POLICY AND PROCEDURES FOR HUMAN GENOMIC DATA SHARING

Policy Statement

The NIH Genomic Data Sharing ("GDS") Policy requires the sharing of large-scale human genomic data generated from NIH-funded research, as well as the use of these data, to an NIH-designated data repository when such sharing is compatible with the consent provided by the participant. The Policy applies to all competing NIH grant applications, proposals, and contracts submitted to NIH after January 25, 2015. All submissions of large-scale human genomic data to an NIH designated data repository, whether required or voluntary, must include a certification by the responsible Institutional Official(s) of the submitting institution that the submission of data to the repository is appropriate and consistent with the NIH GDS Policy. At Harvard, the Institutional Official of the relevant IRB provides this certification. This policy and procedures guidance outlines the responsibilities of the Harvard investigator, the relevant IRB, and the Institutional Officials in submitting, reviewing and providing the required Institutional Certification for a submission to an NIH–designated data repository. Where no Institutional Certification is required, Harvard investigators need not meet all the elements of the procedures for Institutional Certification outlined below. PIs may, however, consult with the IRB regarding whether genomic data sharing is appropriate in such instances.

Applicability

The NIH GDS Policy applies to all Harvard investigators who:

- Submit NIH grant applications or proposals after the January 25, 2015 deadline to conduct research that will generate large-scale human genomic data or will use these data for subsequent research; or
- Plan to voluntarily submit any large-scale human genomic data into one of the NIH-designated data repositories, even if the research itself is not NIH-supported; or
- Plan to voluntarily submit human genomic data where Institutional Certification is otherwise required by NIH Program Officers, or other external obligations.

The NIH GDS Policy also applies to the relevant Harvard IRBs and Institutional Officials.

Definitions

NIH Genomic data sharing policy: ("GDS Policy") is the policy that large-scale human and non-human genomic data generated from NIH-funded research should be shared broadly and responsibly, through a central repository. In the case of human genomic data, the GDS Policy requires that the IRB of a submitting institution review the informed consent under which human material was collected to

determine whether it is appropriate for the data to be shared for secondary research use. Further information on the GDS Policy can be found at: http://gds.nih.gov

Under the GDS Policy, an **NIH-designated data repository** is any data repository maintained or supported by NIH either directly or through collaboration, including any of the following:

- Database of Genotypes and Phenotypes ("dbGaP"),
- Gene Expression Omnibus ("GEO"),
- Sequence Read Archive ("SRA"), or the
- Cancer Genomics Hub.

The **human genome** is all of the DNA contained in a human cell, including both the DNA comprising chromosomes within the nucleus and the DNA in mitochondria.

Institutional Officials: At Harvard, the Institutional Officials include the University Chief Research Compliance Officer for the Harvard University Area IRB and the Harvard Faculty of Medicine IRB ("HMS/HSDM IRB"), and the Harvard T.H. Chan School of Public Health ("Harvard Chan School IRB") Associate Dean for Regulatory Affairs and Research Compliance for the Harvard Chan School.

Large-scale genomic data is defined by the NIH GDS policy as including genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data. Examples of research that are subject to the GDS Policy include, but are not limited to, projects that involve generating the whole genome sequence data for more than one gene from more than 1,000 individuals, or analyzing 300,000 or more genetic variants in more than 1,000 individuals, or sequencing more than a 100 isolates of infectious organisms such as bacteria. Additional examples may be found in the Supplemental Information to the NIH Genomic Data Sharing Policy at: https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf. Generally, sequencing the genomes of fewer than 100 human research participants, RNA-seq data from cadaveric tissues from less than 100 donor samples, or single-cell RNA seq data from less than 100 donor samples are NOT considered large scale genomic data.

Phenotype data: Data on health conditions, behavioral characteristics, or measurable observable traits (such as blood pressure, alcohol consumption, cholesterol, or eye color) that are obtained during physical or psychological examinations and maintained in a medical or research record. Phenotype data may also include information about medical treatments, drug tolerance, and family medical history as well as responses to questionnaires.

Relevant IRB: The Harvard IRB that reviews research proposals from the PI's School. The Harvard University Area IRB serves as the Institutional Review Board for Schools located on the Cambridge and Allston campuses at Harvard. The two Harvard Longwood Medical Area (LMA) IRBs serve as the Institutional Review Boards for Harvard T.H. Chan School of Public Health (HSPH) and Harvard Faculty of Medicine (HMS/HSDM).

Submitting Institution: The institution listed on the deposit application is the submitting institution. This is the institution responsible for the deposition of the data into an NIH designated repository. Factors used to consider whether Harvard is the depositing institution include: was a Harvard IRB the IRB of record for

a collection study; was Harvard the primary recipient of a federal award that funded the generation of the genomic data; was the genomic data generated at Harvard; and/or will the genomic data continue to be stored at Harvard after the project is over.

Procedures for Submission of Data Generated from Materials that are to be Prospectively Collected

For Institutional Certification of a data submission to an NIH-designated data repository, a Harvard Principal Investigator ("PI") must first submit to the relevant IRB:

- A description of all data fields (genotype and phenotype) being submitted to the NIH-designated data repository.
- A copy of the consent form(s) used to enroll participants and collect their samples and phenotype data. (Note: Investigators are advised to always use the most up to date informed consent template in ESTR Library).
- A description of the PI's plan for de-identifying datasets for transmission to the NIH-supported data repository and how the key linking the identity of each study participant will be maintained. Investigators should de-identify human genomic data 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.
- A written statement by the Harvard PI that the key to identifiers of study participants will never be shared by the Harvard investigator with NIH.

IRB Review and Verification: For any proposed submission of phenotype data and genotype data to an NIH-designated data repository, the relevant IRB will review the documents provided by the PI to determine whether:

- 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.

IRB determination: After reviewing the IRB submission, data sharing plan, and the relevant consents, if any, the IRB may determine the following:

- Submission of the specified data to an NIH-designated data repository is appropriate;
- Submission of the specified data to an NIH-designated data repository is not appropriate
- Submission of the specified data to an NIH-designated data repository is not appropriate, but that an attempt to cure the deficiency is appropriate. Therefore, as a condition for submission, the IRB may:

- Require revision of the consent form(s) to be consistent with a submission of data to the NIH-designated repository with re-consent of the research participants;
- Permit the submission, but subject it to certain restrictions or limitations on use, as the relevant Harvard IRB may specify; or
- Request additional information, as necessary.

Procedures for Submission of Data Generated from Materials that Have Already Been Collected or Collected by Another Institution

For Institutional Certification of a data submission to an NIH-designated data repository, a Harvard Principal Investigator ("PI") must first submit to the relevant IRB:

- A description of all data fields (genotype and phenotype) being submitted to the NIH-designated data repository;
- A copy of the consent form(s) used to enroll participants and collect their samples and phenotype data.
    - For data submissions where a Harvard IRB did not review the protocol under which the original samples and phenotype data were collected:
        - The relevant IRB may accept certification from the IRB or other ethical oversight body of record, or from the biospecimen bank/repository from where the samples were received, in lieu of the original consent forms, or
        - The relevant IRB may request the original consents and review the consent form in accordance with the procedures outlined below, in the section **IRB Review and Verification**, or
        - The relevant IRB may request documentation that the original collection was consistent with 45 CFR 46.
    - If no consents are available, the Harvard PI must:
        - Request written confirmation from the provider of the original samples that submission to an NIH-designated repository is consistent with the NIH GDS Policy, or
        - Consult with the relevant IRB to determine whether data may be submitted to an NIH-designated repository.
- A description of the PI's plan for de-identifying datasets for transmission to the NIH-supported data repository and how the key linking the identity of each study participant will be maintained. Investigators should de-identify human genomic data 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.
- A written statement by the Harvard PI that the key to identifiers of study participants will never be shared by the Harvard investigator with NIH.

IRB Review and Verification: For any proposed submission of phenotype data and genotype data to an NIH-designated data repository, the relevant IRB will review the documents provided by the PI to determine whether:

- 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;
    - For studies using data generated from human material collected **after January 25, 2015** (the effective date of the GDS policy), the consent form, must:
        - Include a discussion of future research uses and broad sharing of genotypic and phenotypic data.
        - Address whether participants' individual-level data will be shared through unrestricted or controlled-access repositories, and
        - Describe the likelihood of re-identification in the future.
    - For studies using data generated from human material collected **before January 25, 2015**, the IRB should review the consent form, if any, to ensure that the data submission is not inconsistent with the informed consent provided by the research participant.
    - If there is a consent form, the IRB may consider the following questions:
        - Do the consent documents identify future research use of the specimens or data?
        - Do the consent documents address disposal of the specimens or data?
        - Do any of the consent forms contain statements such as "your data/specimens will not be shared" or "will only be seen by the research team?"
        - Did the original consent forms limit future use to specific projects, conditions, disease states, or to non-commercial research?

    - For studies using data generated from materials where information about the consent process or the consent forms are not available[1], Institutional Certification will be based upon the IRB's determination of whether:
        - Submission of the data to an NIH depository would be appropriate;
          The protocol for the collection of genomic and phenotypic data is consistent with 45 CFR Part 46;
          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

[1] This is only applicable to the submission of genomic data generated from deceased individuals (e.g. data generated from cadaveric tissue) or data generated from samples collected before January 25, 2015, where information about the consent process or the consent forms are not available. The NIH GDS policy indicates that NIH-designated depositories will accept genomic data if the individuals from whom the data was derived are deceased and information about the consent process or the consent forms are not available and the submitting institution, in concert with its IRB and/or privacy board, finds that submission of the data to the NIH would be appropriate (e.g., in cases where there is consent, the consent form does not preclude data sharing) and meets the other expectations of the Institutional Certification specified within the NIH GDS policy, then such data would be accepted by an NIH-designated data repository. Although NIH-designated data repositories do not currently involve human subjects research under 45 CFR 46, the criteria defined within 45 CFR 46.116(d) might provide a useful framework to institutions, IRBs, and privacy boards in considering submission of data from deceased individuals.

        - The investigator's plan for de-identifying datasets is consistent with the standards outlined in NIH GDS Policy.
- 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 NIH GDS Policy.

IRB determination: After reviewing the IRB submission, data sharing plan, and the relevant consents, if any, the IRB may determine the following:

- Submission of the specified data to an NIH-designated data repository is appropriate;
- Submission of the specified data to an NIH-designated data repository is not appropriate
- Submission of the specified data to an NIH-designated data repository is not appropriate, but that an attempt to cure the deficiency is appropriate. Therefore, as a condition for submission, the IRB may:
    - Permit the submission, but subject it to certain restrictions or limitations on use, as the relevant Harvard IRB may specify; or
    - Request additional information, as necessary.

Institutional Certification

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, including whether genomic summary results should be provided only through a controlled access data access request, are delineated;
- The identities of research participants will not be disclosed to NIH-designated data repositories; and
- Verified that the relevant IRB has reviewed and documented the submission according to the Policy.

For NIH funded genomic research, Institutional Certification is required before data can be shared with an NIH-designed data repository. The Institutional Certification (for sharing human data), should also be provided to the funding NIH Institute or Center prior to award, along with any other Just in Time information (for extramural researchers) or at the time of scientific review (for intramural researchers). Institutional Certification is issued by the relevant IRB after its review is complete. The relevant IRB may use either the available Institutional Certification forms (available at https://osp.od.nih.gov/scientific-sharing/institutional-certifications/) or rely on its own internal templates. The relevant IRB will indicate the issuance of an Institutional Certification in their system of record, Electronic Submission, Tracking, and Reporting (ESTR).

Non-NIH Repositories: The focus of this guidance is the NIH GDS Policy and submission to NIH-designated data repositories. For non-NIH designated data repositories such as the European Genome-Phenome Archive ('EGA"), PIs are advised to comply with the specific requirements of such data repositories. If no IRB approval or Institutional Certification is required, all elements of the procedures outlined above are not required. PIs may, however, consult the IRB regarding whether genomic data sharing is appropriate.

Receipt of Genomic Data: Requests of genomic data from NIH designated data repositories or other data repositories requiring institutional approval or signature must comply with the process established in the University's Data Use Agreements Guidance.

Contacts and Resources

The Office of the Vice Provost for Research: http:///www.vpr.harvard.edu
CUHS IRB: http://cuhs.harvard.edu/
Harvard LMA IRBs: http://www.hsph.harvard.edu/ohra

NIH Genomic Data Sharing FAQs

NIH Points to Consider for IRBs and Institutions

HU Data Use Agreements Guidance

Appendix A
NIH GDS Policy: Guidance for PIs

- **How do I determine whether my project involves Large Scale Genomic Sequencing, subject to the NIH GDS Policy?**

The NIH GDS policy defines large-scale genomic sequencing as including genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data. Examples of research that are subject to the GDS Policy include, but are not limited to, projects that involve generating the whole genome sequence data for more than one gene from more than 1,000 individuals, or analyzing 300,000 or more genetic variants in more than 1,000 individuals, or sequencing more than a 100 isolates of infectious organisms such as bacteria. Additional examples may be found in the Supplemental Information to the NIH Genomic Data Sharing Policy at: https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf. Generally, RNA-seq data from cadaveric tissues from less than 100 donor samples, or single-cell RNA seq data from less than 100 donor samples is NOT considered large-scale genomic data.

- **My research is not NIH-funded. Do I need to comply with the NIH GDS policy?**

Yes, if you decide at some point to voluntarily submit large-scale genomic data to an NIH-designated depository. Even if you do not currently anticipate depositing genomic data or human biospecimen into NIH-designated repositories, you should consider the NIH GDS Policy FAQ sections on Submission of Data to Controlled-Access Repositories and Consent for Broad Sharing found at: https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/ and Harvard University's Policy and Procedures on Genomic Data Sharing. Your plans for depositing may change in the future and we strongly suggest maximizing the downstream use of biospecimen and/or data by including all the required elements of consent initially or ensuring that samples used to generate genomic data meet the requirements of the NIH GDS Policy.

- **My genomic research project involves biospecimen and data that is not individually identifiable. Do I need to submit an IRB protocol for this research?**

Yes, the NIH GDS Policy applies even if the research is considered exempt or "not human subjects research." Therefore, an IRB review is needed for Institutional Certification. This may include (but is NOT limited to): purchased samples, pre-existing cell lines, or cadaveric tissue.

- **What information is required for IRB review and Institutional Certification?**

    - Copies of all consent forms for samples that will be sequenced or have been sequenced for this project.
        - For data submissions where a Harvard IRB did not review the protocol under which the original samples and phenotype data were collected:
            - The relevant IRB may accept certification from the IRB or other ethical oversight body of record, or from the biospecimen bank/repository from where the samples were received, in lieu of the original consent forms, or

        - The relevant IRB may request the original consents and review the consent form in accordance with the procedures outlined below, in the section IRB Review and Verification, or
        - The relevant IRB may request documentation that the original collection was consistent with 45 CFR 46.
    - If no consents are available, the Harvard PI must:
        - Request written confirmation from the provider of the original samples that submission to an NIH-designated repository is consistent with the NIH GDS Policy, or
        - Consult with the relevant IRB to determine whether data may be submitted to an NIH-designated repository.

- An IRB submission in Electronic Submission, Tracking, and Reporting (ESTR) that includes:
    - A description of all data fields (genotype and phenotype) being submitted to the NIH-designated data repository
    - Whether the data will be deposited into a controlled-access or open-access database
    - A description of the PI's plan for de-identifying datasets for transmission to the NIH-supported data repository and how the key linking the identity of each study participant will be maintained.