

De-identification of Protected Health Information

De-identification is the process of removing identifiers from protected health information (PHI), which includes common identifiers such as name, age, address, and SSI, protected by the HIPAA Privacy Rule. Once information has been de-identified, it is no longer considered PHI and is not protected by the HIPAA Privacy Rule. [1]

The de-identification standard is provided by Section 164.514(a) of the HIPAA Privacy Rule, which outlines two methods for de-identification: Expert Determination, and Safe Harbor.

By the Safe Harbor method, 18 types of identifiers are removed, including names, geographical information, numbers and accounts associated with the individual, biometric identifiers, images, electronic records (IP address numbers, emails, and URLs), and, as a catch-all, any other unique identifier. The entity collecting the information must have no “actual knowledge” (defined as clear and direct knowledge) that any remaining disclosed information could be used alone or in combination with other available information to identify an individual. [1, 2, 3]. A complete list of direct identifiers follows this portion of the paper [4]

Alternatively, covered entities may use the Expert Determination method, whereby an expert in scientific and statistical principles, whose expertise is verified by OCR (Office of Civil Rights) review, must determine that the information used presents only a very small risk that a subject could be identified, either alone or in combination with other publicly available information. The expert reviews the information under several principles (replicability, data source availability, distinguishability), and assesses the risk (high or low). For example, an expert might determine that the results of a blood glucose test present low risk of identification because glucose levels vary, while an individual's name and birth date are high risk information because they are relatively stable [1]. The expert must document the statistical methods and analysis used to determine that data has been de-identified, and must keep the certification for a minimum of six years. This method is preferable for scientific studies because it does not require removing all 18 types of identifiers. [5]

De-identification strategies are typically selected in order to minimize the inevitable loss in information and avoid limiting the usefulness of information as much as possible. As de-identification may present problems for statistical analysis by removing gender, race, and age, and other [1]

Similar to de-identified data is coded data, which is data that has been stripped of subject identifiers but for which there exists a code that is designed to enable re-identification. This code, which may be held by a member of the study team, such as a PI, links limited identifiers (such as non-identifying subject IDs) with direct identifiers. The code *is* considered PHI and is under the protection of the HIPAA Privacy Rule.[1, 4]

One recent issue that may become relevant to de-identification in the future is the sharing of genomic data. An increasing number of studies generate large quantities of human genomic data. As a result, the NIH has established the Genomic Data Sharing (GDS) Policy for responsibly sharing genomic data applying to grants submitted after January 25, 2015. [5]

A complete list of direct identifiers covered by the Privacy Rule [4]:

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.
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.
4. Telephone numbers.
5. Facsimile numbers.
6. Electronic mail addresses.
7. Social security numbers.
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.
16. Biometric identifiers, including fingerprints and voiceprints.
17. Full-face photographic images and any comparable images.
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.

1. U.S. Department of Health and Human Services, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Retrieved June 20, 2015 from <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html>
2. University of Miami Miller School of Medicine, de-identified health information (HIPAA). Retrieved June 20, 2015 from http://privacy.med.miami.edu/glossary/xd_deidentified_health_info.htm
3. Johns Hopkins Medicine Definition of de-Identified Data. Retrieved June 20, 2015 from http://www.hopkinsmedicine.org/institutional_review_board/hipaa_research/de_identified_data.html
4. Health Sciences institutional Review Boards, the difference between "de-identified", "anonymous", and "coded" data. Retrieved June 20, 2015 from <https://kb.wisc.edu/hsirbs/page.php?id=25351>
5. HIN Genomic Data Sharing Policy. Retrieved June 20, 2015 from <http://gds.nih.gov/03policy2.html>
6. Atherosclerosis Risk in Communities Study (ARIC). Retrieved June 20, 2015 from <https://biolincc.nhlbi.nih.gov/studies/aric/>
7. BioLINCC Frequently Asked Questions. Retrieved June 20, 2015 from <https://biolincc.nhlbi.nih.gov/faqs/>