

Stronger Protections for, and Encouraging the Use of, De-Identified (and "Anonymized") Health Data

June 26, 2009

Policy Posts are in-depth analyses on current tech policy issues from CDT experts. Sign up to receive the latest Policy Posts here:

[1\) The Importance of De-Identified Health Data](#)

[2\) De-Identification, Limited Data Set Requirements of the HIPAA Privacy Rule](#)

[3\) Why a Re-Examination of De-Identification Policy is Needed](#)

[4\) Some Recommendations for Reform](#)

1) The Importance of De-Identified Health Data

The trend towards adoption of health information technology (health IT) offers substantial benefits not only to individuals in terms of improving health care quality and increasing efficiency, but also to medical research, public health and other functions that derive value from large sets of health-related data.

At the same time, increased electronic flows of health data pose significant risks to privacy. Among the many challenges that will require attention as health IT is promoted through implementation of the stimulus legislation and other means is how to strip health data of personal identifiers in order to eliminate or reduce privacy concerns, while still retaining useful information.

Numerous public and private entities currently use de-identified health data. Among the most widespread applications of de-identified data are quality improvement, public health (including syndromic surveillance), research (including clinical and epidemiological research), and a variety of commercial uses, such as improving the efficiency of operations, and understanding risks to patients.

[Health Privacy Project De-identification Paper](#) [1] June 2009

2) De-Identification and Limited Data Set Requirements of the HIPAA Privacy Rule

Under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, health data is categorized in one of three ways: protected health information, de-identified data, and the limited data set. Health data that is fully identifiable--that contains patient names, addresses or other identifiers--is "protected health information" and is subject to some restrictions on access, use, and disclosure.

Two additional classes of data are stripped of identifiers and are either exempted from, or treated differently under, the Privacy Rule. First, "de-identified" data has been so stripped of common identifiers that there is no "reasonable basis" to believe it can be traced back to the subject. Data that qualifies as "de-identified" under the Privacy Rule is not regulated at all; there are no restrictions on who can acquire it or the purposes for which it can be accessed, used, or disclosed.

A "limited data set," by contrast, is stripped of many categories of identifying information but retains

information often needed for public health and research (such as birth dates, dates of treatment, and some geographic data). Entities covered by HIPAA may share a limited data set for research, public health and health care operations purposes permitted by the Privacy Rule, so long as all recipients are bound by a data use agreement with the originator of the data.

3) Why a Re-Examination of De-Identification Policy is Needed

Although the intentions underlying the Privacy Rule's three-part approach (protected health information, de-identified data, and the limited data set) were laudable, the framework has been rendered less satisfactory as a result of technology changes and a growing sophistication in the use of data. At least three challenges arise. First, not all uses of de-identified health data or a limited data set require identical levels of identity masking. Ideally, a broader spectrum of data "anonymization" options would better meet the needs of different contexts and ensure that data is accessed or disclosed in the least identifiable form possible for any given purpose. We use the term "anonymized" to refer to data that is intended to be anonymous to data recipients.

Second, the Privacy Rule, by permitting use of fully identified data for treatment, payment, health care operations, and a range of other health-related activities, provides little incentive for covered entities to use data that is less than fully identifiable. Of particular concern is the category of health care operations, which includes some tasks that arguably could be fulfilled with data that is less than fully identifiable. Covered entities are required under the Rule to use the "minimum necessary" amount of data needed to accomplish health care operations tasks, but CDT is unaware of any circumstances in which this standard has been expressly interpreted to set limits on the identifiability of data used for a particular function.

Third, the de-identification provisions of the Privacy Rule may no longer be as effective as they once were at protecting privacy. Changes in society and technology, including a vast explosion in the volume of digital data, have made the re-identification of health information easier and cheaper than ever before. In addition, the Privacy Rule has never included mechanisms for holding all recipients of de-identified data accountable for re-identification.

4) Some Recommendations for Reform

CDT proposes several ways to strengthen the Privacy Rule's de-identification standards and to encourage the use of de-identified data through complimentary policies. We also recommend that the Department of Health and Human Services (HHS) consider creating additional data anonymization options (beyond de-identification and the limited data set), either by regulation or through guidance on how to apply the minimum necessary standard to routine uses and disclosures of data beyond treatment. We offer the following specific recommendations to balance the twin interests of flexibility and data protection:

- **Strengthen accountability by requiring data use agreements**

HHS should consider requiring HIPAA covered entities to enter into data use agreements with recipients of de-identified data. Such agreements need not rise to the level of business associate agreements, which are needed to protect fully identifiable data. Instead, they can be more limited in scope, similar to those used for limited data sets. In addition, HHS and Congress should consider how to hold entities disclosing and/or receiving de-identified data accountable when data is inappropriately re-identified.

- **Expand data anonymization options under the Privacy Rule**

Different levels of data protections are appropriate in different contexts. Providing only two options for masking data may limit the value that can be derived from data, leaving researchers and others seeking aggregate data with few alternatives beyond the use of fully identifiable data. HHS should consider developing additional data anonymization options that can be used for a broader range of activities and that are appropriately protected against re-identification.

- **Provide incentives to use less than fully identifiable data for certain purposes**

Fully identifiable data may not be needed to accomplish all of the activities currently included in the Privacy Rule under "health care operations." Ideally, the degree of protection for data should increase with its degree of identifiability. While drafting specific yet sufficiently flexible rules to accomplish a sliding scale of protections will be a challenge, the limited data set may serve as a good model for handling data that is not fully identifiable. At a minimum, protections to ensure that data is not inappropriately re-identified are critical. CDT encourages HHS to incorporate these points in the guidance on the Privacy Rule's minimum necessary standard that the stimulus legislation requires be issued by August 17, 2010.

- **Provide support through "Centers of Excellence" in de-identification**

Given that many HIPAA covered entities do not have sufficient in-house expertise to de-identify data using sophisticated methodologies, HHS should consider designating certain existing or new organizations "Centers of Excellence" with respect to data de-identification. Covered entities seeking to release de-identified data could be required or given incentives to consult with these entities to gain the necessary expertise, or could outsource the work of de-identification to them. In developing this process, HHS should consider partnering with the National Institute for Standards and Technology (NIST), which has significant expertise on data anonymization techniques.

- **Require or encourage the use of limited access datasets and other technical solutions**

HHS should consider requiring or encouraging the use of innovative technical solutions to protect data. One promising approach is the use of "limited access datasets," which give users access only to aggregate data that is relevant to specific questions they pose. Information that is not essential to a particular inquiry, including patient identifiers, is never shared, making it much more difficult to associate data with a particular individual. Such technical solutions should be applied to existing databases rather than creating new ones. Allowing data to remain in its place of origin and bringing critical research questions to the data, is the most efficient and effective way to meet the diverse needs of our health system while protecting privacy and security.

- **Require education and training of staff de-identifying data**

Any staff involved in de-identifying health data or working with it should participate in basic training about how best to use procedural and technical means to protect privacy and security. Staff training, perhaps supported by the Centers of Excellence described above, would help to minimize the likelihood of breaches and other misuses of data.

- **Consider increasing public transparency regarding uses of de-identified data**

As discussed, data that has been de-identified according to the Privacy Rule's provisions is free from restrictions on use and disclosure. If it is rigorously de-identified and sufficiently protected against re-identification, such data does not raise a privacy risk to individuals. However, some have expressed other policy concerns about the ways that de-identified data is currently being used. To address these concerns, policymakers could encourage or require a greater degree of public transparency about how data (including de-identified data) is used. Greater transparency could contribute to the development of guidelines regarding data use.

Copyright © 2013 by Center for Democracy & Technology. CDT can be freely copied and used as long as you make no substantive changes and clearly give us credit. [Details](#).

Source URL:

<https://cdt.org/policy/stronger-protections-and-encouraging-use-de-identified-and-anonymized-health-data>

Links:

[1] http://www.cdt.org/healthprivacy/20090625_deidentify.pdf