

NIH-GDS: Genomic Data Sharing

National Institutes of Health

Data type

Explain whether the research being considered for funding involves human data, non-human data, or both.

Information to be included in this section:

- Type of data being collected: human, non-human, or both human & non-human.
- Type of genomic data to be shared: sequence, transcriptomic, epigenomic, and/or gene expression.
- Level of the genomic data to be shared: Individual-level, aggregate-level, or both.
- Relevant associated data to be shared: phenotype or exposure.
- Information needed to interpret the data: study protocols, survey tools, data collection instruments, data dictionary, software (including version), codebook, pipeline metadata, etc. This information should be provided with unrestricted access for all data levels.

Data repository

Identify the data repositories to which the data will be submitted, and for human data, whether the data will be available through unrestricted or controlled-access.

For human genomic data, investigators are expected to register all studies in the database of Genotypes and Phenotypes (dbGaP) by the time data cleaning and quality control measures begin in addition to submitting the data to the relevant NIH-designated data repository (e.g., dbGaP, Gene Expression Omnibus (GEO), Sequence Read Archive (SRA), the Cancer Genomics Hub) after registration.

Non-human data may be made available through any widely used data repository, whether NIH-funded or not, such as GEO, SRA, Trace Archive, Array Express, Mouse Genome Informatics, WormBase, the Zebrafish Model Organism Database, GenBank, European Nucleotide Archive, or DNA Data Bank of Japan.

Data in **unrestricted-access** repositories (e.g., The 1000 Genomes Project) are publicly available to anyone. **Controlled-access** data (e.g., data in dbGaP) are made available for secondary

research only after investigators have obtained appropriate approval to use the requested data for their proposed project.

Data submission and release timeline

Provide a timeline for sharing data in a timely manner.

In general, NIH will release human genomic data no later than six months after the data have been submitted to NIH-designated data repositories and cleaned, or at the time of acceptance of the first publication, whichever occurs first, without restrictions on publication or other dissemination of research findings. Investigators should make non-human genomic data publicly available no later than the date of initial publication. However, availability before publication may be expected for certain data, projects (e.g., data from projects with broad utility as a resource for the scientific community such as microbial population-based genomic studies), or by the funding NIH IC.

Note: The **Supplemental Information to the GDS Policy** provides expectations for the timelines of data submission and release based on the level of data processing, and additional information about the data levels.

IRB assurance of genomic data sharing plan

State whether an Institutional Review Board (IRB) or analogous review body has reviewed the genomic data sharing aspects of your project, or provide a timeline for such review.

IRB review of the investigator's proposal for data submission is an element of the Institutional Certification which assures that the proposal for data submission and sharing is appropriate. Please keep in mind that an Institutional Certification is generally required for extramural investigators prior to NIH grant award along with other Just-in-Time information or finalization of a contract. For NIH intramural investigators, an Institutional Certification memorandum should be completed and sent from the SD, or delegate, to the IC Genomic Program Administrator (GPA) before research is begun, whenever possible.

Points to Consider for Institutions and Institutional Review Boards in Developing Institutional Certifications for Submitting Human Data under the Genomic Data Sharing Policy.

Appropriate uses of the data

Describe the appropriate use of the data.

Under the GDS Policy, data is expected to be shared for broad research purposes. If such use of the data is not appropriate, as expressed in informed consent documents of the research participants whose data are included in the dataset, any limitations on the data use should be described in the Institutional Certification. NIH provides standard language (see Links tab for URL) to guide the development of data use limitations.

Request for an exception to submission

Explain why in the genomic data sharing plan and describe an alternative mechanism for data sharing .

If submission of human data generated in the study would not be appropriate because the Institutional Certification (see Links tab for URL) criteria cannot be met, the investigator should explain why in the genomic data sharing plan and describe an alternative mechanism for data sharing. If the funding IC grants an exception to submission, the research will be registered in dbGaP and the reason for the exception and the alternative sharing plan will be described. For NIH intramural studies, the NIH Deputy Director for Intramural Research will make the final decision on the exception request, after the IC has made its determination.