ADVANCEHIT

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

---

Thanks to

Margo Seltzer

and

the TaPP '10 Program Committee

and
my 2-year old son's learning model

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.

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

Describe applications of provenance in the NHIN.

Sweeney, L. *Provenance and the Nationwide Health Information Network*. Advance HIT Project. White Paper 1004. Harvard University. 2010.
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.

Buneman et al. Why and Where: A Characterization of Data Provenance. *ACM SIGMOD*, 316-330, London, UK, Jan 2001.
Green et al. Provenance semirings. *PODS* '07, 31--40, Beijing, China, 2007.

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

Describe applications of provenance in the NHIN.

Vision is to share health data widely.

Vision is to share health data widely.

Evidence exists that doing so can offer significant improvements to patient care and dramatic reductions in costs.

flow device needed.

Chaudhry et al., Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine*, Vol. 144, No. 10, 2006.

Given actual and proposed nationwide health data sharing and research on provenance,

Describe applications of provenance in national health data sharing.

Health Data Divide

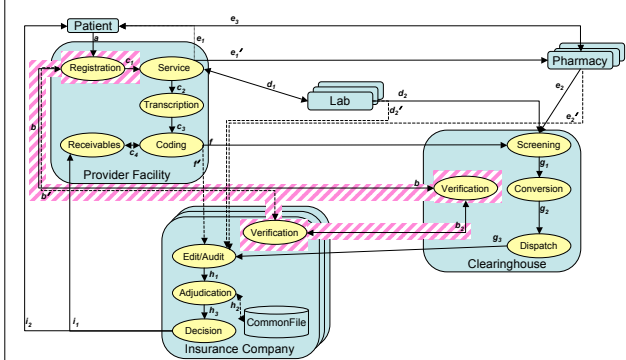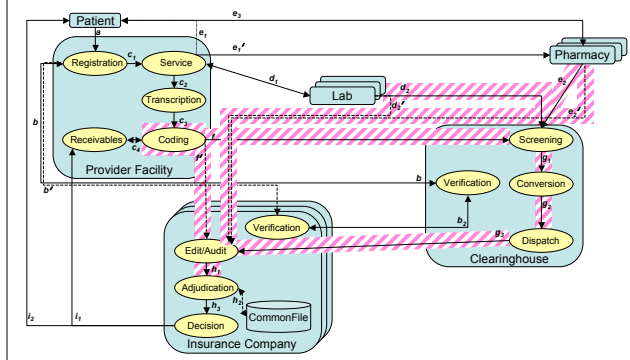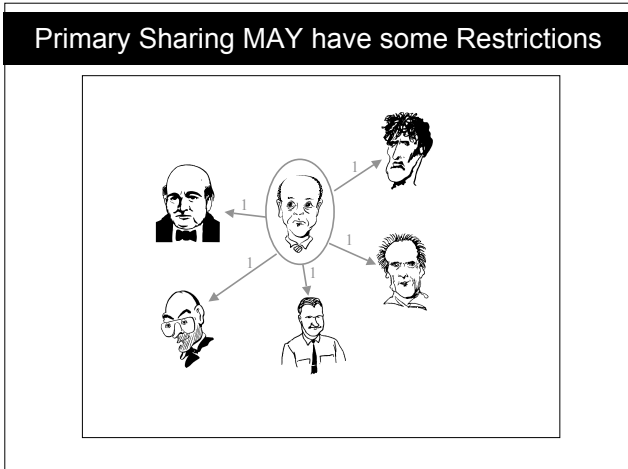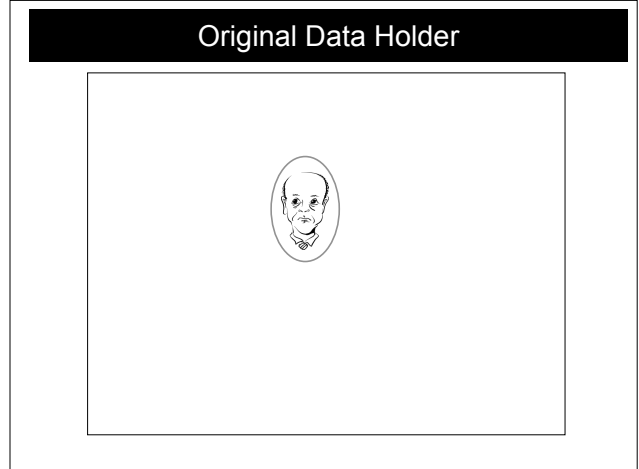| Claims	Clinical
Sharing: insurers & others not include clinical data.	Sharing: providers & others not include all local data.
Optimized for income.	Documents patient care.
National connectivity processing billions of claims a year.	Few, disparate local areas processing select data from some patient visits.

Medical Billing Network
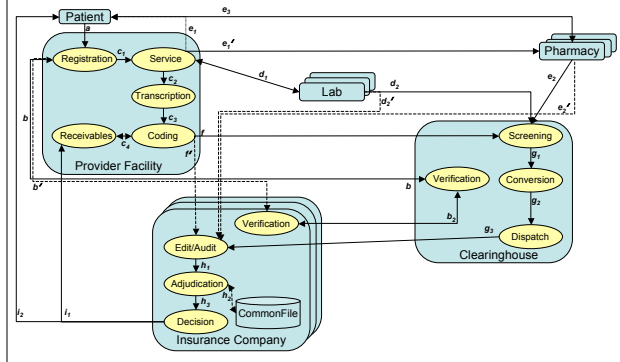
Medical Billing Network (Verify)

Medical Billing Network (Payment)

Medical Billing Network (Datasets)

Original Data Holder

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]

Sweeney, L. Information explosion. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, Urban Institute, Washington, DC, 2001.

Why: provide clinical info for billing claims & public health.

Clinical Data Sharing

Where: reuse test results from reputable sources.

How: reconcile patient information across providers.

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

Clinical Data Sharing

American Recovery and Reinvestment Act of 2009 ("ARRA") Pub. L. 111-5.å

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).

Federal HIT Policy Committee, Final Meaningful Use Objectives and Measures: 2011-2013-2015. At <healthit.hhs.gov>

Regional Exchanges

Connectivity? Data flow? Work flow?

National Committee on Vital and Health Statistics. *Assuring a Health Dimension for the National Information Infrastructure: a concept paper*. Presented to the U.S. Department of Health and Human Services Data Council. October 14, 1998. www.ncvhs.hhs.gov/hii-nii.htm

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, consolidated and provided prior to the point of service?
4. How are data audits conducted?
5. How is data provenance...
6. How are ...
7. ... and sharing
8. ... to the infrastructure and how are th...
9. How ... ional health quality measures assessed?
… patient empowerment? … privacy?

Gap in readiness! More about design solutions later in this talk.

Sweeney, L. *Background for Competitive Designs for the National Health Information Infrastructure*. Advance HIT Project. White Paper 1002. Carnegie Mellon University. 2009.

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

Patient1
Patient2
Patient3
Patient4
Patient5

Provider1
Provider2
Provider3
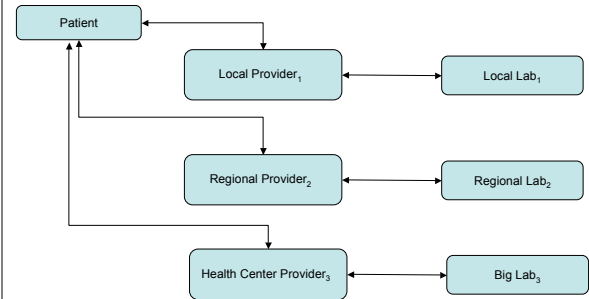
Region 1
Region 2
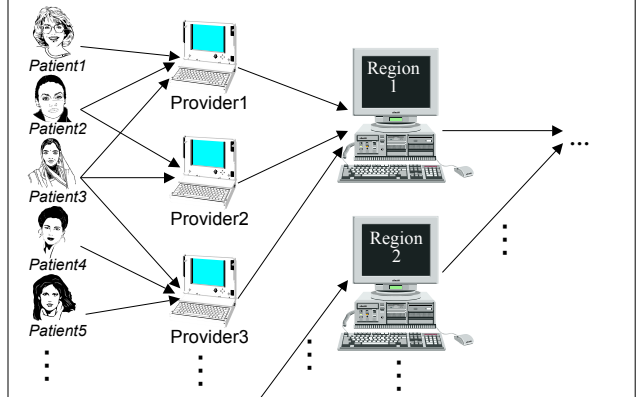
Deduplicated Accounting

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

Patient — Local Provider$_1$ — Local Lab$_1$
Regional Provider$_2$ — Regional Lab$_2$
Health Center Provider$_3$ — Big Lab$_3$

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

Date	Description
10/27/2005[2]	Pharmacy filled prescription for penicillin.
10/28/2005	Emergency room visit: diagnosis and procedures show allergic reaction to penicillin with an immediate hypersensitivity reaction.
1/3/2009	Diagnosis of pregnancy
…	… visits related to pregnancy

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.

7

Direct Care Improvement

Aim (How)

Goal is 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"

Sweeney et al. *Trustworthy Designs for the NHIN*. Advance HIT Project. White Paper 1003. Harvard University. 2010.
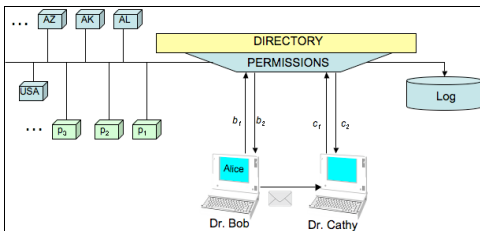
Design 1: Global Query
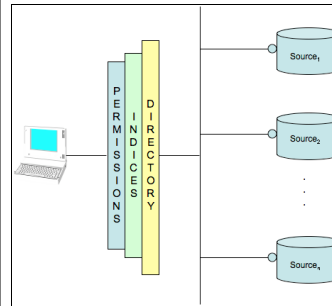
Answer queries across a collection of repositories in real-time, providing the same answer as if all data were centralized.

Analogy: web searching.

Also: numeric results, e.g. counts, regressions, and percentages. Another form of query is data extraction.
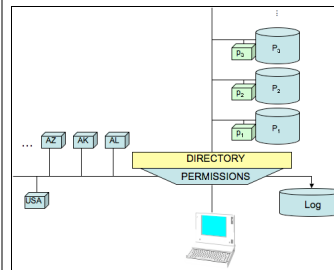
Design 2: Certified Delivery

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

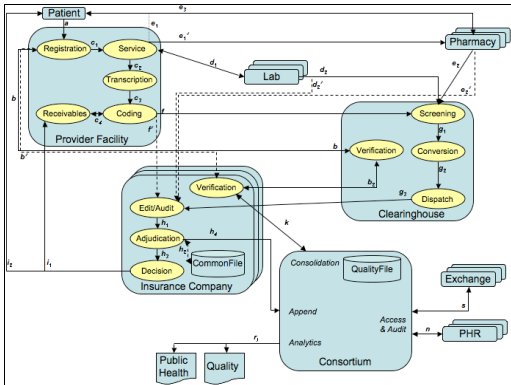Centralized (design 1) or de-centralized services (above).

Design 3: Patient Central

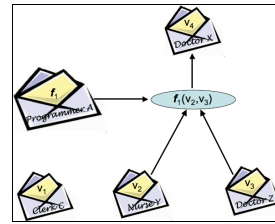Uses globally available storage of patient information arranged by patient.

Analogy: file cabinet.

An EMR not "patient health record" (PHR).

Design 4: Medical Billing Backbone
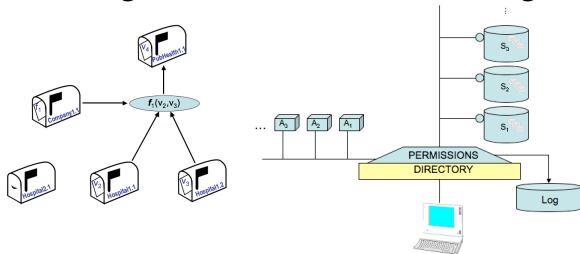

Design 5: Heavy Data

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 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.

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