informatics professionals; and
- Ensure interoperability across NIH-funded repositories.

4. Implement flexible, risk-aligned security requirements

While AMIA supports strong data protections, applying uniform and prescriptive standards, such as broad application of NIST SP 800-171A, risks creating unintended disparities. Under-resourced institutions, including community-based organizations and emerging research programs, may face significant barriers to compliance, limiting participation and exacerbating inequities.

Moreover, applying such requirements only to NIH-funded datasets creates a fragmented security landscape that may not meaningfully advance national privacy or security goals. NIH should instead adopt a flexible, risk-based security framework aligned with data sensitivity and use, coupled with implementation support to ensure broad participation.

5. Clarify alignment with existing regulatory frameworks

NIH should provide clear guidance on how controlled-access requirements interact with existing regulations, particularly the HIPAA Privacy and Security Rules.5,6 In many cases, HIPAA already establishes robust protections for human participant data. Duplicative or misaligned requirements risk increasing administrative burden without improving security outcomes. Greater clarity will reduce compliance complexity and support efficient data sharing.

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

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

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
AMIA supports NIH's efforts to harmonize the GDS Policy with the broader DMS framework. To strengthen implementation, we recommend:

1. Address governance of multi-modal and linked datasets

Genomic data are increasingly integrated with phenotypic, environmental, and real-world data sources. NIH should explicitly address governance for linked and derived datasets, including privacy-preserving linkage methods and secure analytic approaches such as federated and enclave-based models.

2. Strengthen participant trust through transparent data stewardship

Participant trust is foundational to data sharing. NIH should provide model consent language adaptable to evolving research contexts, clarify expectations for legacy datasets, and promote transparency mechanisms that inform participants about downstream data use and governance safeguards.

3. Promote equity in access and participation

Controlled-access processes that are overly complex, costly, or time-consuming risk disadvantaging early-career investigators, community-engaged researchers, and under-resourced institutions. NIH should assess approval timelines, administrative burden, and repository access costs to ensure equitable access to publicly funded data resources.

4. Preserve responsible international collaboration

Genomic research relies on global infrastructure and partnerships. Restrictions that limit the use of international platforms or resources could hinder scientific progress and reduce the competitiveness of U.S.-led research. NIH should ensure policies enable responsible international collaboration while maintaining appropriate safeguards.

Artificial Intelligence and Reproducibility

As controlled-access environments increasingly support artificial intelligence (AI) and advanced analytics, NIH policies should explicitly enable reproducible and secure AI development. AMIA has emphasized the importance of interoperable standards, reproducibility, and responsible AI integration within biomedical research ecosystems.

NIH should clarify whether secure computational access models, such as data enclaves and federated analysis, satisfy policy requirements and support consistent reproducibility standards in controlled-access environments.

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/NIH-RFI-Draft-NIH-Controlled-Access-Data-Policy-and-GDS-Policy-AMIA-Comments.pdf

138. College of American Pathologists

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Suanna Steeby Bruinooge

Name of Organization: College of American Pathologists

Type of Organization: Professional Organization/Association

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/CAP-Response-NIH-Datasharing-RFI-March-2026.pdf

Description: I am attaching a comment letter sent on behalf of the College of American Pathologists.

139. Multicenter Perioperative Outcomes Group

Submit date: 3/18/2026

I am responding to this RFI: On behalf of an organization

Name: Sachin Kheterpal

Name of Organization: Multicenter Perioperative Outcomes Group

Type of Organization: Non-profit Research Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
In general, the policy is too restrictive without adding meaningful increased privacy protections. Existing HIPAA and other regulatory requirements, when enforced, achieve the necessary privacy controls.

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

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

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

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

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/MPOGNIHComment.docx

140. Global BioData Trust (GBDT)

Submit date: 3/19/2026

I am responding to this RFI: On behalf of an organization

Name: Elizabeth Dreicer, Chair

Name of Organization: Global BioData Trust (GBDT)

Type of Organization: Other

Type of Organization - Other: Patient Advocacy, Academic, Think Tank, Industry Organization

Role: Organizational Official

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
1A. REPOSITORY DESIGNATION SHOULD REQUIRE A FIDUCIARY OBLIGATION TO PARTICIPANTS, NOT ONLY TO NIH.

The proposed CAD policy establishes security requirements, access control mechanisms, and data use agreement standards for designated repositories. It does not require that designated repositories carry a fiduciary duty to the participants whose data they hold. This is the central structural gap.

Under the current framework, a designated repository's primary accountability runs to NIH as the funder and policy authority. The participant is a data source, not a principal. The policy requires that the repository protect the data. But it does not, as currently written: require the repository to act in the participant's best interest when NIH's and participant's interests diverge, notify the participant when custody changes, or maintain custodial continuity when funding ends.

This gap is not hypothetical. It is the condition that produced the 23andMe bankruptcy outcome, the All of Us OIG audit findings, and the ARPA-H EVIDENT program's absence of a post-grant custodial framework. In all three cases, participant biology is collected under governance frameworks that give participants no structural authority and created no permanent obligation to them. The court-appointed Consumer Privacy Ombudsman in the 23andMe bankruptcy found that safeguards "depending on consumer action are structurally underprotective for sensitive biological data." A fiduciary obligation, by contrast, does not depend on consumer action. It runs structurally from the custodian to the individual.

The Don't Sell My DNA Act is a step in the right direction and GBDT supports this with refinements to establish statutory language to include an Independent Fiduciary Custodian: an entity whose obligation to the participant is legally defined, permanent, and not subject to commercial revision. NIH's CAD policy should create a parallel provision for designated repositories: a tiered designation in which repositories that carry a formal, enforceable fiduciary duty to participants who qualify for a preferred designation, with access to the most sensitive dataset categories.

GBDT's Recommendation: NIH should add a fiduciary duty to participants as a criterion for repository designation under the CAD policy. At minimum, NIH should require that designated repositories carry a legal obligation to act in the best interest of participants when data use decisions are made; notify participants when custody of their data or samples is proposed for transfer; maintain a post-grant custodial framework specifying how participant data and biospecimens will be governed if the

repository's funding ends or its institutional host changes; and support participant revocation or modification of consent authorizations on a continuing basis. NIH should additionally create a preferred designation tier for repositories that meet a higher fiduciary standard, analogous to the Independent Fiduciary Custodian concept reflected in the latest draft of the Don't Sell My DNA Act.

This is not a new governance model. The organ transplant network established by Congress in 1984 operates on this architecture: hospitals hold biological material temporarily, the coordinating network carries fiduciary obligations to donors and recipients, and no institution owns the biology. Congress extended that model to human organs because it recognized that some biological material requires governance that commercial property rights cannot provide. Human genomic data and biospecimens warrant the same recognition.

Additionally, NIH should require an end-of-grant/program custodial transfer and long term independent custodian.

1B. NIH'S STEWARDSHIP OBLIGATION IS CATEGORICAL AND EXCEEDS WHAT EO 14117 REQUIRES.

We agree with GA4GH that the CAD policy's broad scope and the DOJ Final Rule implementing Executive Order 14117—which focuses on access restrictions for covered persons associated with designated countries of concern—are potentially misaligned.The DOJ rule is a national security transaction-control instrument. NIH's stewardship responsibility is categorical: a permanent obligation across the full data lifecycle to the participants who contributed their biology. These are different frameworks serving different purposes, and they should be implemented independently.

NIH may, and in our view should, adopt lifecycle protections that are stronger than what the DOJ rule requires, including stronger repository security, enhanced auditing, downstream-use restrictions, and post-grant custodial requirements, regardless of whether a given dataset crosses the DOJ bulk-data thresholds. The alternative—calibrating NIH's participant protections to match DOJ's national security transaction thresholds—would mean that a dataset of 99 individuals carries no controlled-access obligation simply because it falls below the DOJ's 100-person threshold. That is not a participant-protection standard. It is an inverse of one.

GBDT’s Recommendation: NIH should clarify in the final CAD policy that its stewardship obligations are independent of, and not bound by, the DOJ/EO 14117 framework. The CAD policy should state explicitly that NIH's lifecycle protections for participant data reflect its obligation as a research steward, not merely as an instrument of national security policy.

1C. THE NATIONAL SECURITY THREAT SURFACE IS BROADER THAN DOJ THRESHOLDS REFLECT.

The risks of AI-related re-identification including model inversion, membership inference, and multimodal data linkage are significant. GBDT offers analysis here from the technical perspective of physicians and medical informaticists who have worked on biosurveillance and biodefense.

The U.S. Intelligence Community has issued repeated warnings about foreign acquisition of biomedical and genomic data at scale. The ODNI/NCSC advisory on the People’s Republic of China’s genomic collection describes how large-scale biomedical data acquisition can create strategic advantages in biotechnology, pharmaceuticals, and population-level intelligence. Health data now sit at the

intersection of national security and biomedical innovation in ways the original genomic data sharing policies did not anticipate.

The threat surface has expanded beyond foreign government actors. The Change Healthcare cyberattack in February 2024 compromised the personal health data of nearly 200 million Americans—the largest health data breach in U.S. history, documented in testimony before the U.S. Senate Committee on Health, Education, Labor & Pensions (HELP Committee) on March 5, 2026. The Project Nightingale case established that a major health system transferred roughly 50 million patient records to a technology company under HIPAA's health-care operations provisions, without patient consent or notification. Contract research organizations have demonstrated that supposedly de-identified medical records can be re-identified with high accuracy using external data sources. And genomic, phenotypic, biometric, and geolocation-linked datasets can, at scale, enable foreign intelligence profiling of sensitive populations, coercion of individuals with government or strategic roles, and AI-enabled inference capabilities developed by adversarial states.

The architecture response to these threats requires more than access controls. It requires, as Dr. Martin described on March 12, a health-data infrastructure built around "post-quantum cryptographic security anchored in biological uniqueness, end-to-end encryption, and zero-trust architecture." Identity claims can be stolen. Biological uniqueness cannot. Modern post-quantum cryptography allows access control to be anchored in cryptographic systems that are resistant to future cyber attacks while binding records to biologically verified individuals. NIH-designated repositories should be required to satisfy, or provide a roadmap to meet, these technical standards.

GBDT'S RECOMMENDATION: The CAD policy should explicitly recognize that designated datasets present national security risks beyond individual privacy, and that repository security standards should be calibrated accordingly. NIH should require that operators of designated repositories receive specific guidance on the national security dimensions of the data they hold, implement or adopt a roadmap toward post-quantum cryptographic security standards, and deploy zero-trust architecture for access management.

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

2. NIH'S FLAGSHIP RESEARCH PROGRAM DOES NOT CURRENTLY MEET THE STANDARDS THIS POLICY PROPOSES.

The RFI asks for feedback on the availability of established repositories for implementing the proposed CAD policy. We raise a foundational concern that must be addressed before any repository designation framework can be credible. Namely, underlying security risks are going unresolved before the CAD policy is implemented.

The NIH All of Us Research Program is the most prominent repository of the kind the CAD policy would govern. More than 832,000 Americans have enrolled, contributing biospecimens, whole genome sequences, electronic health records, wearable data, and surveys. The program runs entirely on annual federal appropriations with no permanent custodial framework specifying what happens to 586,000 banked biosamples and 832,000 genomic records if funding priorities change.

In December 2025, the HHS Office of Inspector General audited the All of Us Data and Research Center and found: no controls in place to prevent unauthorized downloading of detailed participant data, despite program policies prohibiting such downloads; failure by NIH to communicate national security concerns about genomic data to the award recipient; security and privacy weaknesses not remediated within federally required timeframes; and the ability of authorized users to access information systems from foreign countries without approval controls. The OIG's conclusion was that these conditions created an increased risk of participant data being accessed, downloaded, and misused by bad actors, including foreign adversaries.

THE CAD POLICY IS BEING BUILT ON TOP OF INFRASTRUCTURE THAT DOES NOT MEET ITS OWN PROPOSED STANDARDS.

There is a second dimension to the All of Us failure mode that is less discussed but equally important. All of Us was designed as a program of longitudinal partnership with participants. Instead, it has adopted the same extractive architecture it implicitly criticizes in commercial vendors: collecting biological data and samples from participants at scale, making them broadly accessible to third parties, including commercial researchers, offering no fair compensation for that access, and leaving no permanent custodial framework for what happens when appropriations change. Participation has stalled roughly 170,000 participants short of its one-million target after seven years of recruitment. The populations most critical to health equity research, those who have historically been exploited by research institutions, were asked to participate in a program without structural accountability. Senator Murkowski, in the March 5, 2026, Senate HELP hearing, raised the same structural concern about tribal communities specifically: historically, much of the data from tribal communities has been collected and managed by federal agencies or outside entities without clear tribal governance. The same architectural condition applies across indigenous research populations, rural underserved communities, and every group that has been recruited into federal research programs without a permanent, enforceable fiduciary obligation running to them as participants. Extractive models ultimately shrink the participation base. Science suffers.

GBDT'S RECOMMENDATION: Before designating any repository under the CAD policy, NIH should require that the repository demonstrate compliance with the policy's security and access-control standards through independent audit. NIH should develop a, pre-grant, grant and post-grant custodial standard specifying how designated repositories must plan for the long-term governance of participant data and biospecimens when funding cycles end. The All of Us program should be brought into compliance with these standards as a condition of continued designation. NIH should consider whether a permanently endowed, consent-governed, participant-principal trust model could serve as a custodial vehicle for biological materials collected in federally funded programs, ensuring those materials are safeguarded for individuals.

3. Feedback on the appropriateness of the protected data types designated to be controlled-access:
3A. GENETIC INFORMATION IS CONSTITUTIONALLY DISTINCT: IT CANNOT BE DEIDENTIFIED.

We support GA4GH's nuanced observations on summary statistics risk, rare variant re-identification, and the contextual factors, including cohort size, ancestry composition, and disease rarity, that affect disclosure risk.

Genetic information is not merely sensitive data that happens to be hard to anonymize. It is, by its nature, permanently identifying. The GBDT proposed refinements to the Don't Sell My DNA Act provides the most precise available formulation: genetic information "is not able to be deidentified or anonymized." A genome is not a data point that can be stripped of its identifier and rendered safe. It is the identifier. This distinction has practical consequences for policy.

The Nature Communications study published in 2019 found that 99.98% of Americans could be correctly re-identified in any dataset using just 15 demographic attributes, and concluded that even heavily sampled anonymized datasets seriously challenge the technical and legal adequacy of the de-identification release-and-forget model. As AI-powered re-identification tools improve, de-identification is better understood as a temporary legal fiction than a durable technical guarantee. The CAD policy should reflect this. Controlled-access status for genomic data should not depend on a determination that the data is sufficiently de-identified. It should attach to the data type itself.

3B. PHYSICAL BIOSPECIMENS WARRANT THE SAME LEGAL PROTECTION AS GENOMIC DATA, AND CURRENTLY HAVE NONE.

GBDT emphasizes that physical biospecimens held in NIH-funded repositories deserve equal privacy protections as genomic data.

Safe deposit box contents are legally classified under U.S. law as bailed property: they are the personal property of the box holder, not an asset of the bank, and not subject to liquidation when the institution fails. Physical biological samples held in federally funded repositories carry no equivalent legal status. The 23andMe bankruptcy demonstrated this concretely: physical biospecimens were among the corporate assets proposed for transfer to a bankruptcy acquirer, not because anyone designed that outcome, but because no legal framework prevented it.

GBDT'S RECOMMENDATION: NIH should address the legal status of physical biospecimens in the CAD policy framework or in a parallel policy instrument. Specifically, NIH should classify physical biospecimens held in federally funded repositories as the property of the contributing participant, not as an asset of the holding institution; require that repositories notify participants and obtain renewed consent before any proposed transfer of physical biospecimens to a new custodian; and develop a post-grant disposition standard specifying how biospecimens must be handled when a repository's funding or institutional host changes.

4. Feedback on any aspect of the Proposed Revisions to the NIH Genomic Data Sharing Policy:
4. THE CONSENT PROBLEM REQUIRES STRUCTURAL INFRASTRUCTURE, NOT BETTER FORM LANGUAGE.

It is clear that any consent language should be updated to explicitly address AI model training and downstream model reuse. But updating consent language alone simply builds on the existing notice-and-consent model and is thus insufficient, because the model itself is the real problem.

The notice-and-choice consent model is the governance foundation of American health data, including the GDS Policy's consent framework. Its central assumption—that participants read, understand, and freely agree to terms governing their most sensitive biological information—is contradicted by extensive peer-reviewed evidence and confirmed in federal bankruptcy proceedings.

The court-appointed Consumer Privacy Ombudsman in the 23andMe bankruptcy found: of the 18 million customers whose data the company holds, nearly one-third had not logged in since before the June 2022 privacy policy change that authorized bankruptcy data transfer. His conclusion: safeguards depending on consumer action are structurally underprotective for sensitive biological data.

Federal clinical trial consent forms average 12 pages, written at a reading level exceeding the comprehension of more than half of U.S. adults. None of the COVID-19 vaccine trial consent forms reviewed by Mayo Clinic researchers met the FDA's recommended seventh-grade readability standard. Research published in the Journal of Patient Safety found that patients do not read these forms because they trust their physicians, not because they have evaluated and accepted the terms. That physician-patient trust is the actual mechanism of consent, not the signature.

Beyond the research context, the commercial health data system has developed a specific three-step architecture for extracting consent without genuine agreement, which our March 12 convening documented in detail.

Step one: declare patient data to be vendor property. Epic's MyChart Terms of Service state that all information within or made available through the portal belongs to the provider-vendors, reframing ownership before the patient encounters HIPAA at all.

Step two: condition access on a waiver of rights. Epic's September 2025 update requires patients to waive class-action enforcement of HIPAA rights as the price of accessing their own records. Epic's maximum liability for mishandling patient health information under this agreement is fifty dollars.

Step three: route through trust. Patients sign at check-in because they trust their physician, not because they have read or negotiated anything. The physician-patient relationship is used as the delivery mechanism for a vendor agreement that the patient never knowingly made.

This three-step mechanism does not violate HIPAA. It is more insidious: it is the legal architecture that makes HIPAA functionally unenforceable for the individuals it was designed to protect. The rights are preserved on paper while the enforcement and technical substrate are stripped. This is the structural condition our Senate HELP Committee QFR submission described as the MyChart problem. It is directly relevant to the GDS Policy because NIH-funded research draws on data collected through precisely this consent infrastructure.

The March 5, 2026, Senate HELP hearing surfaced two additional dimensions of this problem that NIH's policy framework must address. First, ASTP/ONC's National Coordinator, Dr. Thomas Keane, confirmed under questioning from Chairman Cassidy that "a large portion of the apps that people are uploading their data to are not covered entities, and HIPAA doesn't apply to them." When pressed on whether HHS could regulate this through rulemaking, he stated: "I don't think that we are able to regulate data that the patients have consented to be released." This is a direct HHS admission that under the current framework, consent functions as a liability waiver rather than a protective instrument. Once a patient clicks accept on a portal agreement, the data may flow freely to non-covered-entity AI platforms with no HIPAA obligation whatsoever.

Second, Chairman Cassidy articulated the downstream consequence precisely: "Most people are not going to understand the implications of putting my genetic data up there, which then AI could then figure out who all my relatives are, and they could potentially redline my relatives for insurance." This is

the Chair of the committee that oversees the Assistant Secretary for Technology Policy (ASTP) / Office of the National Coordinator for Health Information Technology (ONC) describing the same re-identification and inference risk GBDT has raised. The combination of consent-as-waiver, non-covered-entity AI platforms, and the permanent identifying character of genomic data creates a risk architecture that current policy frameworks were not designed to address. NIH's GDS Policy revision is the appropriate instrument to establish a higher standard for genomic data.

GBDT'S RECOMMENDATION: The revised GDS Policy should move beyond consent-form language reform and establish a standard for longitudinal, auditable, participant-governed consent infrastructure. NIH should require designated repositories to maintain a participant-accessible record of all consent authorizations and subsequent data uses; support participant revocation or modification of consent authorizations on a continuing basis, without loss of access to other program benefits; and demonstrate that consent interfaces meet comprehension standards appropriate to the population being served.

Consent for AI model training should be treated as a distinct authorization, not bundled into general research use consent. NIH should develop a certification pathway recognizing repositories that implement this standard as preferred custodians for federally funded research.

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
5. TECHNICAL ARCHITECTURE STANDARDS MUST MATCH THE AI ERA THREAT SURFACE.

We endorse GA4GH's technical recommendations on imputation server security: mandatory rather than encouraged implementation of privacy-enhancing technologies including differential privacy, secure enclaves, and secure multiparty computation; clear rules on sharing, auditing, and downstream use of imputation panels trained on controlled-access data; and explicit accountability for the genomic information encoded in trained models.

As physicians and informaticists who have built large clinical data systems, GBDT offers important insight on implementing technical security measures. Imputation servers operating on controlled-access genomic data are part of a broader technical landscape where the relevant security standards have advanced significantly. Four technical requirements should be made explicit in the CAD policy framework for imputation servers and, more broadly, for designated repositories: (1) post-quantum cryptographic security anchored in biological identity verification; (2) clinical natural language processing standards that enable the capture of physician reasoning rather than billing codes, so that shared data actually supports the learning health system mission; (3) uniform clinical meta-ontologies enabling global interoperability across institutions; and (4) end-to-end encryption with zero-trust architecture, designed on the assumption that breaches will be attempted. Encryption at rest and in transit, combined with zero-trust architecture, dramatically reduces the threat surface for ransomware, data exfiltration, and insider misuse.

These standards are achievable. They have been demonstrated in contexts including the Olympic Athlete Management System, which supported secure medical record management across multiple Olympic Games with multilingual ontology architecture. NIH should treat them not as aspirational but as a design floor for any repository or computational infrastructure operating on controlled-access data.

GBDT'S RECOMMENDATION: NIH should require that operators of imputation servers and designated repositories meet or adopt a compliance roadmap toward the four technical standards above. The participant-obligation framework proposed in Section 1 should apply to imputation server operators as well: explicit fiduciary obligation, audit transparency, and a framework specifying how participant data encoded in trained models will be governed when the server is retired or transferred.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/20260318-gbdt_nih-rfi_NOT-OD-26-023_draft-response.docx.pdf

Description: Full Submission, including Statement Overview and Final Conclusion

141. Richard Henson

Submit date: 3/19/2026

I am responding to this RFI: On behalf of myself

Name: Richard Henson

Name of Organization: University of Cambridge, UK

Type of Organization: Academic Institution

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
Further restricting access to data will impede science, including the recent progress on greater reproducibility in science. I have made several important discoveries using neuroimaging data shared by US cohorts. I also run a UK cohort (CamCAN) where we share such data (subject to a DUA) and have seen the huge scientific benefits that have resulted around the world (eg over 300 publications resulting). The proposed restrictions will further damage the US's leadership in science, and researchers will turn to other cohorts like those emerging in China.

2. Feedback on the availability of established repositories for implementing the proposed Controlled-Access Data Policy:
I am familiar with remote access platforms and they involve multiple hindrances that make research harder, as well as being very expensive to setup. While they may be necessary for very sensitive data (eg genetics), they are not necessary for neuroimaging data, where risks are negligible, particularly when participants have given informed consent for such sharing.

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

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

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

142. NIAID Systems Biology Data Dissemination Working Group

Submit date: 3/19/2026

I am responding to this RFI: On behalf of myself

Name: Lars Pache

Name of Organization: NIAID Systems Biology Data Dissemination Working Group

Type of Organization: Other

Type of Organization - Other: Consortium/ working group

Role: Investigator/Researcher

1. Feedback on any aspect of the Draft NIH Controlled-Access Data Policy:
N/A

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

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

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

5. Feedback on the proposed updates to the GDS Policy for Imputation Servers:
N/A

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/ninja-forms/55/Response-to-RFI-NOT-OD-26-023_NIAID-Systems-Biology-Data-Dissemination-Working-Group.pdf

Description: Full comments are attached as a PDF for your review.