Guidance Specifying the Technologies and Methodologies That Render Protected Health Information Unusable, Unreadable, or Indecipherable to Unauthorized Individuals for Purposes of the Breach Notification Requirements under the American Recovery and Reinvestment Act of 2009. [1]

May 21, 2009

Robinsue Frohboese
Acting Director and Principal Deputy Director
Office for Civil Rights
United States Department of Health and Human Services

Dear Ms. Frohboese:

Since 2002, the Markle Foundation's Connecting for Health Initiative has brought together leading government, industry and health care experts to accelerate the development of a health information-sharing environment to improve the quality and cost-effectiveness of health care. The Center for Democracy and Technology, through its Health Privacy Project, promotes comprehensive privacy and security policies to protect health data as information technology is increasingly used to support the exchange of health information. The Center for American Progress has played an active role on a range of health care issues, including a greater focus recently on health IT issues. We, along with Consumers Union, Health Care For All, the National Partnership for Women & Families, Clay Shirky, [2] Jeff Jonas,[3] Deirdre Mulligan,[4] and Peter Swire,[5] jointly submit these comments in response to the guidance published by the Department of Health and Human Services (HHS).

Section 13402 of the American Recovery and Reinvestment Act of 2009 (ARRA)[6] imposes a new duty on HIPAA covered entities[7] and their business associates to notify affected individuals when there has been a breach of protected health information (PHI) that has not been secured through the use of a technology or methodology that renders the information unusable, unreadable, or indecipherable to unauthorized individuals. HHS has recently issued guidance on this issue, providing an exhaustive list of encryption and

[1] Federal Register/Vol. 74, No. 79/April 27, 2009.
[2] Technical Lead for Markle Connecting for Health.
[3] IBM Distinguished Engineer; Chief Scientist, IBM Entity Analytics.
[4] Assistant Professor, UC Berkeley School of Information.
[5] C. William O'Neill Professor of Law at Moritz College of Law of the Ohio State University.
[6] Pub. L. 111-5, 123 Stat. 115 (2009).
[7] Entities subject to the requirements of the privacy and security regulations under the Health Insurance Portability and Accountability Act (HIPAA).

destruction technologies and methodologies that meet these criteria for the purposes of this provision.[8]

HHS has issued this guidance at a critical moment. Through ARRA, Congress and the Administration have made an unprecedented public investment in health IT to improve quality and reduce costs in the health care system. The success or failure of this endeavor will depend in no small measure on the degree to which patients and consumers, as well as health industry stakeholders, trust that health information will be protected from inappropriate use and disclosure. Building and maintaining this trust will require an ongoing commitment from policymakers and industry stakeholders to develop, implement and enforce effective privacy and security policies. Approaches to privacy and security will need to evolve as new protective technologies and threats emerge.

This guidance centers on just one component of a full set of privacy and security policies needed to foster public trust and support health IT efforts. The selected technologies and methodologies listed in the guidance do not by themselves guarantee a secure electronic health information system. Instead, these tools should be viewed only as part of a set of comprehensive information-sharing policies that include strong oversight and accountability mechanisms, adoption of trusted network design characteristics and the implementation of core privacy principles. The core principles must include openness and transparency, purpose specification and minimization, collection and use limitation, individual participation and control, data integrity and quality, security safeguards and remedies.[9] Only by combining all of these elements will we achieve a comprehensive framework that limits unnecessary exposure of personal health information and reduces the risk of inappropriate or unintended uses and disclosures of health data while permitting appropriate sharing and use of health information to ensure patient-centered care, improve health quality and reduce growth in health care costs.

I. Overview of Recommendations:

First and foremost, we want to emphasize that protecting health care data requires vigilant oversight and active monitoring. Methods of securing data that work one year may fail the next, as attackers become more sophisticated and as target data sets proliferate. The privacy risks associated with breached data depend on the data analysis tools and other, related sources of data an attacker can use to access or re-identify breached information. As both tools and available data increase, protective policies, rules and technological solutions must also evolve over time.

The creation and maintenance of an appropriate list of techniques for making data unusable is critical for two reasons. First, individuals should be notified if their health

[8] HHS also issued a request for information to inform the agency's upcoming breach notification regulations on which we have submitted separate comments.

[9] For more information, see the Markle Foundation, *A 21st Century Approach to Private and Secure Health Information Sharing and Improved Quality of Care*, (September 2008), http://www.connectingforhealth.org/resources/20080822_policy_brief.pdf.

data is at risk. Thus, breach notification exclusions should be limited to data that are materially resistant to access by unauthorized parties. Second, the exclusion should provide an incentive for entities holding health care data to use state-of-the-art practices and technologies to protect personal health information. While breach notification rules can be helpful to patients, the notification itself is reactive in nature and does not prevent an actual breach. The real value of identifying strong methodologies for breach notification exclusion is that it encourages the use of those methodologies, ultimately offering greater data protection.

Consistent with this view, and as explained in more detail below, we:

- Support the strong encryption and data destruction standards currently included on the list of technologies and methodologies that render protected health information unusable, unreadable or indecipherable;
- Recommend the addition to the list of accepted technologies and methodologies a one-way hash function, which is particularly useful for comparing population-level data sets without unnecessarily exposing patient data;
- Urge HHS not to add the limited data set to the list of technologies and methodologies because it does not approximate the level of protection achieved through strong cryptography;
- Ask HHS to emphasize that these technologies and methodologies do not supersede or are not a substitute for the requirement to use the minimum amount of data necessary to accomplish a particular purpose;
- Recommend that HHS carefully examine the unintended consequences of adding device access safeguards and drives protected by biometric access protocols before proceeding in this area;
- Recommend that HHS, as part of its study of the HIPAA de-identification standard,[10] consider whether de-identified data should remain outside of regulation under HIPAA, including with respect to breach notification;
- Urge HHS to expressly commit to annually reviewing this guidance and set forth a process for doing so; and
- As part of this annual review, recommend HHS use threat profiles to evaluate the potential of policies, technologies and methodologies to protect and secure PHI.

II. The encryption and destruction standards currently on the list of technologies and methodologies are appropriate at this time.

HHS's exhaustive list of the technologies and methodologies that render protected health information unusable, unreadable, or indecipherable to unauthorized individuals include the following:

- For electronic PHI at rest, data that has been encrypted using a process consistent with National Institute of Standards and Technology (NIST) Special Publication 800-111, Guide to Storage Technologies for End User Devices.

[10] ARRA Section 13424(c).

- For electronic PHI in motion, data that has been encrypted using a process that complies with the requirements of Federal Information Processing Standards (FIPS) 140-2.[11]
- Paper, film or other hard copy media that have been shredded or destroyed such that the PHI cannot be read or otherwise cannot be reconstructed.
- Electronic media that have been cleared, purged or destroyed consistent with NIST Special Publication 800-88, *Guidelines for Media Sanitization*, such that the PHI cannot be retrieved.

We support the inclusion of the items on this list as being strong, current data encryption and destruction standards. We note that encryption need not be expensive, so the technology is accessible even by providers with limited resources.[12]

III. One-way hashing should be added to the list of accepted technologies and methodologies.

One-way hashing, when properly implemented, should be included in the list of technologies exempted from breach notification requirements. The hashing process uses an algorithm to irreversibly convert plain text data into unreadable character strings. This technique can be thought of as a special form of encryption that only works in one direction – data can be encrypted ("hashed"), but it cannot be decrypted. NIST has approved five hashing algorithms that make it computationally infeasible to determine the original data inputs from the hashed data alone.[13]

As noted above, hashing is especially useful for comparing population-level data sets without needlessly exposing patient data, offering critical potential uses in public health, health quality improvement, comparative effectiveness research and performance measurement. Adding the one-way hash to the list of approved methodologies will limit data exposure in population-level research by allowing linking or eliminating duplications across data sets without exposing the underlying identifiable personal health information. Hashing options are low-cost and some NIST-approved algorithms are publicly available, which makes them accessible even for providers and institutions with limited resources.

---

[11] These include, as appropriate, standards described in NIST Special Publications 800-52, *Guidelines for the Selection and Use of Transport Layer Security (TLS) Implementations;* 800-77, *Guide to IPsec VPNs;* or 800-113, *Guide to SSL VPNs,* and may include others which are FIPS 140-2 validated.

[12] For example, the NIST-approved hash algorithm SHA-256 can be downloaded for free from Softpedia: http://webscripts.softpedia.com/scriptDownload/JavaScript-SHA-256-Download-46166.html.

[13] Federal Information Processing Standard (FIPS) 180-3, *Secure Hash Standard*, Pg. iv (October 2008). Specifically, HHS should consider the SHA-2 family of hash functions. There is an effort underway to create a SHA-3 family of hashes but HHS should not delay encouraging the use of one-way hashing in the interim.

Both encryption and one-way hashes can render data indecipherable to persons who do not possess the key. If providers A and B used encryption, they would likely use the key to decrypt each other's data sets and match the plain text to conduct population-level analyses. The added strength of the one-way hash is that it enables the comparison of data sets without needlessly exposing PHI.[14] For example, if health care providers A and B both hashed their data and shared the hashed data with each other, they would be able to find the patients they have in common, but they would not have shared "identifying" information aside from the matching character strings.[15] Matching hashed data sets may also be performed by a trusted third party on behalf of multiple health care entities, adding an additional layer of security.

To qualify for breach notification exclusion, the hashing process should employ one of the five algorithms approved in FIPS 180-3, and the key should be derived using a protected shared key value of sufficient length, per guidance from NIST.[16] At a minimum, the data to be hashed should include the identifiers listed in the HIPAA de-identification rule.[17]

Hashing has imperfections similar to encryption.[18] The data inputs and shared key value must be kept confidential and not disclosed with the data set, just as it defeats the purpose of encryption to disclose the key in the same package as the encrypted data. As a best practice, a new protected key should be issued for each new purpose, sharing party, and data set transfer to reduce the severity of data exposure if the secret key is cracked or

---

[14] Swire, *Application of IBM Anonymous Resolution to the Health Care Sector*, Pg. 11 (February, 2006).

[15] The Record Locator Service (RLS) is another technique for searching for patient records while maintaining anonymity. The RLS acts as an index for sub-networks, enabling entities within that sub-network to know whether other entities hold records related to a particular patient without revealing the contents of the record. *See* Markle Foundation, *Correctly Matching Patients With Their Records*, Connecting for Health Common Framework (April 2006), http://www.connectingforhealth.org/commonframework/docs/P4_Correctly_Matching.pdf.

[16] *See* NIST Special Publication 800-107, *Recommendation for Applications Using Approved Hash Algorithms*, Section 5.2.1 (February 2009). In cryptography, this key is commonly called a salt.

[17] 45 CFR 164.514(b)(2).

[18] We are aware that NIST and the National Security Agency have expressed concern about the use of hashing with respect to health care data primarily due to "man-in-the-middle" attacks. But the real issue is making sure that individuals at the edges can only view the data of individuals that they already have information on. Currently, threats to health data privacy are much likelier to come from the loss of whole data sets, as with stolen laptop or hard drives, or from unauthorized access by legitimate employees, as with curious file clerks accessing VIP records, than from an attacker looking at files in transit.

inadvertently disclosed. Also, encryption is always changing as computational power grows: what was considered unbreakable ten years ago is now often easily defeated. It is therefore important that approved cryptography standards—for both encryption and hashing-- be updated periodically as HHS revisits the breach notification issue annually.[19]

The likelihood of failure for different security systems is often unknown because the system can be attacked by a variety of means. A determined attacker with massive computational power may find a gap in the defenses. The most robust security systems will use a layered approach to data security that incorporates hashing, encryption, and limits on the amount of data that is disclosed. Data at rest should be encrypted or hashed, and the digital connection along which the data is shared should also be encrypted. Whenever possible, PHI should be subjected to a one-way hash prior to being shared with another party. Parties that seek data, like researchers, should specify the purpose to which the data will be put and collect only data needed to accomplish that purpose.

IV. The limited data set should not be added to the list of accepted technologies and methodologies.

We urge HHS not to add the limited data set to the list of technologies and methodologies that can be used to secure data. Significant questions have been raised about whether the de-identification standard, which is even further stripped of patient identifiers, provides sufficient anonymity to data.[20] Re-identification research revealed nine years ago that more than three-quarters of the population can be uniquely re-identified through publicly available population registers using zip code, gender and date of birth.[21] Those items are

---

[19] ARRA Section 13402(h).

[20] One group of pharmacy researchers tested a set of data de-identified under the safe-harbor method for potential for re-identification. Because the de-identified data contained many unique combination opportunities, the researchers determined that "anticipated [data] recipients, such as physicians, nursing agencies, pharmacies, employers, and insurers…could re-identify their members in the study data set with a moderately high expectation of accuracy." Clause, Steven L., et al, "Conforming to HIPAA Regulations and Compilation of Research Data, *American Journal of Health System Pharmacy*, (61) (2004), 1025-1031, at 1029. See also Bradley Malin and Latanya Sweeney, "How (Not) to Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-identification to Evaluate and Design Anonymity Protection Systems," *Journal of Biomedical Informatics* 37 (2004), 179-192; Latanya Sweeney, "Computational disclosure control, a primer on data privacy protection," (2001) available at http://www.swiss.ai.mit.edu.proxy1.library.jhu.edu/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf; Virginia de Wolf et al., "Part II: HIPAA and Disclosure Risk Issues," 28 *IRB: Ethics and Human Research* 6-11 (2006).

[21] Malin and Sweeney, *Determining the Identifiability of DNA Database Entries.* Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc. Nov 2000; 537-541.

not among the identifiers removed from a limited data set under HIPAA.[22] It makes little sense to grant safe harbor treatment to the more identifiable limited data set, which is still considered to be PHI under HIPAA. Information is increasingly difficult to classify as "identified" or "de-identified," particularly as it is copied, exchanged, or recombined with other information. With rapidly evolving technologies and databases, it is more appropriate to describe a spectrum of "identifiability," rather than a binary classification of information as identifiable or not. The question could then become not whether de-identified information might be made re-identifiable, but rather which entities would be able to re-identify the information, how much effort they would have to expend, and what limits are placed on their doing so[23].

Therefore, absent some other safeguard, patient information in limited data set form is not unusable, unreadable or indecipherable, and it does not approximate the same level of protection as strong cryptography (i.e., encryption and hashing). Including the limited data set as another safe harbor equivalent would likely discourage health care entities from encrypting or hashing data, even though encryption and hashing offer much stronger protection.

Further, although use of a limited data set is preferable to using fully identifiable data, encouraging the use of a limited data set by adding it to the list of secure technologies will drive entities to use more data than is necessary in some cases. How much data are needed to accomplish a particular purpose is contextual. If the limited data set is the default, covered entities and business associates will have little incentive to exercise greater discipline in implementing the minimum necessary standard and use only the data needed to accomplish a particular purpose. Granting "safe harbor" status to the limited data set would then put more data at risk, not less.

V. HHS should emphasize that these technologies do not substitute for complying with collection and use limitation principles.

We also urge HHS to ensure that the incentive to use encryption or one-way hashing to protect information is not interpreted in a way that overrides or undercuts the requirement under HIPAA's minimum necessary standard that health care entities use the least amount of data needed to accomplish a particular purpose. A policy of data limitation and purpose specification should be strongly favored as a general matter. This "minimum necessary" principle is already central to the HIPAA Privacy Rule. Health care entities should be encouraged to share the minimum amount of data necessary to complete the immediate task for which the data is needed. Information gatherers, such as researchers, should likewise clearly specify the purpose to which the data will be put and collect only enough data to accomplish that purpose. Collecting, maintaining and sharing any data

[22] 45 CFR 164.514(e)(2).
[23] *See* Markle Foundation, *CT4: Limitations on Identifying Data*, Connecting for Health Common Framework for Networked Personal Health Information (June 2008), http://www.connectingforhealth.org/phti/reports/ct4.html.

that is more extensive than necessary increases both the risk and severity of breach. HHS should make clear in its guidance and regulations that the use of encryption or one-way hashing does not override an entity's obligation to specify their purpose and use only the minimum amount of data necessary to accomplish that purpose.

VI. Device access safeguards should not be added to the list of accepted technologies and methodologies.

Device access safeguards should not be included on the safe harbor list without a more thorough examination (ideally conducted by, or in conjunction with, NIST) of whether they offer the same level of protection as encryption or hashing the data itself. Instead, health care entities should consider device access restrictions as another layer of security, in combination with encrypting or hashing, and sharing the minimum amount of information needed for a specified purpose.

HHS should similarly view the use of biometrics as a device access key with caution. User authentication is certainly an important component of any security system, but strong authentication does not offer protection equivalent to strong encryption.[24] Biometric authentication, including fingerprint-based access to USB drives, will create new security problems for those individuals whose biometrics are used. According to a 2009 study, 45% of data breaches in the health care sector occur as the result of lost hardware, like laptops or USB drives.[25] Because so many breaches result from lost or stolen devices, if biometrics become a standard protection mechanism for devices in the health sector, then biometric patterns may become a top target for unauthorized parties seeking to gain access to health information. Once compromised, an individual's biometric identifier cannot be replaced.

---

[24] Rather than examine biometrics in isolation, HHS should review authentication holistically. Authentication raises multiple difficult issues that require careful review before any one technique, including the use of biometrics, should be considered for a breach notification exclusion. The Markle Foundation addressed many of these issues as part of its Connecting for Health Common Framework. *See*, Markle Foundation, *Authentication of Consumers*, Common Framework for Networked Personal Health Information (June 2008), http://www.connectingforhealth.org/phti/#guide. *See also*, Markle Foundation, *Authentication of System Users*, Connecting for Health Common Framework, (April 2006), http://www.connectingforhealth.org/commonframework/docs/P5_Authentication_SysUsers.pdf.

[25] Curtin and Ayres, *Using Science to Combat Data Loss: Analyzing Breaches by Type and Industry*, Interhack, Pg. 36 (April 21, 2009), http://web.interhack.com/publications/interhack-breach-taxonomy.pdf.

VII. HHS should reconsider the appropriateness of the notification exemption for de-identified data.

We acknowledge that Congress applied the breach notification provisions to unsecured "protected health information" as that term is currently defined in the HIPAA Privacy Rule.[26] Thus, the guidance makes it clear that the notification requirements do not extend to de-identified data as defined in HIPAA. As HHS may be aware, de-identification, particularly through the removal of specific categories of identifiers (commonly referred to as the safe harbor method), does not guarantee anonymity.[27] Consequently, consumers and patients may still be at risk if de-identified data is breached – risks that would be minimized if such data were subject to breach notification requirements that provided covered entities an incentive to protect it at rest and in motion with encryption or one-way hash. Ideally, we believe that, at a minimum, if there is evidence that de-identified data has been breached in plain text form, individuals whose information was part of the data set should be notified.

There may be limits on what HHS can do in this guidance to address this issue. However, we urge HHS to use the de-identification study mandated by Congress,[28] as well as its general HIPAA oversight authority, to assess the potential for re-identification of de-identified data and to ensure that entities that disclose or access such data are held accountable for complying with baseline privacy and security protections. HHS should also explore ways to require recipients of de-identified data to execute data use agreements where they contractually commit not to re-identify the data subjects.

VIII. HHS should use threat profiles to evaluate the potential of policies, technologies and methodologies to protect and secure PHI.

HHS should consider using threat profiles or models to evaluate whether particular policies, technologies and methodologies sufficiently protect health information (including, but not limited to, assessments of whether other technologies and methodologies should be added to the list in the future and whether those already on the list should remain). Threat profiling or modeling involves ongoing assessment of the various threats to health data that exist in the environment; considering whether current policies and security requirements effectively mitigate those risks; and if not, appropriately modifying such policies and requirements to ensure they are sufficiently robust to constitute a comprehensive framework of privacy and security protections for health information.

---

[26] ARRA Section 13402(a).

[27] *See* Comments of Latanya Sweeney to the Department of Health and Human Services On "Standards of Privacy of Individually Identifiable Health Information," (April 26, 2002), http://privacy.cs.cmu.edu/dataprivacy/HIPAA/HIPAAcomments.html; see also footnote 23.

[28] ARRA Section 13424(c).

Of note, NIST has used this process to evaluate the effectiveness of safeguards on electronic voting systems.[29] In the health care context, HHS could, for example, evaluate current policies, technologies and methodologies with respect to whether they meet threats posed by unscrupulous users, loss or theft of data on unencrypted data stores, assembly of identified records from nominally de-identified records, and the rapid innovations in tools and technologies used to break into once-secure systems.

We recommend HHS work with NIST and security experts both inside and outside the health field to develop and implement a threat profile process, adjusting it over time to reflect changing threat patterns. As noted above, the array of threats and protections for data is always changing. Thus, a threat profile for a particular policy or technology would likely become outdated over time. HHS should not rely on any one profile indefinitely; rather, HHS should update threat profiles to keep up with the latest challenges and incorporate the latest developments in privacy and security policy and technology.

IX. Conclusion

In summary, we:

- Support the inclusion of the current technologies and methodologies on the list
- Encourage the addition of PHI that has been hashed through one of the five approved algorithms in FIPS 180-3, where the key is derived using a protected shared key value, and both the key and the underlying data inputs are segregated from the hashed data set;
- Urge HHS not to add the limited data set to the list of technologies and methodologies that can be used to secure data;
- Ask HHS to state clearly that these new security standards do not supersede the obligation to use the minimum necessary amount of data to accomplish a particular purpose;
- Recommend against adding device access safeguards to the list without further review by, or in conjunction with, NIST;
- Request that HHS re-evaluate the current standards for "de-identified" information by assessing the risks presented by the potential to re-identify data; and
- Recommend HHS adopt threat profiles to annually assess the robustness of privacy and security policies and technology practices.

[29] *See* NIST, *Developing an Analysis of Threats to Voting Systems,* (October 2005), http://vote.nist.gov/threats/papers.htm.

We appreciate the opportunity to provide these comments on the HHS guidance on technologies and methodologies that render health data "unusable, unreadable or indecipherable to unauthorized individuals." Please let us know if you have any questions or need further information.

Sincerely,

Markle Foundation

Center for Democracy & Technology

Center for American Progress

Consumers Union

Health Care For All

National Partnership for Women & Families

Jeff Jonas
IBM Distinguished Engineer
Chief Scientist, IBM Entity Analytics.

Deirdre Mulligan
Assistant Professor, UC Berkeley School of Information

Clay Shirky
Technical Lead, Markle Foundation, Connecting for Health

Peter Swire
C. William O'Neill Professor of Law at Moritz College of Law of the Ohio State University