Human Subjects Protection Program
Investigator Guidance
July 29, 2024

GUIDANCE FOR FUTURE USE AND REPOSITORIES

Overview

The purpose of this document is to provide investigators with an overview of the key considerations for future research use of data and/or biospecimens and offer best practices for ensuring compliance with applicable regulations and protecting participant privacy and confidentiality.

Definitions

- **Identifiable**: Data and/or specimens are directly labeled with a subject's identifying information (e.g., name, social security number, medical record number, etc.) so that they can be readily connected to a specific subject. **Example**: A blood specimen is labeled as Rose Smith, MRN #007. This method of labeling does the least to protect subjects from a breach of confidentiality. As such it is rarely approved by the IRB, and only if the research aims cannot be achieved by other means AND the investigator has put protections in place to secure the data.

- **Coded**: Data and/or specimens are labeled with a unique number or code (e.g., Study ID). A separate link (key) is kept which connects this ID number to a patient identifier (e.g., name, medical record numbers, etc.). Data/specimens are usually coded if an investigator anticipates that they may need to gather additional data or verify subject data at more than one point over the life of the study. As long as a link exists, data are considered indirectly identifiable and not

anonymous, anonymized or de-identified. **Example**: The data collection form or specimen is labeled as Study ID #001, and a separate link is kept which correlates this number to the subject's identity (e.g., Study ID #001 Rose Smith).

Anonymized or De-identified: A record and/or specimen from which identifying information (e.g., Study ID, name, medical record number) is removed. For a data set or specimen to be considered de-identified, a key code must not exist and/or the data/specimen must be stripped of any indirect identifiers (study ID) or direct identifiers. The remaining data cannot include any information that could have the potential for deductive disclosure.

- **Anonymous**: Identifiers were not collected at any point in the research and cannot be retrieved by the investigator. **Example**: There are no identifiers (direct or indirectly linked via a code) on the data collection form or associated with the biological specimens.

- **Future Use of Data/Specimens:** "Future use" refers to the potential use of research data/specimens for purposes other than the original study for which they were collected. The use of data/specimens for future research purposes should be specified in a consent form of the original study and may require additional approval or oversight.

- **Repository**: The purpose of a research repository is to provide a means for researchers to share and access resources, to promote reproducibility and transparency in research, and to facilitate future research studies.
  Repository activities involve:
  1. collecting data/specimens,
  2. storing and managing the data/specimens, and
  3. distributing the stored data/specimens to recipient investigators.

- **Biobank**: A collection of human specimens (e.g., blood, urine, tissue, DNA samples) and associated information (e.g., EMR data, study data) stored and organized in a systematic way for research purposes.

- **Database**: A collection of data stored and arranged for ease of search and retrieval. Databases can be used to store a variety of research-data. **Examples:** Subject information, research instruments, and study results.

- **Registry**: A registry is a database maintained for future use that collects and stores data on individuals with a particular condition, disease, or exposure. Registries are often used to track the natural history of a disease, monitor the safety and effectiveness of treatments, identify potential study participants, and inform healthcare policy.

- **Repository Manager/Gatekeeper(s)**: An individual or group of individuals responsible for the security and disposition of the data/specimens stored in the repository.

- **Honest Broker**: An honest broker is a person that acts as a trusted intermediary in the management and handling of research data/specimens. They are responsible for ensuring that identifying information of research participants is kept confidential and that access to the data is only provided to authorized individuals. Honest brokers typically have no direct involvement in the research project, and their primary responsibility is to protect the privacy and confidentiality of research participants.

"Umbrella" Protocol Policy

Umbrella protocols, also known as general or master protocols, are not permitted. An umbrella

protocol typically refers to a single protocol that encompasses multiple sub-studies or research aims under one overarching framework.

To ensure thorough and focused reviews, each study must be submitted as a separate and distinct protocol. This policy guarantees that each research study undergoes an independent review process, addressing specific aims, methodologies, and ethical considerations unique to that study.

Researchers planning to utilize data or specimens for future use must prepare and submit individual applications and protocols for each intended study. This approach ensures compliance with ethical standards and regulatory requirements and maintains the integrity of the review process.

Planning for Future Research Use of Data/Specimens

In some cases, research data/specimens may be used for purposes other than the original study. The requirements for IRB oversight for the use of stored research data/specimens depends on whether the data/specimens can be linked to individually identifiable information (i.e., coded, or directly identifiable) and the terms of the informed consent under which the data/specimens were originally collected.

Below are questions to help you determine whether your protocol, consent form(s), and iStar application need to reflect future research use.

- Are you planning to use coded or identifiable data/specimens for future research purposes other than the original study?

  If yes: The protocol, consent form(s), and iStar application (as applicable) should reflect future use.
  - See section **"Creating a Repository at CHLA**" for further guidance if you intend to create a repository under the original study for which the data/specimens were collected.
  - See section **"Submitting to an Existing Repository"** for further guidance if you intend to submit data/specimens to an existing repository.

- Are you planning to use anonymous or de-identified/anonymized data/specimens for future research purposes other than the original study?

  If yes: The protocol, consent form(s), and iStar application (as applicable) should reflect future use. See section "**Future Use of Data/Specimens"** for further guidance.

- Are you unsure about whether you will keep data/specimens for future purposes other than the original study (i.e., future research studies)?

  If yes: It is recommended that even if there is no clear intention for future research use of the data/specimens, researchers should reflect this possibility in their protocol, consent forms, and iStar application. This is to ensure that participants are informed of the potential future uses of their data/specimens and that the necessary ethical and regulatory approvals are obtained prior to any future use.

Notes:
Failure to disclose potential future use to subjects can limit your options for the future use of data/specimens.

When coded or identifiable data and/or specimens will be released from CHLA, the investigator should contact CHLA Contracts & Clinical Research (CCR) at CHLAclinicalresearch@chla.usc.edu to discuss the CHLA agreement requirements (e.g. Data Use Agreements, Materials Transfer Agreements, etc.) for the study. The CHLA agreement(s) **do not** require submission to the CHLA IRB but must be collected and retained in the investigator's study records.

Creating a Repository at CHLA

Creating a research repository involves careful planning and adherence to best practices to ensure the effective organization, management, and accessibility of data/specimens. A protocol that involves the creation of a repository at CHLA should reflect the following elements:

1. **Objectives:** The objectives of the research repository should be clearly defined in the protocol. This should include the types of data/specimens that will be collected, the purpose of the repository, and any research questions or hypotheses that will be addressed using the data/specimens.
2. **Data/specimen collection:** The protocol should describe the procedures for collecting, processing, and storing the data/specimens. This should include details about the types of data/specimens that will be collected, the sources of the data/specimens, and the methods for processing and storing them.
3. **Informed consent:** Describe the informed consent process for participants whose data/specimens will be included in the repository. This should include details about the risks and benefits of participation, the procedures for obtaining consent, and any options for withdrawal of consent for use and retention of data/specimens.
4. **Privacy considerations:** The protocol should describe details about how participant confidentiality and privacy will be protected, how data/specimens will be de-identified or anonymized, and any measures that will be taken to ensure the security of the data/specimens.
5. **Data/specimen sharing:** The protocol should describe the procedures for sharing data/specimens from the repository with other researchers. This should include details about any restrictions on data/specimen sharing, the criteria for determining who will have access to the data/specimens, and any procedures for reviewing use requests. For example, IRB approval is required to release data/specimens. This should also include documentation and tracking details for where repository data/specimens have been distributed.
6. **Governance:** The protocol should describe the repository manager/gatekeeper and honest broker (if applicable).

In addition to what is described above, researchers should also complete the "Creation of a Research Repository Worksheet (For Repositories Located at CHLA)" located on the HSPP website and upload it to your iStar application.

Submitting to an Existing Repository

If your research study involves submitting data/specimens to an existing repository, then information must be provided outlining the nature of this process.

Investigators should do the following:

1. Check "Submission of Data or Tissue to an Existing Repository" in item 9.2 of the iStar

application.
2. Complete the “Submission of Data or Tissue to an Existing Repository” section of the iStar application located on the HSPP website.
3. If the existing repository is located at CHLA, provide the CHLA IRB# in the repository section of the protocol document.

Consent Requirements

When your research involves the collection of identifiable or coded data/specimens, then the consent form(s) must include information about whether data/specimens will be used in future research. The consent form(s) should include the following information (as applicable):

1. Describe whether it is possible that data/specimens may be used in future research use.
2. Provide details about the specific types of future research that may be done with the data/specimens if known (e.g., studies related to lupus only, etc.). If it is unknown how the data/specimens may be used in future research, then this should also be clearly stated.
3. Describe how the data/specimens will be labeled (e.g., coded, etc.) and who will have access to the key to the code (if applicable). Please see the definitions above for a detailed description of labeling methods.
4. Describe who may use the data/specimens in future research studies (e.g., the researcher conducting the study, the study sponsor, other researchers at CHLA or elsewhere, etc.).
5. Describe how the data/specimens will be shared with others for future research use (e.g., coded, de-identified, etc.).
6. Indicate whether subjects may be re-contacted in the future to obtain their permission (consent) for the future research use of their data/specimens.
7. State whether it is possible that future research may involve whole genome sequencing.
8. If it is possible that data/specimens may be used in future research that could result in new discoveries, state whether the subjects will share in any profits.
9. State whether subjects can withdraw their data/specimens and, if so, describe how this may be done.

See the “CHLA Consent Form Standards and Sample Language” document on the HSPP website for sample language.

Use of Data and/or Biospecimens Obtained from a Repository

CHLA IRB approval is required for CHLA investigators who wish to conduct research with data/specimens from a database or repository. The CHLA IRB requires submission of a separate study application for each research project that will involve the use of data and/or biospecimens from a repository, even if the information/material has been de-identified.

Below are guidelines to follow if you intend to ***use*** data/specimens collected for other purposes in your research:

Clearly outline the plan for using existing data/specimens in the research protocol, consent forms, and other relevant study documents. The protocol and iStar application must include the following information:

1. The source of data/specimens, including any applicable CHLA IRB#.
2. A description of how the data/specimens will be labeled (coded, de-identified/anonymized, etc.).
3. A comprehensive outline of the intended use of the data/specimens, including any genetic

testing.

NIH Genomic Data Sharing (GDS) Policy

The NIH GDS policy became effective on January 25, 2015, and applies to all NIH funded studies (e.g., grants, contracts, cooperative agreement, or intramural research) generating large-scale human or non-human genomic data, regardless of funding level, as well as the use of these data for subsequent research. Large-scale genomic data include genome-wide association studies (GWAS), single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data.

When the NIH GDS policy applies, the protocol and/or iStar application must include the following information:

1. Explicitly state that the data will be sent to the NIH for GDS and describe the genotype and phenotype data to be sent.
2. The source of the data/specimens, including any applicable CHLA IRB#.
3. State whether you are requesting CHLA to certify that the proposed study meets the requirements of the NIH's Final Genomic Data Sharing Policy or if the lead/coordinating site will certify that the proposed study meets the requirements of the NIH's Final Genomic Data Sharing Policy on CHLA's behalf.
4. Identify the NIH databases/repositories where the data will be stored (e.g., database of Genotypes and Phenotypes [dbGaP], Gene Expression Omnibus [GEO], Sequence Read Archive [SRA], Cancer Genomic Hub, etc.).
5. State whether the data will be submitted to an **unrestricted-access** (i.e., data made publicly available to anyone) or **controlled-access** (i.e., data made available for secondary research use only after investigators have obtained approval from NIH to use the requested data for a particular project) NIH database/repository. If submitting to a controlled-access NIH database/repository, also specify one of the data use limitations below for appropriate secondary use.
    a. General Research Use: Use of the data is limited only by the terms of the Data Use Certification. This data will be added to the dbGaP collection.
    b. Health/Medical/Biomedical: Use of the data is limited to health/medical/biomedical purposes and does not include the study of population origins/ancestry.
    c. Disease-specific: Use of the data must be related to the specific disease. Indicate the specific disease.
    d. Other: Describe the other use.
6. The plan for maintaining confidentiality (e.g., data assigned a Study ID, identifiers will not be sent to the NIH, key to the code will be retained at CHLA, etc.).
7. A description of risks to broad sharing, risks to individuals, families, or groups.

The NIH's standards for IRB review and informed consent for data acceptance in GDS databases/repositories are stringent. It's important to note that if any of the required GDS Policy elements are absent from the consent form(s), the NIH may either reject the data or restrict its future use.

If you request CHLA to certify that your proposed study meets the requirements of the NIH's Final Genomic Data Sharing Policy, the consent form(s) will undergo review to ensure the inclusion of the required GDS Policy elements. It's important to understand that if the consent form(s) are found to be inadequate, CHLA may be unable to certify the study and it may be determined that re-consent from subjects is required.