

Compliance Today – December 2023



Mark J. Fox (mfox@acc.org, [linkedin.com/in/mark-fox-chc-chpc-chrc-5376559/](https://www.linkedin.com/in/mark-fox-chc-chpc-chrc-5376559/)) is Privacy and Research Compliance Officer at American College of Cardiology Foundation in Washington, DC.



Thora A. Johnson (thora.johnson@orrick.com, [linkedin.com/in/thora-johnson-390bba133/](https://www.linkedin.com/in/thora-johnson-390bba133/)) is a Partner at Orrick, Herrington & Sutcliffe LLP in Washington, DC.

It is de-identified, or is it?

by Mark J. Fox, CHC CHPC CHRC, and Thora A. Johnson

How many of us have heard one of our stakeholders assert that their data set is de-identified? As privacy compliance professionals, we are often tasked with critically evaluating data sets and guiding stakeholders to fully understand what is and what is not considered de-identified. Another challenge is that stakeholders may not conduct de-identification often and, therefore, may experience definition drift.

This article will explore the two permissible methods of de-identification under HIPAA. We will focus primarily on the safe harbor method as this is the most frequently used method, including exploring special considerations for zip codes, dates, and the “18th identifier”; however, we will touch on the expert determination method as well. Remember, once a data set meets the HIPAA parameters for de-identification, it is no longer subject to HIPAA. We will also differentiate a limited data set from a de-identified data set, as stakeholders often confuse these terms.

Definition of protected health information

Before discussing de-identification methods, let’s review the definition of protected health information (PHI) as described by HIPAA. PHI is defined as individually identifiable health information, including demographic data, that relates to the individual’s past, present, or future physical or mental health condition; the provision of healthcare to the individual; or the past, present, or future payment for the provision of healthcare to the individual, that identifies—or which it could reasonably be believed to identify—the individual in the hands of a covered entity or a business associate.^[1] This definition is very broad, and it is important that the team performing de-identification fully understands this definition.

Data classification

A mechanism that aids in de-identification is data classification. In instances where you are subject to HIPAA and the patient data is in a structured data set, three simple classifications are recommended. First is a fully identifiable data set. A fully identifiable data set contains one or more direct identifiers. Direct identifiers are those from the subsequent list of 18 identifiers that cannot be included in a limited data set, for example, name and address. Second is a limited data set. A limited data set does not contain any direct identifiers but only includes indirect identifiers. The limited data set classification should conform to the HIPAA definition of a limited data set. Third is a de-identified data set. This data set does not contain any direct or indirect identifiers and conforms to the requirements of safe harbor de-identification under HIPAA.

Limited data sets

A limited data set is a data set that removes all elements of PHI except for dates such as admission, discharge, procedure, date of birth, date of death, city, state, zip code, and ages in years, months, days, or hours. Recipients who receive a limited data set must sign a data use agreement and must comply with the terms and conditions of the data use agreement.

A data use agreement limits the use of the limited data set to the following uses: healthcare operations, public health, and research. The data use agreement should limit uses to those negotiated between the organization holding the data set and the recipient.

Know your audience

Individuals often don't understand the definitions of limited data sets and de-identified data sets. To assist in understanding, consider distributing a list of data elements or a data collection form to potential recipients and ask them to highlight the requested fields. Never assume a customer knows these definitions.

This document is only available to members. Please [log in](#) or [become a member](#).

[Become a Member](#) [Login](#)