Penn | Human Research Protections Program
University of Pennsylvania

Institutional Certification Requests for Genomic Data Sharing

Contents

Overview

This guidance applies to the NIH Genomic Data Sharing (GDS) Policy. This guidance supersedes the prior guidance: Summary of IRB Policies, Procedures and Considerations When Reviewing Requests for Certification of Data Submission for Sharing of Data in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).

NIH has strict standards for IRB review and informed consent for the data they will accept for inclusion in GDS data repositories. The IRB is required to review investigators' requests to submit data to the NIH data repositories and must also certify that the informed consent that was obtained from participants was consistent with NIH requirements for sharing genomic data.

In addition, grant applications that request NIH funding for genomic research must develop and provide a plan for sharing genomic data as a part of the required Data Management and Sharing Plan under the NIH Policy for Data Management and Sharing. The genomic data sharing plan must be consistent with NIH GDS policy.

This guidance provides (i) instructions on how to submit requests for Institutional Certification, (ii) information regarding how the IRB will consider these requests and (iii) guidance on how studies can be drafted and submitted in a manner that includes sufficient detail to ensure that Institutional Certification can be granted.

When does the GDS Policy apply?
The GDS Policy applies to the following:

- NIH-funded research that generates human or non-human genomic data (e.g. SNP arrays, genome sequencing, RNA sequencing, transcriptomic, metagenomics, epigenomic and gene expression data, GWAS studies) from more than 100 individuals.
- Studies that are not NIH-supported but plan to submit genotype/phenotype data to one of the following NIH Supported repositories:
    - Database of Genotypes and Phenotypes (dbGaP)
    - Gene Expression Omnibus (GEO)
    - Sequence Read Archive (SRA)
    - Cancer Genomics Hub (CaHUB)

When does the GDS Policy not apply?

- When the genomic data is generated without NIH funds (unless the researcher voluntarily requests submission to one of NIH-supported repositories)
- When NIH-funded projects involve instrument calibration exercises, statistical or technical methods development, or the use of genomic data for control purposes, such as for assay development
- When the following funding is requested: Institutional Training Grants (T32s, T34s, T35s, and TL2s), Career Development Awards (Ks), Individual Fellowships (Fs), Resource Grants and Contracts (Ss), linked awards derived from previously reviewed applications, or facilities or coordinating centers funded through related initiatives to provide genotyping, sequencing, or other core services in support of GDS

How to Submit a Request for Institutional Certification

The request for an Institutional Certification should be sent to the IRB via email. The request should be sent to the general IRB inbox address (PROVOST-IRB@pobox.upenn.edu) and to Patrick Stanko (pstanko@upenn.edu).

The email should also include:

- A cover letter addressed to Patrick Stanko;
- A copy of the informed consent form and, if applicable, the HIPAA authorization form; and
- A GDS certification template completed and signed by the Principal Investigator. Certification letter templates are available from the NIH website. Please ensure that the signature dates are current.

Please ensure that the cover letter and the text of the email both include the protocol number associated with the IRB approved protocol that obtained consent from subjects for the collection of data and/or specimens for future research. If data and/or specimens were provided from a non-Penn study team or entity, please include information regarding that outside entities' collection procedures and consent processes.

The cover letter should serve as a summary of the request for the Institutional Certification. It should clearly state the request for Institutional Certification to share genomic data. The summary should also include justification for why the IRB can appropriately assure that the criteria for certification have been met for the specific study. The criteria for certification are:

- The data submission is consistent with all applicable laws and regulations, as well as institutional policies;
- The appropriate research uses of the data and the uses that are specifically excluded or limited by the informed consent documents are delineated;
- The identities of research participants will not be disclosed/shared; and
- The sharing of data and subsequent secondary research uses are consistent with the informed consent of study participants from whom the data were obtained;
- The investigator's plan for de-identifying data is consistent with the standards outlined above;
- It has considered the risks to individuals, their families, and groups or populations associated with data sharing; and
- The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

Once received by the IRB, the submission will be processed in a manner similar to expedited modifications. An IRB staff member will review the request and, if necessary, contact the study team to obtain additional information. The final decision will then be made by the IRB's signatory official for Institutional Certification.

If the IRB approves the request, the certification letter will be sent to the study team via email. If the IRB determines that certification cannot be given, the study team will receive, via email, written notice of the decision and the rationale for why submission was not approved.

IRB Considerations of GDS Requests

After receiving the Institutional Certification request, the IRB will conduct a review of the study records to confirm that the study has been conducted in accordance with federal and state regulations and institutional policies.

The IRB will review the request summary, study protocol (if available) and protocol summary (or online application) to determine that genomic data sharing is consistent with applicable laws and regulations and that an appropriate plan to protect the identities of participants and to de-identify the data is in place.

In addition, the study will be reviewed to ensure that genomic data sharing will not pose inappropriate risk to participants, their families, and groups or populations associated with the data. Consideration will be given to whether a Certificate of Confidentiality has been obtained, and if not, if one is necessary prior to genomic data sharing. For additional information on Certificates of Confidentiality, please visit the NIH website (https://grants.nih.gov/policy/humansubjects/coc.htm).

There are numerous issues that determine the appropriateness of informed consent for submission of data. These issues are discussed below. Please note that the GDS policy applies to data collected both prospectively and retrospectively and different considerations will be applied depending on the type of study.

IRB Considerations for studies *initiated and conducted prior to* implementation of the NIH GWAS registry and GDS policies (January 25, 2015):

If all data and/or specimens were collected prior to January 25, 2015, please ensure you that you select the appropriate GDS certification letter template from the NIH website. In addition, please include a statement in your cover letter confirming that all data and/or specimens were collected prior to the January 25, 2015 effective date.

The IRB must determine whether the existing genetic materials and data were obtained in manner that is consistent with the requirements for genomic data sharing. Therefore, the manner in which participants were consented for research activities must have appropriately allowed for genetic research and analysis, future use of samples and data for genetic analysis, and the broad sharing of samples and data with investigators outside of the University of Pennsylvania.

The IRB understands that prior to the implementation of the 2018 Common Rule, there will be a great deal of variety in the consent forms used by researchers at Penn to obtain consent for collection and/use of specimens for genetic research. Below are some of the questions the IRB will consider when reviewing consent forms to determine if genomic data sharing is appropriate:

- Does the consent form either allow or preclude:
  - genetic research or analysis;
  - future use and broad sharing of the participant's coded phenotype and genotype data for research;
  - submission of the participant's coded phenotype and genotype data to a government health research database for broad sharing to qualified investigators?
- Does the consent form have any restrictions, such as:
  - types of subsequent research using the participant's phenotype and genotype data;
  - location of such research;

- types of medical conditions or diseases studied;
- duration of storage and use of phenotype and genotype data;
- limitations on who can use the participant's phenotype and genotype data (e.g. some consents may state that only non-commercial researchers can use the data)?

If the IRB determines that genomic data sharing is appropriate, the IRB may delineate in the Institutional Certification any uses of the data that are specifically excluded by the informed consent documents.

If the IRB determines that the consent form is not adequate for certification, the investigator has the option, if feasible, to obtain explicit consent from the research participants for the sharing of their data. If it is infeasible or inappropriate to obtain explicit consent, the IRB will not issue certification, and the decision will be conveyed in the manner previously described.

Please note that per the NIH policy, the IRB cannot waive the criteria for informed consent for genomic data sharing.

IRB Considerations for studies *initiated and conducted after* the implementation of the NIH GWAS registry and GDS policies:
All studies that anticipate sharing genomic data and are ongoing or have not begun are expected to have appropriate de-identification plans and consent forms. Standalone protocols and/or HSERA applications should detail the expected disclosures as outlined below. Likewise, the consent forms should cover the elements outlined below.

For studies that are currently ongoing that do not comply with this policy, modifications should be submitted to bring the study into agreement with the GDS policy. If appropriate, the IRB may require that explicit consent be sought from all participants previously consented.

New studies that have not been submitted for IRB review should detail the anticipated genomic data sharing in their standalone protocols and/or HSERA applications, as outlined below. In addition, the consent form should cover the elements outlined below.

What Should be Covered in IRB Submissions

Protocol / Application

Include the following information with your application, either in the HSERA application or referenced standalone protocol (either as part of an initial submission or modification):

Data Disclosures section

- Indicate that data will be sent to NIH for genomic data sharing
- Indicate whether purpose will be for broad use (i.e., unrestricted or open access) or limited to specific disease or conditions (i.e., controlled access);

- Specific sources of the data to be submitted (e.g., all participants in the study, a specific subset of individuals, participants from all sites, if a modification: specify whether data will be from previously enrolled participants, etc.)
- Type of data that will be shared [i.e., type of genomic data (genotype and phenotype), relevant associated data, and information necessary to interpret the data];

Confidentiality Plan section

- **Plans for maintaining data confidentiality:** Specify that a random, unique code will be assigned to the data sent to NIH to protect participant privacy and confidentiality, and that identifiers will not be sent to the NIH.
  Note: The identities of research participants cannot be disclosed to NIH data repositories; only coded data with all 18 HIPAA identifiers removed will be accepted.

Risks section

- **Possible consequences to participants resulting from a breach of confidentiality:** Describe risks of broad sharing, risks to individuals, families and groups and how those risks will be minimized (i.e., plan for removing identifiers from the data to be provided)

Consent Form Language

Ensure the following information is covered in your consent form.

- Allows for genetic research or analysis
- Allows for future use and broad sharing of the participant's coded phenotype and genotype data for research
- Allows for submission of the participant's coded phenotype and genotype data to a government health research database for broad sharing to
- Discusses risks of broad sharing of phenotype and genotype data
- Discusses privacy risks of data sharing (e.g., the possibility that the coded data may be released to members of the public, insurers, employers, and law enforcement agencies)
- Discusses the risks of computer security breaches relevant to maintaining data in an electronic format
- Discusses relevant risks to relatives or identifiable populations or groups
- Describes how individual privacy and data confidentiality will be protected
- Indicates that identifiers will not be provided to government database
- Discusses that potential benefits may accrue broadly to the public through the advancement of science and understanding of health and disease, rather than resulting in direct benefits to individuals
- Indicates either that research results will not be returned, or only returned in rare instances, and describe the conditions under which this could occur
- Indicates that a participant can withdraw his/her data from future research use. Instruct participant that if they decide to withdraw permission, to notify the Penn investigators in writing. Inform participant that in this case, their data will not be

used for future research but that data that had already been distributed to researchers cannot be retracted
- Allows or precludes commercial use of participant's phenotypic and genotypic data
- Indicates whether the database is unrestricted or controlled access (e.g., researchers must submit requests to access it)
- Discusses any a limit of use (e.g., data may only be used for research on a specific disease)

Suggested Template Language for Genomic Data Sharing

Below is suggested language that may be used to cover all of the above elements, along with broad consent elements outlined by the 2018 Common Rule requirements. Study teams do not have to use this explicit language, but it is provided for investigator convenience and to ensure conciseness in consent forms. Alternative language may be appropriate.

The information collected about you on this study may be shared broadly in a coded format for future research purposes. This may include genetic information (also known as genotype data) and the medical record data (also known as phenotype data).

There are no plans to tell you about any of the specific research that will be done. Possible future research may include future genetic research or analysis. This future research can be done without seeking your consent in the future, as permitted by law. It is possible that you may have chosen not to participate in these future research studies, had you been approached for participation. [Include if applicable: state any limits of use (e.g., data may only be used for research on a specific disease)]

Although your identifiers (name, address, phone number, or any other identifiable information) will not be shared outside of Penn, the following identifiers will be retained by the study team: [detail the identifiers that will accompany data/specimens during storage and sharing, e.g., the linking ID code, date of birth, etc.]. Your information may be stored and used for future research purposes for an indefinite amount of time [or state alternate timeline].

We may share your coded information with other researchers within Penn, or other research institutions, as well as pharmaceutical, device, or biotechnology companies. We may share your coded information broadly in government health research databases such as NIH supported databases [include all or select as applicable]:
- Database of Genotypes and Phenotypes (dbGaP)
- Gene Expression Omnibus (GEO)
- Sequence Read Archive (SRA)
- Cancer Genomics Hub (CaHUB).

[Include if applicable:] Researchers must apply to use the data shared. Special review committees will look at these applications to decide whether or not to share the data.

Researchers must agree to keep data safe and use the data only for the purpose approved by the NIH.

You will [will/will not] be given the results from testing that may be performed on your identifiable specimens as a part of future research. [if results will be returned, please specify the conditions]

There is a risk of breach of confidentiality (unintentional hacking or release of your information). We will do our best to make sure that this doesn't happen. However, we cannot guarantee total privacy. Study data will be physically and electronically secured. We will protect your confidentiality during storage and sharing by [detail confidentiality measures, e.g., encryption, access controls, etc.].

There are potential risks with broadly sharing genetic information. Even without your name or other identifiers, your genetic information is unique to you. The researchers believe the chance that someone will identify you is very small, but the risk may change in the future as people come up with new ways of tracing information. Very rarely health or genetic information could be misused by employers, insurance companies, and others. We believe the chance these things will happen is very small, but we cannot make guarantees. A federal law (Genetic Information Non-Discrimination Act, GINA) helps reduce the risk from health insurance or employment discrimination. The law does not include other types of misuse by life insurance or long term care insurance. Taking part in a genetic study may also have a negative impact or unintended consequences on family or other relationships. It is also possible through these kinds of studies that genetic traits might come to be associated with your group. In some cases, this could reinforce harmful stereotypes.

You will likely not directly benefit from future research with your information. Genetic information that results from this study does not currently have medical or treatment importance. Research with your information may help others by improving our understanding of health and disease, improving healthcare and making safer or more effective medical therapies, and developing new scientific knowledge. If the data or any new products, tests or discoveries that result from this research have potential commercial value, you will not share in any financial benefits.

If you have questions about the storage of your information, or have changed your mind, you can contact [Name or office] at [Phone Number]. If you change your mind, any remaining data will be destroyed. However, in some cases this may not be possible. For example, it may not be possible for de-identified data shared with other researchers to be destroyed because we would not be able to identify it as your data.