A Privacy Strategy for the United States Healthcare Industry

Using data segmentation to enable healthcare privacy

May 21, 2013

Abstract: This document describes a general approach to enable privacy throughout the healthcare clinical automation environment.

Table of contents

Intended audience

Robust privacy represents a critical prerequisite for the United States effort to achieve health information exchange through mechanisms such as the eHealth Exchange. If eHealth Exchange cannot assure appropriate privacy it will not be trusted by either patients or providers. If it is not trusted by these key stakeholders the exchange network will not be used and all the work that has gone into its creation could be wasted.

This white paper explores a data segmentation-based privacy paradigm. Some may feel that it is not appropriate to pursue data segmentation because it is not feasible. **This white paper demonstrates that data segmentation is not only feasible, but it is simple to implement across virtually every electronic health record (EHR) system, offers robust privacy protections, and can support a wide variety of privacy use cases.** The white paper is directed to any stakeholder interested in the privacy issue – whether patient, provider organization, individual caregiver, vendor, administrator or government entity.

One key objective of this paper is to initiate a discussion on patient privacy issues. Some of the issues that need thought and an open discussion of differing views are included in the Discussion section late in this paper.

Defining the need

Clinical information represents the intellectual basis on which physicians and other caregivers deliver safe, effective medical care to a patient. A list of a patient's diagnoses, medications, allergies, medical procedures, clinical findings, etc. help a caregiver determine what actions to take to treat a patient's diseases and conditions and to maintain their health. There is a clear benefit to making sure that this information is made available as widely and as accurately as possible so that every caregiver who encounters this patient can provide the best care in an efficient manner.

This need for information transparency collides, however, with patient expectations and legal requirements to preserve privacy. Clinical care deals with a variety of potentially sensitive areas. Psychiatric care, treatment of juveniles, cancer therapy, obstetrics and gynecology, genetics information, VIP hospitalizations, episodes of sexually transmitted diseases, and a host of other examples represent situations where the patient or legal entities may insist that this information is not publicly available. In many cases these privacy requirements run directly counter to the information sharing needs of caregivers attempting to provide optimal clinical care[1]. This conflict, combined with the lack of appropriate automation in the healthcare system[2] means that healthcare is experiencing significant lost productivity, increased costs, and risks to patient safety and the quality of care. Clinical care also may suffer because a patient may feel they have no alternative but to withhold information from their physician in order to preserve privacy. This white paper describes a privacy paradigm that allows patients and their caregivers to achieve an optimal balance between the need to share clinical information for purposes of medical care and the need to maintain privacy in order to meet the expectations of patients and the requirements of legal entities.

The privacy paradigm described here allows patients, their caregivers, and other stakeholders to make individualized decisions about how to balance the often conflicting requirements of clinical information exchange and patient privacy protections. These decisions must be controlled by the patient working in conjunction with their caregiver(s) within the constraints of local, state, and federal regulations. The strategy must be simple to implement, easy to explain to the patient, readily implementable by clinical information systems and sufficiently flexible to support the complex requirements of healthcare. Furthermore, the system must be able to evolve over time as patients' needs change, privacy requirements mature, and the regulatory environment evolves.

Robust patient privacy cannot be achieved without accurate patient identification. If patient identification is uncertain then accurate linkage of the patient's information across healthcare organizations cannot be relied upon and their expressed privacy consents cannot be respected. The analysis in this white paper assumes that an effective patient identification system is in place. The basic requirements for this patient identification system are listed in Table 1 in Appendix 1 below.

---

[1] Poor Prognosis for Privacy, Wall Street Journal, May 2, 2013, http://online.wsj.com/article_email/SB10001424127887323798104578454793056230984-lMyQjAxMTAzMDAwMTEwNDEyWj.html?mod=wsj_valettop_email#articleTabs%3Darticle.

[2] Ponemon Institute, The Economic and Productivity Impact of IT Security on Healthcare , http://www.imprivata.com/Ponemon-Economic-Impact-Study.

Privacy requirements

It is impractical to attempt to list all the requirements concerning healthcare privacy in any single document. Nevertheless, here is a brief list of privacy requirements to serve as a back drop for the remainder of this white paper.

- Any proposed healthcare privacy paradigm must be vendor neutral so that it can be implemented by a wide variety of vendors and applications.
- The decision to impose a set of privacy restrictions must result from a voluntary choice by the patient to do so.
- The resulting capability must empower patients and their caregivers to simply, rapidly and consistently implement a privacy strategy for any particular patient at any point in their healthcare experience.
- The result must be a privacy system that provides demonstrable protection of patient privacy and that can be applied across a variety of domains including clinical care, medical research, medical education, public health reporting, and various legal and regulatory requirements.
- This privacy capability must be able to evolve over time as changing patient and industry requirements emerge. There must be defined mechanisms to support exceptional circumstances where privacy must be suspended and to then restore the patient's privacy once those circumstances no longer apply. For example, there needs to be a break-the-glass mechanism to enable the provision of life-saving emergency care and a way to restore privacy once that event has ended.

Describing the approach

Data segmentation is the fundamental approach implemented under this privacy paradigm. Under patient and physician control a specified set of clinical information can be associated with a specific privacy identifier. That privacy identifier may be an "open" identifier signifying that the patient wishes this data to be shared with all of their caregivers; or it may be a "private" identifier indicating that there are restrictions concerning to whom the data should be made available and under what circumstances. When an identifier is created it is provided directly to the patient in the form of some token such as a bar-coded identity card, a smart card, a cell phone app, a USB dongle, or an RFID device. The patient is then free to use that identifier to restrict access for any set of clinical encounters they wish by presenting that token as their identification when they register for a subsequent visit.

Note that the fundamental 'granularity' of this data segmentation approach is that of a single clinical encounter (or a series of encounters for a common purpose.) However, there is nothing which prevents an organization or information system from supplying additional privacy constraints if finer granularity is necessary. In particular the Data Segmentation for Privacy (DS4P) work sponsored by the Office of the National Coordinator for Health IT (ONC) can coexist with the data segmentation paradigm described here. There is nothing in this approach which *requires* this additional granularity but it can be fully supported.

Data segmentation

A data segmentation approach enables patient-controlled privacy by providing each interested patient with the tools needed to subdivide their clinical data into independent data sets. The patient is then responsible for deciding which sets of data are revealed to each of their caregivers by making the

corresponding identifier(s) known to them. Typically a patient has one "open" set of data that they wish to be available to all of their caregivers. In addition they may have one or more "private" data sets that represent domains which they consider to be sensitive. The number of these private data sets is determined by the patient's particular situation so the privacy constraints placed on each data set may be different.

Note that the decision on whether to have any private data sets at all is up to the patient and this decision may change over time. Data segmentation is most efficient when the decision to segment a set of information is made proactively, i.e. before that clinical information is created. For example, a patient preparing for their initial visit to a psychiatrist might decide in advance that they wish their psychiatric data to be kept private.

> The patient has an open identifier that is associated with all of his encounters with his family practitioner. He acquires a private identifier when he registers for his first appointment with a psychiatrist, and this is used to "tag" the information associated with that and all future encounters with this new provider.

This proactive data segmentation approach has the advantages of being simple and straightforward. Note, however, that retrospective data segmentation is also possible (see illustration 4 below). It involves more work and has at least two serious limitations at the present time. First, a retrospective privacy decision cannot undo any data exposure that occurred prior to that point in time. Second, current EHR systems have limited data segmentation capabilities which make it a challenge to achieve patient specific data segmentation retrospectively. Presumably this will become more robust over time as EHRs mature with respect to data segmentation capabilities.

A frequent criticism of the data segmentation approach is that it requires the patient to take an active role in managing the privacy of their clinical information. This is a valid criticism but it should be noted that establishing privacy constraints on clinical information is a voluntary patient activity. Many patients (probably a majority) will never choose to establish any patient-specific privacy constraints on their information. They will obtain one single open identifier and use it for all of their data for their entire life. Those that do choose to establish some constraints are precisely the population that is motivated to actively manage their own privacy. As long as the system provides them with reasonably simple methods to manage these privacy decisions and a straightforward mechanism to make changes should an error occur, this patient population will have little difficulty maintaining the privacy protections that they require.

Implementing data segmentation

Virtually every EHR uses a unique and different data structure to support its functionality[3]. This lack of a common architecture makes privacy schemes difficult to implement because they must be implemented differently for each EHR system. However, all EHR system schemes do have one aspect in common. Every EHR assigns each individual some internal identifier that is unique to the patient. This identifier is also unique to that vendor in terms of its structure and operation in uniquely identifying every individual

[3] Healthcare vendors use a variety of data base architectures including hierarchical (MUMPS), relational (SQL), object oriented (Postgres) and other approaches.

in that EHR. The data segmentation approach described in this white paper takes advantage of this commonality that extends across every vendor's EHR architecture.

By building on this common feature, the privacy paradigm described in this white paper gains several key advantages.

- The changes needed to a vendor system are very consistent across vendors. There is no need to vary the implementation due to the "downstream" differences in that particular vendor's data schema. This avoids complexity, implementation cost, and increased management resources for the vendor.
- Second, each identifier is automatically linked to all data in the resulting open or private data segment. This means it is 'easy' for the vendor to segment an entire encounter or series of encounters such as those described in the legislation mandating withholding data for encounters paid out-of-pocket[4].
- Third, this implementation approach yields elegant simplicity to the data segmentation effort. Because the identifier controls a clinical episode and is directly provided to the patient, the patient readily can have an intuitive grasp of how their privacy works and how they control it.

Selected illustrations

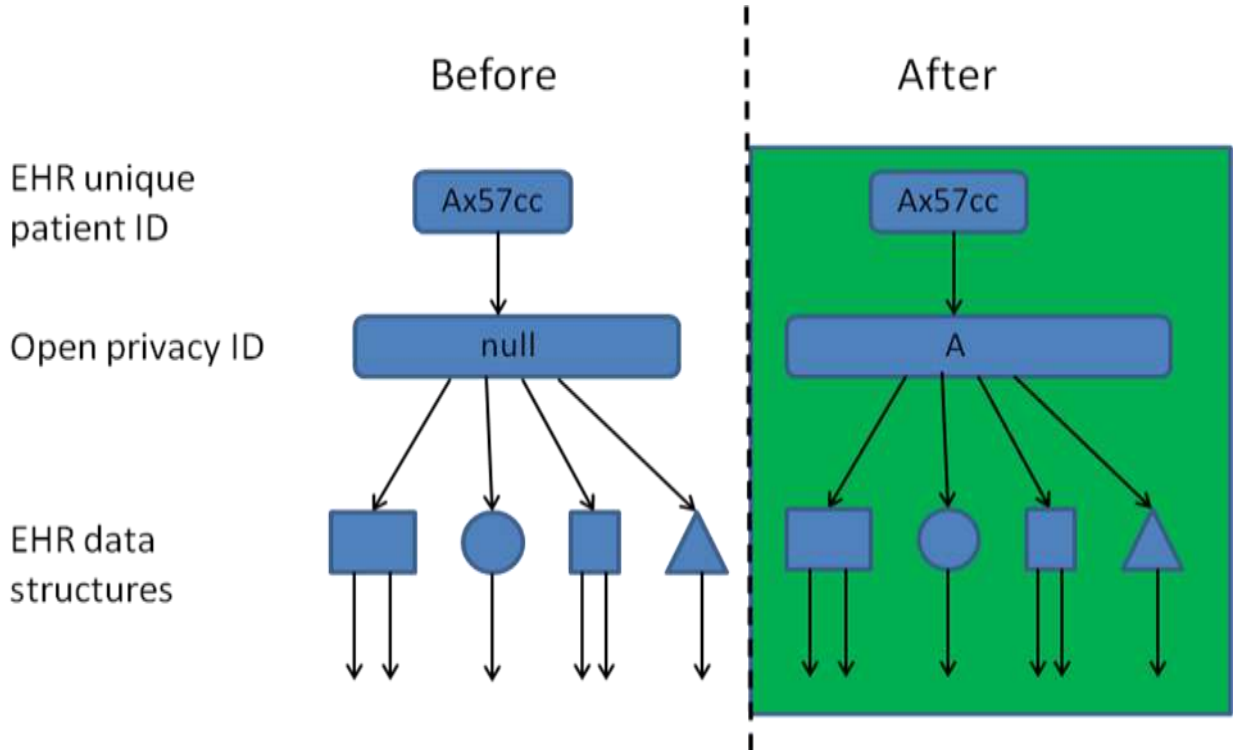
This section 'dives down into the weeds' to provide four illustrations of how the data segmentation privacy paradigm works. If you are the kind of person who enjoys moderate detail then read on, otherwise skip to the Discussion section that follows this.

For purposes of these selected use cases a capital letter (e.g. D) is used to indicate an open identifier. An open identifier is used to designate a set of data that a patient wants to be available to all caregivers. For many patients this may constitute a patient's entire medical record. However, some patients may want to keep portions of their medical information private. An underlined capital letter (H) is used to indicate a private identifier. A data set labeled with a private identifier is meant to be anonymous. In general, only (1) the patient, (2) the provider where the patient received the anonymous identifier and (3) the information system that requested the identifier are aware of its association to the patient. Of course, the patient retains the right to reveal this private identifier to other caregivers if they wish. Thus a private identifier for psychiatric information might be shared by the patient with all the members of a multi-disciplinary psychiatric team that is treating the patient.

Figure 1 illustrates the insertion of an open privacy identifier into a patient's record. Once this has occurred the patient may use their open identifier to provide caregiver access to this information during the registration for any medical encounter.

[4] Rule 45 CFR section 164.522(a)(1)(iv).

Figure 1 – Assigning a Privacy ID to a Patient



The schema is modified to insert a place for a unique privacy ID immediately subordinate to the EHR's unique identifier for the patient. For a patient that does not have a privacy ID this location is set to null. When an open privacy ID (A) is assigned it is stored in this location in the schema. The green rectangle indicates that the data is now part of an open data set.

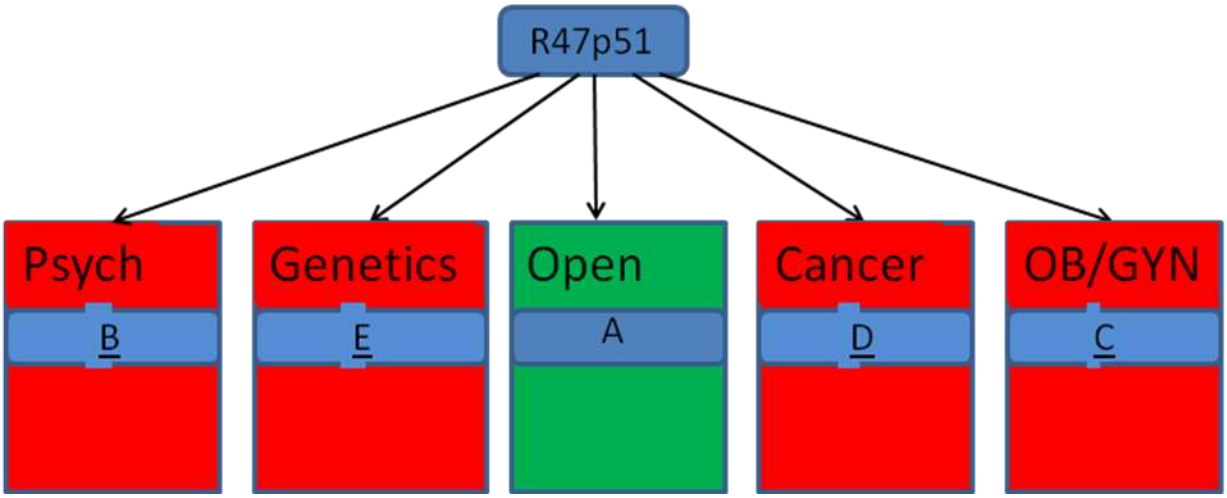
Figure 1 illustrates how unique patient identifiers can be inserted into any EHR's data schema immediately subordinate to the vendor-specific unique identifier for that patient. This permits that individual's medical data to be segmented according to the schemes associated with their privacy identifiers. By taking advantage of this common aspect of EHR architectures, the privacy system enables a simple and robust schema modification to occur across a wide variety of EHR data architectures. Not only does this lead to simplicity but it also enables consistent privacy performance across a wide variety of different EHR systems. Because privacy is a voluntary activity, some patients may not have an assigned open privacy ID.

The workflow to establish an open data set for a new patient is:

- Register the patient in the EHR and obtain the unique EHR identifier
- Acquire an open privacy identifier "A"
- Print an open ID card and give it to the patient[5]
- Establish the basic data structure for the encounter
- Collect information during the encounter linked to "A"

[5] Note that "print a card" could mean "issue a smartcard" or "update a smartphone app", etc.

Figure 2 – A Patient with Multiple Privacy Sets



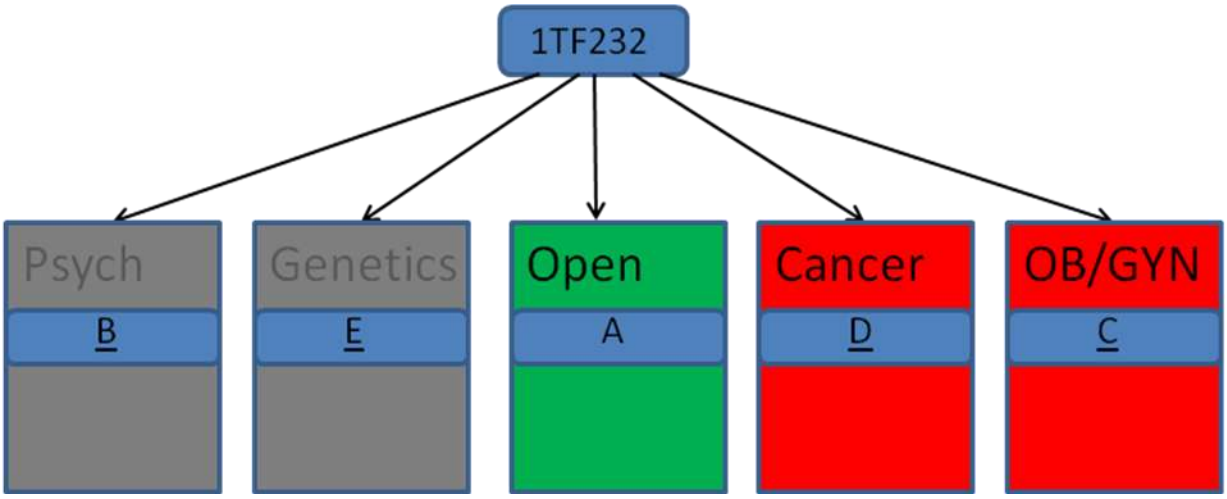
A patient may have as many privacy domains as is required by their particular circumstance. This patient has one open data set and four private ones. Some or all of the private domains may be made accessible to a specific caregiver depending on privacy decisions made by the patient.

Figure 2 shows a patient that has created one open and four private data sets. The patient controls the exposure of the private data sets by choosing to provide or withhold the associated private identifier(s) while registering for a medical encounter.

The workflow to add a private psychiatry data set is:

- Acquire a private privacy identifier "B"
- Insert it into the EHR data structure
- Acquire psychiatry data under identifier "B"
- Print a private ID card and give it to the patient
- Patient can use this ID card for identification on any subsequent psychiatric visits
- Patient and psychiatrist jointly decide what data from this visit is open and what is private
- Psychiatrist sees information in segments A and B

Figure 3 – Gynecologist's View of a Patient



The three colored data sets have been made visible to the physician but the grayed out data sets have not and remain private.

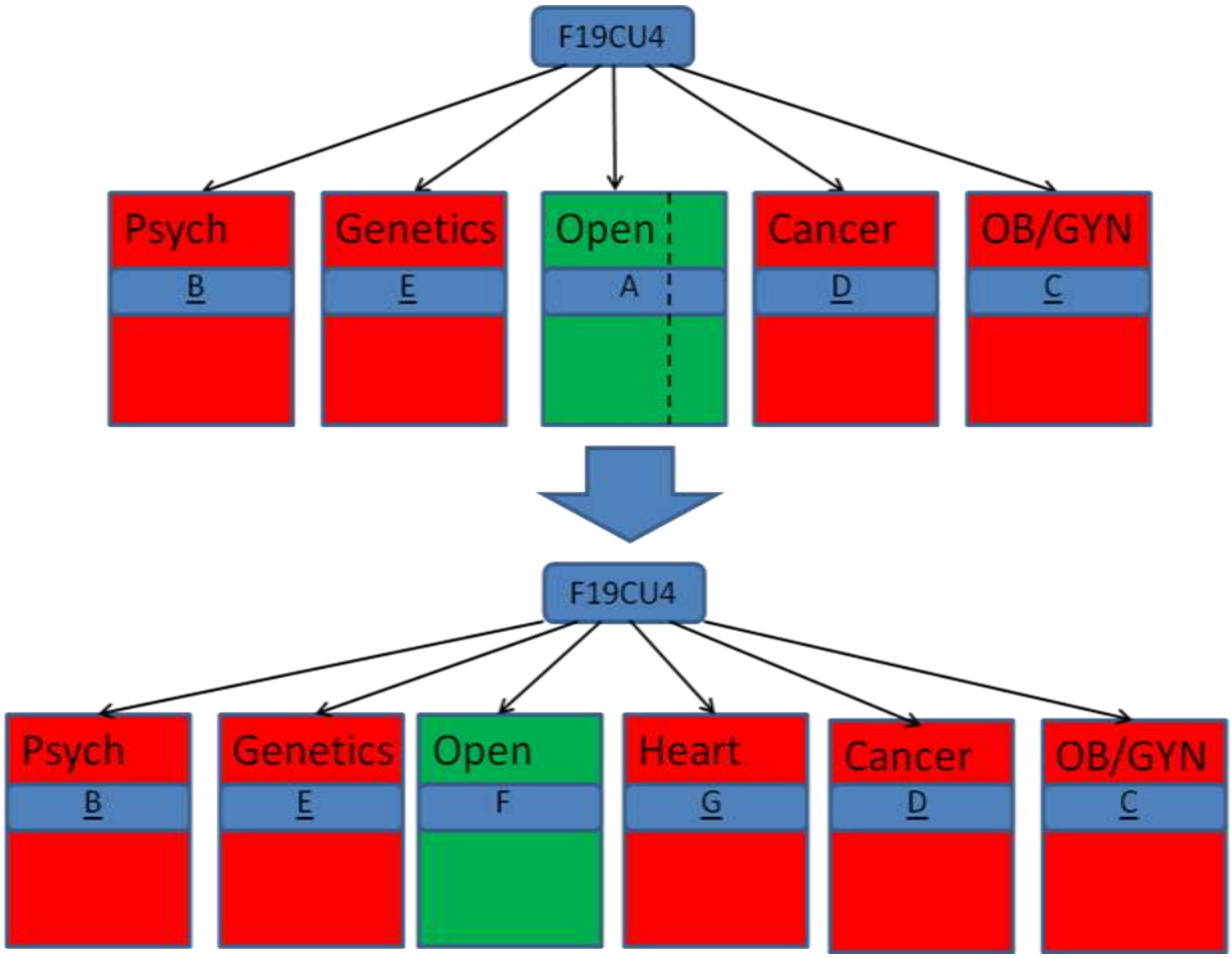
In Figure 3 the patient is visiting a gynecologist. By providing her physician with private identifiers C and D the patient has made those two private data sets as well as her open data set visible to the gynecologist. The patient has not granted access to either the B or E data sets. This illustrates how the patient is empowered for each clinical situation to control what information is revealed. Note that the EHR system may still make all 5 data sets visible for purposes of clinical decision support[6].

The workflow for using multiple private data sets is:

- When a patient is preparing for a clinical encounter they need to determine what data they wish to keep private and what they wish to share
- In this case at registration for the gynecology clinic the patient has provided private identifiers C and D to the clinic registration staff
- (Note that because A is an existing open identifier its associated data set will automatically be included in the data made visible to the gynecologist)
- The information associated with B and E will not be revealed to the gynecologist but still may participate in automated clinical decision support to help ensure patient safety

[6] See discussion topic 1 below.

Figure 4 – Subdividing an Open Data Set



The patient has decided that some of her previously public cardiac information should be made private as indicated by the dotted line subdividing the A open data set. Once new open and private identifiers are obtained the cardiology data can be assigned to the new G private data set. *Note particularly that identifier A is deactivated* and the remainder of A's data is transferred to new open identifier F.

Making a portion of a data set that was previously open into a private data requires that the EHR have the ability to separate out the proper data sets. Given that this capability exists, this use case demonstrates how cardiac data could be moved into a new private data set.

The workflow to make part of an open data set private is:

- Identify the data elements that need to be made private
- Obtain new open privacy identifier F and new private privacy identifier G
- Transfer the cardiac data from A to G
- Transfer the remaining data from A to F
- Deactivate identifier A
- Print out new F and G identifier cards for the patient
- Have the patient destroy identifier card A
- Send out a network-wide directive to replace A with F

Discussion

The privacy requirements of the US healthcare system are sufficiently complex that no single approach can be readily shown to solve all privacy problems. However, the data segmentation paradigm does offer solutions to some of the most serious and persistent privacy conundrums. Here are very brief explanations of how to address some of these perennial privacy problems.

1. ***Data segmentation has traditionally been viewed as being dangerous from a clinical care perspective. A physician treating a patient without full knowledge of their clinical profile is at risk of doing harm through unintended consequences such as giving a medication that could have a negative interaction with another of the patient's medications that the physician is not aware of.*** Note that this situation frequently occurs in today's privacy environment. The patient may intentionally not tell the physician about all their medications or diagnoses, or simply forget one or more.

Frequently debates concerning the merits of data segmentation have been carried out in an environment where only two possible conclusions have been considered.

- The first considers patient privacy to be so important that it justifies the risk of adverse patient outcomes due to physician ignorance of the patient's full clinical picture. According to this extreme it is better to let the patient make their privacy decision and deal with the clinical consequences of their decision.
- The second considers patient safety to be so important that it overrules any attempt by the patient to preserve their privacy. It is better for patients' therapies to be safe and they simply have to forego any desire to maintain privacy.

The privacy paradigm described in this white paper enables a series of intermediate implementations that offer a spectrum of choices between these two extremes. One sample approach described in this white paper is to mitigate the safety risk by making more effective use of automated clinical decision support.

While a physician may, for patient privacy reasons, be shielded from full knowledge of some of the patient's medications, there is no requirement that the clinical decision support (CDS) system be so restricted. The CDS system can be implemented to have visibility to the medications associated with all of the patient's open and private identifiers. The system can, for example, issue a warning to a physician concerning a possible drug-drug interaction. Privacy constraints may prevent the CDS system from displaying the details of the problem. This is not as desirable as the physician having full knowledge of the patient's condition but at least it offers the opportunity to avoid serious known complications. If the physician receives such an opaque warning, that event can trigger a discussion with the patient about the advisability of revealing more to the caregiver concerning their full medical condition.

Other potential compromises to address this problem also exist such as revealing the patient's full clinical picture to one particular physician that the patient chooses as their 'guardian'. It is not the role of this white paper to determine which of the various options listed above represents an appropriate compromise between patient privacy and patient safety. But it is important that a process be initiated to analyze these trade-offs leading to a decision on how to resolve this conflict.

2. ***In an emergency it may be necessary for a physician to exercise a "break-the-glass" option in order to have full and complete knowledge of the patient's situation.*** Because the patient's open and private identifiers are subordinate to the EHR's unique identifier for that individual, it is a relatively straightforward task to establish a procedure by which an emergency room physician can gain access to all of a patient's clinical information. The client organization must establish appropriate policies determining how this is authorized. Once this authorization is achieved the system can then reveal all data sets linked to this EHR identifier thus making all information on the patient available for the emergency care episode. Once the break-the-glass episode has been completed, it is possible to restore the patient's privacy by reinstating the separation between the various privacy domains. A further safeguard for this restoration can be implemented by replacing each of the patient's existing private identifiers with new, independent ones. This ensures that even a person equipped with the now compromised private identifiers will not be able to access these data sets going forward since each one now has a new, independent private identifier.

3. ***A frequent criticism of the data segmentation approach to privacy is that it places a requirement on the patient to manage their various privacy identifiers.*** This is a valid concern based on how the data segmentation paradigm is deployed. Note, however, that enforcing privacy constraints is a voluntary choice by each individual patient. It is likely that less than half of the patient population will ever see the need for such restrictions. Those that do see such a need will be patients who are motivated to manage their own privacy. The data segmentation paradigm thus becomes a significant expression of patient empowerment. It reinforces the reality that patient control of the privacy of clinical information is inextricably tied to patient responsibility to properly manage that control.

4. ***There are many existing and emerging situations where a patient may wish to "donate" some of their clinical data for purposes such as research, education, public health, or population health.*** For these types of situations the data set must be extracted from the patient's data. This extract must then be anonymized and then associated with a new private identifier dedicated to that public purpose. This approach allows a patient's information to participate in these "public good" activities while at the same time preserving the privacy of that patient's information when used for clinical care.

Conclusion

If properly implemented, data segmentation can be a powerful tool to enable privacy in healthcare. It constitutes a vendor-neutral approach that is relatively easy and straightforward to implement across a wide variety of clinical automation systems. The resulting privacy system is reasonably flexible and offers robust protection for data that should be protected. The data segregation approach is easy to explain to a patient. Furthermore it places privacy control directly in the hands of the patient and their physician(s). It thus represents a significant example of the patient empowerment that is widely agreed-upon as a goal for the US healthcare system.

Perhaps the strongest reason to consider a data segmentation approach to health care privacy is its simplicity. Physicians, nurses, medical technicians and other personnel can readily comprehend how storing clinical data under different privacy identifiers makes it possible to restrict access to that data. Even more important, however, is the fact that *patients* can readily understand how to use this approach to achieve the privacy protection they desire. Because the patient is in direct control of each privacy identifier they choose to obtain, the system represents an easily comprehensible approach to

privacy that the patient has a high likelihood of using appropriately. When errors are made, as they inevitably will be, it is reasonably simple for the patient and his/her providers to take corrective action by replacing a compromised identifier with a new one.

There is no reason that this privacy paradigm cannot coexist with other privacy strategies. By offering different privacy options it will be possible for the healthcare industry to build consumer/patient confidence by letting patients choose the approach that seems most appropriate for them. Due to its simplicity, flexibility, patient empowerment and ease of vendor implementation; the data segmentation approach described in this white paper should have broad appeal.

Appendix 1 – Accurate patient identification

Understanding the challenges related to accurate patient identification is a prerequisite to achieve patient privacy. If there is any confusion concerning the identity of an individual then it is impossible to be certain that the healthcare system is properly respecting privacy requirements established by that person. Nor is it possible to correctly aggregate all of the information that constitutes that person's medical record. Table 1 lists the patient identification requirements that are assumed as a background to this privacy white paper. It is assumed that this patient identification system results in the creation and use of a set of privacy identifiers that meet the following requirements.

Table 1
Patient identification requirements

Property	**Definition**	**Description**
Must be unique	Not duplicated	The same identifier is never assigned to more than one patient
Must be accurate	Represents the 'right' person	Authentication
Must be permanent	Available for a lifetime	No arbitrary duration limits
Must be abstract	No embedded data	Any identifier can be assigned to any individual
Multiple IDs may be issued to one patient	As many as needed	Number is driven by the patient's situation
Must be readily available	When and where needed	Assigned in real time when a need is identified
Open & private IDs both must be offered	Specifies the nature of the associated data	Open-freely sharable, private-restricted
May be terminated	Can be invalidated	By the patient or an authorized agent (fraud)
May be anonymous	Does not identify the individual	To preserve the privacy of the data set
Must be under patient control	Patient empowerment	The identifier is made directly available to the patient
May be voluntary	Patient/provider choose to participate	Privacy is a choice by the individual patient
Replacement must be possible	Substitute a new ID	Ability to deactivate an ID and substitute a new one

Appendix 2 – VUHID in support of privacy

As noted in the preface, functional privacy cannot be achieved in the absence of highly accurate patient identification. The Voluntary Universal Healthcare Identifier (VUHID) system offered by Global Patient Identifiers, Inc. (GPII) offers full support for both the patient identification and privacy functions described in this white paper. It is based on an ASTM International/ANSI standard that defines a standardized healthcare identifier that contains an embedded privacy class code as well as a series of check digits that prevent counterfeiting. This privacy class determines, among other things, whether the identifier is an open or a private identifier. Each VUHID identifier is globally unique and meant to be permanently assigned to a patient. There is no embedded patient identification information in VUHID identifiers which means they can be assigned when and where needed to enable privacy functions. Multiple identifiers can be assigned to an individual, for example to support the data segmentation paradigm described here. The VUHID system enables creation and management of identifiers on demand. It supports identifier data location, identifier termination and network-wide identifier replacement functions. More information is available at www.gpii.info.

Barry Hieb, M.D.
Chief Scientist
520.320.6220
520-342-8457 (mobile)

Rob Macmillan, CEO
520-449-9840 (mobile)

Global Patient Identifiers, Inc. at www.gpii.info
Sponsor of the Voluntary Universal Healthcare Identifier Project