# De-Identified Data

**De-Identified Data** is health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual.

**CODED DATA IS NOT THE SAME AS DE-IDENTIFIED DATA:** **Coded Data** is data in which identifying information (such as name or social security number) has been replaced with a number, letter, symbol, or combination thereof (i.e., the code); and a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens.*

**Protected Health Information (PHI)** is individually-identifiable health information maintained in any form or medium. HIPAA outlines 18 PHI identifiers:

1. Names (of patient, provider, relatives, caregivers, legal representatives)
2. Any geographical 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.
3. Any elements of dates (except year) for dates directly related to an individual, including date of birth, date of death, **admission date, discharge date, test/procedure date**; 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 ***(All dates are considered identifiable.)***
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social security numbers ***(VA considers scrambled SSNs identifiable.)***
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers ***(IP addresses can make internet survey results identifiable.)***
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code

---

* In general, OHRP considers private information or specimens to be individually identifiable when they can be linked to specific individuals by the investigator(s) either directly or indirectly through coding systems unless:

- The information/specimens were not collected specifically for the currently proposed research project through an interaction or intervention with living individuals; and
- The investigator(s) cannot readily ascertain the identity of the individual(s) to whom the coded data pertains because, for example:
    - the investigators and the holder of the key enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (note that the HHS regulations do not require the IRB to review and approve this agreement); or
    - there are IRB-approved written policies and operating procedures for a repository or data management center that prohibit the release of the key to the investigators under any circumstances, until the individuals are deceased; or
    - there are other legal requirements prohibiting the release of the key to the investigators, until the individuals are deceased.

**References:**


1. The Health Insurance Portability and Accountability Act of 1996
2. VHA Handbook 1200.12 - Use of Data and Data Repositories in VHA Research
3. VHA Handbook 1605.1 – Privacy and Release of Information