CASC | Coalition for Academic Scientific Computation

Coalition for Academic Scientific Computation (CASC) Response to NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing Policy

Published Document: 2025-23246 (90 FR 59131)
Deadline: March 18, 2026

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.

1. ***Availability of established repositories for implementing the proposed Controlled-Access Data (CAD) 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-pra…) 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.*

   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.

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.

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).

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).*

   CASC has no comments with regards to this element of the request for information.