Provenance and the Nationwide Health Information Network

Latanya Sweeney, PhD
Harvard University, MIT, Carnegie Mellon University

AdvanceHIT produces visionary papers and in-depth analyses to support HIT innovation. latanya@seas.harvard.edu AdvanceHIT.org

Disclaimer

The views and opinions in this presentation represent my own and are not necessarily those of HHS or the Obama Administration. These views are for the benefit of public discourse and public education, and are not necessarily an opinion regarding any position I may take on related issues decided by the HIT Policy Committee.

Thanks to

Margo Seltzer
and
the TaPP ‘10 Program Committee
and
my 2-year old son’s learning model

Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance, describe applications of provenance in the NHIN.


In progress and under advisement! latanya@seas.harvard.edu AdvanceHIT.org
Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance, describe applications of provenance in the NHIN.

**Provenance**

*Why*
- Describes why a particular piece of data is in the result set of a query.

*Where*
- Describes the source(s) from which a particular data element is derived.

*How*
- Describes how tuples in *where* provenances interact to produce a result.

---

Vision is to share health data widely. Evidence exists that doing so can offer significant improvements to patient care and dramatic reductions in costs.


Given actual and proposed nationwide health data sharing and research on provenance, describe applications of provenance in national health data sharing.
Health Data Divide

Claims
Sharing: insurers & others not include clinical data.
Optimized for income.
National connectivity processing billions of claims a year.

Clinical
Sharing: providers & others not include all local data.
Documents patient care.
Few, disparate local areas processing select data from some patient visits.

Medical Billing Network

Medical Billing Network (Verify)

Medical Billing Network (Payment)
Medical Billing Network (Datasets)

Primary Sharing MAY have some Restrictions

Secondary and Alternative Sharing Unbounded

Data sharing is hidden.
Patients have no control.
No basis for correcting values.
Potentially a billion dollar market [PWC 2009]

Clinical Data Sharing

Why: provide clinical info for billing claims & public health.
Where: reuse test results from reputable sources.
How: reconcile patient information across providers.

ARRA ("the stimulus bill") about $19-40B.

Clinical Data Sharing

Meaningful Uses
- CBOE for medication, laboratory, diagnostic, imaging, etc.
  CMS quality measure: % of orders so entered.
- Implement drug-drug, drug-allergy, drug-formulary checks.
- Maintain and up-to-date problem, medication, and allergy list
- Check insurance eligibility.
  CMS measure: % patient with insurance eligibility confirmed.
- Submit claims electronically.
  CMS measure: % claims submitted electronically to all payers.
- Exchange clinical information (e.g. medication list, allergy list).
  CMS measure: exchange clinical info (e.g. medication list).

Regional Exchanges
Connectivity? Data flow? Work flow?


1. How are records for the same patient identified as belonging to the same person?
2. What information is made available to which providers?
3. How is relevant patient information determined and consolidated and provided prior to the delivery of service?
4. How are data audits conducted across the network?
5. How is data provenance and integrity assured?
6. How are corrections made, identified, and replicated?
7. What incentives keep providers reporting and sharing information?
8. How do providers connect to the infrastructure and how are they authenticated?
9. How are national health quality measures assessed? … patient empowerment? … privacy?

Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance, describe applications of provenance in the NHIN.

Complex Provenance

I naively introduce the term complex provenance to describe problems whose aim involves one kind of provenance but whose solutions tend to introduce unwanted side effects in another kind of provenance.

In the next slides, I introduce 3 sample health data sharing applications involving complex provenance.

Deduplicated Accounting
**Deduplicated Accounting**

**Aim (Why)**
Goal is to get a distinct global count of patients matching a given criteria. Validation seeks to describe why a particular patient is in the result set.

**Complication (Where)**
Common approach is to provide identifiers (e.g., patient SSNs) rather than counts so recipient can perform deduplication. This generates a privacy concern by explicitly describing source(s).

---

**Testing & Liability**

**Aim (Where)**
Goal is to reuse test results. In order to do so, must make sure source is reputable and image is not modified and belongs to the correct patient and is accurate. (“digital signatures”)

**Complication (How)**
Providers are concerned about liability. Credentials are not the same as a trustworthy relationship. If a provider acts on results, even in part, he increases his liability.

---

**Direct Care Improvement**

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/27/2005</td>
<td>Pharmacy filled prescription for penicillin.</td>
</tr>
<tr>
<td>10/28/2005</td>
<td>Emergency room visit: diagnosis and procedures show allergic reaction to penicillin with an immediate hypersensitivity reaction.</td>
</tr>
<tr>
<td>1/3/2009</td>
<td>Diagnosis of pregnancy</td>
</tr>
</tbody>
</table>

Dr. Faye recommends an endocarditis prophylaxis and prescribes Biocef, orally. Life-threatening complications result because Eve did not remember, and Dr. Faye did not know, that Eve has a penicillin allergy with an immediate hypersensitivity reaction.
Direct Care Improvement

Aim (How)
Goal is to use consolidated information over time and from different providers to improve decision-making for the patient. Meaningful uses include patient reminders and drug-allergy interactions.

Complication (Why)
Need to decide whether a particular piece of data should be in the result set or whether it should be valued, or whether data may be missing.

Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance,

Describe applications of provenance in the NHIN.

-3 complex provenance applications
-2 provenance components

Data Segmentation

Goal is to provide patients with some privacy protection by allowing some data not to be shared.

How
Generates an “inference problem” about data relations that may allow the missing information to be learned.

Corrections

In the data sharing environments described so far, there is no mechanism for propagating corrections or updating information.

Why
Can generate false information when incorrect data is in the result set.
Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance, describe applications of provenance in the NHIN.

“resolving problems through the design of the NHIN”


Design 1: Global Query

Answer queries across a collection of repositories in real-time, providing the same answer as if all data were centralized.

Analogy: web searching.

Also: numeric results, e.g. counts, regressions, and percentages. Another form of query is data extraction.

Design 2: Certified Delivery

Point-to-point delivery of health information with verifiable and accountable endorsements of sender and receiver per contents. Analogy: email and fax.

Centralized (design 1) or de-centralized services (above).

Design 3: Patient Central

Uses globally available storage of patient information arranged by patient.

Analogy: file cabinet.

An EMR not “patient health record” (PHR).
Design 4: Medical Billing Backbone

Tethers or embeds meta-information along with data values. Meta information travels with data.

“Dependency tracking model” adds retrospective accountability. (shown)

“Sticky policies” affixed to data values to represent data sharing allowances, prohibitions and consents.

Design 5: Heavy Data

Design 6: Pointer Addressing

Share published network addresses of data rather than data values themselves.

Pointer addressing (a) and its use with heavy data using grid computing technologies (b).

Base Case Designs

Fax and Email
Common method currently done.

NHIN Connect
Software operations: (1) patient lookup; (2) document query; (3) document retrieval; (4) audit log query; (5) authorized case follow-up; and, (6) event messaging.

Secure Email
Provides send email over secure channels to combat eavesdropping.
Given proposed meaningful uses for the nationwide health information network (NHIN) and research on provenance, describe applications of provenance in the NHIN.