PUBLICATION

Guidance on De-Identified Protected Health Information Offers In-depth Instruction on Technical Issues [Ober|Kaler]

January 06, 2013

The HITECH Act required the Secretary of Health and Human Services to publish a number of "Guidance" documents to inform the health care industry and its advisors about practical aspects of HIPAA compliance and HITECH implementation. At the end of November 2012, the Secretary published one such document Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act Privacy Rule. The Guidance does not break any new ground, but it does provide a practical instruction on how to take advantage of de-identification to make the essence of large data bases of protected health information available for secondary use. These information caches are increasingly valued for public or other population-based analytics purposes such as epidemiology and private purposes such as business planning and, in some instances, marketing or fund-raising.

The HIPAA privacy rule has always provided for two methods for de-identifying protected health information and thus removing it from the scope of HIPAA: a safe harbor and a process of manipulation approved by a statistician or other person with knowledge and experience in rendering data not individually identifiable. The newly published guidance provides additional details regarding the applicability and practical use of both methods.

The safe harbor requires masking or eliminating eighteen different data elements, some direct such as name or address and some indirect such as medical record numbers. The eighteen listed data elements include a catch-all "any other unique identifying number, characteristic, or code," so providers seeking to use this method still must exercise some independent judgment in identifying the data elements for masking or removal. The safe harbor method also requires that the covered entity not have "actual knowledge" that the information, once the elements are eliminated or masked, could be used alone or in combination with other information to identify an individual subject. The Guidance provides examples of the catch-all "other unique identifier," generally in either technical terms, such as an "Identifying Code" ("a value that is derived from a non-secure encoding mechanism") or in common sense terms, such as an identifying characteristic (the patient is the "current President of State University"). Perhaps more interesting is the Guidance's discussion of when a covered entity has actual knowledge that the de-identified data could be used in combination with other information to identify the individuals.

A lot has changed since the privacy rule was published. Sophisticated methods have developed for taking data from a variety of sources and comparing characteristics and linking to an acceptable degree of probability the data in the various sets to one individual. The Guidance makes it clear that knowledge that such sophisticated statistical linkages are possible does not meet the actual knowledge threshold. Actual knowledge, in the words of the Guidance, must be "clear and direct." Any lesser standard "would not be consistent with the intent of the Safe Harbor method, which was to provide covered entities with a simple method to determine if the information is adequately de-identified."

The bulk of the Guidance deals with the more technical "expert determination" method of de-identification. If a person with appropriate knowledge and experience with generally accepted statistical principals and methods determines that the risk is "very small" that the information could be used, alone or in combination with other

information reasonably available to the anticipated recipient, to identify the individual and documents that determination, the information may be considered de-identified and outside the scope of HIPAA. The key differentiating point is, of course, that de-identification under a method approved by a qualified expert can retain some of the data elements that would otherwise have to be masked or eliminated under the safe-harbor method.

The theme of the Guidance with reference to expert determinations is flexibility. No specific professional degree or certification is required for an individual to be considered a de-identification “expert”: “From an enforcement perspective, OCR would review the relevant professional experience and academic or other training . . . as well as actual experience of the expert using health information de-identification methodologies.” Similarly, a “very small” risk is defined by the expert, based on the ability of the anticipated recipient to identify an individual. A detailed work-flow is provided to illustrate how this determination may be made. Interestingly, the Guidance specifically approves the possibility of providing a single recipient with interlocking data sets, which if combined, could clearly identify the individuals, for example, one set contains detailed geocodes and generalized age ranges and the other set contains generalized geocodes and “fine-grained age” such as days from birth, if there is a data use agreement that prohibits the recipient from combining the two data sets.

Much of the discussion in the Guidance of the expert determination method of de-identification is technical, at least from a lay-person's point of view. For example, the section on approaches that an expert may use to mitigate the risk of identification of an individual in health information contains a three-page, extensively footnoted discussion of “suppression,” “generalization” and “perturbation” as possible methods. Readers are cautioned that “Table 6, as well as a value of k equal to 2, is meant to serve as a simple example for illustrative purposes only.” A good indication as to whether one may be considered an “expert” for de-identification purposes may well be whether one considers this detailed, technical guidance both accessible and useful.

Ober|Kaler's Comments

In the end, while not necessarily tasty fodder for most lawyers and privacy officers, the Guidance is quite valuable for experts, especially statisticians and others who will be asked to develop and certify an Expert Determination. As secondary uses of data derived from protected health information increase in public and private importance, the flexibility of the expert determination approach (even given its technical complexity), as opposed to the mechanical nature of the safe harbor, may become increasingly valuable as a means of de-identifying data while maintaining important data points for statistical analysis.