NIH GENOMIC DATA SHARING (GDS) POLICY

The NIH GDS Policy replaces the Genome-Wide Association Studies (GWAS) policy and became effective with NIH grant applications submitted on or after the January 25, 2015 due date and thereafter. The GDS Policy applies to all NIH-funded research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research (see examples).

WHAT THE REVIEWER NEEDS TO EVALUATE

NIH expects investigators and their institutions to provide basic plans for following this policy in the “Genomic Data Sharing Plan” located in the Resource Sharing Plan section of funding applications. Any resources that may be needed to support a proposed genomic data sharing plan (e.g., preparation of data for submission) should be included in the project's budget. A more detailed genomic data sharing plan should be provided to the funding IC prior to award. Note: Current guidance is that applicants who generate GDS data must include a GDS plan; those using human genomic data from NIH-designated data repositories need not submit a GDS plan.

Until the implementation of this policy (see “NIH GDS Policy” below) is finalized, the role of the reviewer is as follows:

• If GDS is applicable to the application, the reviewer should provide comments regarding the proposed GDS plan or indicate the need for a GDS plan if not included in the application.
• The reviewer should not include an assessment of the plan in the determination of the application’s Overall Impact/Priority score.
• However, some special initiatives may contain additional requirements related to GDS; in these cases, the adequacy of the GDS sharing plan may be an additional review criterion, and the reviewers would then factor an evaluation of the plan into the overall evaluation of scientific merit, i.e., the Overall Impact/Priority score.

At a minimum, the Resource Sharing Plan section of the grant application should include:

• Type of data that will be shared (i.e., the type of genomic data, relevant associated data, and information necessary to interpret the data);
• The data repository to which the data will be submitted;
• The timeline for the data to be shared;
• Any limitations on the secondary research uses of the data, if the study involved human data; and
• Acknowledgement that the Institutional Certification will be submitted and assurance by the Institutional Review Board (IRB) that the data can be shared through NIH-designated data repositories, consistent with data sharing under the NIH GDS Policy.

Investigators should provide a justification for an exception to the data sharing expectation in the grant application, prior to the grant award.

NIH GDS POLICY
The GDS Policy applies to all NIH-funded research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data, irrespective of funding level and funding mechanism (e.g., grant, contract, cooperative agreement, or intramural support). The Supplemental Information to the GDS Policy provides examples of genomic research projects that are subject to the Policy.

The GDS Policy has no direct cost threshold associated with it and applies only to grant activities requesting support for research, such as:
- Research project grants (Rs);
- Program projects (Ps) and SCORs (Ss);
- Cooperative agreements for research (Us);
- Individual career development awards (Ks) that include a research component;
- S activities that include a research component; and
- All other activities that include a research component.

The GDS Policy does not apply to:
- Institutional training grants (T32s, T34s, T35s, and TL2s);
- K12 career development awards (KL2s);
- Individual fellowships (Fs);
- Resource grants and contracts (Ss);
- Linked awards derived from previously reviewed applications (KL1, KL2, RL1, RL2, RL5, RL9, TL1, UL1);
- Facilities or coordinating centers funded through related initiatives to provide genotyping, sequencing, or other core services in support of GDS.

Examples of such research include, but are not limited to, projects involving:
- Sequence data from more than one gene or region of comparable size in the genomes of more than 1,000 human research participants.
- Sequence data from more than 100 genes or region of comparable size in the genomes of more than 100 human research participants.
- Data from 300,000 or more variant sites in more than 1,000 human research participants.
- Sequence data from more than 100 isolates from infectious organisms.
- Sequence data from more than 100 metagenomes of human or model organism microbiomes.
- Sequence data from more than 100 metatranscriptomes of human or model organism microbiomes.
- Whole genome or exome sequence data of more than one model organism species or strain.
- Comprehensive catalog of transcripts and non-coding RNA from one or more model organism species or strains.
- Catalog of more than 100,000 SNPs from one or more model organism species or strains.
- Comparisons of genome-wide methylated sites across more than 10 cell types.
- Comparisons of differentially methylated sites genome-wide at single-base resolution within a given sample.

**GDS SHARING PLAN REQUIREMENTS**

Applicants preparing grant applications that include GDS research are expected to state in the cover letter that the studies proposed will generate large-scale human and/or non-human genomic data, and they should include a GDS plan. When the applicant is unable to submit GDS data (due to
informed consent issues, local laws and limitations, concerns about harms to individuals and groups, or other limitations), the applicant must describe why sharing (in part or in full) is not possible. Exceptions to NIH expectations for data submission to an NIH-designated data repository will be considered on a case-by-case basis by the NIH funding Institute or Center (IC).

The sharing plan should contain details about the submission of relevant data to the NIH GDS data repository including:

**For Non-human Genomic Data Sharing Plans, information should include:**

1. **Data Submission Expectations and Timeline**

Large-scale non-human genomic data, including data from microbes, microbiomes, and model organisms, as well as relevant associated data (e.g., phenotype and exposure data), are to be shared in a timely manner. Genomic data undergo different levels of data processing, which provides the basis for NIH’s expectations for data submission. These expectations are provided in the Supplemental Information (see “Supplemental Information” link at the end of this chapter). In general, investigators should make non-human genomic data publicly available no later than the date of initial publication.

2. **Data Repositories**

Non-human data may be made available through any widely used data repository, whether NIH-funded or not, such as the Gene Expression Omnibus (GEO), Sequence Read Archive (SRA), Trace Archive, Array Express, Mouse Genome Informatics (MGI), WormBase, the Zebrafish Model Organism Database (ZFIN), GenBank, European Nucleotide Archive (ENA), or DNA Data Bank of Japan (DDBJ) (see GDS policy link for more repository information). NIH expects investigators to continue submitting data types to the same repositories that they submitted the data to before the effective date of the GDS Policy (e.g., DNA sequence data to GenBank/ENA/DDBJ, expression data to GEO or Array Express). Data types not previously submitted to any repositories may be submitted to these or other widely used repositories as agreed to by the funding IC.

**For Human Genomic Data Sharing Plans, information should include:**

1. **Data Submission Expectations and Timeline**

Investigators should submit large-scale human genomic data as well as relevant associated data (e.g., phenotype and exposure data) to an NIH-designated data repository in a timely manner. Investigators should also submit any information necessary to interpret the submitted genomic data, such as study protocols, data instruments, and survey tools.

Genomic data undergo different levels of data processing, which provides the basis for NIH’s expectations for data submission and timelines for the release of the data for access by investigators. These expectations and timelines are provided in the Supplemental Information (see “Supplemental Information” link at the end of this chapter). In general, NIH will release data submitted to NIH-designated data repositories no later than six months after the initial data submission begins or at the time of acceptance of the first publication, whichever occurs first, without restrictions on publication or other dissemination.

2. **Data Repositories**
Investigators should register all studies with human genomic data that fall within the scope of the GDS Policy in dbGaP by the time that data cleaning and quality control measures begin, regardless of which NIH-designated data repository will receive the data. After registration in dbGaP, investigators should submit the data to the relevant NIH-designated data repository (e.g., dbGaP, GEO, SRA, the Cancer Genomics Hub). NIH-designated data repositories need not be the exclusive source for facilitating the sharing of genomic data, i.e., investigators may also elect to submit data to a non-NIH-designated data repository in addition to an NIH-designated data repository. However, investigators should ensure that appropriate data security measures are in place, and that confidentiality, privacy, and data use measures are consistent with the GDS Policy.

3. **Tiered System for the Distribution of Human Data**

Respect for, and protection of the interests of, research participants are fundamental to NIH’s stewardship of human genomic data. The informed consent under which the data or samples were collected is the basis for the submitting institution to determine the appropriateness of data submission to NIH-designated data repositories and whether the data should be available through unrestricted or controlled access. Controlled-access data in NIH-designated data repositories are made available for secondary research only after investigators have obtained approval from NIH to use the requested data for a particular project. Data in unrestricted-access repositories are publicly available to anyone (e.g., The 1000 Genomes Project).

4. **Informed Consent** (see also Guidance on Informed Consent link below)

For research that falls within the scope of the GDS Policy, submitting institutions, through their Institutional Review Boards (IRBs), privacy boards, or equivalent bodies, are to review the informed consent materials to determine whether it is appropriate for data to be shared for secondary research use. Specific considerations may vary with the type of study and whether the data are obtained through prospective or retrospective data collections. NIH provides additional information on issues related to the respect for research participant interests in its Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications.24

For studies initiated after the effective date of the GDS Policy, NIH expects investigators to obtain participants’ consent for their genomic and phenotypic data to be used for future research purposes and to be shared broadly. The consent should include an explanation about whether participants’ individual-level data will be shared through unrestricted- or controlled-access repositories.

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.

5. **Institutional Certification**- Needs to be submitted just-in-time to the Program Officer

6. **Justification for any Exceptions to Data Submission Expectations**

7. **Data Withdrawal**
Submitting investigators and their institutions may request removal of data on individual participants from NIH-designated data repositories in the event that a research participant withdraws or changes his or her consent. However, some data that have been distributed for approved research use cannot be retrieved.

Additionally, the applicant must address any special requirements related to GDS data sharing and access that may be present in the Funding Opportunity Announcement.

**International Collaborations and Foreign Grants**
If the U.S. institution is the primary grantee, then the domestic institution is responsible for its sub-grantee or subcontract arrangements and is expected to ensure that this policy is adequately addressed in the application.

For more information, see the following NIH websites:


