**Responsibilities of Investigators Submitting Genomic Data**

***Summarized from NIH Genomic Data Sharing Policy NOT-OD-14-124***

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html#sthash.MuceJv8h.dpuf>

**A. Genomic Data Sharing Plans**

Investigators seeking NIH funding should contact appropriate IC Program Official or Project Officer as early as possible to discuss data sharing expectations and timelines that would apply to their proposed studies. 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 and proposals. 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. The Institutional Certification (for sharing human data), should also be provided to the funding IC prior to award, along with any other Just-in-Time information. NIH expects intramural investigators to address compliance with genomic data sharing plans with their IC scientific leadership prior to initiating applicable research and are encouraged to contact their IC leadership or the Office of Intramural Research for guidance. The funding NIH IC will typically review compliance with genomic data sharing plans at the time of annual progress reports or other appropriate scientific project reviews, or at other times, depending on the reporting requirements specified by the IC for specific programs or projects.

**B. Non-human Genomic Data Sharing Plans**

***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. In general, investigators should make non-human genomic data publicly available no later than the date of initial publication. However, earlier availability (i.e., before publication) may be expected for certain data or IC-funded projects (e.g., data from projects with broad utility as a resource for the scientific community such as microbial population-based genomic studies).

***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). 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.

**C. Human Genomic Data Sharing Plans**

***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 repository48 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. 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.

Investigators should de-identify human genomic data that they submit to NIH-designated data repositories according to the standards set forth in the HHS Regulations for the Protection of Human Subjects to ensure that the identities of research subjects cannot be readily ascertained with the data. Investigators should also strip the data of identifiers according to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. The de-identified data should be assigned random, unique codes by the investigator, and the key to other study identifiers held by the submitting institution.

Although the data in the NIH database of Genotypes and Phenotypes (dbGaP) are de-identified by both the HHS Regulations for Protection of Human Subjects and HIPAA Privacy Rule standards, NIH has obtained a Certificate of Confidentiality for dbGaP as an additional precaution because genomic data can be re-identified. NIH encourages investigators and institutions submitting large-scale human genomic datasets to NIH-designated data repositories to seek a Certificate of Confidentiality as an additional safeguard to prevent compelled disclosure of any personally identifiable information they may hold.

***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, that is, 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***

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.

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.

For studies using data from specimens collected before the effective date of the GDS Policy, there may be considerable variation in the extent to which future genomic research and broad sharing were addressed in the informed consent materials for the primary research. In these cases, an assessment by an IRB, privacy board, or equivalent body is needed to ensure that data submission is not inconsistent with the informed consent provided by the research participant. NIH will accept data derived from de-identified cell lines or clinical specimens lacking consent for research use that were created or collected before the effective date of this Policy.

NIH recognizes that in some circumstances broad sharing may not be consistent with the informed consent of the research participants whose data are included in the dataset. In such circumstances, institutions planning to submit aggregate- or individual-level data to NIH for controlled access should note any data use limitations in the data sharing plan submitted as part of the funding request. These data use limitations should be specified in the Institutional Certification submitted to NIH prior to award.

***5. Institutional Certification***

The responsible Institutional Signing Official of the submitting institution should provide an Institutional Certification to the funding IC prior to award consistent with the genomic data sharing plan submitted with the request for funding. The Institutional Certification should state whether the data will be submitted to an unrestricted- or controlled-access database. For submissions to controlled access and, as appropriate for unrestricted access, the Institutional Certification should assure that:

The data submission is consistent, as appropriate, with applicable national, tribal, and state laws and regulations as well as relevant institutional policies;

* Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated;
* The identities of research participants will not be disclosed to NIH-designated data repositories; and
* An IRB, privacy board, and/or equivalent body, as applicable, has reviewed the investigator’s proposal for data submission and assures that:
* The protocol for the collection of genomic and phenotypic data is consistent with 45 CFR Part 46;
* Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
* Consideration was given to risks to individual participants and their families associated with data submitted to NIH-designated data repositories and subsequent sharing;
* To the extent relevant and possible, consideration was given to risks to groups or populations associated with submitting data to NIH-designated data repositories and subsequent sharing; and
* The investigator’s plan for de-identifying datasets is consistent with the standards outlined in this Policy (see section IV.C.1.).

***6. Exceptions to Data Submission Expectations***

In cases where data submission to an NIH-designated data repository is not appropriate, that is, the Institutional Certification criteria cannot be met, investigators should provide a justification for any data submission exceptions requested in the funding application or proposal. The funding IC may grant an exception to submitting relevant data to NIH, and the investigator would be expected to develop an alternate plan to share data through other mechanisms. For transparency purposes, when exceptions are granted, studies will still be registered in dbGaP, the reason for the exception will be included in the registration record, and a reference will be provided to an alternative data-sharing plan or resource, if available. More information about requesting exceptions is available on the GDS website.

***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.

See: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html#sthash.MuceJv8h.dpuf>