The New Health Information Architecture:
Coping with the Privacy Implications of the Personal Health Records Revolution

A. Michael Froomkin*
Professor, University of Miami School of Law
Project HealthDesign ELSI Group

The rapid evolution of computer technology combined with ambitious projects to standardize the exchange of health information has produced rapid changes in health information architecture. Unsurprisingly, national health policy has struggled and sometimes failed to keep up. It's hard to deal revolutionary changes in quick succession. And nowhere is this gap between rapidly changing practices and slowly changing policy more noticeable than in the context of privacy policies.

The health community is still coming to terms with an earlier revolution in health information: the introduction first of information technology and, second, of Electronic Health Records (EHRs). As a result of this evolution, "[c]linically rich information is now more readily available, in a more structured format, and able to be electronically exchanged throughout the health and healthcare continuum."[1] But alongside these positive developments have come new issues about privacy, about liability, and about the appropriate level of government regulation, issues which take on new urgency as non-health data about the person increasingly is being linked to health data.

And now, a new revolution is under way: the introduction of Personal Health Records (PHRs) -- often patient-created, and perhaps also patient-centered or even patient-maintained, health records that are stored and queried online and even shared via informal social networks.

Today, health providers and payers largely dominate and control the health information architecture. Patients have a limited role in controlling the release and management of their health information: signing waivers without reading them and serving as a "sneakernet" by hand-carrying records between health care providers who do not have

---

 Support for work on this paper was provided by a grant from the Robert Wood Johnson Foundation® in Princeton, New Jersey.

[1] National Committee on Vital and Health Statistics, Enhanced Protections for Uses of Health Data: A Stewardship Framework for 'Secondary Uses' of Electronically Collected and Transmitted Health Data (2007) at p.4.

more direct or reliable means of communication. Patients have a right to demand a copy of their records and have some say as to whether and with whom to share it. And patients are of course free to form support or other groups that involve sharing and using their health data.

Odds are, however, that in the near future, the center of gravity for health information management will move towards the patient, or towards an online agent of the patient's, or to a patient-initiated (but perhaps not patient-controlled) data repository. Providers and payers will still have an important role, but the entrance of important new data sources and data managers will complicate health information architectures -- and pose difficult new challenges for privacy policies. This should be a serious concern given that existing health privacy policies appear inadequate for even the existing, simpler, health information architecture.

It is important to understand that both the social and technical elements of this new architecture pose challenges for policy-makers. The introduction of PHRs probably facilitates but certainly coincides with the introduction of deep changes in both the creation and management of health data. Increasingly, patients -- or at least devices under their control -- will be the authors of massive amounts of health information, only some of which will actually have treatment implications. (These data also may be linked to the growing amount of non-health-related information available in distributed databases.) The output of monitoring devices will not inevitably have be tied to a PHR – these new devices can beep, phone home, or do their own analysis of the data they capture, but the temptation to tie them in to networked data repositories promises data synergies and faster treatment response. The challenge of sorting this information -- and the danger of missing something relevant[2] -- will add to the pressures on any attempt to build privacy in from the ground up.

Meanwhile, patients will be offered new opportunities to manage their own care by consulting online information sources -- and, eventually, diagnostic tools -- and through the proliferation of online or meshed[3] support groups or user communities.

An additional complication arising from the introduction of PHRs is that they are unlikely to replace EHRs: we are seeing a supplementation rather than a transformation, and it is

---

[2]Cf. Shana Campbell Jones, Joseph McMenamin & David C. Kibbe, *The Interoperable Electronic Health Record: Preserving its Promise by Recognizing and Limiting Physician Liability*, 63 Food & Drug L.J. 75 (2008).

[3]"Meshed" groups are localized digital communities who communicate through an intranet or more commonly via short-range wifi. An example of meshed communications is a group of hand-held devices (e.g. Gameboys) in close proximity communicating via infrared, bluetooth or wifi connections. Although the communications protocols are similar to those used on the internet, depending on how the devices are designed, and whether any of them are linked to the greater internet, the internet need not be involved.

likely that both EHRs and PHRs will co-exist in the information ecosystem. Indeed, it can be argued that a good part of the appeal of PHRs is that they address the failure of EHRs to be networked and shared. To the limited extent, however, that we have adequate privacy rules in place for EHRs it is not at all clear that these suffice to deal with the new challenges arising from fully networked EHRs, much less a world of distributed and shared "PHRs on a stick."

It may be that the most promising possible privacy solutions to the risks posed by PHRs, at least in the near term, are primarily technical rather than legal. Unfortunately, the technologies with the most promise from a health privacy point of view are justly controversial for other reasons. Although they may offer some privacy benefits, the likely gains suffice neither to justify the adoption of widespread Digital Rights Management (DRM) and Trusted Computing, nor of a national patient ID system. However, were either of those proposals to become realities, they may offer at least partial solutions to the privacy vulnerabilities caused by the widespread deployment of PHRs.

1.0 Charting Revolutions

Every patient interaction with the modern health care system involves the creation of new health records. The first digital health records revolution -- the introduction first of information technology and, second, of Electronic Health Records (EHRs) -- were and are primarily concerned with computerizing and rationalizing the flows of these data. The introduction of another revolution -- Personal Health Records (PHRs) -- means that patients and new third parties will increasingly create health care data (or link other data to health data) without the participation of either health care providers or insurers: the patient will create the data, or devices under the patient's control will create a continuous stream of data, or Internet-enabled third parties will create and process the data.

1.1 First Step: Information Technology

At first, health information revolution involved little more than the introduction of computers. Computerizing health records makes them more useful. Even if the information is not shared more widely, computerization makes possible a number of enhancements in patient care such as patient safety alerts and health maintenance reminders. Sharing the information more broadly improves -- or, at least, should improve -- coordination of care among disparate providers; in the best case the information may also be available to first responders and other emergency health care providers.

Most of the first-order benefits of computerized records could be described as doing the same types of things that had been done for years, only better. Thus, while data flows increased, the fundamental architectures of information flows changed only a little -- and mostly in response to the demands of the health insurance/payment systems.

Graphically, the information architecture looked something like a series of these simplified flows, replicated for each doctor or other health services provider:

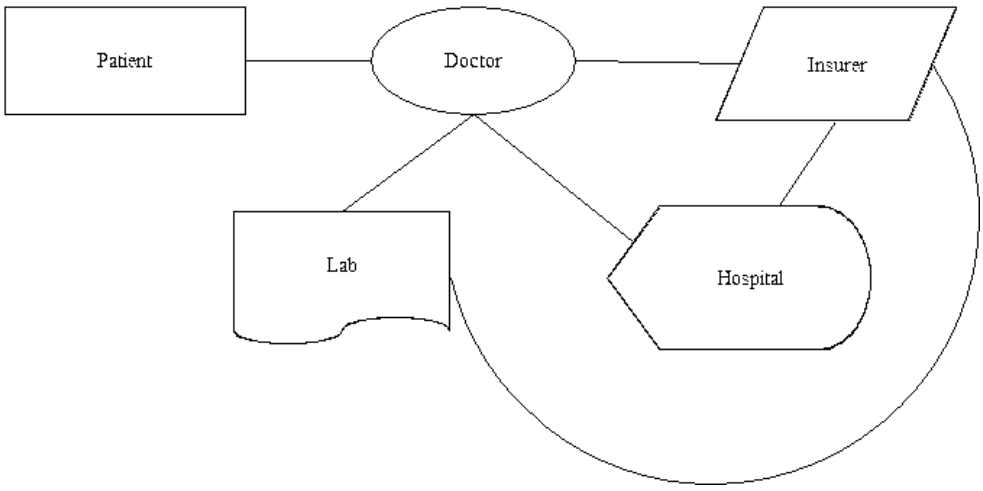

Fig 1:Simplified Data Flow for EHRs

In this information architecture, systems are frequently separate, especially at the provider level. Data can be shared between these otherwise separate systems, particularly via the patient or the insurer, but this is not inevitable. For example, lab reports go to a doctor; the patient knows the test was done, but may not be aware of all the detailed results. The fact of the test is shared with the insurer, but the results need not go to the insurer either. Furthermore, different health services providers tend to be separate from each other. In the ordinary case, there is no information flow between a doctor and a dentist, nor even necessarily between different specialists such as an ENT and an orthopedic surgeon. For one doctor to benefit from data collected by another required either that they have some means of communication or, most commonly, that the patient set one up or deliver the data -- often hand-carrying records from one place to another.

1.2 Second Step: EHRs

Electronic Health Records (EHRs) offer the prospect of rationalizing -- and expanding -- the flow of health-related information. An EHR is intended to be a comprehensive health record about a person consisting of data gathered from multiple sources, not simply from health and payment professionals.

Thus, instead of multiple, essentially separate, copies of the information flow in Figure One, the goal is to have a single, more unified, data architecture, a simplified version of which would look like Figure Two:

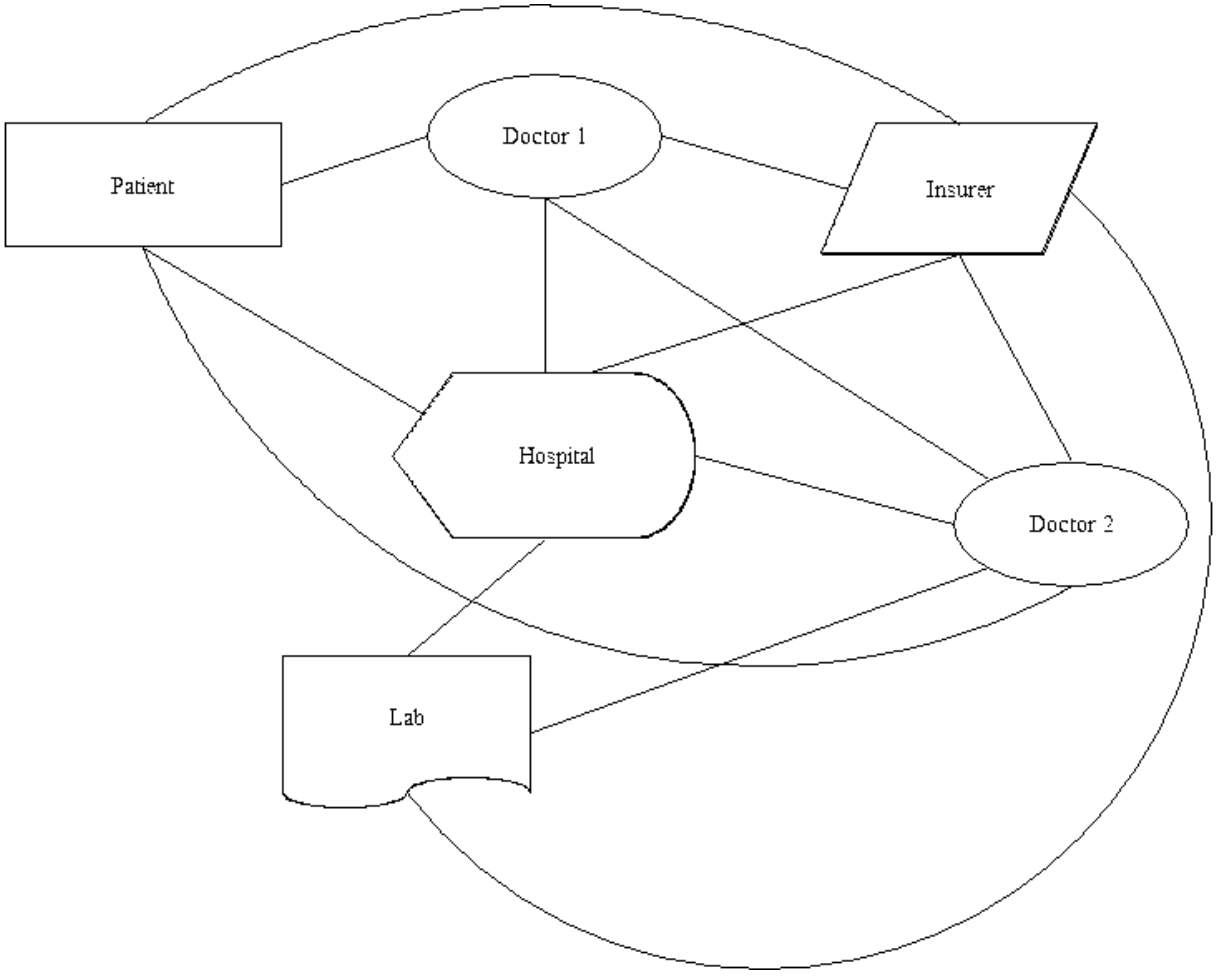

Fig 2: Simplified Data Flow for EHRs

EHRs offer the prospect of substantial efficiencies and better care by better organizing and managing existing data: They change the flow of data, eliminate both duplication and ignorance. Like the computerization of records that predates them and makes them possible, EHRs do not in the main create new data so much as organize vast amounts of existing data in a new way. But the quantity of linked data are potentially vast. The records contemplated,

> would not merely replicate, in electronic form, today's patient record, but could include, in addition to the individual's medical history, other information such as his or her family medical history, as well as genomic,

> pharmacogenomic, and nutrigenomic data, environmental exposures, dietary and exercise practices etc.[4]

Of course, having all this patient data available also means it can be used in new ways.

> It would be the key to "empower individual patients to assume a much more active, controlling role in their own health care; improve access to timely, effective, and convenient care; improve patient compliance with clinician guidance; enable continuous monitoring of patient conditions by care professionals/care teams; and enable care providers to integrate critical information streams to improve patient-centered care, as well as to analyze, control, and optimize the performance of care teams."[5]

Thus, the patient's data becomes not only a way to better serve this patient, but to collect information about the health care delivery system itself:

> For quality measurement, reporting, and improvement, fully automated data collection processes provide for more efficient access to more comprehensive databases for benchmarking, as well as identification of new opportunities for improvement in care delivery. The ability to mine more comprehensive databases makes knowledge discovery more readily available for continuous quality improvement. [Health Information Exchange (HIE)] technologies that enable virtual aggregation of data and enhanced data linkage, such as individual person matching algorithms, support longitudinal data collection to improve future care of an individual and quality outcomes analysis.[6]

Perhaps this new ocean of information also will facilitate the collection of large-scale research data relating to clinical populations:

> Clinical and population research can be strengthened. Identification and participation of candidates for clinical trials across a larger geographic area

---

[4] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 29 (2008), http://www.ced.org/docs/report/report_healthcare2007dcc.pdf.

[5] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 29 (2008).

[6] National Committee on Vital and Health Statistics, Report to the Secretary of the U.S. Department of Health and Human Services on Enhanced Protections for Uses of Health Data: A Stewardship Framework for "Secondary Uses" of Electronically Collected and Transmitted Health Data (October 21, 2007) at 6.

> enables more comprehensive cohorts for testing hypotheses. Health services and other population-based research is aided through the availability of large databases. As a result, hypotheses can be tested or complications detected more rapidly. [7]

Perhaps -- but we are not nearly there yet. Achieving all these disparate goals requires many technical changes in the ways that the US health care systems handles patients and their data. Among the requirements are:

- Unique and consistently used patient identification, such as an ID number
- Standardized data formats
- Technologically compatible hardware and software at each point in the chain of information creation and sharing
- Compatible policies on information disclosure and usage

Not all of these changes are proceeding at the same pace--some, such as the patient identifier, are far more advanced than the deployment of widely compatible hardware and software, which lags badly.[8] Furthermore, as noted below, the implementation (or in some cases the prospect) of these technical changes has spurred the demand for regulatory changes which may in time require further changes in information sharing practices, particularly as regards patient privacy. As one recent report put it tactfully, "EHRs might be characterized today as an extremely slowly developing success story."[9]

We are only beginning to understand the social and regulatory challenges of EHRs, and already the next revolution is beginning.

1.3 Third Step: Patient-Centered and Patient-Created Health Data

Until recently, it was generally assumed that EHRs would be produced, collated, stored, and disclosed by health services providers and by health services payment providers such as insurance companies and the government. This followed the typical model of health

---

[7] National Committee on Vital and Health Statistics, Report to the Secretary of the U.S. Department of Health and Human Services on Enhanced Protections for Uses of Health Data: A Stewardship Framework for "Secondary Uses" of Electronically Collected and Transmitted Health Data (October 21, 2007) at 6.

[8] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 30-31 (2008).

[9] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 30 (2008).

care which was highly provider-centric: patients went to professionals for treatment, and their task was to follow instructions.

Indeed, despite the claim, noted above, that the deployment of EHRs would "empower individual patients to assume a much more active, controlling role in their own health care" there has in fact been little about EHR deployment that has empowered patients. Instead, the process of EHR deployment has been primarily a 'back office' development, largely invisible to the patient except to the extent it has required signing a new raft of unintelligible consent forms. The use of electronic communications technologies to share records is of value, but is not very visible; automated reminders to patients to refill their prescriptions are certainly of value, but can hardly be called "empowering."

The PHR revolution threatens to provide the most revolutionary change of all, genuine patient empowerment as regards medical information. But under the current terms of trade the PHR revolution offers this empowerment at a price that may be too high in privacy terms; indeed it remains unclear whether the net effect from the individual point of view will be empowering or not.

If existing EHRs are subsumed into larger and more comprehensive PHRs, then the new information architecture (potentially) provides a significantly greater role for the patient -- and for the healthy consumer, who also becomes involved as producer and consumer of medical information. The creation and deployment of *personal* health records (PHRs) involves four elements, each of which offers the prospect of major changes to the health information ecology: (1) viewing the patient, and devices controlled by the patient, as important sources of health-related data; (2) giving the patient much greater control over health information; (3) moving personal data storage and/or queries based on personal data towards internet-based applications; and (4) permitting -- even encouraging -- patients to share health data via informal social networks either online or using mesh technologies. Meanwhile, there is every reason to believe that one of the trends which began earlier -- the linking of non-medical information to the health data -- will continue apace.

(1) Patient-created information

The introduction of PHRs dovetails nicely with the deployment of new portable or home-based health sensors and other devices designed to have patients record their own health information. This patient-generated information becomes part of the PHR, and becomes available for the patient to share with medical personnel or with third-party sources of medical advice such as internet-based support groups.

The patient has the greatest control over this personally created information. Conversely, some of this information can be among the most intrusive if it is widely shared. Blood sugar records may or may not be a great privacy concern depending on the patient, but a

health diary with discussions of the patient's moods, sex life, drug use, or interactions with family members surely will be.[10]

(2) Patient control over information created by others

The PHR model allows data to be mobile, and patient-centered. Having people carry their medical information on a card or chip, as in Germany,[11] offers a solution to the problem of access for emergency medical responders faced with an unconscious patient. Wherever the data may reside, putting the patient in charge of it offers the prospect of allowing the patient to control disclosure, thus setting privacy levels to the patient's preferences; reality, alas, is not so tidy as health providers and insurers tend to require substantial disclosures as a condition of treatment and payment.

Although the PHR ideal permits a radial information system with the patient at the center, the reality is likely to be far more complex, with information flowing in all directions:

---

[10]Note the "core services" identified by Project HealthDesign's Functional Requirements for PHR "Building Blocks":

> * Medication list management—Record, manage, share, and provide advice to consumers based on a list of the specific medications they are taking.
>
> * Calendaring—Track, share, and remind consumers of scheduled events relevant to the management of their health and their lives.
>
> * Observations captured in the course of daily living—Manage health-related information captured outside of the health care system. For example, acknowledging the important fact that your blood pressure measurement may be very different at home than at the doctor's office; the idea is to track information where people are…at home, work, or school, in transit, at the park, etc.
>
> * Identity management—Manage user authorization and authentication and allow consumers to monitor and control access to their own health data.

Project HealthDesign Releases Functional Requirements for PHR "Building Blocks", Feb. 25, 2008, http://projecthealthdesign.typepad.com/project_health_design/2008/02/project-healthd.html.

[11] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 38 (2008).

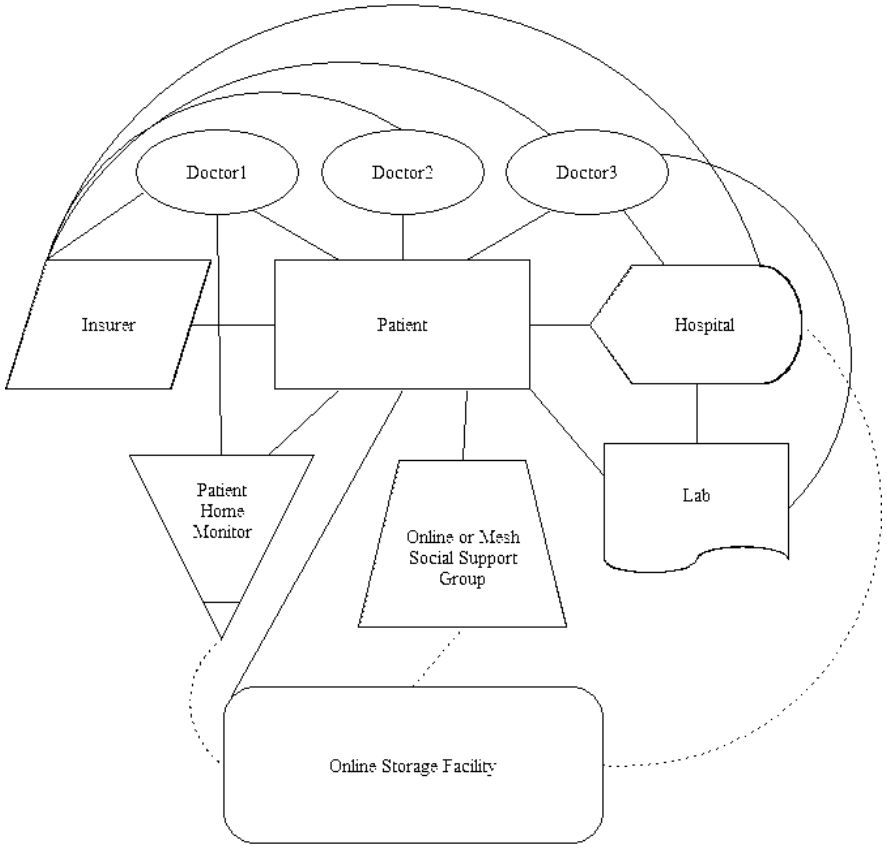

Figure 3: Dotted Lines Indicate Optional Flows

Even if the patient is at the center of the information flows, whether patients will have the canonical (authoritative) copy of their health records remains debatable. There are good arguments for three significantly different information architectures, as each maximizes a different set of values. Insurers and providers are likely to resist any information architecture in which they do not both control their data and have means to authenticate patient information. Indeed, the authentication issue is particularly critical as it is the basis of reliable diagnosis -- and protection against liability for mis-diagnosis. Citizens, on the other hand, may be understandably reluctant to surrender control over a wider and deeper system of records.

It is important to note that the more the patient is able to prevent the dissemination of the data in the PHR, the more this control may frustrate some of the objectives of the earlier health data revolution. For example, the goals noted above of "fully automated data collection processes" that would "provide for more efficient access to more comprehensive databases for benchmarking" and would allow extensive data mining "for continuous quality improvement" are each potentially hampered by a regime in which the patient can refuse to allow personal data to be harvested and analyzed. Researchers have suggested that protocols for anonymizing patient data should serve to alleviate privacy concerns. Unfortunately, it is increasingly clear that any substantial quantity of data carrying the

granularity researchers increasingly desire carries with it a very substantial risk that the data can be reverse-engineered to become patient-identifying information.

Patient as Adjunct. The least disruptive model would essentially leave the architecture growing out of the introduction of EHRs intact, and treat patient-created data as no more than an additional input into the existing system. Patients might control the records they create -- devices need not automatically "phone home" to the medical professional whose job it was to analyze the data -- but the device's retention of data would ordinarily be a caching function rather than as permanent or canonical storage. Basically healthy patients might, for example, upload their data to their doctor the week before their checkup; on the other hand, very ill patients might prefer to have a data stream continually flowing to a monitoring service.

In this model, patients might carry a copy of their master records in a device under their direct control, but this would primarily serve as a convenience. The master copy of records created by health professionals would remain under their control, and the patient's copy would be accepted as authoritative only if authenticated by digital signatures or other cryptographic techniques and if it carried a sufficiently recent and cryptographically controlled timestamp.

The primary advantages of this model are (1) it is least disruptive; and (2) it addresses the needs of medical providers and insurers for assurance that data is not subject to deletion by the patient. (Note that the danger of data *alteration* by the patient is relatively easily solved by the adoption of cryptographic authentication techniques such as digital signatures which make data forgery extremely unlikely. Proper use of digital signatures would require a larger-scale deployment of a public key infrastructure than currently exists; the primary challenges here, however, are social and financial, not technical.)

The primary disadvantage of this model is that the patient is not really at the center of the architecture at all, even if the patient is given a copy of all relevant information. Other actors create and amend the canonical copy of the patient's data, and there is nothing in the model which actually requires that the patient be in the information loop for any data held by others.

Patient as Digital Fortress. At the other extreme, one can imagine an architecture in which the canonical copy of information is that held by the patient and the patient controls who it is shared with. As noted above, the dangers of digital manipulation by the patient can be limited by the use of cryptographic signing technologies. If the patient holds the information on a physical storage device in the patient's control, then the patient has the ability to prevent the dissemination of that information -- at least until it is shared once. At that point, the digital fortress has been breached, and whether the information becomes more widely shared is no longer a technical issue, but rather a question of of law, of contract, or of medical ethics. Unfortunately, history teaches us the wisdom of Benjamin Franklin's observation that, "Three can keep a secret -- if two of them are dead."

A difficult question is whether anything can and should be done in this scenario to give all recipients of health information notice that the patient has chosen not to share some information. It is possible to have the PHR contain an index digitally signed by some trusted third party. If the patient did not make the index available to, say, a doctor, this could serve as a red flag that the patient is holding back some information created by another. (Similar, more complex, techniques might be available for patient-created data.) Although the use of such an index protects the doctor from being manipulated by the patient, in many cases it also greatly limits the patient's practical ability to refuse to share data since the counter-party will have notice that the data exist.

Patient Embedded in Digital Network. There are many reasons why patients may not choose to hold their data in a physical device they control. Devices -- even a modest USB drive -- cost money. They break or wear out. They can be lost or unavailable in emergencies.

Online PHR data vaults offer a number of possible advantages including: availability wherever there is Internet access; formatting data in standard formats; connections to third-party providers of medical information and other value-added services; and, easy interface with relevant online (or even local electronic) support groups.

The challenge and opportunity of this version of the information architecture is that it may put the provider of the online service at the center of the information network. Some of the implications of this shift are discussed below.

(3) Growth of internet-based health applications

The Internet has already become the leading source of self-help medical information for millions of people. This phenomenon can only grow, particularly as both Google and Microsoft are moving aggressively into health search and health information provision via Google Health and Microsoft HealthVault respectively.

Both companies allow users to create a PHR by uploading electronic medical records. Neither claims an ownership interest in the data. Both stress in their advance publicity that the user will be able to decide whether to share any or all data in their PHR with others, and to exclude anything they wish when sharing.

Google says that they will not sell patient data, nor will they carry targeted advertising on Google Health. They also say they will not enter data into a PHR themselves. Thus, if users want to link data held by a medical provider, the user must initiate the linkage by finding the providers' name on a Google Health pull-down menu. Google will then send the user to provider's web site, where the user will be required to complete a form requesting the provider to copy the information to the user's Google site.

Microsoft HealthVault stresses information sharing among firms. Microsoft hopes to be the intermediary allowing companies to link their health data together, allowing the user to have a PHR on any third-party site (even Google).

(4) Establishment of informal social networks either online or using mesh technologies

The Internet allows access to another source of information—peer groups made up of individuals who share an interest in the same medical condition:

> Even more important for group participants than the information provided may be the sense of connection to others facing similar problems—others just like oneself. The information and support are particularly helpful for patients with less-common conditions where an individual's caregiver may have encountered the condition rarely, if at all. In one well-designed web-based group for sufferers from rare carcinoid cancer, for example, a healthcare expert offers scientifically validated information that helps patients separate fact from fiction.[12]

Having individuals share their health data with others is, paradoxically, both the most empowering and dis-empowering thing they can do. It is empowering because support groups have well-documented health benefits, and because of the 'many hands' phenomenon of group information discovery and dissemination. Alas, there are also dangers to information promiscuity: under current legal and technical conditions, data shared with one's peers is data over which one no longer has control, for there is little other than social pressure to keep peers from passing that data on, whether wittingly or accidentally.

2.0 Policy Issues

The policy response to these revolutions -- and especially the introduction of PHRs -- remains far behind the facts being made on the ground. In part, this may be the familiar phenomenon that rules and standards lag technical change. But maybe not: another part of the problem is that the issues are very difficult. Worse, the different participants in the information exchanges described above have conflicting interests.

It is true that putting the patient at the center of the health information ecology upsets most of the little we thought we knew about how to manage patient health information.

---

[12] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 38 (2008).

It is also true that existing rules, notably the Health Insurance Portability and Accountability Act (HIPAA),[13] are not designed for a world in which people hold their own data: HIPAA's assumptions are very much those of the first medical data revolution, of computerized records and EHRs and will not, do not, map well onto a world in which PHRs become the norm.

And it also true that the invitation to patients to contribute their own data to a PHR vastly increases the -- already significant -- possible intrusiveness of access to those records. Whether it is medical devices that record physiological information (heartbeat, blood sugar), or devices recording patient behavior (did you take your pill), or an invitation to self-report (a diary), linking this information to a master record that is uniquely identified to the patient changes the character of the personal information recorded there.

The problem is only compounded when one imagines how health data might be linked to other data lifestyle and transactional information already being collected both on and off line. "Quantity has a quality of its own," and our privacy calculus will need to change to reflect this.

That change will be difficult because the need to fund treatment makes the universe of traditional medical records bi-polar: financial realities mandate access not just for health professionals but for insurance and other fee-payers. But this access has costs:

> Erosion of trust in the healthcare system may occur when there is a divergence between what the individual reasonably expects health data to be used for and uses made for other purposes without the knowledge and permission of the individual. Compromises to health care may result when individuals fail to seek treatment or choose to withhold information that could impact decisions about their care because either they do not understand or do not trust how their data might be used or their identity protected. Risk for discrimination, personal embarrassment, and group-based harm may be amplified as there is greater ability to compile longitudinal data, re-identify data that have been de-identified, and share data through HIE.[14]

---

[13]Health Insurance Portability and Accountability Act ("HIPAA") of 1996, Pub. L. No. 104-191, § 261, 110 Stat. 1936 (1996), codified at 42 U.S.C. § 1320d (2000) et seq.

[14]National Committee on Vital and Health Statistics, Enhanced Protections for Uses of Health Data: A Stewardship Framework for 'Secondary Uses' of Electronically Collected and Transmitted Health Data (2007) at p.5.

2.1 Conflicting Interests

That patients, providers, payers, and potential providers of online services have divergent interests only complicates the policy environment. There have been a number of excellent reports on medical privacy or personal health records in the last few years which touch on the subject with varying degrees of frankness. Even the most frank, however, do not claim the issues are near to resolution.

A recent federal panel simply punted on the issue:

> ...testimony also indicated that there are growing uses of identifiable personal health information that fall outside of the HIPAA chain of trust (or other regulations, such as those covering research on human subjects). For example, when an individual supplies personal health information to a personal health record (PHR) web site not sponsored by a covered entity or business associate, the personal health information is not protected under HIPAA.
>
> Testifiers observed that there will be increasing challenges with respect to HIPAA and chain of trust with hybrid PHRs, in which both covered entity-supplied and individual supplied health data are collected.[15]

The Markle Foundation's recent reports propose a "federated" model of data management set out in its "Common Framework."[16] This model decentralizes data and puts rule-making in hands of the data-holder, which seems to anticipate the impact of the PHR revolution. But so far -- it remains a work in progress -- the Markle proposal has yet to answer the hardest questions about who gets access to what when.

A similar frankness about the magnitude of the problem appears in the comprehensive report on "Harnessing Openness to Transform American Health Care" recently issued by the Committee for Economic Development:

> But agreement on principles still leaves many difficult issues to be resolved. Rules must be developed about who is allowed to have access to what

---

[15]National Committee on Vital and Health Statistics, Report to the Secretary of the U.S. Department of Health and Human Services on Enhanced Protections for Uses of Health Data: A Stewardship Framework for “Secondary Uses” of Electronically Collected and Transmitted Health Data (December 19, 2007) at 18-19.

[16]Connecting for Health Personal Health Technology Council, Connecting Americans to Their Health Care: A Common Framework for Networked Personal Health Information (2006), http://www.connectingforhealth.org/commonframework/docs/P9_NetworkedPHRs.pdf

> information and under what conditions. Then the system must be able to verify that the party requesting access is authorized to have access, and can be identified and authenticated as the appropriately authorized party. ...
>
> There are obviously a myriad of other questions that will have to be decided. Will there be national standards for privacy and security preempting state rules or will national standards create baselines for privacy and security protections? How will the system deal with circumstances that do not readily allow a patient to authorize access to information? (Studies on how to improve emergency care show how contentious issues of consent can be.) What, if any, are the appropriate limits on patient control of access? How will exceptions be dealt with? How will disputes be resolved? How will the system be structured so that the patient-centered processes for controlling access to information do not impede the delivery of services—so that practitioners, wary of anything that gets in the way of their providing quality patient care, will not reject or undercut the system? How will public health needs, such as in the case of a pandemic, be balanced against patient privacy rights? What will be done in the case of unauthorized access to patient information? Will patients be able to opt out of the system, or will the system, as one leading expert suggests, gain support by requiring that patient's opt-in? And given researchers' concerns (it has been argued that the famous Framingham Heart Study could not be conducted now under today's less rigorous HIPAA regime), will a system designed to protect patient privacy be flexible enough to allow the use of EHRs for research purposes? The questions go on and on.[17]

The problem may be even worse than this summary makes it seem, as some of these goals simply may not be achievable, notably the creation of a perfectly de-identified system.[18]

2.2 Inadequacies in Current Approach

HIPAA is the major federal statute governing the flow of health information. In broad terms, HIPAA incorporates something of a fiduciary model. Duties of confidentiality run from health care providers to patients. When health care providers share data with health care plans, or health care "clearinghouses" (processors of data created by another), the

---

[17] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 34 (2008). (footnote omitted)

[18] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 34 (2008).

data held by those entities too become subject to HIPAA's strictures, such as they are. The analogy to a fiduciary model is not precise because HIPAA defines the types of entities that become subject to its rules. And that coverage has gaps.

Notably, HIPAA does not cover many parties likely to have access to PHRs.[19] A patient, even one who holds data created by her doctor, is not a fiduciary to herself, and thus HIPAA does not impose privacy duties on her. Nor, in most cases, do HIPAA's duties extend to third parties with whom the patient may share medical data unless they belong to one of the classes of people defined by HIPAA itself.[20]

As Robert Gellman notes, health records in a PHR may lose their privileged status in a large number of circumstances:

> • PHR records can be more easily subpoenaed by a third party than health records covered under HIPAA.[[21]]
> • Identifiable health information may leak out of a PHR into the marketing system or to commercial data brokers.
> • In some cases, the information in a non-HIPAA covered PHR may be sold, rented, or otherwise shared.
> • It may be easier for consumers to accidentally or casually authorize the sharing of records in a PHR.
> • The linkage of PHR records from different sources may be embarrassing, cause family problems, or have other unexpected consequences[22]

There is an ongoing debate about the extent to which HIPAA will cover third parties such as Google Health who are holding data on a patient's behalf. These parties would be covered if they held information for medical providers because we know that HIPAA covers health care "clearinghouses" (processors of HIPAA-covered data created by another).[23] On the other hand, entities which neither "furnish, bill or receive payment for, health care

---

[19] DIGITAL CONNECTIONS COUNCIL OF THE COMMITTEE FOR ECONOMIC DEVELOPMENT, HARNESSING OPENNESS TO TRANSFORM AMERICAN HEALTH CARE 32 (2008).

[20]See Definitions of a Covered Entity, 45 C.F.R. 164.501.

[21]See generally, A. Michael Froomkin, Project HealthDesign ELSI Group, Forced Sharing of Patient-Controlled Health Records (2007), http://www.projecthealthdesign.org/media/file/Forced-sharing.pdf.

[22]Robert Gellman, World Privacy Forum, Personal Health Records: Why Many PHRs Threaten Privacy (2008), http://www.worldprivacyforum.org/pdf/WPF_PHR_02_20_2008fs.pdf.

[23]See Definitions of a Covered Entity, 45 C.F.R. 164.501.

in the normal course of business" nor "process, or facilitate the processing of, health information from nonstandard format or content into standard format or content or from standard format or content into nonstandard format or content"[24] are not subject to HIPAA. Thus, it would appear that so long as these online health data providers are getting health data from the patient, they would not be covered under HIPAA. It follows that patient-generated health data shared with an online site, such as data from a daily heart or blood sugar monitor, is also excluded from HIPAA's coverage.

The consequences of these perhaps understandable lacunae in HIPAA become far more serious if the patient becomes the center of the health records nexus.[25] In the ordinary course, one entity, such as a data repository, becomes subject to HIPAA when it gets its data from a covered entity, such as a hospital or insurance company. But what happens when the data is passed from the hospital to the patient to the online repository? Since the patient is not a HIPAA "covered entity" although the hospital is, the data is not -- from a HIPAA point of view -- being held on behalf of a "covered entity" and thus the online repository is neither a "covered entity" nor a covered "clearinghouse". That's something of a strange result, but it's also an understandable one. Were the rule so inclusive that it swept up all data sent by a patient that happened to include medical information, it would sweep far too broadly. But the consequence of this exclusion in a patient-centered information regime is that the exception threatens to swallow the rule.

Part of the gap may be filled by state law. Currently, the California Medical Information Act (CMIA)[26] likely has the broadest reach and seems likely to apply to many PHR products. Thanks to a recent amendment, this statute now applies to "[a]ny business organized for the purpose of maintaining medical information in order to make the information available to an individual or to a provider of health care at the request of the individual or a provider of health care, for purposes of allowing the individual to manage his or her information, or for the diagnosis and treatment of the individual."[27] Previously, the CMIA applied only to firms with the "primary purpose" of making the information available for purposes of diagnosis or treatment. The new definition sweeps far more widely: by including all firms that intend to make information "available to an individual ... for the purpose of ... managing his or her information."

---

[24] Cf. HHS, Are You a Covered Entity? online at http://www.cms.hhs.gov/HIPAAGenInfo/06_AreYouaCoveredEntity.asp.

[25] Cf. Laura Dunlop, *Electronic Health Records: Interoperability Challenges Patients' Right to Privacy,* 3 SHIDLER J. L. COMM & TECH. 16, (2007), http://www.lctjournal.washington.edu/Vol3/a016Dunlop.html ("EHRs could create the potential for privacy violations on an unprecedented scale")

[26] California Civil Code, sections 56 - 56.37.

[27] California Civil Code Section 56.06(a).

It is important to note, however, that even when the California approach reaches PHRs held by third parties such as Google Health, the law does not cover information residing on physical devices controlled by the individual (for example, the "PHR on a stick"). Nor would it cover disclosures by the patient to third parties neither involved in treatment nor holding the data for the patient. Thus, for example, in the absence of specific contractual confidentiality agreements there appears to be little to prevent members of a support group who acquired a person's personal medical data from passing it on to marketers or others.

2.2 Possible Solutions

If it is not yet clear what will work to help patients keep control over the spread of their health information, it is at least clear that some seemingly promising approaches are at best insufficient. A workable solution, if one exists, will require something genuinely novel.

HIPAA, at least in its current form, offers at most modest privacy benefits. Even as regards EHRs -- the data and entities undeniably within its scope -- the statute seems to have created paperwork for relatively little result. Complaints of data breaches are legion, prosecutions are so rare as to be almost non-existent, in part because the government interprets the criminal penalties in HIPAA to apply only to covered entities themselves and not to their employees or officers.[28] And, as we have seen, as it now stands HIPAA is even more toothless when applied to PHRs when the information is held by individuals or their non-HIPAA online agents. (Note, however, that HIPAA already applies to information uploaded from patient devices to health providers, although only to the provider's copy of the information, not the patient's.)

Given HIPAA's rather modest privacy success to date as regards EHRs, one can expect at most modest privacy gains from expanding HIPAA's reach to emerging health information intermediaries. There may be some substantial value in revising HIPAA's security requirements to extend to online services such as Google Health and Microsoft's HealthVault, but the worth of the privacy protections is less evident.

And whether or not HIPAA is extended to health intermediaries, it cannot practically be extended to health and health-related data emanating from the patients themselves. It would not be practicable to change HIPAA's orientation in the hopes of making an information privacy requirement into a legal servitude running with health information, somewhat like easements run with real property. This approach faces a number of

---

[28] *See* Peter Swire, Center for American Progress, *Justice Department Opinion Undermines Protection of Medical Privacy* (June 7, 2005), http://www.americanprogress.org/issues/2005/06/b743281.html.

problems, not least the First Amendment.[29] Even restricting any rule to commercial applications -- which lessens but does not eliminate the First Amendment problem -- health or medical information covers too much data, and too many types of data, to permit easy definition. And these data can be shared in so many legitimate ways that as yet no one has worked out how to craft a meaningful regulation navigating between over and under inclusiveness.

But unless this problem can be solved, at some point it may become necessary to ask whether PHRs have such toxic privacy side-effects that they are a cure worse than the diseases they are designed to cure.

How then to solve it? If the PHR privacy problem is to be solved, it will only be by the introduction of radical approaches that break with the HIPAA framework. The most promising avenues may be technological rather than legal.

(1) Exploiting Mass Digital Rights Management

One possible -- if not very attractive -- model involves a technology called Digital Rights Management (DRM). Heavily promoted by the music and film industries who see it as a way of preventing unlicensed online file-sharing, DRM technologies attempt to control use of digital media by preventing access, copying or conversion by end users to other formats. These technological limitations are typically enforced by encrypting the data. In order to access the content (e.g. play a song or movie) the consumers must have access to the decryption method. And the content providers only give access to licensed players, whether software or hardware; to get the decryption technology, designers must promise to limit the consumers' use of their purchased media.

DRM, when it works properly, offers the promise of not only policing access to data but also enforcing recordkeeping. By forcing accounting of access, DRM systems can be designed to enforce not just record-keeping but disclosure. Furthermore, if copies are limited to trusted systems, then the restrictions imposed by DRM will follow the data wherever it goes. DRM is not limited to music and film: so-called "enterprise digital rights management" (E-DRM or ERM) -- increasingly called IRM (Information Rights Management) -- systems apply DRM technology to the control of access to documents and other data.

Whatever it is called, DRM technology is controversial for a number of reasons. Among them are (1) the limitations it places on the platforms consumers can use to play content they have purchased; (2) the tendency of content providers to seek not only to block copyright violations but also to place technological limits on legal uses such as resale

---

[29]Cf. Pamela Samuelson, *Questioning Copyrights in Standards*, 48 B.C. L. Rev. 193 (2007), but see Molly Van Houweling,*The New Servitudes* (forthcoming Geo. L.Rev (2008)), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1028947.

under the "first sale doctrine" and fair use;[30] (3) the difficulty of moving content from one platform to another, (4) the political and economic attempts by content providers to encourage or mandate the use of "trusted" hardware platforms engineered for copy protection -- which means that, for example, a user's PC would have to pay for extra hardware (with possible security holes[31]) designed to work against the consumer for the benefit of others, and (5) the ease with which DRM systems that "phone home" to check whether a user has a valid license can be adapted to become full-fledged spyware systems.

Both hardware and especially software DRM systems can be reverse engineered, and the news that a system such as CSS or, more recently, Blu-Ray BD+,[32] has been 'cracked' is becoming almost routine. The sorts of DRM technology currently being deployed to protect movies and music are just not sufficiently secure and reliable to be adopted for the long-term protection of health records.

Nevertheless, were a sufficiently secure and ubiquitous DRM architecture deployed nationally as part of a campaign to prevent digital 'piracy', this DRM architecture might be adapted to solve the PHR privacy problem. In this highly imaginary system, the creator of each type of data could define which recipient would be allowed to access a given datum, and whether they would be allowed to keep copies[33] or share the data with others. The widespread deployment of so-called "trusted" hardware would allow additional refinements: for example, the system could be configured to require notice to the patient, or even real-time authorization by the patient, before allowing access to certain data.

In fact, however, the large-scale deployment of reliable "anti-piracy" DRM is unlikely, which is just as well given the all-too-real dangers noted above that "anti-piracy" DRM would be built around a framework of anti-privacy spyware.

---

[30]For a legal analysis see Rajiv Batra, John Padro, Seung-Ju Paik & Sarah Calvert, *The (Potential) Legal Validity of E-book Reader Restrictions*, -- Columb. Sci. Tech L. Rev -- (2008) (quoted in Gizmodo, *Amazon Kindle and Sony Reader Locked Up: Why Your Books Are No Longer Yours* (Mar. 21, 2008), http://gizmodo.com/369235/amazon-kindle-and-sony-reader-locked-up-why-your-books-are-no-longer-yours .

[31]See, e.g., Bruce Schneier, *Palladium and the TCPA*, Crypto-Gram Newsletter (Aug. 15, 2002), http://www.schneier.com/crypto-gram-0208.html#1.

[32]See AnyDVD HD now with BD+ support, http://forum.slysoft.com/showthread.php?t=14786 (March 19, 2008).

[33]Note that preventing the creation of local copies is difficult due to the existence of the so-called "analog hole" -- screens of data can be copied (by hand if necessary).

(2) Purpose-Built DRM Systems

Thus, if some sort of robust DRM is needed to protect health records from unauthorized access and unwelcome uses, much of it would have to be something designed specifically for medical purposes. Designing bespoke hardware and software has both advantages and disadvantages. The advantages are that the platform and the software can be tailored to the particular needs of the participants in the health records ecosystems. And, in theory, systems can be hardened against the dangers that particularly threaten those records. On the other hand, bespoke systems require much larger investments in both software and hardware than would be needed if the health records system were able to piggyback on a deployment driven by another industry's felt needs.

The needs of a health records DRM differ substantially from that of the software, film or music industries. In the 'creative' worlds, the authors (content providers) are small in number compared to the readers. The content providers are not, in the main, interested in permitting the consumers to modify their content. Audit trails are of interest only to the extent that they identify persons who make unlicensed copies of content. The system exists primarily to ensure that content providers are paid -- whether per copy or per use.

The needs of the health system are different and in many ways more complicated. Everyone is potentially an author. Permissions are not binary -- different persons in the system have varying rights to read, store, or make entries in a PHR. Audit trails that identify who had access to content, and who authored content (and when) may be essential. There needs to be an emergency override system for first-responders who constitute a large and a priori unknown class of people. Patients have a presumptive right to information about them, but there must be an override system for minors, incompetents, and perhaps others. Relatives need access to some or all of the information in some cases, but the patient needs a way to block it in others. Both care providers and payers need classes of information to effectuate their roles and to protect their legal and financial interests.

Researchers have not been blind to these issues, but the complexity of the problem means that solutions remain far away; nothing currently deployed comes close to doing what is required for the coming world of internet-based health applications and informal social networks for the exchange of health information. Consider, for example, the very admirable efforts of the Indivo project.[34] The goal of the Indivo project is to design a framework that would allow patient control and ownership of medical information, with granular access by a diverse group of parties, and the ability to have routine updates from

---

[34] *See* Kenneth D. Mandl, William W. Simons, William C.R. Crawford & Jonathan M. Abbett, *Indivo: a Personally Controlled Health Record for Health Information Exchange and Communication*, 7 BMC MEDICAL INFORMATICS AND DECISION MAKING 25 (2007), http://www.biomedcentral.com/1472-6947/7/25; Invido Health: Concept and Research, http://indivohealth.org/pages/concept.

both human and mechanical data sources.[35] The system provides a high degree of flexibility and patient control -- but only so long as the data stays within the system. In terms of the nomenclature used above (which differs somewhat from Indivo's terminology), Indivo solves or promises to solve, not just the privacy and information control problems of EHRs, but even those of patient-controlled and patient-centered records -- a form of PHRs. Thus, patients have control over information about them created by others, and over the dissemination of information they create, but again, only so long as the information remains within Indivo. What is more, the records in the Indivo system are "complements to, rather than replacements for existing healthcare information management systems." As those existing systems are run by health care providers, they contain the data to which HIPAA applies most directly. Indivo thus offers a solution -- or at least a partial solution -- to privacy issues relating to the patient-controlled health data that HIPAA cannot reach.

Unfortunately, it remains unclear how an Indivo-like system would interface with internet-based information systems, or with a multitude of handheld devices that allow patients to share health information with others in a health maintenance community. Worse, there seems no way that a comparable level of information control and security could easily be maintained once information is copied outside the system unless the devices to which the data were sent had been designed to be secure and to comply with rigorous information-sharing polices. Indivo is impressive, and seems designed to extend to accommodate many types of information, but the very sobering challenge will be how to replicate its virtues -- which depend on the centrality of the Indivo server in an information architecture -- to a multipolar world of multifarious devices and online services each of which sees itself as a data center.

The good news is that the threat model faced in the health privacy arena differs in one critical particular from the problem that record publishers and movie studios think they face: in the case of the .mp3s and DVDs the publishers believe they must guard against potentially malevolent users who want to break the copy protection in order to make unlicenced copies. In the case of PHRs, we can assume that the original user, at least, has an interest in maintaining control over personal health information, and indeed has some incentive to use privacy enhancing technologies (PETs) although we cannot assume that users have any technical sophistication, and must presume they are error-prone. The

---

[35]"Indivo record owners can subscribe to data updates from hospital information systems, practices, and regional health information organizations (RHIOs) also known as subnetwork organizations (SNOs). Indivo records can also be registered with regional record locator services making their data available to institutions within the RHIOs/SNOs." Kenneth D. Mandl, William W. Simons, William C.R. Crawford & Jonathan M. Abbett, *Indivo: a Personally Controlled Health Record for Health Information Exchange and Communication*, 7 BMC MEDICAL INFORMATICS AND DECISION MAKING 25 (2007), http://www.biomedcentral.com/1472-6947/7/25

bad news is that getting a wide group of third parties to accept PETs that cost money and may mostly benefit someone else may not be simple.

(3) National Patient ID Number

A different method to build in technological privacy protections would combine technology and law. Instead of building a complex infrastructure that sought to build fences around each datum, the key would be to create duties that would bind anyone making commercial use of the information, or linking it to other data via a nationally defined patient ID number. A standardized patient ID number formed part of the original HIPAA proposal but was eliminated due to public opposition. If, however, systems such as Google Health or Microsoft's HealthVault become popular, users of those systems will have a de facto national patient ID number whatever it is actually called.

So long as firms find it beneficial to use a national ID system for authentication and especially for data indexing and matching, they should be willing to accept a degree of expense or constraint regarding the way that they manage the information created, verified, and indexed thanks to this new technology.[36] Thus, it could be politically feasible to condition the use of the new national index number on adherence to national data protection and privacy rules. The ownership and dissemination of private sector data would remain a matter of contract and state law as it is today,[37] but would be constrained by the third party's duty to adhere to government-defined data protection rules when using the federally owned ID number to index data, or even when using any data that had been so indexed.[38]

The linchpin of this approach is to have the government own both the national ID numbers themselves and the standard by which the information is readable from the card. By creating a *sui generis* property interest in the number that it would hold, the government would give itself the leverage for a deal: firms that wished to avail themselves of the cost-

---

[36]This and the next paragraphs are adapted from A. Michael Froomkin, *Creating a Viral Federal Privacy Standard*, 48 B.C.L. Rev. 55, 75-78 (2007), available at http://law.tm/docs/virtial-privacy-standard.pdf.

[37]As a general matter, and absent duties of confidentiality which fall primarily on professionals such as lawyers and doctors, the facts of an economic transaction belong jointly and severally to the parties. *See* A. Michael Froomkin, *The Death of Privacy?*, 52 STAN. L. REV. 1461, 1502 (2000), available at www.law.miami.edu/~froomkin/articles/privacy-deathof.pdf.

[38]The obligation to comply with data protection rules would thus run with the data, just as do the obligations under the European Data Protection Directive. On the Directive see generally, JOEL REIDENBERG & PAUL SCHWARTZ, DATA PRIVACY LAW (1996).

saving benefits that using and relying on the new cards might bring -- or to get government health benefits --- would have to agree to be bound by a specific data privacy rules, and at the very least would also have to agree to only share their data with firms who had agreed to be bound by the same rules. Preferably, the duty to observe the privacy rules would be made 'viral' -- they would run with the number.[39]

Any sort of national ID system has many unattractive features. But if the private sector is building a de facto national health ID system anyway, a federal system with privacy built in from the ground up might be a preferable alternative to an ID system built without it.

3.0 Conclusion

As the health system has moved from paper records to computerized records, to EHRs and now to PHRs, the architecture of health information has undergone a rapid series of changes. Relevant public policy has not adapted as quickly.

PHRs, like EHRs before them, promise many benefits, some of which will even be realized. Along with these benefits, however, come substantial privacy risks. Current policy, including HIPAA, fails to address the privacy issue. Nor are there any obvious solutions to these dangers currently on offer either in the policy space or in the technical realm. Certain technical solutions based on DRM or on a national health ID system have some promise, but whatever privacy gains they will bring seems insufficient to offset the other problems they likely create. Creating a digital rights management system specifically for health data might have the fewest side-effects, but no such system currently exists, and thus of the technological alternatives it would require the most substantial research and testing. In addition, because the very significant costs of deploying a purpose-built system would not be defrayed by its adoption for some other purpose, cost will be an additional barrier to adoption.

[39] A. Michael Froomkin, *Creating a Viral Federal Privacy Standard*, 48 B.C. L. REV. 55, 75-78 (2007), available at http://law.tm/docs/virtial-privacy-standard.pdf.