

Definitions

Anonymous data – Data that do not include identifiers (personal identifiers, indirect identifiers, audio recording, video recordings) are considered anonymous

Coded – Information is considered to be coded if it meets the following criteria:

- Identifying information (such as name or medical record number) 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 (e.g. the code); and
- A key to decipher the code exists, enabling the linkage of the identifying information to the private information or specimens.

Note: The code should not contain any combination of information related to the individual, such as initials, date of birth, social security (whole or partial).

De-identified – All direct personal identifiers are permanently removed; no code or key exists to link the information or materials to their original source and the remaining information cannot reasonably be used to anyone to identify the source

Health Insurance Portability and Accountability Act (HIPAA) Identifiers:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, 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.
 - b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.
 - c. All elements of dates (except year) 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.
3. Telephone numbers.
4. Facsimile 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 audio or video recordings, and fingerprints.
16. Full-face photographic images and any comparable images.
17. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.

Human Subject - According to the Code of Federal Regulations (CFR), a human subject is “a living individual(s) **about whom** an investigator (whether professional or student) conducting research obtains:

1. Data through intervention or interaction with the individual, **or**
2. Identifiable private information.” [45 CFR 46.102(e)(1)]

Identifiable Private Information - Information that has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public or accessed for research purposes. Includes (but is not limited to):

1. Information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place.
2. Information that has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public.
3. Private information or information of a nature which is individually identifiable (i.e., the identity of the subject is or may readily be ascertained).
4. Information of a nature such that the identity of the subject is or may readily be ascertained by the investigator or may be associated with the information.
5. Information which was collected specifically for the proposed project through intervention or interaction with living individuals and is of a nature that the investigator can readily ascertain the identity of the individual.

Indirect Identifiers are instances where a researcher does not collect or use a data set with personal identifiers, such as names, but combines enough information that someone could potentially identify them. Indirect identifiers include, but are not limited to:

- Age
- Ethnicity
- City or state of residence

- Occupation or role
- Job function or title
- Specific time, event, context or occasion

Limited Data Sets are only for purposes of research, public health, or health care operations. Data is still PHI but has had certain identifiers removed; as such, Privacy Regulations require covered entities to enter into data use agreements with recipients of “limited data sets”. A limited data set can contain:

- Dates such as admission, discharge, service, DOB, DOD
- City, state, five digits or more zip code; and
- Ages in years, months or days or hours.

All the following identifiers must be removed in order for health information to be a limited data set:

- Names
- Street addresses (other than town, city, state and zip code);
- Telephone numbers;
- Fax numbers;
- e-mail addresses;
- Social security numbers;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate license numbers;
- Vehicle identifiers and serial numbers, including license plates;
- Device identifiers and serial numbers;
- URLs;
- IP address numbers;
- Biometric identifiers (including finger and voice printers); and
- Full face photos (or comparable images)

Publicly Available means that the general public can obtain the data. Data are not considered “publicly available” if access to the data is limited to researchers.

Public Use Data Sets are data sets prepared with the intent of making them available for public use. The data available to the public are not individually identified (are de-identified) nor maintained in a readily identifiable form. Data shared informally among colleagues does not constitute public use data. Public Use Data Sets (de-identified) are

not considered human subject research (as it does not involve access to identifiable private information about the persons from/about whom the data were collected).