**Guidance for Institutions Submitting Grant Applications and Contract Proposals under the NIH Genomic Data Sharing Policy for Human and Non-Human Data**

**Purpose:**

This document provides guidance to institutions and organizations regarding the expectations outlined in the Genomic Data Sharing (GDS) Policy. Institutions and organizations may use this document to develop their own internal guidance or institutional policy documents.

**Background:**

Genomic research advances our understanding of factors that influence health and disease, and sharing genomic data provides opportunities to accelerate that research through the power of combining large and information-rich datasets. To promote robust sharing of human and non-human data from a wide range of genomic research and to provide appropriate protections for research involving human data, the National Institutes of Health (NIH) issued the *NIH Genomic Data Sharing Policy* on August 27, 2014 in the *NIH Guide for Grants and Contracts* (available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html), and in the *Federal Register* (available at https://federalregister.gov/a/2014-20385) on August 28, 2014.

**GDS Policy:**

The GDS Policy and supporting documents set forth expectations and responsibilities for investigators and grantee institutions that ensure broad and responsible sharing of large scale genomic data in a timely manner.

This policy pertains to all NIH-funded research generating large-scale ***human*** or ***non-human*** **genomic data** and the use of these data for subsequent research. Examples of large-scale genomic data may be found within the Supplemental Information to the GDS Policy and include, but are not limited to, Genome-Wide Association Studies (GWAS); single nucleotide polymorphism (SNP) arrays; and genome sequence, transcriptomic, epigenomic, and gene expression data. This Policy applies to all funding mechanisms (grants and contracts) without a threshold for cost.

NIH strongly encourages all investigators conducting research involving large-scale genomic data to comply with the expectations outlined in the Policy and as described in NOT-OD-14-111: Implementation of the NIH GDS Policy for NIH Grant Applications and Awards:

- Previously unfunded grant applications involving large-scale genomic data are expected to address genomic data sharing if submitted for the **January 25, 2015** due date or thereafter**.**
- For studies funded prior to January 25, 2015, investigators:
  - are encouraged to comply with the expectations outlined in the GDS Policy
  - should provide an updated genomic data sharing plan to the funding NIH Institute or Center (IC) in the next submission of the research performance progress report, and
  - should transition to an informed consent for broad sharing and future research purposes, if possible, particularly for new or additional collections of specimens, if the studies involve human participants.

**Responsibilities of Investigators and Institutions in Submitting and Accessing Data Under the GDS Policy:**

1. *Grant Applications and Contract Proposal*

a) *Data Sharing Plan*. Information regarding how a research project will share large-scale genomic data should be described in the Resource Sharing Plan section of the grant application or as per instructions in a Request for Proposals for a contract. Certain NIH Funding Opportunities and Notices may specify different instructions.

At a minimum, such data sharing plan 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
- 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

Examples of data sharing plans can be found at: http://gds.nih.gov/06researchers1.html.

In cases where data submission to an NIH-designated repository is not appropriate, that is, if the criteria of the Institutional Certification cannot be met, investigators should provide a justification for an exception to the data sharing expectation in the grant application, prior to the grant award, or in the contract proposal. The NIH funding IC may grant an exception to the expectation for submission of data, and the investigator would be expected to develop an alternate plan to share data through other mechanisms. Exceptions to submission of data under the GDS Policy are made by the NIH funding IC; the study will still be registered in the database of Genotypes and Phenotypes (dbGaP) and the exception publicly explained there.

b) *Budget*. The application should propose any appropriate funds needed for data sharing activities.

c) *Human Subjects*. Investigators must address relevant human subjects' protection issues in the Human Subjects section of the application or proposal.

Prior to award, investigators must submit an Institutional Certification that the data are appropriate for sharing and consistent with the informed consents of the study participants. The Institutional Certification is submitted by the Institutional Signing Official as part of the Just-in-Time process and specifies terms and conditions under which human data may be shared. 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;[1]

[1] For the submission of data derived from cell lines or clinical specimens lacking research consent that were created or collected before the effective date of this Policy, the Institutional Certification needs to address only this item.

- Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated;[2]
- 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;[3]
  - Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;[4]
  - 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 the GDS Policy (see section IV.C.1. in the GDS Policy).

Investigators are encouraged to obtain a Certificate of Confidentiality to help protect against compelled disclosure of any personally identifiable information.

Investigators should include as part of their annual research reporting requirement (e.g., Type 5) an update on progress made toward their genomic data sharing plan.

2. *Informed Consent*

Respect for, and protection of the interests of research participants, are fundamental to the NIH's stewardship of human genomic data. 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 informed consent should state whether the data will be shared via unrestricted (data accessible to anyone via a public website) or controlled-access (data are available if certain stipulations are met) repositories. Also, for studies proposing to use genomic data from cell lines or clinical specimens that were collected or created after the effective date of the Policy, NIH expects that consent for future research use and broad sharing will have been obtained, even if the cell lines or clinical specimens are de-identified.

For studies proposing to use genomic data from cell lines or clinical specimens[5] 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-

[2] For guidance on clearly communicating inappropriate data uses, see NIH Points to Consider in Drafting Effective Data Use Limitation Statements, http://gwas.nih.gov/pdf/NIH_PTC_in_Drafting_DUL_Statements.pdf.
[3] 45 CFR Part 46. Protection of Human Subjects. See http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
[4] As noted earlier, for studies using data or specimens collected before the effective date of this Policy, the IRB, privacy board, or equivalent body should review informed consent materials to ensure that data submission is not inconsistent with the informed consent provided by the research participants.
[5] Clinical specimens are specimens that have been obtained through clinical practice.

identified. However, 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.

3. *Data Submission*

Investigators with questions about whether the Policy applies to their current or proposed research should consult the relevant NIH Program Official or Project Officer or the IC's Genomic Program Administrator (GPA).

The GDS Policy expects that large-scale genomic research data from NIH-supported studies involving human specimens will be submitted to an NIH-designated data repository. Non-human data may be submitted to any widely used repository or to the same repositories they submitted specific types of data to previously. NIH has provided examples of relevant databases on the GDS website.

Expectations and timelines for data submission and release vary according to the type of data to be submitted and the level of data processing.

- For non-human genomic data, data should be submitted no later than the date of initial publication. However, an earlier release date may be expected for certain data types or projects with broad utility as a resource for scientific community.
- For human genomic data, data submission expectations are generally based on the processing that data must undergo prior to submission. Expectations for human genomic data preparation and submission can be found in the Supplemental Information to the GDS Policy.

Human genomic data are to be registered in dbGaP, but then submitted to the appropriate NIH-designated repository. Due to the potentially sensitive nature of these data, steps must be taken to protect the interests of the research participants. Before data are submitted to NIH-designated repositories, submitting investigators are expected to de-identify the data per HIPAA regulations and assign a random, unique code to the data to protect participants' privacy and confidentiality. The key to the code will remain at the institution and will not be shared with NIH and dbGaP staff.

Information on how to submit genomic data into dbGaP can be found at:

- http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=HowToSubmit.pdf
- http://youtu.be/gIjN0Er39us

Submitters may request removal of data from NIH-designated data repositories in the event that a research participant withdraws his or her consent. However, data that have been distributed for approved research use cannot be retrieved.

*4. Data Access for Human Genomic Data*

To request access to any individual-level human datasets within the controlled-access of dbGaP or other NIH-designated repositories, investigators and the Signing Official at the requesting investigator's institution co-sign a data access request (DAR), which will be reviewed by an NIH Data Access Committee (DAC).

Although data are de-identified, approved users of controlled access data are encouraged to consider whether a Certificate of Confidentiality could serve as an additional safeguard to prevent compelled disclosure of any genomic data they may hold.

*a) Terms and Conditions for Research Use of Controlled-Access Data*

The data requestor and the requestor's institution sign a Data Use Certification (DUC) agreeing to the NIH terms of access for the dataset. DACs make decisions based primarily on conformity of the proposed research use to the data use limitations specified in the Institutional Certification provided by the submitting institution. If the DAR is approved, the requestor is granted access to a dataset for one year. At the end of the one year period, users can request access for an additional year through a renewal process, or close out the project.

Investigators and their institutions that are approved to download data from NIH-designated repositories are expected to abide by the NIH Genomic Data User Code of Conduct through their agreement to the DUC.

Human Genomic Data is accessed through dbGaP:
https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login

Information on how to access genomic data in dbGaP can be found:

- https://dbgap.ncbi.nlm.nih.gov/aa/dbgap_request_process.pdf
- http://youtu.be/-3tUBeKbP5c

*i) Acknowledgements*

By agreeing to the terms of the data use certification, investigators and their institutions agree to acknowledge in all oral or written presentations, disclosures, or publications the contributing investigator(s) who conducted the original study, the funding organization(s) that supported the work, the specific dataset(s) and applicable accession number(s), and the NIH-designated data repositories through which the investigator accessed any data.

*ii) Data Security*

NIH expects that investigators who are approved to use controlled-access data will follow guidance on security best practices that outlines expected data security protections (e.g., physical security measures and user training) to ensure that the data are kept secure and not released to any person not permitted to access the data.

*b) Conditions for Use of Unrestricted-Access Data*

Investigators who download unrestricted-access data from NIH-designated data repositories should:

- not attempt to identify individual human research participants from whom the data were obtained
- acknowledge in all oral or written presentations, disclosures, or publications the specific dataset(s) or applicable accession number(s) and the NIH-designated data repositories through which the investigator accessed any data

5. *Intellectual Property*

The NIH GDS Policy considers basic sequence data (e.g., genotypes, haplotypes, p-values, allele frequencies) to be "pre-competitive", therefore these data and any conclusions derived from them should remain freely available without licensing requirements.

**Additional Information/Resources**

- NIH Genomic Data Sharing Website
- NIH Genomic Data Sharing Policy
- NIH Points to Consider for Institutional Review Boards and Institutions in their Review of Data Submission Plans for Institutional Certifications Under NIH's Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) Points to Consider Data Use Limitations
- NIH Guidance for Developing Data Sharing Plans
- NIH database of Genotypes and Phenotypes (dbGaP)
    - NIH database of Genotypes and Phenotypes (dbGaP) Submission System
    - dbGaP Access System
    - YouTube Tutorials for dbGaP:
        - Data Submission: http://youtu.be/gIjN0Er39us
        - Data Access: http://youtu.be/-3tUBeKbP5c
- eRA Commons
- National Human Genome Research Institute (NHGRI) Consent Form Examples and Model Consent Language for Genomics Research
- NIH Genomic Data Sharing Website Data Repositories