Proposal for How to Support Consumer Choice
Re: Privacy issues and PHI Redaction:

Date: December 8, 2010
Version 1.8 (original version from 2001)
Document editor: John Mattison (john.e.mattison@kp.org) feedback appreciated.

Statement of the Problem:
The rights of consumers to control sensitive clinical information sometimes conflict with the needs of providers to deliver high quality and safe care. If providers do not have complete information on an individual patient, their recommendations for care may not be safe. Providers strongly believe that patient safety and quality of care suffer when data is unpredictably redacted from a health record, and yet consumers often prefer to remove this information, and they have a legitimate right to withhold information at their discretion. This document attempts to address both concerns and propose a solution set that can best accommodate the needs of consumers with the needs of providers.

Background:
There will always be a need for prosaic free text notes to capture the nuances of the human health experience. Most digital health records today contain extensive information in the form of free text. Attempts to eliminate all free text notes have been consistent failures, and it is likely that the prosaic component of health records will remain prominent for many years. Many would argue that the hybrid nature of health data will persist in perpetuity. As a consequence, any systematic method of redaction in health records must be effective for both discrete data and for free text entries.

"Reliable redaction" of sensitive data can be safely achieved in summary tables involving discrete machine-readable data, e.g. problems, meds, diagnoses, encounter summaries etc. In addition, when these sensitive data are contained within discrete note types, e.g., notes in sensitive services (mental health, alcohol and chemical dependency, genetic counseling, etc.), those note types can be redacted in totality.

The same precision of redaction cannot be achieved in the redaction of free text. Even with advanced natural language processing, (NLP) it is not possible to efficiently and effectively parse free text progress notes for this type of information. To use a common example, an HIV patient typically has numerous references to that illness in almost every medical encounter they have. Even if advanced NLP could redact all references to HIV (which is nearly impossible), it would render many encounter notes into a swiss cheese of information that will be as dangerous as it is incomplete. It would not be safe to base clinical decisions on information from a chart redacted in such a fashion.

Historical proposals/debates on this topic:

1) Explicit Definition of What is "Sensitive" Information:
Legal attempts to create canonical definitions of what is "sensitive" invariably discover that there is a huge range of what consumers consider sensitive information. Some general categories are most often considered sensitive by a majority of consumers, where there is widespread social stigma or clear risk of harm. Common examples include revealing domestic violence reports, chemical dependency programs, and mental health visits. Unfortunately, aside from a few common examples, the variation in what people consider sensitive is highly variable. This is not simply a matter of "where do we draw the line", because some people consider innocuous

diagnoses like "kidney stones" as very "sensitive" to them yet are very comfortable with disclosing their diabetes.

The reverse perspective is equally likely in a different person with the same two diagnoses. Furthermore, what an individual consumer considers "sensitive" changes over time. Common examples include an initial diagnosis of HIV, Cancer, or Diabetes, which may be initially very sensitive, but over time that same individual often becomes more accepting of the diagnosis as a reality that they are comfortable sharing with any provider. Many consumers are quite comfortable with all their health information being shared with any treating physician, whereas nearly every health condition imaginable is considered sensitive by some considerable subpopulation of consumers. This prospective and prescriptive definitional approach to what is "sensitive" is not practical.

State legislative attempts to define "sensitive" data have historically created mostly impractical and ineffective solutions. There is also added complexity imposed on interstate HIE based upon the diversity of statutes among different states. There is no "correct canonical solution" to this problem, so any federal attempts to prescribe any such data set, or to pre-empt state efforts in this space have almost no chance of addressing the core issues. Existing attempts to legally define sensitive information approach the unimplementable, e.g. in California Health and Safety Code 123148 states "*none of the following clinical laboratory test results and any other related results shall be conveyed to a patient by Internet posting or other electronic means: (1) HIV antibody test. (2) Presence of antigens indicating a hepatitis infection. (3) Abusing the use of drugs. (4) Test results related to routinely processed tissues, including skin biopsies, Pap smear tests, products of conception, and bone marrow aspirations for morphological evaluation, if they reveal a malignancy."* How does one define and enforce *"and any other related results*". This issue rapidly becomes unmanageable in a disease like HIV that is systemic and affects the most fundamental elements of a record, namely problems, medications, allergies and results of innumerable diagnostic tests.

2) Prospective Labeling by Providers of "Sensitive Information" During Data Capture:
One historically proposed solution is that consumers declare what they consider sensitive at the time they are revealing that information to a provider. The provider would then "label" those elements as "restricted", and they would only be available to that provider but to no other, either within their institution or through HIE mechanisms. The problem with this approach is that a consumer's view of what is "sensitive" changes over time, and five years later they might have a very different view of what is sensitive information. A typical example is the new diagnosis of breast cancer, which is a crushing blow to any woman and many prefer to initially consider this diagnosis as highly sensitive. Over time most women become much more open about sharing this information. Relying on a realtime attribution of "sensitive status" assumes that all data whether so labeled or not, is static and will never change. Unfortunately consumer views on what is "sensitive" changes over time, and this solution falls far short of what is necessary to address those changing views of consumers over time, and any such approach would not obviate the requirement for an opportunity to reconsider all such sensitive attributions at **every** subsequent point of disclosure. This provider-based approach to labeling sensitive information is impractical at best.

1) **Opt-in vs. Opt-out**:
Allowing a consumer to formally either Opt-in or Opt-out of HIE has been considered the ultimate form of redaction, but if any consumer opts out (or fails to opt in), it doesn't obviate the need for an alternative mechanism that allows them to share their personal medical information with a more effective mechanism for redaction. Whether Opt-in or Opt-out is used is immaterial, and either option can achieve desired results if implemented thoughtfully. We do not believe that any federal efforts to specify whether all HIEs should use an opt in vs. an opt out mechanism would address any of the primary issues. In fact, there is a good argument for any organization to pursue an "Opt In" approach initially, and convert to an "Opt Out" approach once the mechanisms for secure HIE (and more widespread consumer trust of the process) are well-established. Federal guidance on this issue might be

useful as informational, but it is unlikely that regulatory prescriptions will accelerate the adoption of safe and effective HIE.

2) **Restriction of HIE to Discrete Data**:
We have found this approach useful for our initial work with the NHIN mechanisms. However, as soon as practical we will begin exchanging full text documents (encounters, operative notes, consultations, etc.). While it is logical to start simple and expand as appropriate, it is very difficult to service information needs of providers by restricting electronic exchange to discrete data. There is a wealth of information that resides only in free text documents. In the pre-HIE world, both paper records and faxes provide full text documents, and these two mechanisms still exist today. It would be inadvisable to restrict HIE to discrete data only as a way to address the challenges to safe and effective redaction of free text.

The novel proposal below is an attempt to address the apparently conflicting needs of consumers and providers around redaction of sensitive information.

Proposed Solution:

Allow "tiered options" for patient consent

Option 1: Consumer authorizes the exchange of all information, with no redactions. This approach would allow for full exchange of free text data without any attempt at redaction. (within the constraints of existing statutes and several universally recognized restrictions).

Option 2: Consumer authorizes exchange of discrete data only (problems, meds, diagnoses, encounter histories) These discrete data can be sent with redaction performed using automated tools. This option would not include the exchange of any free textual data (aside from demographic identifiers). Progress notes, admission notes, encounter summaries, consultation notes, etc. would not be exchanged under this option. There are serious restrictions to quality of care when documents are restricted in this manner.

Option 3: Consumers opt out of HIE mechanisms (provider to provider) for either discrete data or free text. Patients will have full access to their personal health record information using standards-based formats which can be managed directly by the patient, or an untethered[1] PHR of their choice, e.g. Google, Microsoft, etc. Patients could redact these records in any way they feel appropriate to protect their privacy. In options 2 and 3, consumers must be advised of the risks of incomplete information sharing, so they have informed consent for the clinical risks and burdens of not disclosing all information. There is no need to assign any value judgments or pejorative language to these options, but there is a need to educate and inform their consent options.

Practical Implications:

1) Consumers will need clear education as to the implications of these options, and specifically they should understand that redaction of free text can result in harm to them. It also places a larger burden on both the consumer and the provider to carefully review and update all information. (This is qualitatively the same as the legal burden today, but does amplify this burden quantitatively.)
2) Consumers will need to be educated that there is no mechanism for safe "automated redaction" of the free text components of their records, which is why the redaction of free text can only occur under their control, or by a proxy they assign to do that work.
3) If consumers consent to information exchange beyond discrete summary data, they have two options:
a) accept the risks of sending free text notes intact using HIE mechanisms, or
b) assume full accountability for maintaining a personal health record (PHR), and providing redacted versions of their PHR to whom, when, what, and where they choose.

Implications for addressing the core problems:

1) **Providers** will have confidence that they are receiving complete and reliable narratives whenever consumers authorize information exchange of all data including classical HIE (provider to provider

exchange of information). (That is the often the case in the current model of exchange of paper records and faxes today).

2) **Providers** will be subject to unpredictable redaction when relying on untethered PHRs for clinical information.
3) **Providers** will also acquire an obligation to interview the patient much more deliberately about key issues when using PHR as source information, recognizing the risk of undisclosed and idiosyncratic sensitivities and redactions.
4) **Consumers** can choose the privacy model that works for them.
5) **Consumers** must understand the risks to themselves when they redact sensitive information.
6) Existing models of authorization and exchange of paper information and data redaction will continue using current mechanisms.
7) Consumers could request the right to redact only a sub-segment of their record, which through its transmission as a PHR source would clearly declare "caveat emptor" for that component of their record, while preserving the usual presumed integrity of the remainder of their record.

Outstanding Issues:

1) This solution set depends on readily available PHRs that meet two specific requirements:
   a) common interoperability standards between PHRs and EHRs
   b) support for the individual consumer's ability to redact their information within a PHR that they maintain.
2) This solution set introduces the risk of a consumer "showing" PHR information to a provider and declining to allow the provider to keep a copy of that information in their EHR. This creates a legal liability for the physician in subsequent quality reviews and medico-legal disputes. Specifically, if they use information to make a clinical decision based on what they could view at that time, and they cannot reconstruct that information subsequently, their decision may appear to be indefensible.
3) This solution set will be difficult to implement unless it achieves universal support and becomes part of the NHIN framework.
4) Technical semantic support for provenance of data at an atomic level will be required to support the comingling of EHR and PHR data and subsequent transmission to 3rd and subsequent recipients of these same data. (The collateral benefit of this atomic-level provenance is the ability to propagate corrections of erroneous data across the HIE networks and all persistent datastores of those data. Absent this solution, error-correction of propagated and comingled data is impractical.)

[1]Tethered vs. untethered PHRs:
Tethered PHRs are direct derivatives from EHRs, and their contents represent a proper subset of the EHR itself. **Untethered PHRs** can be aggregated from multiple different sources and edited by individual consumers to varying degrees. There are different types of redaction which can occur. In general redaction and consumer-directed editorial maintenance of health records is better facilitated on untethered PHRs, but can be permitted and managed in either tethered or untethered forms. It is important to recognize that even with tethered PHRs, if the patient is allowed to both "add" and to "delete" data into their PHR, these features can be used to explicitly falsify the record as overcompensation for any sensitivity (e.g. declaring an HIV negative screening when the patient is known to be HIV+). Whenever the consumer has the right to "delete" autonomously, the data integrity of that record becomes an issue. Clearly, it is ethical for patients to withhold information initially, and to designate specific information they choose not to share freely. Consumers have done so since the time of Hippocrates. What is challenging is not so much the ethics of any deliberate misrepresentations, but the potential harm to that individual from inappropriate treatment resulting from that incomplete or falsified record, and hence the need to support the provenance that would warrant more of a "caveat emptor" attached to specific "at risk" elements of a record, whenever the provenance reveals passage through an unregulated environment such as an untethered PHR.