KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

De-Identification of Confidential Health Data: Policies & Issues

Deven McGraw
Director, Health Privacy Project
June 25, 2012

Health Privacy Project at CDT

- Project's aim: Develop and promote workable privacy and security policy solutions for personal health information.
- Workable means privacy as *enabler*. Privacy is not the endpoint – it is the means to building trust in data sharing, including for secondary purposes.
- Without privacy protections, people will engage in "privacy-protective behaviors" to avoid having their information used inappropriately.
- Public trust in electronic data infrastructures depends on keeping data confidential & protected from unauthorized access, use and disclosure

cdt KEEPING THE INTERNET OPEN·INNOVATIVE·FREE

Federal (HIPAA) Policy on "De-identification"

- "De-identified data" = data that meets HIPAA standard for deidentification
- Data that meets the HIPAA de-identification standard is not PHI and largely not regulated by HIPAA
- De-identification standard = no reasonable basis to believe the data can be used to identify an individual (45 CFR 164.514(a))

De-identified Data under HIPAA

- Two methods may be used to de-identify:
  - Statistical method requires someone with statistical expertise to determine that the risk is “very small” that the information, on its own or in combination with other reasonably available information, could be used by an anticipated recipient to identify an individual (164.514(b)(1))
  - Safe harbor requires the removal of 18 specific data elements; in addition, data holder must not have actual knowledge that the data, either alone or in combination with other data, could identify an individual.(164.514(b)(2))
- Entities may assign a code to allow de-identified data to be re-identified as long as code is not shared. (164.514(c))

De-identification Policy Challenges (1)

- Risk is contextual:
  - What other data does the data recipient have access to
  - What is the recipient's motivation to re-identify or use inappropriately
- HIPAA safe harbor approach assumes a static environment and concludes that data can be deemed to raise a very low risk without consideration of this context
- However, HHS wanted to create an easy-to-use, cookbook approach that would encourage de-identification

De-identification Policy Challenges (2)

- Ensuring very low risk of re-identification – particularly through safe harbor standard – may be getting more difficult due to increased availability of data
- Statistical method for de-identification is meant to be flexible over time – but robustness depends on quality of statistical analysis
- Safe harbor (removal of 18 specific data elements) may lose its potency over time
- De-identification means low risk, not no risk - we still need policies to address

Less Identifiable = Less Risk

- Zero risk cannot be achieved – but this does not mean all data present equal risk
- De-identifying or removing identifiers from data, or shielding identity through use of technology, provides additional protections for confidentiality and (ideally) preserves utility
- HIPAA requirements to use the minimum necessary amount of information needed to accomplish a particular purpose arguably applies to data identifiability
- Data minimization is a widely accepted fair information practice principle

CDT Recommendations on De-identification

- Review de-identification safe harbor standard on regular basis to bolster its efficacy
  - Expand safe harbors?
- Process for vetting statistical de-identification
- Strengthen accountability for re-identification of de-identified data
- Consider whether health data should ever be made publicly available (vs. solely through data use agreement)

cdt KEEPING THE INTERNET OPEN • INNOVATIVE • FREE

CDT Recommendations on De-identification(2)

- Designate de-identification “Centers of Excellence”
- Consider increasing public transparency re: uses of de-identified data
- Require recipients of de-identified data to adopt security protections

Encouraging use of less identifiable data

- HIPAA permits use of fully identifiable data where “less identifiable” data would suffice
  - Health care operations, for example (quality assurance, credentialing, business analytics)
- De-identified data is often not useful for research, public health, and quality purposes because too much data is removed
- Limited data set (LDS) preserves more data – but still rigid and may not be useful for many important purposes

CDT on “Deidentification”

- White Paper (June 2009): “Encouraging the Use of, and Rethinking Protections for, De-Identified (and “Anonymized”) Health Data”: http://www.cdt.org/files/pdfs/20090625_deidentify.pdf
- Policy Post (shorter version of above) (6/26/09): http://www.cdt.org/policy/stronger-protections-and-encouraging-use-de-identified-and-anonymized-health-data
- iHealthBeat Perspectives (even shorter) (7/30/09): http://www.ihealthbeat.org/perspectives/2009/anonymized-medical-data-protects-privacy-improves-care.aspx