UNG UNIVERSITY of NORTH GEORGIA™

Institutional Review Board (IRB)

IRB Guidelines 1.5

Working with Existing Data

Version 1.1: Created 4/20/2016

Overview

The IRB realizes that some research involving existing data sets and archives may not meet the definition of human subjects' research and therefore would not require IRB review; however, some may meet definitions of research and may require IRB review and approval. To get a better sense of how the IRB deals with these nuances, please read the document carefully.

Basic Terms and Definitions

The policy as well as definitions are informed by federal policy documents (e.g. 45 CFR 46, OHRP guidance letters) and institutional best-practices at other US universities

Human subject: A term used to describe a living individual about whom an investigator conducting research obtains either (1) data through intervention or interaction and/or (2) private identifiable data.

Private Information: Human Subject information that is created and/or occurs in a social context where the individuals has a reasonable expectation of privacy (e.g. no observation and/or recording takes places without his/her consent) is called private information. This also means that individual information that has been provided or shared with the expectation that it will not be made public (e.g. medical information) falls into this category.

Data: Includes private information (e.g. individuals answers to interview questions, medical records etc.) and biological specimen (e.g. individuals' tissue samples, bodily fluids etc.).

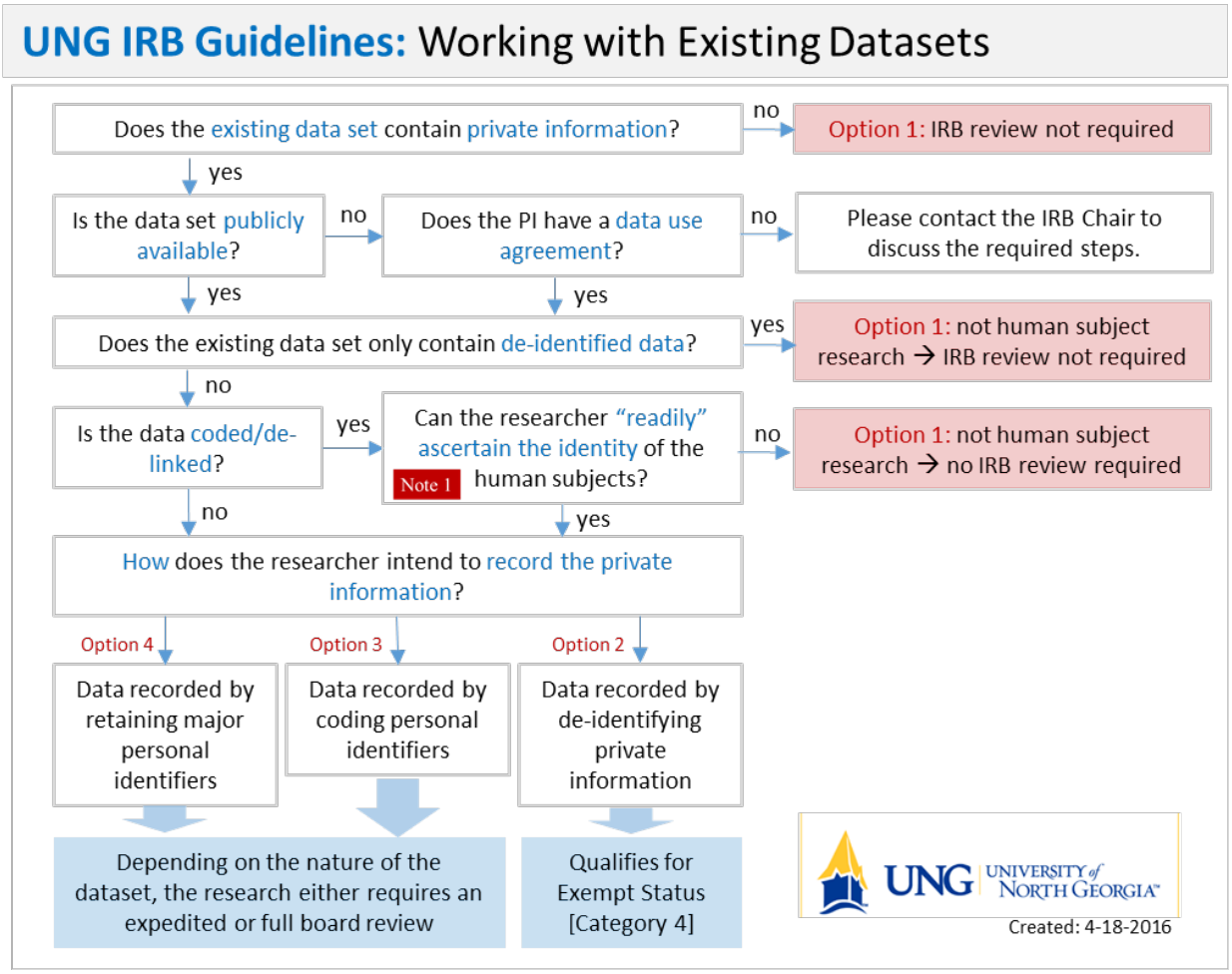
Identifiable Data: This type of data contains direct links between the data and the individuals who provided them (e.g. an opinion on an issue is linked to a real name of the person). There is no expectation of confidentiality as to how the researcher will use data in their work and/or publications. Example: an oral history of Gulf War Veterans.

Coded and De-Linked Data: *Coded data* and *de-linked data* are terms that for the purpose of this document can be used interchangeably. *Coded or de-linked data* simply means that private identifiable information (e.g. a participant's name, social security data) has been replaced by a code (e.g. a number, letter, and/or symbol) alongside with a master key to decipher that code. Master keys are stored separately (to guarantee confidentiality) and usually destroyed after the study's completion (to finish the de-linking process).

De-Identified Data: Private information is recorded in such a way that the source of the information (e.g. the person) cannot be identified either directly or indirectly. Hence, coded and de-linked data should NOT be treated as de-identified UNLESS all means of identification have been destroyed (e.g. the master key or other deciphers).

Personal Identifiers: Personal identifiers are data elements that allow someone to directly or indirectly identify a human subject in a dataset. Examples: direct identifiers such names and/or social security numbers; indirect identifiers such as having access to a combination of data

elements such as race, age bracket and sex that may allow someone to identify a person from a dataset even without having access to direct identifiers.



Note 1: The answer to this question should be "yes" if the investigator can readily ascertain the identity of the individual(s) to whom the identifiable private data pertains. The answer should only be "no" in cases where the investigator does NOT have access to the key under ANY circumstances, until the individual(s) is/are DECEASED. For example, (1) the investigator and the holder of the master key enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (IRB does not need to review this agreement), (2) there are IRB-approved written policies and operating procedures for a data repository or data management center that prohibit the release of the key to the investigator under any circumstances, until the individual is deceased, and (3) there are other legal requirements prohibiting the release of the key to the investigators until the individuals are deceased.

The UNG IRB Position Dealing with Existing Data Sets

Existing Data Sets: Option 1

De-identified data - whether publicly available or accessible through data use agreements – do not constitute human subjects research as defined at 45 CFR 46.102. UNLESS a project merges multiple data sets and in so doing enables the identification of individuals, the project *does NOT require IRB review*. Many studies utilize data made available through large data consolidation bureaus and consortiums. Here is a list of data holders whose archives include publicly available, de-identified data.

1. Inter-University Consortium for Political and Social Research (ICPSR)
2. U.S. Bureau of the Census
3. National Center for Health Statistics
4. National Center for Education Statistics

Existing Data Sets: Option 2

Research involving the analysis of publicly available data or non-publically available data (with data use agreement) containing private identifiable information that will not be recorded by the researcher in a manner that allows the direct or indirect identification of individuals qualifies for *exempt review* (45 CFR 46.101(b)4). Investigators submitting protocols involving these research procedures are asked to provide the following information to aid the IRB in making a determination of exemption.

1. description of data set and availability;
2. description of data to be accessed for analysis; and
3. copies of data use agreements required by data holder (non-publicly available data)
4. description of the de-identification process

Existing Data Sets: Option 3

In cases where researchers code private identifiable information from publically available and/or non-publically available data sets, *expedited or full IRB review is required*. For studies involving the analysis of coded private information or the analysis of private information by a third party on behalf of a research team, the IRB requests that investigators submit the following information to the IRB:

1. description of data set and availability;
2. description of data to be accessed for analysis; and
3. description of the data access or security plan to be implemented (including any requirements required by the data holder).

Existing Data Sets: Option 4

Research involving the analysis of publically and non-publicly available data that contains private identifiable information about living individuals is considered by the IRB to constitute human subjects research that is not exempt from 45 CFR 46 and IRB review requirements. This review is conducted under *expedited or full IRB board review* procedures in accordance with the IRB's review policies. As a part of the IRB application, investigators submitting protocols involving these research procedures will be asked to include the following information in their protocol submissions to aid the IRB in its review.

1. description of data set and availability;
2. description of data to be accessed for analysis; and
3. copies of data use or security agreements required by data holder; and

4. description of data security and access procedures to be implemented (including any requirements required by the data holder.

References

- The Common Rule 45 CFR 46: http://www.hhs.gov/ohrp/humansubjects/commonrule/
- HHS Guidance Document: http://www.hhs.gov/ohrp/policy/cdebiol.html
- https://research.missouri.edu/irb/files/secondary_data_projects.pdf

If you need this document in another format, please email irbchair@ung.edu ***or call 706-867-2969.***