

Patient Linking Hearing

HIT Policy Committee - Privacy & Security Tiger Team

December 9, 2010

Sean Nolan, Chief Architect

Microsoft Health Solutions

Introduction

Thank you for the opportunity to participate in this discussion, one that we agree is of significant importance for the nation. Microsoft Health Solutions delivers healthcare products across the care continuum, every one of which is dependent on patient linking at some level. For example:

- **Amalga**, our real-time enterprise data insight system, aggregates patient information from dozens of systems within and between institutions.
- **Vergence**, our enterprise context management system, exists entirely to support effective management of identities across workflows.
- **HealthVault**, our consumer health platform, links consumer records across the incredibly fragmented data silos that together represent the totality of an individual's care.

While there is much to be discussed regarding advanced techniques for computational matching, it is crystal clear that there will never be a "perfect" solution to this problem. Even ignoring their political implications, unique identifiers do not provide a complete answer --- ID cards are lost, humans make mistakes and care must be delivered nonetheless.

We believe that what is most important to the national dialogue is to recognize that different use cases require different levels of confidence. It is important that we build this recognition into standards and regulatory requirements so that we do not inadvertently suppress innovation that could otherwise significantly advance the state of care.

In my testimony today, I will highlight two examples where different approaches to patient linking support real care scenarios, and look forward to the discussion and questions that they may elicit.

Differentiated Use Cases in the Enterprise

A key tenet of the Amalga system is to use the right information at the right time to answer questions. This demands that we retain a great deal of information *about* the information we manage – for example, the *degree of confidence* we have in any given patient match computation. With this metadata available, we can do the right thing in different situations. For example:

- When matching biometric readings such as an EKG, an extremely high degree of confidence is required --- if unique identifiers do not produce a "perfect" match, human review is almost certainly required before clinical action can be taken, despite the labor costs involved.

- When presenting a list of patient allergies, it may well be appropriate to accept a higher “false positive” rate in order to reduce the chances of “false negatives” --- because the ramifications of missing a potential allergy tend to be far more severe than those caused by “treating around” one that doesn’t actually exist.

In these cases, a system might show all matches with greater than 80% confidence, perhaps with an associated “confidence” user interface element that can alert the provider to look more closely if appropriate. This same approach may be taken in computation of drug interactions, or other similar safety-related use cases.

- When reporting aggregated information or performing population analysis, even lower confidence rates, or those computed by more experimental methods may be acceptable. In these cases, the error rate may be determined to statistically “wash out,” enabling useful and important population-level conclusions to be drawn.

In all of these cases, what is most important is not how matches are computed in the whole, but how match-related metadata is used to support a diverse set of use cases, each of which requires a different idea of how “clean” a match must be.

This approach also enables improvement over time as technology advances, supports offline validation of success rates, and is effective in encouraging the “resiliency and recovery” that Rich described so eloquently in his testimony today.

Privacy-Preserving Links for Consumer Records

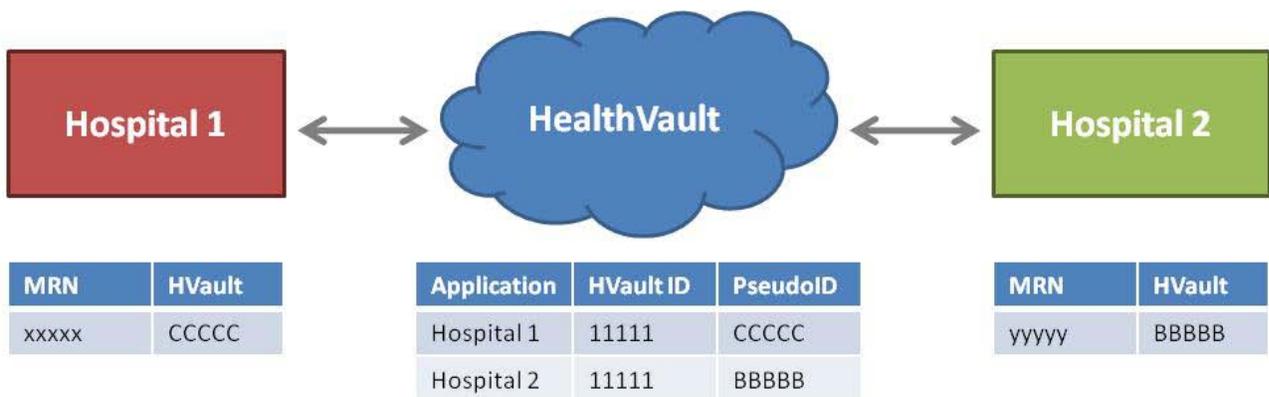
Consumer health platforms such as Microsoft HealthVault introduce another level of complexity into an already complex equation. HealthVault has created a sophisticated approach to matching that addresses and mitigates these challenges:

- Demographics provided through the Internet are generally self-asserted and difficult to trust as input to matching algorithms.
- Anything less than near-perfect match confidence is unacceptable because of the privacy and legal implications of releasing sensitive information inappropriately.
- Sharing unique identifiers across “loosely-coupled” boundaries carries significant privacy risk due to potential “collusion” between third parties. For example, if one service knows that patient “X” has HIV and another service knows that patient “X” lives at a particular address, together significant invasions of privacy become trivial.

To address the first two issues above, HealthVault has rejected traditional patient matching methods altogether. Instead, the system relies on a technique called “dual-credential presentation.” The idea is

that, if an individual can prove that they own a particular HealthVault record *and* that they are a specific patient at an institution, a link between HealthVault and the institution can be inferred. Each connecting institution decides how best to identify its patients – from conservative in-person proofing to more online-friendly credit-report-style questionnaires – and combines that with a HealthVault login to complete a link that complies with their specific policies.

The privacy risk identified in the third bullet above is addressed by giving each external system its own, randomly-generated identifier for connected HealthVault records. In the figure below, you can see that although both hospitals have relationships with the individual “11111” in HealthVault, each “knows” that patient by a different “pseudo-identifier.” In this way, HealthVault is able to act as a de-facto EMPI between loosely-coupled systems that have no knowledge of each other, encouraging reliable exchange without risk of collusion between them.



Thank You

The work that has been done by the Tiger Team since its formation earlier this year has been instrumental in accelerating real-world techniques for healthcare information exchange. I and Microsoft appreciate the opportunity to contribute as we collectively learn how to best continue to make progress.

Appendix: Direct answers to posed questions

In many cases, the responses for the questions differ significantly between our consumer and enterprise businesses. Where that is the case, I have split our response into two sections.

What are your standards for identifying individuals?

HealthVault

we never directly rely on traditional patient-matching technologies and techniques to connect individuals with their health information. Instead, we use the “dual-credential presentation” approach described above, in which both proof of ownership of a HealthVault record *and* proof of identity at each

connected institution must be provided together. Because the presenting individual has access to both systems and claims a linkage, we accept the claim as valid.

Amalga and Vergence

In our enterprise systems, we use a more traditional patient-matching approach that combines deterministic, probabilistic and manual matching as appropriate for each institution. We will often incorporate existing in-situ EMPI technologies into our products on site to accomplish this as well.

How can you be sure that you are accurately linking a patient with his/her data?

HealthVault

The decision to link is encapsulated within the institution's policies and techniques for releasing information to patients (as part of the dual-credential presentation approach). No decisions or linkages beyond this are explicitly drawn within HealthVault; all work done to verify linkage happens at the connecting institution.

Amalga and Vergence

The appropriate level of confidence for matching differs from use case to use case. In general, matching is "spot-checked" periodically using manual sampling techniques. For cases that cannot tolerate error, manual steps can be injected into clinical workflow where humans can verify or resolve matches. In these cases, our tools such as HealthVault Community Connect provide "side-by-side" comparisons of linked records to help ease the labor burden of inspection.

What problems are you having with patient-matching?

HealthVault

The burden of matching at each institution is higher than ideal, and certainly the ability to prove possession of a unique national identifier would simplify connections. However, this comes with a set of privacy implications that may not justify the benefit, especially recognizing that even such a system would still carry a non-zero rate of error.

Amalga and Vergence

There is a high engineering cost to effectively managing the "merge" and (in particular) "un-merge" events that can occur due to identified errors in patient matching. Data structures are more complex and costlier to maintain as a result of these relatively infrequent events. Consistent adoption and handling of internal identifiers across modalities could dramatically improve this situation.

What level of accuracy do you establish for patient matching?

In all cases (directly for enterprise and indirectly for consumer), we defer to our customers' policies and chosen technologies for establishing matches.

What lessons learned do you have from solving this problem?

We have addressed two of the key lessons for matching in our formal testimony above. Beyond these, it is clear that all data must be resilient to errors discovered after-the-fact. While this introduces an ongoing engineering cost, it is far less than the cost of dealing with errors post-hoc.

What are the cost implications of various solutions?

Any solution involving manual intervention has direct and obvious cost implications. Whenever possible, we attempt to avoid the requirement of manual intervention, and create user interfaces such as “side-by-side” comparison that increases the efficiency of manual work when it is required.

What should ONC do to address patient matching problems in information exchange?

We believe it is important to recognize that different use cases require different levels of confidence for patient matching. ONC should ensure that standards and regulations it contributes to recognize this situation and build into exchanges a sharing of metadata about matching --- how it was performed, confidence levels in objective form, source systems used to generate matches, and so on.

By doing this, ONC will help ensure that the maximum use cases will be enabled over time.

What are the solutions? What is the status of those solutions for healthcare?

What are the gaps?

There are a plethora of matching techniques being used in the market today, and very limited adoption of existing standards for matching. Entire companies are built around sophisticated “EMPI” technology suites, and they are doing an increasingly good job of optimizing probabilistic matching and manual workflows.

Within this rapidly evolving landscape, it is Microsoft’s belief that the appropriate action right now is to encourage the market to continue its progress. The likelihood of large-scale political changes such as development of a national patient identifier is remote at best, and as discussed would not completely solve the problem in any case.

By embracing this innovation and using a metadata model to communicate the state of patient matching during information exchange, standards and regulations can evolve in parallel with technology and best practices in this area.