



HEALTH IT LAW & INDUSTRY



REPORT

Reproduced with permission from Health IT Law & Industry Report, 2 HITR 18, 04/26/2010. Copyright © 2010 by The Bureau of National Affairs, Inc. (800-372-1033) <http://www.bna.com>

Data Security

BNA INSIGHTS: The Difficult Task of Setting Standards for the De-Identification of PHI

By JENNIFER S. GEETTER

The Health Information Technology for Economic and Clinical Health Act (HITECH Act), enacted as part of the American Recovery and Reinvestment Act of 2009, requires the Department of Health and Human Services (DHHS) to issue guidance on methods for de-identification of protected health information (PHI) as designated in the Privacy Rule of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The Privacy Rule is part of a suite of regulations that includes the Security Rule and the Standards for Transactions and Code Sets (45 CFR § 160 & 162), which have been enacted pursuant to HIPAA. The HITECH Act requires the Secretary of DHHS, in consultation with stakeholders, to issue guidance on how to

Jennifer S. Geetter, a partner in the Washington office of McDermott Will & Emery, focuses on emerging health issues, including those involving electronic health records and data strategy initiatives. She has been recognized by Chambers USA as a leading individual in health care in Washington. She can be contacted at jgeetter@mwe.com or (202) 756-8205.

best implement the Privacy Rule's current standards for the de-identification of PHI set forth at 45 CFR § 164.514(a)-(b).

Privacy Rule and the HITECH Act

The Privacy Rule sets forth the permissible uses and disclosures of protected health information (PHI) that is held or transmitted by covered entities and their business associates. When health information does not identify an individual, and there is no reasonable basis to believe that it can be used to identify an individual, it is considered to be "de-identified" and thus outside the definition of PHI.

The Privacy Rule designates two ways that a covered entity can determine that health information is de-identified. The first is the so-called "safe harbor" approach by a covered entity, which renders information de-identified by removing 18 enumerated identifiers (for example, names, dates of service and birth, and certain address information) and has no knowledge that the remaining information could be used either alone or in combination with other information to re-identify someone.

Second, a covered entity may engage a qualified statistical or scientific expert to render the data set de-identified through accepted analytic techniques, pro-

vided that the risk the information could be re-identified is very small.¹

The stakes of non-compliance with de-identification are now even higher. First, the HITECH Act has intensified civil monetary penalties for improper uses and disclosures. Second, OCR issued an Interim Final Rule, effective September 2009, setting forth regulations for breaches of unsecured PHI. These additional regulations may require notice to affected individuals and the Secretary of DHHS.

Expanded Penalties

Section 13402(h) of the HITECH Act defines “unsecured protected health information” as “protected health information that is not secured through the use of a technology or methodology specified by the Secretary” of DHHS such that it is unreadable, unusable or indecipherable. DHHS guidance has listed the two acceptable technologies and methodologies for rendering PHI secure—encryption and destruction—and included specific definitions for each. For instance, paper, film or other hard-copy media are considered “destroyed” only if they are shredded or altered so that they cannot be read or otherwise reconstructed, and redaction is specifically excluded as a mode of destruction.

A covered entity is not required to render PHI secure. However, if there is an improper use or disclosure of unsecured PHI, then the covered entity must determine whether this use or disclosure has caused a significant risk of harm to the affected individuals. The results of this risk assessment will determine whether the covered entity is obligated to notify the affected individuals and the Secretary of DHHS.²

Prior to the enactment of the HITECH Act, covered entities were subject to HIPAA civil money penalties of up to \$100 per violation, with an annual cap of \$25,000 for identical violations within a calendar year. The Interim Final Rule, published on October 30, 2009, preserves this scheme for violations occurring prior to February 18, 2009. For violations occurring on or after February 18, 2009, the Rule amends the HIPAA enforcement regulations to include the imposition of tiered ranges for civil money penalty amounts based upon the increasing levels of culpability associated with such violations. Under the new scheme, the range of minimum penalty amounts for each offence increases from \$100 for first-tier violations to \$50,000 or more for fourth-tier violations. Similarly, the penalty amount available in a calendar year for identical violations is substantially increased from \$25,000 to \$1,500,000.

In increasing severity, the tiers of culpability under the Interim Final Rule are:

1. Entity did not know (and, by exercising reasonable diligence, would not have known) that it violated the applicable provision.
2. Violation is due to reasonable cause and not to willful neglect.
3. Violation is due to willful neglect and was corrected during the 30-day period beginning on the

first date the entity knew, or, by exercising reasonable diligence, would have known that the violation occurred.

4. Violation is due to willful neglect and was not corrected during the 30-day period beginning on the first date the entity knew, or, by exercising reasonable diligence, would have known that the violation occurred.

The maximum \$1.5 million calendar year penalty applies in all categories. In the first three categories the maximum penalty per violation is \$50,000. In the fourth category \$50,000 is the *minimum* penalty per violation.³

OCR Conference: Themes

The HITECH Act also requires that OCR revisit certain bedrock HIPAA concepts to determine whether they remain sufficient in today’s data environment. One such concept under re-examination is de-identification. To meet the mandate of the HITECH Act, OCR held a two-day open-house on March 8 and 9, 2010, in Washington, D.C. to hear from a wide range of informed stakeholder regarding de-identification approaches, best practices for implementation and management of the current de-identification standard, and potential changes to address policy concerns. OCR states that it will “synthesize the input from workshop panelists and general comments to incorporate into guidance . . . for public comment.”⁴

Panel participants were biomedical, policy, legal and statistical stakeholders who addressed the current safe harbor and statistical de-identification standards and whether such approaches are adequate in today’s data environment. HHS has indicated that it will publicly communicate the substance of conference proceedings as part of the process to set final guidance. Attendees’ comments and concerns indicate the immense task still before OCR in promulgating and implementing de-identification standards under the HITECH Act.

Much has changed since the Privacy Rule went into effect. It has been 14 years since HIPAA passed, and we live in an age where digitalization of data is rapidly the norm. The Internet and other easily, publicly searchable sources mean that more information moves more quickly. In addition, the amount and variety of “auxiliary data” involving large, publicly available data sets typically unregulated by HIPAA are multiplying. This auxiliary data can be used to triangulate data sets and to re-identify individuals described within them. This new data environment, combined with the HITECH Act’s intention to accelerate the digitalization of medical information through electronic health records (EHRs) has raised concerns that the existing rules may no longer be adequate.

In addition, the conference confirmed that there is a definite and widespread sentiment among medical researchers that the general public and political leaders may be underestimating the nature of existing privacy

¹ U.S. Department of Health and Human Services, “HIPAA Privacy Rule’s De-Identification Standard,” <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/deidentificationworkshop2010.html>

² Federal Register, 8/24/09, <http://edocket.access.gpo.gov/2009/pdf/E9-20169.pdf>

³ McDermott Will & Emery, “HHS Issues Interim Final Rule Conforming HIPAA Civil Money Penalties to HITECH Act Requirements,” http://www.mwe.com/index.cfm/fuseaction/publications.nldetail/object_id/ae68626d-301b-4aa7-9a20-911cbe1b1f4a.cfm

⁴ HHS, “De-Identification Standard,” note 1. A webcast of the conference is available at <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/deidentificationworkshop2010.html>.

protections, at least with respect to being able to balance the desire for a high level of privacy reassurance with other public mandates (for example, research that can produce disease or health condition treatment breakthroughs).

Finally, the great unknown is what genetic information means for de-identification. As Supreme Court Justice Sandra Day O'Connor once observed in evaluating another legal standard, the concept of de-identification may be on a "collision course with itself" as it takes as a premise that de-identification is, itself, possible and a distinguishable data event. The inclusion of genetic information in a data set, however, poses a futuristic challenge to the binary split between identifiable and de-identified information. Although our ability to re-identify someone from genetic information right now is minimal because we typically do not have an identifiable, matching sample in a database, as genetic information proliferates this may no longer be the case.

Against this backdrop, panel speakers discussed a number of very specific topics, but their comments were generally motivated by a number of over-arching general themes, including:

- How do we balance the privacy harm risks to the public from insufficient de-identification rigor against the treatment and research harm risks to the public from insufficient information to power medical and research decision-making? The conference was marked by a constant push-pull between those advocating for stricter de-identification standards (including those proposing methodologies that effectively reduce the risk of re-identification to zero) and those who respect privacy concerns but worried that an exclusive focus on privacy will undermine other social priorities, principally, quick and accurate access to medical information for treatment and robust data sets for biomedical research. There was a definite sense among conference attendees with a biomedical innovation and research mission that current data security provisions already hamper medical research efforts. Many researchers felt that limited data sets were insufficient to do the type of research necessary to support biomedical research. These speakers contended that the Privacy Rule already is a significant impediment to biomedical research and should be relaxed with respect to biomedical research (if in fact research should not be carved out of the Privacy Rule entirely). Some researchers were concerned that the de-identification (and limited data set) restrictions make it difficult to conduct peer review of medical research since it is not always possible to review details of the underlying data source used to conduct the original study. Similarly, researchers found the Privacy Rule to make very difficult the process of linking, in which records are taken from a variety of different providers and sources to create a fuller data picture in order to do personalized medicine research or other research where a full clinical picture is critical. Concern was expressed that the lack of a national patient identifier exacerbates this problem. Especially vulnerable is research to support personalized medicine, defined as genetically-based identification and treatment of specific patient subgroups, which may in many cases be able to identify individuals for whom a new treatment is highly effective even where the treatment shows no benefit in the general population. Personalized medicine has strong support in the Obama Administration as a means of controlling health costs while advancing innovative treatment, both of which could be jeopardized without adequate data access.
- Do we need to reduce the risk of re-identification to zero? If not, when is information de-identified "enough" so that the risks of re-identification are acceptably low? For example, panelists considered whether actual harm resulting from the improper use or disclosure should be a factor in considering de-identification standards. Panelists suggested that in some cases the simple fact that individual data has been re-identified may not be a cause of specific harm. Other speakers, however, felt that even if the individual did not suffer reputational or identify harm, suffer discrimination or other loss of opportunity, or suffer personal embarrassment as a result of the improper use and disclosure, that person's autonomy—the ability to control how information is used and disclosed within a set of established parameters—had been violated.
- Even if re-identification is technically possible, should the worst case scenario, for example, a malicious hacker accessing PHI, drive general standards? For example, several panelists queried how real are widespread efforts to re-identify PHI. From this perspective, even if re-identification is technically possible using available secondary data sources (for example, publicly available disease state registries and voter registration logs), there was considerable debate as to the frequency with which a third party would go to the trouble and expense of seeking to re-identify the health information. Some panelists felt that hackers and other unwanted third parties were more interested in obtaining easily identifiable financial information than in re-identifying health information simply because it was possible to do so. Panelists explored whether theoretical risks proven in academic circles but not seen, so far, to be prevalent should be the principal driver in setting standards of general applicability.
- Should there be different rules for data sharing with different types of data recipients (for example, a researcher versus a for-profit third party)? Panelists explored whether OCR should further diversify the rules to account for different types of data recipients. Currently, the Privacy Rule does not recognize certain types of privileged data recipients, notably, researchers can receive a limited data set (which contains PHI) for research purposes without subject authorization; however, in most cases, the Privacy Rule treats all data recipients the same. One potential approach could be to distinguish between different types of PHI recipients and to allow for greater data access for more trusted users. Complementing this belief is the viewpoint that the risk of re-identification should be considered in light of the entire data release "context." In this view, focusing on the data included is insufficient for assessing privacy compliance guidelines. Instead, the nature of the data recipient, the motivations of the recipient, the data

release method, and similar criteria should all be considered in assessing re-identification risk. Levels of “trust” could well be seen as a legitimate means of setting de-identification standards, and overly strict standards may not be warranted for trusted users, particularly those pursuing biomedical research efforts.

Where Do We Go From Here?

There was no shortage of ideas expressed about how OCR could or should address these privacy concerns. Speakers from a variety of scientific and industry disciplines suggested these as possible approaches to deal with the difficulties of de-identification:

- Establishment of “centers of excellence” or other OCR-approved centers that would govern de-identification approaches. This was suggested as a means to increase consistency and give greater reassurance of expertise in de-identification. These centers might make available statistical de-identification models to smaller covered entities that might not otherwise have the in-house or budgetary resources to avail themselves of commercially available methods.
- Creation of a “differential privacy” standard, which would demonstrate mathematically that information cannot be re-identified under any circumstances (essentially, a zero-tolerance standard).
- Accounting of de-identified information in a publicly available resource as a mechanism to sufficiently inform members of the public how their data has been used to give them the means of assessing whether they were “harmed” by PHI use.
- Establishment by OCR of the means by which sensitive health information about groups could be protected even if de-identified.
- Greater clarity on how to address “free-text” areas in medical records and other data sources. While the information contained in such fields might not be one of the 18 safe harbor identifiers, many speakers felt that it was sufficiently rich in

anecdotal and other personal details that it could be used for re-identification.

- Incentivizing “keeping the data local.” This approach would involve having third-party researchers submit their analytical questions to the data source rather than requesting a de-identified or limited data set on which the researchers would themselves run the analytical queries. The benefits of this approach would be that it would reduce the number of quasi-public and public data sets and would reduce the risks of errors in creating the data sets that could result in re-identification. If this approach is endorsed, it suggests that there may be a fundamental shift in how data is organized and aggregated, and that entities with a head start in creating large, searchable, integrated databases may come out ahead.

Final Thoughts

It is difficult to predict where OCR goes from here. Some speakers felt that there had not been sufficient opportunity to create a “community of stakeholders” to consider all of the risks and benefits of revising de-identification standards. Others felt that risks of re-identification had become sufficiently severe that additional safeguards were a must. Others still felt that the risks to medical treatment and research were not being given their due.

Although it is by no means clear how OCR is leaning, given the overall general HIPAA atmosphere evidenced by the issuance of stricter standards throughout 2009, covered entities and business associates should be prepared for stricter standards. Of course, it is also possible that OCR will focus on specific, smaller tweaks rather than a fundamental overhaul. No matter the outcome, de-identification remains a core tool of covered entities to comply with HIPAA and to subsequently use and disclose data.

Given the penalties and the new unsecured PHI rules, covered entities and business associates would be wise to revisit their approaches to de-identification and to make necessary adjustments.