Comments of the World Privacy Forum and the Electronic Frontier Foundation

Regarding

Department of Health and Human Services, Agency for Healthcare Research and Quality, National Health Data Stewardship Request for Information

VIA email to *steward@ahrq.hhs.gov*

P. Jon White, MD
540 Gaither Road
Rockville, MD 20850

August 9, 2007

The World Privacy Forum and the Electronic Frontier Foundation are pleased to have this opportunity to offer comments to the Agency for Healthcare Research and Quality (AHRQ) in response to its June 4, 2007, Federal Register notice.

The World Privacy Forum is a non-partisan, non-profit public interest research organization.[1] Our focus is on conducting in- depth research and analysis of privacy issues, including issues related to health care. The Electronic Frontier Foundation (EFF) is a nonprofit organization that works to protect civil liberties, privacy, and consumer rights in the digital age.[2] Founded in 1990, EFF has more than 13,000 members throughout the United States. EFF is concerned that the digitization of health records will unnecessarily increase privacy and security risks to personal medical information, and that public/private partnerships in the area of electronic health records will be unduly shaped by commercial interests at the expense of patient privacy.

I. General Comments about the proposed National Health Data Stewardship Entity

We are deeply concerned about the general trend in HHS to view national databases and large networks of personal health information as a solution to many different problems. In AHRQ's June 4 Request for Information regarding a National Health Data Stewardship, AHRQ contemplates a "public/private," national database of health care data as a solution for various health purposes. AHRQ has not offered sufficient justification for this database, nor specific information supporting the utility of this project.

We are skeptical about the proposed AHRQ National Health Data Stewardship project because the amount of data is – notwithstanding the power of modern computers – unmanageable to a

[1] <http://www.worldprivacyforum.org>.
[2] < http://www.eff.org>.

significant degree. Many existing large computerized record systems with information on most Americans – and we point here to credit bureau records as a good example – are filled with inaccuracies, cross-matched individuals, out-of-date information, and incomplete data. To continue with our example, the credit reporting system is also characterized by a procedure for data correction that does not work well and that does not always prevent the reentry of data already determined to be wrong. The lives of some individuals are affected adversely in a material way every day by the use of these records. We note that the credit bureau system is regulated, and that the Fair Credit Reporting Act/ FACTA gives data subjects certain rights, including rights of correction and deletion of false or incorrect information. We note that the proposed AHRQ system does not appear to contemplate comparable legislation for any of the likely data subjects, including health care professionals and organizations. Even if HIPAA applied to the system, HIPAA's correction and deletion rights are not as broad as those provided for credit records in the FCRA.

The presumption that these massive amounts of personal data can be collected, linked, stored, used, and disclosed in a manner that is fair to the data subjects and that properly protects their privacy and other interests is not one to which we subscribe. The inability of our society to manage enormous volumes of personal data fairly is a constant. It does not matter how valuable the stated objective of the database may be. The realities of actual data management make those objectives only partially attainable at best, with those partial benefits often only available after years of effort and great expenditures. Whether the delayed benefits ever catch up to the current and future costs (adjusting both for the time value of money) is always a question.

Further, any new compilation of personal data raises the serious possibility that the data will be used for secondary purposes. The use of data collected for one purpose for a wholly independent purpose is such a common event that it has its own terminology: *data creep*. This is one of our principal concerns, and we will return to it later in these comments.

As a preliminary matter, we think that it would be useful to the discussion and debate if AHRQ could point to any existing model for the proposed National Health Data Stewardship that manages comparable volumes of data, that includes data about comparable numbers of individuals and health care providers, that is of comparable complexity in terms of goals and analysis. We certainly do not expect to find a model in the American health care world. We note, for example, that by comparison, the HHS CMS systems would, in the end, likely be much smaller than the proposed AHRQ National Health Data Stewardship entity or system. It would inform decision makers and stakeholders if they could identify a fair, functional, and cost-effective parallel activity from another sphere.

If there is nothing of comparable scope, then that alone suggests the need to proceed with extra caution. Reports from the Government Accountability Office and from agency Inspectors General are filled with examples of ambitious new federal databases that cost huge amounts of money and failed to get out of the starting blocks.

II. Critical Policy Issues Relating to the Proposed Program

Our concerns regarding the proposed National Health Data Stewardship entity focus principally on the privacy consequences of the National Health Data Stewardship activities. Because of the shortness of time, we are able to offer only a few preliminary observations now. As the debate on this subject continues and expands, we will have more to say in the future, and we expect to be included in stakeholder conversations about this system.

A. Questions about ownership and management of the proposed national health database

One of the significant questions that must be answered is: *who will run the database*? This question brings forward extremely complex issues relating to privacy concerns. We see three broad possibilities, each posing a different set of difficulties.

If a federal agency runs the database (or even if it only directs the maintenance of the database as a government function), the Privacy Act of 1974 will apply. Like all privacy laws, the Act is a mixture of good, bad, and other protections for individuals. We note, however, that Privacy Act data is generally available for law enforcement use and for several other purposes specified in subsection (b) of 5 U.S.C. § 552a. We raise the question whether the use of the data for any of these purposes is consistent with the objective of the National Health Data Stewardship. If a new database serves to expose personally identifiable health data to new uses and disclosures, much more attention needs to be paid to the rights and remedies of individuals. The Privacy Act of 1974 may not offer suitable protections for individuals in this proposed new context.

Another possibility is that the database may be operated by an entity that is subject to the HIPAA privacy and security rules. Whether this is a likely or even possible outcome is far from clear. We observe here too in passing that the HIPAA privacy protections are a mixed bag from a privacy perspective. The inclusion of patient data in a database far removed from the original record keepers may expose individuals to greater legal, bureaucratic, and personal hazards, even if HIPAA applies.

The third possibility is that the new National Health Data Stewardship will be operated by an entity or entities that are not subject to the Privacy Act of 1974, HIPAA, or any other existing health or other privacy law. State health privacy laws could be irrelevant to the maintenance of a patient health database not operated by health care providers or insurers. From a privacy perspective, this outcome is likely to be far worse than the others are.

We could carry on this analysis at greater length, but the point should already be clear. Basic choices about the structure and operation of the National Health Data Stewardship can have enormously different consequences for privacy. What is needed is a complete and independent analysis of the choices and the consequences by someone without a financial stake in the outcome.

If a National Health Data Stewardship can be justified at all, we would like to offer a guess that the best choice here would be for the data to be locked up against secondary uses through specific legislation that defines all permissible data processing activities for identifiable data.

Whether our guess is correct certainly remains to be seen, and the analysis we just suggested would likely offer an answer.

B. Issues relating to identifiability

Identifiability is a notion that is in flux, in that it is becoming more and more difficult to de-identify data thoroughly. Some of the issues posed by AHRQ in the Federal Register notice encompass data aggregation, and this leads us to the identifiability issue. One way to balance privacy and other interests has traditionally been to aggregate, encrypt, or de-identify data in a way that avoids or minimizes the possibility of identification of individual patients.

While the use of de-identified data is an extremely attractive solution, it is a particularly complex one. We remind AHRQ that there are no clear existing standards for when data is actually de-identified. We challenge AHRQ to answer the question: what constitutes de-identified data? We would like to see an answer to this question using nationally or internationally accepted and peer-reviewed standards arrived at through a formal process. We are well aware of the HIPAA de-identification standards, but this standard is just one de-identification standard among many, and there is no determination that HIPAA would even apply to this program.

We caution that the enormous amount of personal information available from public and private sources means that the realm of truly non-identifiable personal information is shrinking every day. We do not believe that AHRQ should casually assume that coding, encryption, or even aggregation would provide a broad or permanent solution to privacy concerns raised by a National Health Data Stewardship. See, e.g., Computer Science and Telecommunications Board, National Research Council, *Who Goes There? Authentication through the lens of Privacy*, at 5 (2003). (noting that while public-key cryptography "is often touted as a virtual panacea for . . . authentication and confidentiality challenges," "implementation and deployment details are key to . . . effectiveness, security, usability, and privacy protection.")

We also do not believe that identifiability is a fixed matter – what is identifiable is changing every day as technologies advance and as networks make previously non-identifiable data identifiable. As a starting point for analysis of identifiability issues, we refer you to *Opinion 4/2007 on the Concept of Personal Data* issued recently by the European Union's Article 29 Working Group.
<http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf>.

C. Determining the process by which privacy will be protected in the system

How privacy issues will be raised and resolved in this program is an important part of what needs to be discussed and addressed prior to any program implementation. We support two methods of considering privacy. As the outlines of the National Health Data Stewardship become clearer, we may have more suggestions.

Establishment of a full-time, independent privacy officer: The World Privacy Forum favors the establishment of a full time, independent privacy officer for any major personal data activity.

A National Health Data Stewardship is likely to qualify as both large enough and consequential enough to support a privacy officer. The benefits of a privacy officer will be most fully realized if appointment comes at the earliest possible stage rather than at the end when most structural and institutional decisions have already been made. The privacy officer should be well qualified, and should not have bias either for or against technology. In short, this individual needs to be able to be a genuine steward of privacy interests. This individual should not have conflicts of interest with technology vendors or other relevant institutions.

Privacy impact assessments (PIAs) that go beyond what the law requires: PIAs should be required for significant activities affecting personal privacy, and this program qualifies. Each PIA should be published for public comment, and the comments should be thoughtfully considered by decision makers. The E-Government Act of 2002 requires federal agencies to prepare a PIA under specified conditions. Whether or how the Act may apply to a National Health Data Stewardship should not determine whether a PIA is completed. We recommend that one or more PIAs *with public comment periods* be an integral part of the planning process for any National Health Data Stewardship.

We also recommend that the PIA exceed the statutory requirements of the E-Government Act. The PIA could be conducted by the privacy officer or by a third party with suitable experience and independence from project planners and others with a conflict of interest. The need for independence is crucial. Too often, PIAs are prepared by project managers, contractors, or others with too much of a stake in the project to be objective. For a project of the enormous size and consequence of a National Health Data Stewardship, the PIA process may require public hearings, mandatory public responses from program decision makers, and an iterative process that utilizes multiple reports and responses as the project moves along toward design and implementation.

III. Issues relating to data security and quality

If a database of the proposed magnitude is built, some automatic challenges and difficulties may be expected. We are all familiar with the medical data breaches that have impacted literally tens of millions of individuals. This National Health Data Stewardship system, due to its size and scope, will be subject to an extraordinary level of risk, perhaps even an unprecedented level of risk. We will discuss just a few of these risks.

A. Medical Identity Theft

As the World Privacy Forum has noted in its widely-cited 2006 report on medical identity theft, medical data is among the most sought-after and valuable data in the world.[3] Criminals can steal millions of dollars from public and private insurers using just a few thousand medical identities. As the criminals commit their crimes, they alter patient files, often introducing changes to victims' health records that are potentially life-threatening. Medical identity theft is a problem quite similar in terms of scope to the asbestos problem prior to its recognition. Unfortunately, the

[3] *Medical Identity Theft: The Information Crime that can Kill You*, May 3, 2006, World Privacy Forum. < http://www.worldprivacyforum.org/pdf/wpf_medicalidtheft2006.pdf>.

health care sector has not yet addressed the medical identity theft problem. We pose to AHRQ that it will be necessary to grapple with this issue at some point in time, either now, or later, and that it would be highly preferable and better for patient health care quality to grapple with the medical identity theft issue now.

There are three significant medical identity theft risks for this program:

- Patients who are victims of medical identity theft can have inappropriate and even potentially life-threatening errors in their medical files archived by the proposed AHRQ program with no clear plan of access, correction, or specific risk mitigation for issues stemming from medical identity theft.

- If there is a data breach of the AHRQ database, patients who have been compromised in the breach need robust and practical recourse. They will need to be informed of the breach in a timely manner, they will need to be able to access and correct their files, and they will need to access a complete audit trail to prove the facts of the case. They may also need services such as credit monitoring, which the government should provide. All too often we have seen breaches announced, followed by subsequent announcements that "no incidents have occurred." Unfortunately, medical identity theft is extremely challenging to detect through common means such as credit monitoring, and it can take up to two years or more to come to patients' attention. Innovative systems of notice and monitoring may be needed to fill the gap in existing methods. Data breach of this system and proper notification of victims will need to be accomplished carefully due to these risks.

- Insider access to the files. Most medical identity theft occurs because a trusted insider with appropriate access to health care files has made a bad decision and has chosen to abuse their insider access to steal and disseminate health care records. Health care records are among the most prized data in the world for thieves, as their sale results in high-dollar returns with low risk of being caught. . If AHRQ constructs a centralized database of this size and magnitude, AHRQ can be assured that the thieves may work diligently to hack the system, but they will find it much easier to buy off someone who already has access to the system. This needs to be accounted for, planned for, and mitigated in advance. AHRQ may be reluctant to look at insiders as the most significant potential threats. However, by not doing so the agency would raise the risk level to an unacceptably high point.

Because of the high risks medical identity theft poses, we recommend that AHRQ institute risk mitigation specifically for medical identity theft in this program. We note that both patients and health care providers are at risk for medical identity theft. We do not know if the HIPAA security rule will apply to this system, however, no matter what policy structure applies to the proposed AHRQ program, specific medical identity theft risk assessment needs to be conducted.

The difficulties associated with medical identity theft have not been faced by the US government yet. We refer you to the World Privacy Forum FAQ on medical identity theft<

http://www.worldprivacyforum.org/FAQ_medicalrecordprivacy.html> that contains a legal analysis of patients' rights and availability of recourse to recover from the crime under HIPAA and other laws.

B. Insider Access and Access Controls

We have already mentioned insider access in the context of medical identity theft. But there is a broader issue here. There has generally speaking been a security skew toward securing databases against hackers and against technical mischief. The more difficult problem, of course, is securing data against legitimate insider access. In health care fraud cases, time after time after time prosecution records show that it is a clerk or another insider employee with legitimate access who has created the problem.

With a database of this magnitude, it will be extremely difficult, even in a tiered access system, to keep the data totally secure. The temptation of accessing large amounts of data corralled in one place we predict will prove to be too tempting to resist. We do not believe that AHRQ currently has an answer to the fundamental security issues a vast centralized database introduces, because this problem is an extremely challenging one to solve. When contemplated in the context of a national centralized database containing health care files of potential all patients, the risks and challenges soar.

We refer AHRQ to a National Academy of Sciences report discussing the privacy and security implications of nationwide ID systems. We refer you to this report because this report contemplated a centralized national database on the scale of what AHRQ is proposing. The report is quite pessimistic about multilevel access controls, which are often proposed as solutions for managing insider access to data. "A further complication would result if it were decided that different users should be granted different levels of access to the database . . . . This raises query capability, access control, and security issues. . . . It is a significant challenge to develop an infrastructure that would allow multiple kinds of queries, differing constraints on queries (based on who was making them), restrictions on the data displayed to what was needed for the particular transaction or interaction, and varying thresholds for security based on the requirements of the user." Computer Science and Telecommunications Board, National Research Council, *IDs – Not That Easy: Questions About Nationwide Identity Systems* at 42 (2002).

We also caution about any patient identifiers used in the system. The scale of the AHRQ project means that the patient identifier required by HIPAA – and that the public rejected at the first hint of its development and that was followed shortly thereafter by a congressional prohibition – could come back to life. We refer to the National Academies of Sciences' CSTB analysis of single identifiers used in centralized contexts such as the one AHRQ is planning: "if a single identifier is relied on across multiple institutions, its fraudulent or inappropriate use . . . could have far greater ramifications than if used in only a single institution." Computer Science and Telecommunications Board, National Research Council, *Who Goes There? Authentication through the lens of Privacy*, at 10 (2003). Thus, the report recommended that "[a] guiding principle in the design or selection of authentication technologies should be to minimize the linking of user information across systems unless the express purpose of the system is to provide

such linkage. Id. at 11. Moreover, such systems "should be designed to make secondary uses difficult, because such uses often undermine privacy, pose a security risk, create unplanned-for costs, and generate public opposition to the issuer." *Ibid*.

C. Data Migration and Cleansing

We note that data migration and cleansing will be a profound quality of information issue in this system. Even if there were just one data system in the U.S. for managing health care data, with a system so large, data migration and cleansing issues would still arise. However, given the fractured nature of the data systems, the data migration issues for the proposed system are profound to the point we do not believe they can be readily overcome, if at all, without substantial financial support.

We understand that AHRQ wants to push new standards to mitigate this issue. However, AHRQ should be aware that standards need to be created in a deliberative and open public manner, and cannot be rushed. A rush to weak standards in order to facilitate a national database would be a mistake.

D. Data of non-residents

Something that was not mentioned in the Federal Register notice was how to handle data of individuals who are citizens of other countries. This is an issue that many U.S. systems – such as the proposed NHIN – tend to gloss over. However, this proposed AHRQ system will not be able to easily do so due to the fact that the system is intending to store data for potentially many years, and will potentially expose this data to many secondary uses. This system, if it captures the data of citizens of the European Union, for example, could have to face need to abide by EU standards of privacy and security. If it does not, then needed data may not be available. As many countries around the world have stronger privacy laws than the US, the problems of international transfers of health data needed for oversight, management, or research may require special attention.

E. Access and opt-out for consumers

AHRQ has not raised the issue of consumer participation for this activity. It needs to be. Consumers do not like to hear that they have no choice but to allow their data to be stored, warehoused, accessed, and shared without having any say in the matter. This issue could prove to be a significant public stumbling block, especially if secondary uses of the data are likely. We note that in the RFI, data access, correction, and deletion plans by consumers were not addressed or brought up. Consumers should be the *first* concern.

The system may include health information that individuals – or existing laws – consider to be the most sensitive data available. Some of that data – such as substance abuse treatment records – may be totally unavailable to any National Health Data Stewardship. Other records, including genetic, psychiatric, and AIDS/HIV, may also have special rules. How any national system will

sort out the different rights, procedures, and limitations will not be easy to assess. The role of consumer consent also needs to be addressed.

IV. Importance of the Role of Fair Information Practices

AHRQ should be familiar with the concept of Fair Information Practices from its work with HIPAA. It is my hope that the agency, as it seeks to analyze the privacy impacts of a national health database, will use the most robust possible set of Fair Information Practices as a guide. Fair information practices are not invariant, but most codes of fair information practices include these elements:

1.*Collection Limitation Principle*
There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

2.*Data Quality Principle*
Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

3.*Purpose Specification Principle*
The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

4.*Use Limitation Principle*
Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [the purpose specification principle] except:

a) with the consent of the data subject; or
b) by the authority of law.

5.Security Safeguards Principle
Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.

6. *Openness Principle*
There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

7.*Individual Participation Principle*
An individual should have the right:

a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;

b) to have communicated to him, data relating to him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to him;
c) to be given reasons if a request made under subparagraphs(a) and (b) is denied, and to be able to challenge such denial; and
d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.

8.*Accountability Principle*
A data controller should be accountable for complying with measures which give effect to the principles stated above.
Compliance with all of these basic principles is especially important for a system which will contain and disseminate highly personal information of the utmost sensitivity.

All of the fair information practices need to be fully integrated into the system to address the privacy interests of those whose personal information would be contained in the AHRQ databases and systems. Each principle should be looked at and specifically implemented by AHRQ.

V. Role of the privacy community and public dialog

AHRQ is proposing a significant system in its National Health Data Stewardship. We point for reference purposes to our experience of AHRQ's 2006-2007 Health Information Security and Privacy Collaboration (HISPC) project, where some of the participating states initially did not include consumers in the project teams.[4] As a participant in the HISPC process, the World Privacy Forum found the initial lack of consumer input from some of the state participants to be problematic, especially since one of AHRQ's goals was to "promote stakeholder identification of practical solutions and implementation strategies through an open and transparent consensus-building process."[5] The reality is that there has been scant will to deeply engage all consumers in some of the key health information exchange planning and policy processes out of a fear that consumers' concerns will stop or slow health information exchange.

We point to another significant HHS project, the National Health Information Network (NHIN), where there has also been a disparity of planning that has not represented consumers or patients adequately. The pronounced lack of substantive outreach to and a subsequent lack of input from privacy and consumer groups is marked in the NHIN process. We are aware that there has been outreach to and input from consumer groups representing narrowly focused health care areas. However, there has been a lack of systematic, substantial outreach to the groups focused on

[4] See for example, the Kansas HISPC team, which originally did not include consumer representation. <http://www.csgmidwest.org/About/MLC/AnnualMeeting/2006presentations/braman.pdf> at slide 11, bullets no. 4 and 5.

[5] See AHRQ *Privacy and Security Collaboration* web page, "Goals of the Privacy and Security Project." <http://healthit.ahrq.gov/portal/server.pt?open=514&objID=5562&mode=2&holderDisplayURL=http://prodportallb.ahrq.gov:7087/publishedcontent/publish/communities/a_e/ahrq_funded_projects/rti_public_page/main.html>.

privacy and a lack of focus on reaching the broadest array of *all consumers,* not just groups focused on particular diseases. We are concerned that this proposed National Health Data Stewardship system, which also strives to be a national system aggregating large quantities of patient data, will repeat the problems we have seen with the HISPC and the NHIN processes.

The input of privacy and consumer groups can act as an early warning system of consumer privacy concerns, which can in turn assist in discovering areas in need of further analysis, input, and adjustment early in the planning process. We would like to see AHRQ hold formal public hearings on this system, and we request that any public hearing include a variety of points of view, including robust input from the privacy community. If the National Health Data Stewardship officially moves forward, we request that AHRQ create a neutral and level playing field for participation by all stakeholders by setting aside budget to facilitate participation by a diverse array of non-profit groups; lack of funding often means that non-profits cannot participate in the policy and planning process. The health care industry, vendors, agencies, and others who may seek to benefit financially from a National Health Data Stewardship rarely seem to have resource limitations. And we remind you that the privacy community is not monolithic and has its own diversity of opinions that deserve to be heard along with other perspectives.

Privacy is not a barrier to advancement in science nor is it a barrier to improvements in health care quality. Privacy has too often been shunned by the very professionals and agencies that should be embracing it. Privacy is a core feature of our health care system, not a bug to be evaded, ignored, or overcome. Without adequate attention to privacy, HHS and its partners run the risk of losing the public trust to accomplish new and useful activities that may involve large amounts of health data.

Respectfully submitted,

Pam Dixon
Executive Director, World Privacy Forum

Lee Tien
Senior Staff Attorney, Electronic Frontier Foundation