Comments of the World Privacy Forum to the Federal Trade Commission

regarding

Health Breach Notification Rulemaking, Project No. R91102

Via www.regulations.gov

Federal Trade Commission
Office of the Secretary
Room H-135 (Annex M)
600 Pennsylvania Avenue, NW
Washington, D.C. 20580

June 1, 2009

Re: Health Breach Notification Rulemaking, Project No. R91102.

The World Privacy Forum appreciates the opportunity to comment on the Federal Trade Commission's proposed rulemaking, the Health Breach Notification Rule. The notice of proposed rulemaking appeared in the Federal Register on April 20, 2009 at 74 Fed. Reg. 17914–17925.

The World Privacy Forum is a non-partisan, non-profit public interest research and consumer education organization. Our focus is on conducting in-depth research and analysis of privacy issues, in particular issues related to information privacy, health privacy, and financial privacy. More information about the activities of the World Privacy Forum is available at our web site, <http://www.worldprivacyforum.org>.

I. Comments about Scope

The rule needs to more fully describe its jurisdictional scope. The Commission is aware that Internet activities and Internet service providers operate internationally. Some of the companies providing Internet services, and specifically, Personal Health Record (PHR) services, are multinationals. It is no longer possible to assume that the United States is isolated from the rest of the world and constitutes a single, clearly separate jurisdiction. Both businesses and consumers operate routinely in international space, and rules of U.S. agencies need to address how they apply to consumer and business activities that take place in whole or in part in that space when possible.

We recognize that the jurisdictional issues are challenging. But despite the difficulty, we believe it would be best to further define jurisdiction, if possible.

For example:

- Will the rule apply to a PHR vendor that is located in or that operates in another country?
- Will the rule apply to a PHR vendor located in another country that provides services to an American citizen who lives in that other country and whose records come from both American and foreign sources?
- Will the rule apply to a PHR vendor located in another country whose services include health records about treatment provided in the United States to a foreign national during that foreign national's visit to the United States?
- Will the rule apply to a foreign PHR vendor that obtains consent for disclosure of records from American citizens and that operates wholly offshore?
- Will the rule apply to an U.S.-based PHR vendor who provides services only to foreign nationals?
- Will the rule apply to a foreign PHR vendor that provides services only to citizens of other countries?

The answer to these questions is crucial to implementation of the proposed rule. The answers may also affect how some requirements are carried out. For example, if a Personal Health Record vendor has a significant number of clients living outside the United States, will notices to the media have to be provided to foreign media and in other languages?

II. Comments about Definition of Personal Health Record

The FTC's background comments state:

> "Proposed paragraph (d) defines a ''personal health record'' as an ''electronic record of PHR identifiable health information on an individual that can be drawn from multiple sources and that is managed, shared, and controlled by or primarily for the individual.'' This language is substantively identical to the definition of personal health record in section 13400(11) of the Recovery Act."

We are concerned with the possibility that this definition will allow PHR vendors to evade the regulatory requirement. The Commission needs to address the possibility that vendors will deliberately shape their activities to avoid compliance.

There are several parts of the definition that are readily subject to abuse. These are:

A. "....can be drawn from multiple sources."

This definition can be readily worked around. What if there is only one source for the health information? Suppose that all of a patient's records are drawn from a Health Maintenance

Organization (HMO). If a PHR vendor offers services only for record of that HMO, then there are no multiple sources and the regulation will not apply.
To correct this issue, the Commission needs to emphasize that:

> **1)** The data subject is a source of information, and
>
> **2)** if the source of patient records to the PHR vendor itself obtains health information from multiple sources, then the PHR vendor also draws records from multiple sources.

These small changes could make a big difference.

B. "....is managed, shared, and controlled by or primarily for the individual."

It would not be difficult for a PHR vendor to define its services so that they are not *for the individual* but for the family of the individual.

Similarly, a PHR vendor could define its services so that they serve other purposes (including its own goals of making a profit, providing services for third parties, or providing services to covered entities or other providers of information).

To remedy the potential for mischief here, the Commission needs to clarify that this part of the definition should be read expansively so that the provision of any services to the individual who is the subject of the record brings a PHR provider under the definition.

III. Comments about Definition of PHR-Related Entity

The background comments of the FTC regarding the proposed rulemaking state:

> Proposed paragraph (f) defines the term ''PHR related entity'' to cover the three types of entities set forth in clauses (ii), (iii), and (iv) of section 13424(b)(1)(A) of the Recovery Act. First, the definition includes entities that are not HIPAA-covered entities and that offer products or services through the website of a vendor of personal health records. This definition is substantively identical to the statutory language but also clarifies that HIPAA-covered entities are excluded. This clarification is consistent with the coverage of section 13424, which requires a study and report on the ''Application of Privacy and Security Requirements to Non-HIPAA Covered Entities.'' Examples of entities that could fall within this category include a web-based application that helps consumers manage medications; a website offering an online personalized health checklist; and a brick-and-mortar company advertising dietary supplements online. Consumers interact with entities in this category by clicking on the appropriate link on the website of a vendor of personal health records.

We note with great interest the language that extends the definition to entities that *offer products or services through the website of a vendor of personal health records*. We think that these

entities belong within the scope of the rule, and expressly support the FTC's inclusion of these entities in the rule.

In the proposed rule, the FTC further notes that the category could include: "online applications through which individuals, for example, connect their blood pressure cuffs, blood glucose monitors, or other devices so that the results could be tracked through their personal health records. It could also include an online medication or weight tracking program that pulls information from personal health records." We agree with all of these examples and support their inclusion in the final rule.

A. Recommendation to expressly include search engines accessible through PHR vendors in the rule

One further example of a related entity that should be expressly mentioned by the Commission in the commentary is a search engine or search tool.

A search engine that is accessible through the website of a PHR vendor should be covered by the rule. For example, some large PHR vendors offer users the ability to do health information searches while they are logged in to their PHRs. The content of those searches may or may not be covered under the vendor's PHR-specific privacy policy. In some cases, those searches may be covered under a broader, more general privacy policy applicable to the search engine itself. So for example, if a consumer is using a PHR at a web site, then also uses the available search engine box on or near the PHR page to conduct a search on Huntington's disease, the contents of that search should be included in this rule because it was conducted at the PHR page or area.

The requirement is essential because many search engines collect and retain identifiable information, such as the content of previous searches, IP addresses, and cookies. The ease with which individual consumers (and their medical conditions and doctors) can be identified through the searches they have made at a search engine is well illustrated by the 2006 AOL Data Breach. In a complaint filed by the World Privacy Forum with the Commission following that breach, we stated:

> On August 9, 2006 the New York Times published an article on its front page identifying one of the AOL users by name and with a photo. The reporters were able to identify the AOL user by analyzing the search queries AOL revealed in its posting of the Spring 2006 data. A Washington Post story published August 15, 2006 described a confidential annex submitted in an Electronic Frontier Foundation FTC complaint that identified 10 to 15 individuals from the data. The World Privacy Forum, in its research, was able to identify a probable minor by name and geographic location from the data and tie that data to photos and videos of the probable minor. The World Privacy Forum was also able to identify an individual's medical situation, home address, telephone number, and pediatrician from the data to a high degree of confidence.[1]

[1] World Privacy Forum Petition to the FTC regarding AOL, August 16, 2006. <http://www.worldprivacyforum.org/pdf/WPF_FTCcomplaint8162006fswp.pdf>. Paragraph 37 (footnotes omitted).

We are not asking the FTC to bring all health-related searches at all search engines under the rule. We are asking the FTC to bring searches conducted from PHRs and PHR web sites under the rule in the same way that the other tools are brought in.

B. Recommendation to include advertisers in the rule

The FTC's statements about third parties in its background discussion clarifies that it is very easy for a PHR vendor to provide the opportunity for the sharing of patient information with third parties by allowing those third parties to provide products and services to consumers through links on the vendor's website. In some PHR business plans, these third parties are advertisers who may in fact learn the identity or partial identity of the consumer in some cases. For this reason, we recommend that advertisers should also be included in the FTC's background description.

IV. Comments about Law Enforcement Delayed Notice

The background comments of the FTC regarding the proposed rulemaking state:

> Finally, proposed section 318.4(c) allows breach notification to be delayed upon appropriate request of a law enforcement official. The proposed burden of proof and law enforcement provisions are substantively identical to sections 13402(d)(2) and 13402(g) of the Recovery Act.

We have several recommendations regarding this section.

First, the rule does not define *law enforcement official*. Does it include a school crossing guard, wetlands regulation enforcement officer, or weight and measures examiner? We urge the Commission to tighten this definition.

Second, the proposed rule does not do a clear job of explaining how the requirement for notice without unreasonable delay will fit together with the possibility of a delayed notice. We are concerned about timing issues. For example, how long can an entity with a breach notification obligation wait to provide notice to individuals? Can it wait sixty days while it searches for a law enforcement officer who might ask for a delay? We think that the Commission needs to address who qualifies as a law enforcement official for the purpose of the rule and how the requirement for prompt notices meshes with the possibility of a law enforcement-requested delay.

Third, the Commission needs to define a clearer standard that limits the ability of law enforcement to impose an unlimited delay. We suggest that any notice delayed over 30 days be reported to the Commission and that the Commission reserve the authority to overrule the delay.

We are concerned, among other things, that a law enforcement officer who has no role in pursuing a criminal investigation related to the breach could determine that notice "would

impede a criminal investigation." At the very least, the rule should say: "would impede a criminal investigation being conducted by the law enforcement official who requested a delay."

V. Comments about Substitute Notice

The background comments of the FTC regarding the proposed rulemaking state:

> As to the first method of substitute notice, the Recovery Act states that the posting should appear for a period determined by the Commission and be ''conspicuous.'' The Commission believes that six months is an appropriate time period for posting of the notice and has so specified in the proposed rule. Requiring a six month posting will ensure that individuals who intermittently check their accounts obtain notice, without being unduly burdensome for businesses.

If the PHR vendor knows which individuals have failed to receive notice, then general home page notice is not the proper remedy. The recommendation is premised on the theory that these individuals will visit the website and see the notice. If those individuals visit the website, then the vendor should be required to provide actual, direct notice to the individuals through the website, for example, notice to the customers' log-on screen or to their account area.

The general home page notice is fine, and we support the six-month time frame the FTC has proposed. But it is no substitute for actual, direct, web-based notice to each individual for whom previous notice has failed. If those individuals do not return to the website, no form of website notice will work. But if they do return, then direct notice on a case-by-case basis will surely be more effective than a general notice that is much more likely to be ignored.

For example, we have seen mass breach notices that were posted on web sites conspicuously for a week or a day, then disappeared into a link available from a privacy policy, and so forth. Even when companies had the ability to notify users directly, they did not, and instead chose to post just one mass notice.

"Conspicuous" can have many interpretations by web site operators. A non-avoidable notice directly to each users' account would be a useful and we think important adjunct to a more general notice.

VI. Comments about Access and Acquisition of Information

In its background discussion, the FTC described examples of access and acquisition, and outlined the presumption of its proposed rule:

> Thus, the proposed rule creates a presumption that unauthorized persons have acquired information if they have access to it, thus creating the obligation to provide breach notification. This presumption can be rebutted with reliable evidence showing that the information was not or could not reasonably have been acquired. Such evidence can be

> obtained by, among other things, conducting appropriate interviews of employees, contractors, or other third parties; reviewing access logs and sign-in sheets; and/or examining forensic evidence.

We strongly support this thoughtful analysis of access and acquisition, and believe the FTC has found the right approach on this issue. We also support the specific scenarios the FTC described.

VII. Comments about "De-identified Data" and "PHR Identifiable Information"

With respect to the definition of what constitutes PHR identifiable health information, we have comments in two general areas, de-identified data, and the fact of having an account as PHR identifiable information. We note that issues surrounding de-identified data are complex and challenging.

A. De-identified data (HIPAA standard)

In the background, the FTC states:

> ...if there is no reasonable basis to believe that information can be used to identify an individual, the information is not "PHR identifiable health information," and a breach notification need not be provided. For example, if a breach involves information that has been "de-identified" under HHS rules implementing HIPAA, the Commission will deem that information to fall outside the scope of "PHR identifiable health information" and therefore not covered by the proposed rule. The HHS rules specify two ways to de-identify information: (1) if there has been a formal determination by a qualified statistician that information has been de-identified; or (2) if specific identifiers about the individuals, the individual's relatives, household members, and employers are removed, and the covered entity has no actual knowledge that the remaining information could be used to identify the individual.

The cite the FTC referred to in this first part of the definition is for de-identified sets under HIPAA. The term *de-identified* is expressly defined by the HIPAA Privacy Rule, and the definition is lengthy, much more so than indicated in the paragraph above.

Our first recommendation is that the text incorporate the *full list* of what is removed in HIPAA de-identification instead of just a cite to the definition. This is needed because some smaller commercial entities that are unfamiliar with the HIPAA de-identification standards may remove just the few elements listed in the background discussion from data, thinking it is the entire standard. That these few data elements are not inclusive of all of the data elements that must be removed to fully de-identify under HIPAA should be made more clear.

Our second recommendation is that any entity that has had a breach of HIPAA de-identified information should still be required to make notice to the FTC and to the general public about the breach. The notice would not need to be made to the individuals, but notice would still need to be

made to the public so that a track record regarding data handling may be assessed by the public and regulatory authorities. Also, the public should have the ability to monitor for re-identification themselves if they believe their de-identified data has been breached at a particular company.

Our third recommendation is that the entity that has breached the de-identified data must monitor for any signs of re-identification. If there are signs that re-identification has occurred, then steps will need to be taken to provide much more robust public notice of the breach and level risk. See section B, 1-5 below for a more detailed discussion of this. One of the problems with individual breach notification for data that has been de-identified is that the data has to be re-identified, and that introduces an entirely new set of challenges.

We note that there is a privacy interest in de-identified data, even data de-identified to HIPAA standards, because it is not actually "anonymous" data.[2] For example, Dr. LaTanya Sweeney has famously re-identified patients' medical records that had been scrubbed of identifiers and sold to industry as anonymous records. Dr. Sweeney re-identified records by combining the supposedly "anonymized" records with publicly available voter database records which she purchased for 20 dollars.[3]

B. Other kinds of "non-individually identifiable information"

In the background comments, the FTC notes that:

> There may be additional instances, where, even though the standard for de-identification under 45 CFR 164.514(b) is not met, there is no reasonable basis to believe that information is individually identifiable. The Commission requests examples of such instances.

We believe that what the Commission is discussing here is partially de-identified data, or limited data sets. As defined by HIPAA, a limited data set is Protected Health Information that excludes certain direct identifiers such as name, street address, telephone number, biometric identifiers (including photographs), and other information. Limited data sets are "de-identified" to some extent, but not to the stricter standards set out under 45 CFR 164.514(b). Limited data sets are not anonymous.

As such, limited data sets are considered Protected Health Information under HIPAA because the data can still be re-identified in some circumstances. Limited data sets, under current regulation, may only be disclosed by HIPAA-covered entities for research, public health, or health care

---

[2] See EPIC, *Supplemental Brief for the Electronic Privacy Information Center as Amicus Curiae and 16 Experts in Privacy Law and Technology in Support of Defendant – Appellant Kelly A. Ayotte, urging reversal*, August 20, 2007. *IMS Health Inc. v. Ayotte*, 490 F. Supp. 2d 163, 165-74 (D.N.H. 2007). <http://epic.org/privacy/imshealth/epic_ims.pdf>.

[3] LaTanya Sweeney, *k-anonymity, A Model for Protecting Privacy*, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570. See Figure 1 for how the data overlay of "anonymous" and identifiable data works to identify the data set members. < http://privacy.cs.cmu.edu/people/sweeney/kanonymity.pdf>.

operations, and must be released only subject to a formal Data Use Agreement that complies with HIPAA limited data set data use agreement standards. For more, see CFR § 164.514(e), <http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/adminsimpregtext.pdf>.

The health care entities covered in this rulemaking may be using or releasing limited data sets. The commercial entities covered in this rulemaking may be using something very similar to limited data sets in terms of reduced direct identifiers attached to the data. However, the commercial entities covered in this rulemaking will not have the restrictions of HIPAA-defined limited data sets, for example, commercial entities may have partially de-identified data similar to a limited data set, but they may be using it without any formal data use agreements.

Limited data sets and other partially de-identified data sets pose many challenging privacy issues regarding data breach notification. For the purposes of these comments, we will use the term *limited data sets* to apply only to HIPAA-covered entities, and we will use the term *partially de-identified data sets* to refer to similarly de-identified data sets held by commercial entities not subject to HIPAA.

First, there are privacy interests in limited data sets and partially de-identified data sets, as the HHS proposed data breach guidance notes.[4] Among the greatest risks is that of re-identification, with the potential for unanticipated and unauthorized secondary uses of the data. As technologies have matured and computing power has increased, the risk of re-identification of information contained in limited data sets or partially de-identified data sets is no small matter.

The AOL data breach of consumer information was of a partially de-identified data set. Yet some of those consumers could still be identified.[5] A Netflix data set of its customers' movie preferences was released to the public for purposes of improving the Netflix movie recommendation system. Netflix described its data set as anonymized. Researchers, using the "anonymized" Netflix data, were able to combine the Netflix data with information from other databases and subsequently reliably re-identify significant portions of the data set.[6]

While re-identification of partially de-identified data sets and limited data sets is not always trivial, it is getting much easier. As computing power increases and the amount of public information readily available on individuals expands, the ability to re-identify partially de-identified datasets is a significant risk factor for data breaches of limited data sets and other similarly partially de-identified data. It is again worth noting that even data fully "de-identified" according to strict 45 CFR 164.514(b) HIPAA standards can be re-identified.[7]

---

[4] "Because PHI in limited data set form is not completely de-identified, the risk of re-identification is a consideration in determining whether it should be treated as unusable, unreadable, or indecipherable to unauthorized individuals for purposes of breach notification... ." p. 74 Fed. Reg. 19009.

[5] World Privacy Forum, *Petition to the FTC regarding AOL*, August 16, 2006. <http://www.worldprivacyforum.org/pdf/WPF_FTCcomplaint8162006fswp.pdf>. Paragraph 37, for example.

[6] Arvind Narayanan and Vitaly Shmatikov, *How to Break the Anonymity of the Netflix Prize Dataset*, March 17, 2008, < http://arxiv.org/pdf/cs/0610105v1>.

[7] See LaTanya Sweeney, *k-anonymity, A Model for Protecting Privacy*, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570. LaTanya Sweeney authored a study in which 87 percent of individuals contained in the U.S. Census data could be uniquely identified by date of birth, ZIP Code, and gender. See Latanya Sweeney, *Uniqueness of Simple Demographics in the U.S. Population*, Carnegie Mellon

We acknowledge that many casual recipients of a laptop with a limited data set or a partially de-identified data set would have neither the incentive nor the capability of re-identifying the records. Policy should not be made on the sole assumption that all lost data will end up in the hands of an individual or entity with the skills, equipment, information, and expertise necessary to re-identify data. But neither should policy be made on the sole assumption that no one at any time will have the interest or the means to re-identify the data found in a de-identified data set. As time goes on and technology improves, re-identification of limited data sets and other partially de-identified data sets may become child's play.

Because these risks exist, we do not support a straightforward statement of non-notification for partially de-identified data or limited data sets. We note that in the area of health care-related data, the risk of medical identity theft is of particular concern in a breach situation.

We have some recommendations.

1. In the event of a breach of partially de-identified data or limited data sets, vendors or covered entities should be required to conduct a risk assessment specifically for risk of re-identification

Requiring a thorough risk assessment of the likelihood for re-identification of the breached data is crucial here, and would be based on the amount of data lost (both limited data set and other data that may have been lost with it), the currency of the data, any degree of encryption or coding, and other factors.

2. Monitoring by vendor for evidence of re-identification of data sets

After a breach, the vendor that breached a data set that was partially de-identified should be required to regularly assess whether any evidence on re-identification appears following the breach.

Because the vendor may or may not have the ability to know the identities of the users, it may not be able to evaluate on its own all of the available evidence without assistance. Consultation with the originator of the data set or other parties relevant to the data can be required for a reasonable period of time following the breach, perhaps up to a year. The goal would be to find any evidence of re-identification of the data.

3. In the event of re-identification, notification should be made with controls

Any significant harm to an individual because of a breach of a limited data set or a partially de-identified data set will occur as a result of re-identification of that data. **If there is a significant likelihood that re-identification has, will, or may occur, then individuals should be notified of the breach.**

University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-WP4. Pittsburgh: 2000.

The process for breach notification for partially de-identified data sets can involve more complexity than for notification for other kinds of breaches, and should be done with strict controls, as discussed in D below.

We note that a standard of notification only in the event of re-identification would be inappropriate for data that is already identifiable, where any unauthorized disclosure to any recipient immediately affects the privacy interest of the data subject. We also note that many partially de-identified data sets will likely require individual notification due to risk. But much will depend on the circumstances of the breach.

4. If data sets are to be re-identified to make breach notification, the data should be re-identified with strict controls.

If records must be re-identified for the purpose of making breach notification, the re-identified records should be segregated and used only for the breach notice and follow up. If notification or re-identification tasks are outsourced to a third party, there must be a data use agreement that specifies the rules for maintenance, use, disclosure, storage, and destruction of the records.

There may be some circumstances where a third-party vendor without a direct relationship to the consumer has to make breach notification to individuals. The re-identified individuals' data should not be sold or used for marketing purposes by the company, or subject to any uses beyond breach notification.

5. Mandatory reporting of limited data set or partially de-identified data set breaches to the FTC and the public

If limited data sets or other kinds of partially de-identified data sets are not going to be included in individual breach notice, then we still urge mandatory reporting of the breach. "Mass" or "public" reporting of breaches involving limited data sets is an important step to add, particularly in a digital environment.

Whether or not individual breach notices are required, **all limited data set breaches or partially de-identified data set breaches should be reported to the FTC. The information about the breach should be posted on the FTC web site and made available to the public.**

Consumers should know when a vendor has breached partially de-identified data. The reason for this is the privacy interest in this data, and the potential for re-identification.

C. Fact of having an account as "PHR Identifiable information"

The FTC has included in its definition of PHR identifiable information "the fact of having an account with a vendor of personal health records or related entity, where the products or services offered by such vendor or related entity relate to particular health conditions." (Emphasis original).

The World Privacy Forum supports the FTC's inclusion of this information as PHR Identifiable Information. However, we also believe that HIV status and mental health conditions should not be the only conditions to which this definition applies. In our reading of the FTC language, we believe the FTC intends that the rule applies to HIV, mental health, *and* any other unique or specific condition.

We are concerned that some may read the rule to apply only to the two examples mentioned. Because of this, we request that the rule should make it even more clear that the language includes but is not limited to these specific categories.

D. PHR identifiable health information including information that relates to name and credit card information

The FTC states that:

> First, because the definition of "PHR identifiable information" includes information that relates to the "past, present, or future payment for the provision of health care to an individual," the proposed rule covers breaches of such information.

We support and agree with the FTC's inclusion of this data in breach notification. Further, it is in congruence with a final rule already in place, the Red Flag Rule.

*****

The World Privacy Forum appreciates the opportunity to offer comments. Please contact us with any questions.

Respectfully submitted,

Pam Dixon
Executive Director
World Privacy Forum
www.worldprivacyforum.org
760-436-2489