National Committee on Vital and Health Statistics

Advising the Secretary of Health and Human Services on health information policy since 1949.

Themes from the May 24-25, 2016 Hearing

"De-Identification and the Health Insurance Portability and Accountability Act (HIPAA)"

NCVHS Privacy, Confidentiality, and Security Subcommittee

De-identification Hearing Objectives:

1. Increase awareness of current and anticipated practices involving protected health information such as the sale of information to data brokers and other data-mining companies for marketing and/or risk mitigation activities;
2. Understand HIPAA's de-identification requirement in light of these practices, and
3. Identify areas where outreach, education, technical assistance, a policy change, or guidance may be useful.

National Committee on Vital and Health Statistics

Advising the Secretary of Health and Human Services on health information policy since 1949.

Expert Testimony

Micah Altman, PhD	Director of Research, MIT Libraries
Daniel Barth-Jones, MPH, PhD	Asst. Professor of Clinical Epidemiology, Mailman School of Public Health, Columbia University
Cavan Capps, CISSP	Big Data Lead, U.S. Census Bureau
Sheila Colclasure, MA	Privacy Officer, Acxiom
Jeptha Curtis, MD	American College of Cardiology
Michelle De Mooy	Deputy Director, Privacy and Data Project, Center for Democracy and Technology
Yaniv Erlich, PhD	Assistant Professor of Computer Science, Columbia University
Simson Garfinkel, PhD	Information Scientist, Information Technology Laboratory, National Institute for Standards and Technology

Kimberly S. Gray, JD	Chief Privacy Officer, Global IMS Health
Bradley Malin, PhD	Assoc. Professor of Biomedical Informatics & Computer Science and Director, Health Data Science Center, Vanderbilt University
Jacki Monson, JD	Vice President, Chief Privacy and Information Security Officer, Sutter Health
Jules Polometsky, JD	CEO, Future of Privacy Forum
Ashley Predith, PhD	Executive Director, President's Council of Advisors on Science and Technology
Ira Rubinstein, JD	Senior Fellow, Information Law Institute, New York University School of Law
Vitaly Shmatikov, PhD	Professor, Dept. of Computer Science, Cornell University
Cora Tung Han, JD	Senior Advisor, Federal Trade Commission, Bureau of Consumer Protection

Themes

1. There is a “privacy-data collision” and de-identification is a key factor.
2. The science of de-identification is ahead of current practice.
3. De-identified data may be re-identified.
4. De-identification practice requires understanding of risk assessment and risk mitigation.
5. Policies and procedures for managing de-identified data are weak.
6. Guidance and regulations must address genomics and other emerging issues.

1. The "privacy-data" public policy collision

"The most pressing issue we are facing today is how to protect individuals' privacy and dignity while enabling all the useful services, science and research made possible by large-scale data analysis." Vitaly Shmatidov

- De-identification challenges vary by data set and use.
- Need to consider de-identification as part of a spectrum of disclosure limitations, techniques and tools.
- Trade off between utility and the extent of de-identification
- Promoting public use data, but it carries no restriction on re-identification
- Who is responsible for certifications when multiple data sets are integrated?
- Individuals want greater control over access and use

2. The science of de-identification is ahead of current practice .

De-identification "pragmatist and formalists have shown little inclination to engage in fruitful dialogue…and find ways to resolve their differences." Ira Rubinstein

- HIPAA Techniques
  - Safe Harbor suppresses eighteen kinds of identifying information and requires that the entity attest that it does not have actual knowledge that the data could be re-identified.
  - The Expert Determination or Statistical Standard is used less frequently because it is more expensive and there is a shortage of experts.
- Other Techniques - Field swapping, the addition of "noise" to a data set, use of synthetic data sets.
- Deeper training is needed in this area; need to standardize curricula for de-id practitioners and academia
- Robust research is needed with translation and spread into practice of what is learned.
- Special difficulties of narrative and unstructured data

3. De-identified data does not stay de-identified

"Data that appears correctly de-identified today might be identifiable tomorrow based on a future data release."
Simson Garfinkel

- Need a workable definition of re-identification is needed
- Re-identification research confirms some rate of re-identification.
- Re-identification science has weaknesses that can lead to policy shortcomings.
- Safe Harbor may not be sufficient and may need other restrictions such as adding a provision to contracts and Data Use Agreements prohibiting downstream re-identification
- Consideration of "context" of collection, opt=in/opt-out
  - Under what circumstances should individuals be notified of a new use?
- Real v. perceived harm, re-identification does not necessarily mean actual harm

4. De-identification practice requires understanding of risk assessment and mitigation

"We have a 'fuzzy notion' of the capabilities and motivation of the 'anticipated recipient' for data." Bradley Malin

- De-identification challenges vary by data set and use.
- Risk means something different to everyone
- What constitutes a "small risk" under HIPAA?
- Need more robust risk assessment tools, methods
- Even responsible recipients can be hacked.

5. Policies and procedures for managing de-identified data are lagging.

> "We should expect a continuous evolutionary process in which new privacy protection technologies are developed concurrently and in harmony with the new ways of using the data that enhance human well-being. Vitaly Shmatikov

- Sound policies and procedures are as important as methods and technology; as in security practice, sound process is key.
- Education based on use cases
- Lifecycle/Data Stewardship – appropriate controls for each stage of life cycle
- Educate IRBs on de-identification science
- Lack of resources to audit de-identification

6. Guidance and regulations need updating for a more complex information ecosystem

> "FIPPs should be used as a set of levers, which can be modulated to address big data by relaxing the principles of data minimization and individual control while tightening requirements for transparency, access and accuracy. Jules Polonetsky

- Enhanced de-identification strategy and mechanisms need to be formally evaluated.
- De-identification conflicts with data exchange regulations
- Address de-identification of genomic data; revisit it regularly
- Sector specific laws; data in different environments are treated differently.
- Policy incentives to improve application of de-id techniques and tools
- Expert determination method needs greater transparency and charity around best practices

Areas for Recommendations

For HHS/OCR

1. Support de-identification and re-identification research
2. Provide practical guidance and tools on how to assess risk of re-identification
3. Reinforce that de-identification is one approach to data protection that needs to be bolstered by other mechanisms such data sharing agreements, consent/authorization, encryption, security and breach detection.
4. Establish some form of oversight for de-id data holders and advance legal penalties for unauthorized re-identification.

Areas for Recommendations

5. Establish a clearinghouse for de-identification best practices.
6. Clarify policy for gnomic data sharing.
7. Improved education on de-ID or anonymization.

For covered entities

1. Strengthen data release and data sharing policies and processes.
2. Strengthen accountability for vendors and business associates