4140-01-P

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

AGENCY: National Institutes of Health, HHS.

ACTION: Notice.

SUMMARY: The National Institutes of Health (NIH) is requesting public input on its proposal to establish harmonized and transparent policy requirements for protecting human participant research data. Specifically, NIH proposes (1) establishing policy requirements for which data should be controlled-access under NIH data sharing policies, and (2) revising the NIH Genomic Data Sharing Policy to simplify and harmonize requirements.

DATES: To ensure consideration, comments must be submitted in writing by [**INSERT DATE 90 DAYS AFTER PUBLICATION IN THE FEDERAL REGISTER]**.

ADDRESSES: Comments may be submitted electronically to https://osp.od.nih.gov/comment-form-draft-nih-controlled-access-data-policy-and-proposed-revisions-to-nih-genomic-data-sharing-policy/.

Comments are voluntary and may be submitted anonymously. You may also voluntarily include your name and contact information with your response. Other than your name and contact information, please do not include in the response any personally identifiable information or any information that you do not wish to make public. Proprietary, classified, confidential, or sensitive information should not be included in your response. After the NIH Office of Science Policy (OSP) has finished reviewing the responses, the responses may be posted to the OSP website without redaction.

FOR FURTHER INFORMATION CONTACT: Taunton Paine, Director, Division of Scientific Data Sharing, NIH Office of Science Policy, at (301) 496-9838 or *SciencePolicy@od.nih.gov*.

SUPPLEMENTARY INFORMATION:

Background

NIH serves as the steward of a wide range of research data and continuously works to optimize open sharing with appropriate protections throughout the entire data lifecycle. Given its numerous established data policies, NIH is proposing a holistic update to its data policy framework to strengthen data protections, clarify requirements, and reduce duplicative burden. First, NIH is proposing a new NIH Controlled-Access Data Policy to support the research community in fulfilling NIH data sharing expectations. This proposed policy specifies human participant data types required to be managed via controlled-access and provides criteria for assessing the need for controls for other data types. It also provides a standard set of expectations across NIH Institutes, Centers and Offices to promote maximal responsible human participant data sharing through controlled access while simultaneously responding to emergent privacy and security risks, including those outlined in the following security directives:

1. Executive Order 14117 and the 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 (see: 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) identifying specific data types, along with associated thresholds, that should be protected to mitigate national security risks.
2. The Consolidated Appropriations Act, 2023, requiring updates to genomic data sharing policies and practices to account for national security risks, Public Law No: 117-328. (see: https://www.congress.gov/bill/117th-congress/house-bill/2617/text).

3. The Government Accountability Office report on Human Genomic Data, recommending NIH develop and implement procedures to proactively and comprehensively monitor researcher compliance with data management and security measures for human genomic data, GAO-25-107377 (see: https://www.gao.gov/products/gao-25-107377).

Second, NIH is proposing to revise the NIH Genomic Data Sharing (GDS) Policy to reduce duplicative policy requirements and improve overall performance. The GDS Policy, issued in 2014, promotes broad, responsible, and timely sharing of genomic research data derived from NIH research. As a landmark policy, it has played a crucial role in facilitating rapid access to valuable genomic data while ensuring participant protection through rigorous informed consent and privacy safeguards. Since 2014, NIH has issued the NIH Data Management and Sharing (DMS) Policy as well as streamlined and strengthened its controlled-access practices, including:

- harmonizing the oversight and management of controlled-access data repositories and access management systems (see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-159.html),
- modernizing security standards for controlled-access data subject to the GDS Policy (see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-24-157.html),
- establishing minimum expectations for access to controlled-access data by developers (see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-24-157.html), and
- prohibiting access to NIH controlled-access data repositories and associated data by institutions located in countries of concern (see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-083.html).

NIH is proposing to revise the GDS Policy to enhance efficiency and reduce redundancies with these recent directives and policy developments.

NIH requests public input on both the Draft NIH Controlled-Access Data Policy and the proposed revisions to the NIH Genomic Data Sharing Policy.

NIH Controlled-Access Data Policy

Scope and Applicability

This Policy applies to all NIH-supported research generating human data or deriving data from human data, cell lines, or biospecimens. This Policy applies to the NIH Intramural Research Program and all NIH funding mechanisms (e.g., grants, cooperative agreements, contracts, Other Transactions), regardless of activity code.

This Policy does not apply to NIH research that only involves:

- Generation and sharing of non-human data
- Collection and sharing of human cell lines and biospecimens (see: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-160.html)

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. Additionally, this Policy is not intended to address sharing data for regulatory approval or to comply with regulatory requirements (e.g. submission of data to regulatory approval bodies).

Requirements

Human Data Types Required to be Protected through Controlled-Access (see Appendix for definitions)

The following data types, listed below, must be protected throughout the data lifecycle. Institutions conducting NIH-supported research must ensure that these data types are protected even if not sharing through a controlled-access data repository. These categories are based on the data types provided in 28 CFR Part 202, as well as other data commonly generated or used in NIH-supported research that warrant additional controls. 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. Protected data types include:

- Covered personal identifiers
- Precise geolocation data
- Biometric identifiers
- Genomic data
- Epigenomic data
- Proteomic data
- Transcriptomic data
- Personal health data
- Personal financial data
- Individual level clinical trial data
- Imaging data of the human face or head regions

Requirements for Controlled-Access Data Sharing

Controlled-access data repositories sharing human participant data types identified in this Policy must adhere to security and operational standards appropriate for safeguarding of human data. NIH Controlled-Access Data Repositories (CADRs) subject to the "Required Security and Operational Standards for NIH Controlled-Access Data Repositories" are fully compliant with the requirements of this Policy(https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/requirements). Other controlled-access data repositories can meet the requirements of this policy if, at a minimum, their security and operational standards include:

- 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).

NIH does not consider repositories only requiring user registration and/or provide non-binding guidance on data usage to be controlled-access repositories.

Additional Policy Considerations

Data not explicitly required to be managed via controlled-access under this Policy should be assessed for the need for controls. Criteria for making these assessments include any of the following:

1. Explicit limitations on subsequent use, such as those imposed by laws, regulations, policies, informed consent, and agreements.
2. 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.
3. Lack of adequate data de-identification or the possibility of re-identification cannot sufficiently be reduced.

Proposed Revisions to the NIH Genomic Data Sharing Policy

NIH is proposing several revisions to the GDS Policy to reduce duplicative policy requirements and improve overall performance. Importantly, the core tenets of this policy will remain intact. The effective date for changes will align with the target effective date of the NIH Controlled-Access Data Policy described above.

Proposed GDS Policy revisions:

1. **The scope of the GDS Policy is proposed to be revised as follows:**

- Apply only to human data. The NIH Data Management and Sharing Policy (DMS Policy) encompasses all types of scientific data, including genomic data. The GDS Policy scope will apply specifically to human genomic data to outline the specific protections needed to ensure participant autonomy, privacy, and security. Non-human genomic data will no longer be subject to the GDS Policy.
- Simplify thresholds for "large scale" genomic data. 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. Studies generating and sharing human genomic data below this threshold will be subject to the expectations of the DMS Policy and the proposed NIH Controlled-Access Data Policy described above.
- Establish consistent requirements across NIH. To reduce complexity, NIH Institutes, Centers, and Offices (ICOs) will not be permitted to expand the scope of the GDS Policy through individual program or policy expectations. While ICs may request additional data protections or submission of data to NIH controlled-access repositories in NOFOs, they may not characterize these requests as modifying the expectations of the GDS Policy. These data submissions will be governed by the DMS Policy, the "Required Security and Operational Standards for NIH Controlled-Access Data Repositories," the proposed NIH Controlled-Access Data Policy, and other relevant policies.

2. **Timelines for Data Processing**: Timing of data release will be removed from the GDS Policy as data release should be immediate, depending on repository operations. The levels of data processing previously provided in supplementary information to the GDS Policy will no longer be used. 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).

3. **NIH proposes modernizing the following data submission and sharing practices by:**

- Clarifying expectations for sharing human genomic data openly. Expectations for sharing genomic data (and other omics data) openly or with controls will be entirely governed by the proposed NIH Controlled-Access Data Policy, including expectations for informed consent to share data openly.
- Allowing for HIPAA Expert Determination. Identifiers are allowed to be submitted under the GDS Policy so long as the dataset is de-identified according to Expert Determination and is accepted by the controlled-access data repository. Data derived from human research participants must be de-identified according to the following standards:
    - Identities of research participants cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR 46.102(e); Federal Policy for the Protection of Human Subjects); and
    - Either remove 18 identifiers enumerated at 45 CFR 164.514(b)(2) (the HIPAA Privacy Rule) or meet the Expert Determination standard according to 45 CFR 164.514(b)(1)
- Expanding institutional review capacity. 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. For example, Human Research Protection Programs (HRPPs) would be permitted to review and approve data submissions.

- Strengthening requirements for participant consent. 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, consistent with 45 CFR 46.116(d) (the Common Rule). If consent has not been obtained, the data cannot be shared. NIH accepts data when collected under informed consent for research use from a Legally Authorized Representative, consistent with the Common Rule, as long as the consent meets other expectations of the GDS Policy (e.g., consent is expected to be for future research use and be opt-in, not opt-out). This includes, but is not limited to, the following situations:
    - Consent is obtained from next-of-kin or applicable legal authority in cases where the individual is deceased
    - Consent is obtained from a Legally Authorized Representative, next-of-kin, or other forms of surrogate or proxy decision-making in cases where the individual lacks capacity to consent for themselves
    - Assent is obtained from minors, with parental permission
- Updating expectations for imputation servers. NIH currently allows Approved Users to use imputation servers for imputation with controlled-access data from studies subject to the GDS Policy only in limited circumstances, such as the National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics in Precision Medicine (TOPMed) Imputation Server. Specifically, these servers offer a secure environment for users to upload genotypes, the results are encrypted, and after 7 days the data are deleted from the server upholding the non-transferability agreement in the Data Use Certification (DUC) Agreement. The server operates in an environment consistent with security standards in the NIH Security Best

Practices for Controlled-Access Data Repositories and employs countermeasures to reduce the risks of certain attacks.

NIH currently does not allow users to develop their own imputation panels or servers. NIH has heard interest from the research community in allowing Approved Users to develop their own imputation panels and servers using controlled-access data from studies subject to the GDS Policy. NIH requests input on clarifying that Approved Users may operate imputation servers if they 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.

Request for Input

NIH invites public input on any aspect of the Draft NIH Controlled-Access Data Policy and the proposed revisions to the NIH Genomic Data Sharing Policy. Input is specifically requested on:

1. **Availability of established repositories for implementing the proposed Controlled-Access Data Policy.** NIH has made investments in expanding the capacity of controlled-access data repositories (see: https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/accessing-data/best-practices) and is interested in additional resources that may be needed to meet an anticipated increased demand for storing and managing larger amounts of controlled-access data.

2. **Appropriateness of the protected data types designated to be controlled-access.** The data types subject to the Controlled-Access Data Policy, including whether any should be added, removed, or definitions clarified (e.g., whether NIH should consider adding

thresholds for the number of analytes for particular data types). Additionally, any factors that should be considered when sharing data openly without controls, given the Draft Controlled-Access Data Policy's requirements for informed consent and institutional review. NIH may provide FAQs or additional guidance on data types that typically should not be controlled, such as genomic summary results, summarized result data (including from clinical trials), and specific low-risk components of controlled-access data.

3. **Proposed Updates to the GDS Policy for Imputation Servers.** NIH is interested in options or strategies that maintain the privacy of imputation servers and reference panels, such as technologies that operate servers in secure environments or use privacy enhancing technologies (PETs).

Appendix to the Controlled-Access Policy: Definitions

- ***Genomic summary results.*** The output of analyses of genomic data across many individuals included within a dataset. Includes systematically computed statistics such as, but not limited to: 1) frequency information (e.g., genotype counts and frequencies, or allele counts and frequencies); and 2) association information (e.g., effect size estimates and standard errors, and p-values). These values may be defined and calculated using scientifically relevant subsets of research participants included within study populations (e.g., disease, trait-based, or control populations).
- ***Covered personal identifiers***. Any listed identifier: (1) In combination with any other listed identifier; or (2) In combination with other data that is disclosed such that the listed identifier is linked or linkable to other listed identifiers or to other sensitive personal data. This excludes (1) 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);

and (2) A network-based identifier, account-authentication data, or call-detail data that is linked only to other network-based identifier, account-authentication data, or call-detail data as necessary for the provision of telecommunications, networking, or similar service (28 CFR 202.212).

- ***Listed identifier***. Any piece of data in any of the following data fields: (a) Full or truncated government identification or account number (such as a Social Security number, driver's license or State identification number, passport number, or Alien Registration Number); (b) Full financial account numbers or personal identification numbers associated with a financial institution or financial-services company; (c) Device-based or hardware-based identifier (such as International Mobile Equipment Identity ("IMEI"), Media Access Control ("MAC") address, or Subscriber Identity Module ("SIM") card number); (d) Demographic or contact data (such as first and last name, birth date, birthplace, ZIP code, residential street or postal address, phone number, email address, or similar public account identifiers); (e) Advertising identifier (such as Google Advertising ID, Apple ID for Advertisers, or other mobile advertising ID ("MAID")); (f) Account-authentication data (such as account username, account password, or an answer to security questions); (g) Network-based identifier (such as Internet Protocol ("IP") address or cookie data); or (h) Call-detail data (such as Customer Proprietary Network Information ("CPNI") (28 CFR 202.212).
- ***Precise geolocation data***. Data, whether real-time or historical, that identifies the physical location of an individual or a device with a precision of within 1,000 meters (28 CFR 202.242).
- ***Biometric identifiers***. Measurable physical characteristics or behaviors used to recognize or verify the identity of an individual, including facial images, voice prints and patterns, retina and iris scans, palm prints and fingerprints, gait, and keyboard usage patterns that

are enrolled in a biometric system and the templates created by the system (28 CFR 202.204).

- ***Genomic data***. Data representing the nucleic acid sequences that constitute the entire set or a subset of the genetic instructions found in a human cell, including the result or results of an individual's "genetic test" (as defined in 42 U.S.C. 300gg-91(d)(17)) and any related human genetic sequencing data (28 CFR 202.224).
- ***Epigenomic data***. Data derived from a systems-level analysis of human epigenetic modifications, which are changes in gene expression that do not involve alterations to the DNA sequence itself. These epigenetic modifications include modifications such as DNA methylation, histone modifications, and non-coding RNA regulation. Routine clinical measurements of epigenetic modifications for individualized patient care purposes would not be considered epigenomic data because such measurements would not entail a systems-level analysis of the epigenetic modifications in a sample (28 CFR 202.224).
- ***Proteomic data***. Data derived from a systems-level analysis of proteins expressed by a human genome, cell, tissue, or organism. Routine clinical measurements of proteins for individualized patient care purposes would not be considered proteomic data under this rule because such measurements would not entail a systems-level analysis of the proteins found in such a sample (28 CFR 202.224).
- ***Transcriptomic data***. Data derived from a systems-level analysis of RNA transcripts produced by the human genome under specific conditions or in a specific cell type. Routine clinical measurements of RNA transcripts for individualized patient care purposes would not be considered transcriptomic data under this rule because such measurements would not entail a systems-level analysis of the RNA transcripts in a sample (28 CFR 202.224).
- ***Personal health data***. Health information related to disease, diagnosis, or treatment and indicates, reveals, or describes the past, present, or future physical or mental health or

condition of an individual; the provision of healthcare to an individual; or the past, present, or future payment for the provision of healthcare to an individual. This term includes 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 (28 CFR 202.241).

- ***Personal financial data***. Data about an individual's credit, charge, or debit card, or bank account, including purchases and payment history; data in a bank, credit, or other financial statement, including assets, liabilities, debts, or trades in a securities portfolio; or data in a credit report or in a "consumer report" (as defined in 15 U.S.C. 1681a(d)) (28 CFR 202.240).
- ***Individual level clinical trial data***. Detailed data collected from each participant during a clinical trial, excluding summary results.
- ***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.

Dated: December 11, 2025.

Matthew Memoli,

Principal Deputy Director,

National Institutes of Health.

[FR Doc. 2025-23246 Filed: 12/17/2025 8:45 am; Publication Date: 12/18/2025]