1. PURPOSE

The terminology used to describe research types, biospecimens, coding, and data often have different meanings to different groups. The intent of this guidance document is to standardize the terminology used for NIAID DMID documents (policies, SOPs, protocol templates, manuals, etc.). Where applicable, terms were chosen to match those found in regulations and/or international standards such as ICH.

2. SCOPE

This document applies to biospecimens and data owned and/or stored by DMID including biospecimens and data collected from all DMID sponsored clinical trials.

3. GUIDANCE

3.1 Non-exempt Human Subject Research (HSR) can be divided into two types:

- **Primary Human Subject Research**
  Definition: Research using (all of the following):
  - identified or coded* data or biospecimens from a living individual
  - an intervention or interaction,
  - for the original purpose specified in a protocol.
    (see 3.5 for definition of identified or coded)*

  Primary HSR does not always require a direct interaction with an individual. Primary HSR includes modifying the environment or vectors (e.g., mosquitoes) that indirectly interact with the individual. Primary HSR also includes the collection of information by observation of reports or systems that hold the information (e.g., using medical records, laboratory reports, etc.).

- **Secondary Human Subject Research**
  Definition: The re-use of identified or coded data or biospecimens that were collected from some other ‘‘primary’’ or ‘‘initial’’ research activity. This would include use of data from existing clinical sources (e.g., medical records, clinical biospecimens, or laboratory results). There is no new collection of data or biospecimens from the individual.

  [Note – research using medical records, national death databases, disease and drug registries, claims data, and medical and administrative records from routine medical practice, etc., may be considered primary or secondary, and requires case by case assessment]

3.2 Written protocols

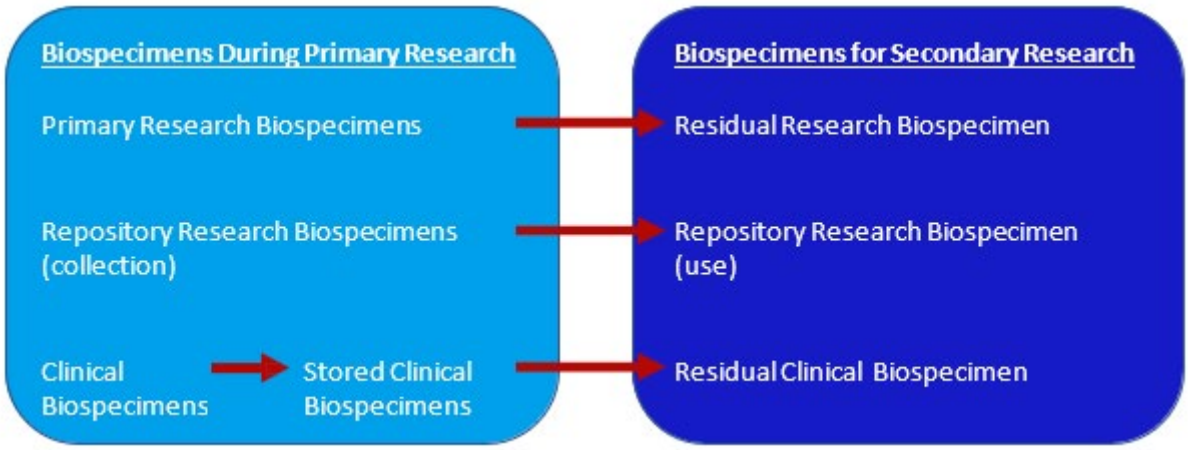
3.3 Following the above, non-exempt Human Subject Research Protocols can be divided into two types:

- **Primary Research Protocol**
  Definition: The document that describes the objective(s), design, methodology, statistical considerations, and organization for the primary human subject research.

- **Secondary Research Protocol**
  Definition: The document that describes the objective(s), design, methodology, statistical considerations, and organization for the secondary human subject research.

3.4 Sample Classification

The words "biospecimens" and "samples" are often used interchangeably. For the purpose of this document, the word "biospecimen" will be used, but this can be interchanged with "samples" as appropriate.



Biospecimens During Primary Research – Any specimens collected as part of primary research fall into one of these categories.

Primary Research Biospecimens

- Definition: Biospecimens collected for research purpose and would not be collected if a person was not participating in the research. The testing of these biospecimens is specified in the primary study protocol.[1]
- Biospecimens may be sent to a Clinical Laboratory or Research Laboratory.
  - Clinical Laboratory: Typically used for diagnostic tests, such as tests for screening, safety, and determination or confirmation of specific conditions (e.g., HLA-typing, blood-typing, infectious-organism presence or absence, assays for an infectious agent, etc.)

- Research Laboratory: A facility that tests specimens but generally does not report patient specific results for the assessment of the health of individuals, or for the diagnosis, prevention, or treatment of any disease.

Clinical Biospecimens

- Definition: Biospecimens collected for clinical purposes where the results are used for medical or clinical care (i.e., the biospecimen was not collected only for the research study).[2]
  - Clinical Laboratories generally are regulated (e.g., CLIA or similar foreign regulations) or otherwise approved by a health care facility to provide testing for routine health care.
  - Results from clinical biospecimens may be used in research protocols.

Stored Clinical Biospecimens

- Definition: Clinical Biospecimens that are retained (e.g. leftover serum, or unopened backup tubes, etc.) after the original testing per the clinical lab policy and procedures. This sample is fully controlled by the clinical lab and not the research team.

Repository Research Biospecimens *(previously referred to as "Future Use")*

- Definition: Biospecimens collected with the intent to store for additional research (i.e., biospecimens collected beyond those needed for primary research). These must be described in the primary research protocol with sufficient detail of how they will be used, and the research subject must give specific consent for their collection.[3]
  - The primary research protocol should specify the collection of these samples, the general types of research that will be performed with the sample, the conditions under which data and specimens will be released to recipient-investigators, and procedures for protecting the privacy of subjects and maintaining the confidentiality of data.
  - The use of these biospecimens is described in a secondary research protocol (or less frequently as an amendment to the primary research protocol).
  - No specific secondary research protocol is developed at the time of collection.

Biospecimens for Secondary Research – Definition: Biospecimens that remain after all testing specified in the primary protocol is complete.

Residual Research Biospecimens *(previously referred to as "Residual biospecimens")*

- Definition: Any leftover Primary Research Sample after laboratory testing is completed per protocol. The protocol should explicitly state these are allowed to be stored and used for secondary research, and the research subject must give specific consent for their retention. [4]

Repository Research Biospecimens

- See definition under "Biospecimens During Primary Research" above.

Residual Clinical Biospecimens

- Definition: Any Clinical Biospecimens (including Stored Clinical Biospecimens) that remain after clinical laboratory testing is completed. [4]

Old DMID definitions (for reference only)

1. Residual biospecimens [Similar to Residual Research Sample]

Definition: Residual biospecimens/specimens are those that are left over after a study has been completed; a study is considered completed when all data analyses under the protocol have been performed and the Clinical Study Report (CSR) has been approved by DMID and, as appropriate, submitted to the FDA. Residual biospecimens cannot be shipped for purposes other than per protocol analysis until the CSR has been approved by DMID and, as appropriate, submitted to the FDA. Citation: DMID OCRR policy

2. Future Use [Similar to Repository Research Sample]

Definition: Biospecimens/specimens collected specifically for future use are additional amounts and/or types of specimens acquired during the study solely for the purpose of future research. The collection of the future use biospecimens has been described in the protocol and informed consent. Unlike residual biospecimens, future use biospecimens may be shipped to recipients for testing at any time. Citation: DMID OCRR policy

3.5 Categories of Data

The words "information" and "data" are often used interchangeably. For the purpose of this document, the word "data" will be used, but this can be interchanged with "information" as appropriate.

Source Data

- Definition: All original clinical findings, observations, or other activities in a clinical trial. Source data are contained in source documents (original records or certified

copies of original records). [5] Source data should be complete enough for the reconstruction and evaluation of the trial.

Primary study data

- Definition: All of the data collected from one or more sources (i.e., source data) and extracted for purposes and analyses specified in the primary protocol.

Secondary study data

- Definition: Data extracted from the primary study data for purposes and analyses specified in a secondary research protocol. [4]

Analyses of primary study data

- Definition: the analyses of the primary study data as stated in a Primary Research Protocol or separate analysis plan associated with the Primary Research Protocol. This includes all primary, secondary, and exploratory endpoints listed in the original protocol.

Analyses of secondary study data

- Definition: the analyses of the secondary study data as stated in a secondary Research Protocol. These data were collected from some other ‘‘primary’’ or ‘‘initial’’ activities.

Metadata

- Definition: a set of data that describes and gives information about other data (e.g. when, how, why, etc. were the data collected).

3.6 Categories for Coding Biospecimens and Data

There are four general categories of coding: identified, coded, anonymized, and anonymous. Coded data or biospecimens can be single or double coded.

Identified Biospecimens and Data

- Identified data and biospecimens are labeled with personal identifiers such as name or identification numbers (e.g., social security or national insurance number). Identified data and biospecimens offer privacy protection comparable to that of general health care confidentiality in everyday medical practice. Clinical biospecimens and source data may be labeled with personal identifiers and require privacy protections at the site. For research purposes, stored biospecimens and data entered in a study database should not contain personal identifiers, while source records are generally identified. [6]

Coded Biospecimens and Data

- Coded data and biospecimens are labeled with at least one specific code and do not carry any personal identifiers. It is possible to trace the data or biospecimens back to a given individual with the use of a coding key. [7]
    - In single coding, the key is usually maintained by the research site.
    - Double-coded are given a second code after the primary coding (i.e., there two sets of codes). The second coding refers back to the first coding. For example, a sample number that refers back to a participant number. It is more difficult to trace the data or biospecimens back to a given individual.
        - DMID sponsored trials typically use biospecimens that are double coded.

Anonymized Biospecimens and Data

- Anonymized data and biospecimens are initially single or double coded but where the link between the subjects' identifiers and the unique code(s) is eliminated. This can occur either by:
    - the code being destroyed, or
    - the coded label is removed from the sample. [8]
- Once the link has been deleted, it is no longer possible to trace the data and biospecimens back to individual subjects through the coding key(s). Anonymization is intended to prevent subject re-identification. Biospecimens can be given broad categories of identifiers (i.e., an influenza specimen collected in a 30-40 year-old from Maryland) but based on categories of data used and prevalence of disease this data is not specific enough to identify an individual.

Anonymous Biospecimens and Data

- Anonymous data and biospecimens are (were) never labeled with personal identifiers when originally collected, and no coding key was generated. Therefore, there is no potential to trace back data and biospecimens to individual subjects.

Avoid the words "identifiable" and "deidentified"

- The word "identifiable" and "deidentified" have different meanings for different people. E.g., some people believe "deidentified" means coded, while others think it means anonymous. "Identifiable" can also be interpreted as "identified" (i.e., labeled with PII) or coded but traceable back to a subject with enough steps even if across institutions. It is suggested these terms be avoided to eliminate confusion.

4. REFERENCES

1. Derived from: CLIA https://www.ecfr.gov/cgi-bin/text-idx?SID=1248e3189da5e5f936e55315402bc38b&node=pt42.5.493&rgn=div5#se42.5.493_13
2. Derived from: CLIA. https://www.ecfr.gov/cgi-bin/text-idx?SID=1248e3189da5e5f936e55315402bc38b&node=pt42.5.493&rgn=div5#se42.5.493_13
3. Derived from: Derived from OHRP Guidance-Issues to Consider in the Research Use of Stored Data or Tissues(1997); and, NCI: Informed Consent https://biospecimens.cancer.gov/bestpractices/elp/ic.asp
4. Derived from: NIH OHSRP secondary research definition.
5. ICH E6 (GCP) R2, section 1.51
6. ICH E15, section 2.3.1
7. ICH E15, section 2.3.2
8. ICH E15, section 2.3.3

5. ADDITIONAL INFORMATION

5.1 Document Lead: DMID Associate Director of Clinical Research

5.2 Posting external: Yes