



Anonymous and Deidentified Data

Protocols must include the description of identifiability of such data by classifying data as either anonymous, de-identified, or coded. Note that with small subject populations, such as those drawn from the University of Montana community, a constellation of characteristics of a population may allow for individuals to be identified regardless of definition.

Anonymous

The DHHS-Office for Human Research Protections (OHRP) describe coded and de-identified data but does not define anonymous in the regulations.

Data are anonymous if no one, not even the researcher or a third party entity (e.g., Qualtrics), can connect the data to the individual who provided it through direct identifiers such as name, address, IP address or any type of identification number or indirect identifiers (i.e., other unique individual characteristics like age, race, socioeconomic level, etc.) that might make it possible to identify an individual from a pool of subjects.

If a survey includes a demographics section or is being collected through a third-party vendor, the researcher may not be able to guarantee anonymity and that term should not be included in recruitment or consent materials.

Please note that the actual administration of anonymous surveys is still considered human subjects research because investigators are interacting with human participants.

De-identified

Data are considered de-identified when all direct or indirect identifiers or codes linking the data to the individual participant's identity are destroyed or broken, such that the investigator no longer has the ability to ascertain the identity of the participants. De-identification can occur by removing the code from the dataset or destroying the file link. After de-identification, even members of the study team no longer possess the means to link the data to participants. For example, data collection sources like MTurk are **not** completely anonymous and their data should be classified as de-identified, as the participant IDs are linked and stored by Amazon.com.

For example, a dataset received from another researcher may contain a code, but it may be considered de-identified provided the original researcher does not use the link to re-identify the individual or disclose the link to the receiving researcher.

De-identified data are sometimes referred to as "anonymized" data but note that de-identified data are not anonymous since investigators had access to identifiers prior to de-identification.



As a reminder for secondary use of data, if an investigator was involved with the original data collection, even if all identifiers have been removed, this is still considered human subject research and does not meet the definition of not human subject research (NHSR).

Coded

Data are considered coded if it meets the following criteria:

1. Identifying information (such as name or address/location) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a number, letter, symbol, or combination thereof (i.e. the code); and
2. A key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens

Coded data allows investigators to discuss a study without revealing their study participants' identities. All identifying information should be stored separately from coded study data, and no one outside of the study team should be able to link a participant's identity to coded data. For example, biospecimen data is collected from participants. Upon enrollment in the study, each participant is given a unique ID code that accompanies the biospecimen. The spreadsheet linking the participant to their ID code is kept password protected on a virtual workstation on a secure server that only the PI and study investigators can access.

Coded data are not anonymous since members of the study team have access to the identifiers. As long as a link exists, data are indirectly identifiable, and not anonymous or de-identified.

The code should not be a combination of information related to the individual, such as initials, date of birth, etc. It can be sequential numbers and/or letters, such as ST01, ST02, ST03, and so on.

HIPAA and De-identification

As per the Health Insurance Portability and Accountability Act (HIPAA), a dataset is considered de-identified if 18 identifiers are removed (45 CFR 164.514(b)(2)). If the dataset is not subject to HIPAA, it is considered anonymous if the identity of the human subjects cannot be readily ascertained. HIPAA identifiers include:

- A. Names;
- B. All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, 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:

- a. The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
 - b. The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
- C. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and 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;
- D. Telephone Numbers;
- E. Fax Numbers;
- F. Electronic mail addresses;
- G. Social security numbers;
- H. Medical record numbers;
- I. Health plan beneficiary numbers;
- J. Account numbers;
- K. Certificate/license numbers;
- L. Vehicle identifiers and serial numbers, including license plate numbers;
- M. Device identifiers and serial numbers;
- N. Web Universal Resource Locators (URLs);
- O. Internet Protocol (IP) address numbers;
- P. Biometric identifiers, including finger and voice prints;
- Q. Full face photographic images and any comparable images; and
- R. Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section; and 45 CFR 164.514(b)(2).