# GUIDANCE ON CREATING RESEARCH DATA AND TISSUE REPOSITORIES

This Guidance has been developed to assist investigators in understanding what approvals must be obtained from the Institutional Review Board ("IRB") to comply with both the Health and Human Services ("HHS") Regulations and the Health Insurance Portability and Accountability Act ("HIPAA") Privacy Rule as they relate to the creation and use of information in and associated with data and tissue repositories for research purposes. This Guidance derives from applicable regulations as well as the guidelines published by the Office of Human Research Protections ("OHRP") in November 1997 and the HHS Guidance on Research Repositories, Databases and the HIPAA Privacy Rule published on January 12, 2004.

## I. What is a Data/Tissue Repository?

Data/Tissue Repositories and Research Data/Tissue Repositories. A Data/Tissue Repository is a collection of health information either standing alone (Data Repository) or in association with collected biological materials such as tissue, blood, or other specimens (Tissue Repository); the health information is generally derived from patient medical records, research records, other hospital records, or through interactions with patients. Data/Tissue Repositories may be used for various purposes, including research, quality improvement and clinical care. There may be circumstances where one Data/Tissue Repository is of a hybrid nature and serves several different purposes at the same time (for example, a database of clinical information about a certain patient population may occasionally be accessed for research purposes under an IRB-approved protocol). This Guidance is intended to describe the requirements around creating or using Data/Tissue Repositories where the *primary purpose* is to compile data and/or specimens with associated data that may be mined or shared for research purposes (each a "Research Repository"). If the primary purpose of a Data/Tissue Repository is clinical care, quality improvement, payment for treatment services, public health, or other non-research activities, IRB approval and oversight is generally not required and this Guidance does not apply. Furthermore, this Guidance is not intended to govern data compiled in a computer database, with or without associated specimens, for recordkeeping or analysis for a single clinical trial or medical records research project; those databases and tissue repositories should be described in the specific protocol for the research study in which they will be utilized as a tool to facilitate data analysis. Tissues that are maintained without *any* associated information (for example, excess clinical specimens or residual specimens from research that do not have any connection to the health information of an individual patient or subject) are also outside the scope of this Guidance, which focuses on protecting the information that is sometimes associated with biological specimens.

Type of Information in Repository. Data/Tissue Repositories may be designed to contain data, with or without associated specimens, in a variety of formats, including data that are:

- individually identifiable (the identity of the data source is or may readily be ascertained by the investigator or associated with the information),
- coded (the data are not individually identifiable, but they contain a code through which they may be linked back to the data source by anyone who possesses the key to that code), or
- anonymized (the data are not individually identifiable, and there is no way to link them back to the data source).

Because OHRP and HIPAA regulations do not completely align with respect to what constitutes identifiable information, it is important to remember that information that is not individually identifiable per OHRP regulations (as defined above) may still be "Protected Health Information"

(PHI) for purposes of HIPAA such that the protections of HIPAA apply. PHI is any data that includes any one of 18 identifiers listed on **Appendix A**.

- For data to be completely de-identified for HIPAA purposes (and thus outside of HIPAA's requirements), all 18 HIPAA identifiers must be stripped.
- Alternatively, researchers may remove only certain direct identifiers, creating what is known as a "limited data set" for HIPAA purposes. See **Appendix B**.

## II. Approvals Needed for Research Repositories

The creation and maintenance of a Research Repository are considered "research" activities, to which both HHS and the HIPAA privacy regulations apply, even before any specific research projects are carried out using the stored data or specimens with associated data. The requirements related to Research Repositories are triggered at two different points in time: first, when the Research Repository is created (information is collected and compiled) and/or maintained (information is updated, managed or organized to facilitate research); and second, when the information stored in the Research Repository will be used or shared for a specific research purpose.

### 1. Creating and Maintaining a Research Repository

### A. Is IRB Approval Required?

IRB review and approval is required prior to the creation of any Research Repository, unless the IRB determines that it does not qualify as "human subjects research" (e.g. no interaction with individuals and the investigators are only receiving anonymized data), or that it does qualify as human subjects research but meets the criteria for exemption (see below). Investigators should submit an application to the IRB, which describes, at a minimum:

- the purpose of the repository;
- the anticipated sources of information and any associated specimens that will be included in the repository;
- how the repository will be maintained; for example, if it is intended to be a living repository how information may continue to be updated and from what sources;
- the process by which such information will be collected;
- the confidentiality protections that will attach to data maintained in the repository;
- whether highly sensitive information will be maintained in the database and, if so, whether a Certificate of Confidentiality will be pursued (see below for a discussion of when a Certificate of Confidentiality may be required by the IRB);
- who is in charge of maintaining the repository and fielding requests to use the data and any associated specimens contained therein;
- what are the criteria that will be used to evaluate requests for data and any associated specimens (for example, will requests be entertained from investigators both inside and outside of the institution, will requests be entertained from commercial as well as non-profit entities, will evidence of IRB approval of the specific research use be required, etc.);
- whether investigators will be granted direct access to the repository to mine certain data and/or utilize associated specimens for research purposes and, if so, how it will be ensured that only the minimum necessary information is available to them during that process, or

whether specific data sets, with or without associated specimens, will be made available to interested researchers; and
- whether data, with or without associated specimens, will be maintained and/or made available to investigators in identifiable format or only as a limited data set or fully de-identified data set.

The way in which data and any associated specimens are collected (whether through interaction with the individual, or by reviewing existing records and using excess biological materials) and the degree to which the data, as collected and once stored, may be used to identify the person from whom it came will impact the requirements that attach to both the creation and use of the Research Repository.

- If researchers will have no interaction with human individuals and will only be collecting data and/or specimens that are already anonymized, the IRB may determine that the proposed activities do not meet the definition of "human subjects research" and no IRB approval or informed consent is necessary.
- The collection of only coded data and/or specimens may also not meet the definition of "human subjects research," so long as the investigator receiving the coded data/specimens set will never be provided with the key to the code or any identifying information and the data (alone or associated with specimens) were not collected for the purposes of his or her research.
- If researchers are not interacting with individuals but are reviewing existing identifiable records or specimens from which they will collect and store data that are *not* individually identifiable, the activity may meet the definition of "human subjects research," but may qualify for an exemption from IRB oversight and informed consent. Investigators should note that IRB review is still required to determine that an exemption applies.
- If researchers are interacting with individuals to collect data/specimens and/or are storing the data and/or specimens in an identifiable format, IRB approval and either informed consent or IRB waiver of the consent requirement will be necessary.

Some of the many possible permutations that arise in repository research and their impact on the requirements for IRB approval, informed consent, and HIPAA authorization are discussed in more detail below and summarized further in **Appendix C** to this Guidance for easy reference.

### B. Is Consent/Authorization Required by the Individuals Whose Information Will be Collected and Stored in the Research Repository?

The collection and storage of *individually identifiable* data in a Research Repository requires informed consent from the individuals whose information will be included, unless the IRB grants a waiver of informed consent on the grounds that (1) the research activity involves no more than minimal risk to subjects, (2) the waiver will not adversely affect the rights and welfare of the subjects, (3) the research could not practicably be carried out without the waiver, and (4) whenever appropriate, the subjects will be provided with additional pertinent information after participation. To the extent investigators will be collecting information directly from individuals for inclusion in the Research Repository (for example, through a questionnaire or at the time of a standard clinical intervention or other research activity), a waiver will not likely be appropriate. However, if the Research Repository will be built using information, with or without associated specimens, previously collected and retained in existing medical or research records or labs, it is possible a waiver of informed consent might be granted.

To the extent the Research Repository will include information that constitutes PHI (e.g. the data contain at least one of HIPAA's 18 identifiers), authorization from the individuals whose information will be included is required by HIPAA, unless the IRB grants a waiver of authorization on the grounds that (1) the research activity involves no more than minimal risk to the privacy of the individuals, (2) the research could not practicably be conducted without the waiver, and (3) the research could not practicably be conducted without access to and use of the PHI. Again, whether the PHI will be collected directly from individuals through an interaction or intervention, or collected from existing sources will be relevant in the IRB's determination of whether a waiver is appropriate for a given Research Repository.

It is possible that the IRB may determine that data, with or without associated specimens, originally created or collected for another research activity (for example, a clinical trial) was provided with research informed consent and authorization that included the storage of the data/specimens for potential secondary research uses and, as such, additional informed consent is not required in order to include the data in the Research Repository. It is within the IRB's discretion to determine whether the scope of an original consent and authorization covers the Research Repository activity, or whether new consent and authorization or waiver of both is required.

## 2. Using Data Stored in a Research Repository

### A. Is IRB Approval Required?

At the time a researcher has a specific research project for which he or she would like to access, or receive an individually identifiable data set or biological specimen from, an existing Research Repository, the researcher must apply to the IRB for approval of the specific project, notwithstanding the fact that the Research Repository is operated under an IRB-approved protocol. To the extent the investigators applying to create a Research Repository also have specific research projects for which they would like to use the Repository, it may be possible for those purposes to be combined in the same application; such circumstances should be discussed with the IRB on a case-by-case basis.

Note that any time a researcher wishes to use existing individually identifiable data or specimens for a specific research purpose, even if those data are maintained in a clinical or quality improvement Data Repository that did not require IRB approval to create and maintain, or the identifiable specimens are excess materials maintained in a clinical lab, the specific research project must be submitted to the IRB for review and approval, unless the IRB determines that it meets the criteria for exemption.

To the extent that the only data or specimens that will be shared with a researcher are not individually identifiable (the identity of the individuals cannot be readily ascertained from the data set or specimen that is shared), IRB review and approval may not be required, either because the IRB determines that the proposed research does not meet the definition of human subjects research or because it qualifies as exempt from review. Note, however, HIPAA authorization (or IRB waiver of the authorization) may nonetheless be required if the data contain, or the specimens are associated with, one or more of the 18 HIPAA identifiers.

### B. Is Consent/Authorization Required by the Individuals Whose Information Will be Used or Shared for Specific Research Purposes?

The use of individually identifiable data or specimens housed in a Research Repository for a specific research purpose is human subjects research for which consent and HIPAA authorization are required. If the data or specimens proposed to be used were originally collected as part of an IRB-approved research protocol with the informed consent and authorization of subjects (including a Research Repository protocol where individuals gave consent and authorization to have their specimens and/or information included), it is possible that additional consent to the specific research use will not be necessary so long as the proposed use is determined by the IRB to be consistent with the uses outlined in the original IRB-approved consent form. If the IRB concludes that the proposed use is not consistent with the uses outlined in the original IRB-approved consent form, or if the data/specimens to be used were originally collected through clinical or other interventions for which a research informed consent was not obtained, the principal investigator must either obtain consent and authorization from subjects whose specimens and/or information will be used or, more likely, apply for a waiver of informed consent and authorization, assuming that the applicable criteria (outlined above) are met. It is within the IRB's discretion to determine whether an individual who affirmatively agreed to participate in the Research Repository gave consent that encompassed a specific proposed research use, or whether additional steps are required before the specimens and/or data may be used for that specific purpose. Note that even where the IRB determines that a prior informed consent covers the specific proposed research use, HIPAA requires a new authorization at each step of the process. Therefore, investigators are nonetheless required to obtain authorization to the specific research use or, more likely, seek IRB waiver of the authorization requirement.

To the extent researchers seek to use only data, with or without associated specimens, that are in the form of a Limited Data Set (data with certain direct identifiers stripped in accordance with HIPAA), informed consent and authorization are not required; however, the recipients of the data are required to enter into a Data Use Agreement through which they agree to protect the confidentiality of the data and limit the uses to those agreed upon by the parties. Requests for Data Use Agreements should be directed to the IFAR Manager of Grants and Contracts.

To the extent researchers seek to use only data, with or without associated specimens, that are completely de-identified under HIPAA (all 18 HIPAA identifiers have been removed), no consent or authorization is required for the specific research purpose. Note that this type of specific research project may not require IRB review and approval, either because it is not human subjects research or on the grounds it is exempt. Again, the IRB must determine whether a specific project is not human subjects research or it otherwise exempt from IRB review; this is not a determination that investigators should make alone.

## III. Certificates of Confidentiality

Research Repositories that include sensitive information may need to obtain a Certificate of Confidentiality from the National Institutes of Health. A Certificate of Confidentiality protects researchers from being required to disclose or produce identifying characteristics associated with the sensitive data, for example in response to a subpoena or in a court proceeding. A Certificate of Confidentiality may be required by the IRB if a **substantial portion of your research proposal involves data, with or without associated specimens, about any of the following activities or topics:**

- HIV testing or other information about HIV status
- Illegal drug use, including the results of illegal drug testing (blood or urine)

- Sexual attitudes, preferences, or practices, STDs, the use alcohol, drugs, or other addictive products and on participants' psychological well being or mental health. (A COC is usually not required when one or two questions regarding alcohol use are directed at adults age 21 or over)
- Genetic studies (including those that collect and store biological samples for future use) or other genetic information that discloses an unknown disease or predisposition which may, if it became known, adversely affect the participant or family members employment or ability to obtain insurance
- The IRB may, in its discretion, request that the Principal Investigator obtain a Certificate of Confidentiality for research activities not on this list.

If you have questions about whether your Research Repository may require a Certificate of Confidentiality and what the appropriate process is for securing one, please contact the IRB. In general, a Certificate of Confidentiality issued for the creation of a research repository will not cover future research conducted with data from the repository. Investigators seeking access to sensitive data in a Research Repository will likely need to seek a new Certificate of Confidentiality to cover the specific research project.

## IV. What Forms to Complete?

Investigators must complete the HSL Human Subject Application (and the additional forms indicated on this form e.g. personnel roster, HIPAA waiver request, etc.) as well as the Research Repository Application – Form H.

## APPENDIX A

Data containing any one of the following 18 identifiers is considered PHI under HIPAA:

(1) Names;
(2) All geographic subdivisions smaller than a State, including:
- street address
- city
- county
- precinct
- zip codes and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly-available data from the Bureau of the Census: (i) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people, and (ii) the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000;

(3) Telephone numbers;
(4) Fax numbers;
(5) E-mail addresses;
(6) Social Security numbers;
(7) Medical record numbers;
(8) Health plan beneficiary numbers;
(9) Account numbers;
(10) All elements of dates (except year) for dates related to an individual, including:
- birth date
- admission date
- discharge date
- date of death
- all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;

(11) Certificate/license numbers;
(12) Vehicle identifiers and serial numbers, including license plate numbers;
(13) Device identifiers and serial numbers;
(14) Web Universal Resource Locators (URLs);
(15) Internet Protocol (IP) address numbers;
(16) Biometric identifiers, including finger and voice prints;
(17) Full face photographic images and any comparable images;
(18) Any other unique identifying numbers, characteristics, or codes.

## APPENDIX B

A limited data set may be created by removing from the individual's PHI the following direct identifiers of the individual or of relatives, employers or household members of the individual:

(1) Names;

(2) Postal address information, other than town or city, State, and zip code;

(3) Telephone numbers;

(4) Fax numbers;

(5) Electronic mail addresses;

(6) Social security numbers;

(7) Medical record numbers;

(8) Health plan beneficiary numbers;

(9) Account numbers;

(10) Certificate/license numbers;

(11) Vehicle identifiers and serial numbers, including license plate numbers;

(12) Device identifiers and serial numbers;

(13) Web Universal Resource Locators (URLs);

(14) Internet Protocol (IP) address numbers;

(15) Biometric identifiers, including finger and voice prints; and

(16) Full face photographic images and any comparable images.