Identifiers Guidance

Protected Health Information (PHI)

PHI 18 identifiers defined under HIPAA Privacy Rule include:

1. Names
2. All geographic subdivisions smaller than a State (*limited identifier), 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:
   - The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
   - The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
3. 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 (*limited identifiers)
4. Telephone numbers
5. Vehicle identifiers and serial numbers, including license plate numbers
6. Fax numbers
7. Device identifiers and serial numbers
8. Email addresses
9. Web Universal Resource Locators (URLs)
10. Social security numbers (includes parts of SS#s)
11. Internet Protocol (IP) addresses
12. Medical record numbers
13. Biometric identifiers, including finger and voice prints
14. Health plan beneficiary numbers
15. Full face photographic images and any comparable images
16. Account numbers
17. Any other unique identifying number, characteristic, or code, (excluding a random identifier code for the subject that is not related to or derived from an existing identifier)
18. Certificate/license numbers

Personally Identifiable Information (PII)

PII is any data that could be used to identify an individual. Examples include:

- Name, such as full name, maiden name, mother's maiden name, or alias
- Personal identification number, such as social security number (SSN), passport number, driver's license number, taxpayer identification number, patient identification number, and financial account or credit card number
- Address information, such as street address or email address
- Asset information, such as Internet Protocol (IP) or other host specific persistent static identifier that consistently links to a particular person or small, well defined group of people
- Telephone numbers, including mobile, business, and personal numbers
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), x-rays, fingerprints, or other biometric image or template data (e.g., retina scan, voice signature, facial geometry)
- Information identifying personally owned property, such as vehicle registration number or title number and related information
- Information about an individual that is linked or linkable to one of the above (e.g., date of birth, place of birth, race, religion, weight, activities, geographical indicators, employment information, medical information, education information, financial information)

The Family Educational Rights and Privacy Act (FERPA) specifically defines PII as including, but not limited to:

1. The student's name;
2. The name of the student's parent or other family members;
3. The address of the student or student's family;
4. A personal identifier, such as the student's social security number, student number, or biometric record;
5. Other indirect identifiers, such as the student's date of birth, place of birth, and mother's maiden name;
6. Other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty; or
7. Information requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates.

Protected Health Information (PHI):

The Privacy Rule (HIPAA) protects all individually identifiable health information held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. Protected Health Information (PHI) is information, including demographic data which relates to:

- the individual's past, present or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or there is reasonable basis to believe it can be used to identify the individual

See OHRP Guidance Regarding Methods for De-identification of PHI in Accordance with HIPAA Privacy Rule:
https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#protected

Personally Identifiable Information (PII)

2CFR200.79 defines PII as information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. Some information that is considered to be PII is available in public sources such as telephone books, public Web sites, and university listings. This type of information is considered to be Public PII and includes, for example, first and last name, address, work telephone number, email address, home telephone number, and general educational credentials. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified. Non-PII can become PII whenever additional information is made publicly available, in any medium and from any source, that, when combined with other available information, could be used to identify an individual.

Coded data:

This refers to data which have been stripped of all direct subject identifiers, but in this case each record has its own study ID or code, which is linked to identifiable information such as name or medical record number. The linking file must be separate from the coded data set. This linking file may be held by someone on the study team (e.g. the PI) or it could be held by someone outside of the study team (e.g. a researcher at another institution). A coded data set may include *limited identifiers under HIPAA. The code itself may not contain identifiers such as subject initials or medical record number. Note: Secondary research with coded private information or coded biological specimens must be submitted to the IRB for review.

De-identified data:

This refers to data which have been stripped of all subject identifiers, including all 18 HIPAA identifiers. This means that there can be no data points that are considered limited identifiers under HIPAA; i.e. geographic area smaller than a state, elements of dates (date of birth, date of death, dates of clinical service), and age over age 89. If the data set contains any *limited identifiers, it is considered a limited data set under HIPAA. If the data includes an indirect link to subject identifiers (e.g. via coded ID numbers), then the data is considered by the IRB to be coded, not de-identified. Please note that data can be considered de-identified under the Common Rule but NOT the HIPAA Privacy Rule (e.g., limited data sets), and vice versa (e.g., no HIPAA identifiers are included but the combination of data points could make subjects identifiable).

Anonymous data:

Essentially the same thing as de-identified data, this refers to data which have been stripped of all subject identifiers and which have no indirect links to subject identifiers. There should be no *limited identifiers in an anonymous data set.