UCLA Research Administration
Human Research Protection Program

Guidance and Procedure: NIH Genomic Data Sharing (GDS) and Genome Wide Association Studies (GWAS) (October 20, 2020)

Introduction

The NIH GDS Policy became effective of on January 25, 2015, and sets forth expectations that ensure the broad and responsible sharing of genomic research data. 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. Information on JIT requests, see UCLA OHRPP Guidance and Procedure: Funding Applications and UCLA IRB Review.

Definitions

The following words and definitions are defined by the NIH policies and are useful to understand the NIH GDS Policy.

- **NIH 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 genomic data** – GWAS, single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomics, epigenomic, and gene expression data. See Supplemental Information to the NIH Genomic Data Sharing Policy for more examples that are subject to this policy.
    - 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
    - Sequence data from more than 100 isolates from infectious organisms

- **Database of Genotypes and Phenotypes (dbGaP)** – Data from the NIH-funded human genomic research will be submitted to an NIH-designated data repository known as dbGaP.

- **NIH GWAS Data Repository** – Also known as dbGaP, a database developed by the National Center for Biotechnology Information (a division of the National Library of Medicine) to archive and distribute the results of studies that have been investigated.

- **Controlled-access** – Data is available to an investigator for a specific project only if certain stipulations are met.

- **Unrestricted-access** – Data is accessible to anyone via public website (previously referred to as “open access”)

- **Coded data** – any identifying information, such as name, that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a number, letter, symbol, or combination thereof (i.e., the code) and a kay to decipher the code exists, enabling linkage of the identifying information to the private information or specimens

- **Deidentified data** – data that has been de-identified according to the following criteria:
    - The identifiers cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR 46.102(f));
    - The 18 identifiers enumerated at 45 CFR 164.514 (b)(2) are removed;
    - AND the submitting institution has no actual knowledge that the remaining information could be used alone or in combination with other information to identify the subject of the data

Protocol or IRB Application Requirements

The protocol or appropriate webIRB sections should describe the following:

- Explicitly state that data will be sent to NIH for GDS

- A description of the genotype and phenotype data to be sent

- A statement indicating whether data will be collected prospectively or if the data has already been collected
    - If the specimens have already been collected, indicate the UCLA IRB number these specimens were collected under.

- A plan for maintaining privacy, e.g., identifiers will not be sent to NIH

- A description of risks to subjects resulting from loss of privacy

Informed Consent Guidelines

Investigators who intend to use cell lines or clinical specimens collected after the NIH GDS Policy effective date, to generate genomic data may only do with consent, even if the data is generated from specimens that 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 the NIH GDS Policy and that lack of consent for research use and data sharing, investigators should provide a justification in the funding request for their use.

For studies using data from specimens collected before the effective date of the NIH 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. UCLA OHRPP will 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 NIH GDS Policy effective date.

UCLA OHRPP are required to review the informed consent to determine whether it is appropriate for data to be shared for secondary research use. NIH has recommended the following information be included in the consent form in order to be compliant with the NIH GDS policy:

- Genomic and phenotypic data, and any other data relevant for the study (such as exposure or disease status) will be generated and may be shared broadly and used for future research in a manner consistent with the participant's informed consent and all applicable federal and state laws and regulations.

- Prior to submitting the data to an NIH-designated data repository, data will be stripped of identifiers such as name, address, account and other identification numbers and will be deidentified by standards consistent with the Common Rule and HIPAA. Safeguards to protect the data according to Federal standards for information protection will be implemented.

- Access to de-identified, individual-level participant data will be controlled, unless participants explicitly consent to allow unrestricted access to and use of their data for any purpose.

- Aggregate study information (including genomic summary results) and study analyses may be shared in scientific literature or through other public scientific resources, such as data repositories or other data sharing resources that provide broad or unrestricted access to the information.

- Because it may be possible to re-identify de-identified genomic data, even if access to data is controlled and data security standards are met, confidentiality cannot be guaranteed, and reidentified data could potentially be used to discriminate against or stigmatize participants, their families, or groups. In addition, there may be unknown risks due to computational methods, analytic technologies, or techniques (e.g., generation of information that could allow participants' identities to be readily ascertained).

- No direct benefits to participants are expected from any secondary research that may be conducted.

- Participants may withdraw consent for research use of genomic or phenotypic data at any time without penalty or loss of benefits to which the participant is otherwise entitled. In this event, data will be withdrawn from any repository, if possible, but data already distributed for research use will not be retrieved.

- The name and contact information of an individual who is affiliated with the institution and familiar with the research and will be available to address participant questions.

- The privacy protections, and limitations of those protections, afforded by a Certificate of Confidentiality to individual-level data do not apply to summary results.

Institutional Certification

The NIH requires that the submitting institution should provide an Institutional Certification consistent with the genomic data sharing plan. The Institutional Certification should state whether the data will be submitted to an unrestricted or controlled-access database.

The Institutional Certification should assure:

- 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 the NIH GDS Policy (section IV.C.1.)

Once OHRPP has verified that the consent and institutional certification requirements have been met, OHRPP will facilitate the Institutional Officials signature required on the institutional certificate.

References and Regulations

- NIH Guidance on Consent for Future Research Use and Broad Sharing of Human Genomic and Phenotypic Data Subject to the NIH Genomic Data Sharing Policy
- NIH Points to Consider for Institutions and Institutional Review Boards in Submission and Secondary Use of Human Genomic Data under the National Institutes of Health Genomic Data Sharing Policy
- National Institutes of Health Genomic Data Sharing Policy