What is de-identified data?

De-identified data are not subject to the requirements of the Privacy Rule because the data are not individually identifiable. There are two ways to de-identify data:

**Safe Harbor Method**—in which all of the following elements are removed from a data set:

- Names
- All geographic subdivisions smaller than a State, including:
  - Street Address
  - City
  - County
  - Precinct
- Zip Codes and their equivalent geocodes, Except for the initial three (3) 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 (3) initial digits contains more than 20,000 people and
  - the three (3) digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000
- Telephone Numbers
- Fax Numbers
- Email Addresses
- Social Security Numbers
- Medical Record Numbers
- Health Plan Beneficiary Numbers
- Account Numbers
- All Elements of Dates (except year) for dates related to an individual, including:
  - Birth date
  - Admission date
  - Discharge date
  - Date of death
  - All ages over 89 and all elements of dates (including year) indicative of such ages and elements may be aggregated into a single category of age 90 or older.
- Certificate/License Numbers
- Vehicle Identifiers and Serial Numbers, including license plate numbers
- Device Identifiers and Serial Numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) Address Numbers
- Biometric Identifiers, including finger and voice prints
- Full Face Photographic Images and any comparable images
- Any other Unique Identifying Numbers, Characteristics, or Codes

If all of the 18 identifiers listed above are removed, the information is no longer
1. individually identifiable,
2. PHI, and
3. Subject to HIPAA’s requirements.

**Statistical Method**—in which certification is provided by “a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable so that there is a ‘very small’ risk that the
What is de-identified data?

information could be used by the recipient to identify the individual who is the subject of the
information, alone or in combination with other reasonably available information.” For more
information see NIH Guidance Protecting Personal Health Information in Research:
Understanding the HIPAA Privacy Rule.
A data set may also be considered de-identified if an expert in statistical and scientific methods
determines and documents that the methods used to de-identify or code the data present a very
small risk that the information can be used alone or in combination with other reasonably
available information to identify an individual.
"Anonymous" data are not necessarily considered de-identified under HIPAA. Anonymity under
the federal Common Rule requires that individuals cannot be readily ascertained by the
investigator and cannot be associated with the data. According to the Common Rule standard,
anonymous data may retain dates of treatment. Under HIPAA's more stringent requirements,
however, such data would be considered identifiable data.

Unique solution ID: #1274
Author: Jennifer Weddle
Last update: 2013-04-23 19:02