Inter-institutional Data Sharing: Policy and Practice

October 26, 2010
Nick Anderson, Ph.D.
Assistant Professor, Biomedical Health Informatics
Adjunct Assistant Professor, Bioethics and Humanities
Associate Director, Biomedical Informatics
Institute of Translational Health Sciences
University of Washington
Overview

• The evolving why and what of research clinical data sharing
• Policy and practice
• Evolution of a data sharing system by example: CICTR/HOMERUN
Intersections: policy and politics

• National expectations to increase data sharing (AHRQ, NIH, ONC HITECH,)
• Privacy disclosures could significantly affect community trust$^1,2$
• Significant new challenges and risks associated with privacy breaches (enhanced HIPAA)


2. Yarborough M, et al. Transforming the culture of biomedical research from compliance to trustworthiness: insights from nonmedical sectors. 2009 Apr;
Resetting (presently low) expectations

• Rare sharing of data led to results on Alzheimers (NYT Aug 12, 2010)
• Data’s Shameful Neglect (Nature Editorial 2009)
• Empty Archives (Nature 2009)
• Desperately Seeking Cures (Newsweek May 2010)
The Future: Open and Share

"Somebody has to share. If we all hold on it, we all lose it."

George Church, Personal Genome Project, May 12, 2009, Chicago Public Lecture
Policy and stewardship

• Share what now?
• Who owns repositories, who owns data, who owns derivative data, who owns obligation to downsteam secondary owners?
• What are the trade-offs between identifiability, protection and research utility?
• Is there a precedent? Does this matter?
• Does the patient/community care or have a say?
• Does the city, state or government care or have a say?
Evolving research data sharing modes

• Forms
  – None (status quo)
  – Inter-investigator
  – Federated inter-institutional
  – Grid-based collaborations
  – Researcher/institution to community/patient
  – Federal mandated (dbGAP + others)

• Drivers
  – Large population sets/new analysis techniques
  – Federal mandates
  – Community pressure
Opposite directions of evolving policy approaches

1. Greater de-identification/ security
   - Large scale efforts
   - Limited understanding of secondary consequences

2. Enhanced interaction with patients
   - More constrained population sizes
   - Patient managed consents
   - Participatory governance
“The design of systems determines the kinds of politics that can take place in them, and designing a system is itself a political act”

Mitch Kapor – Electronic Frontier Foundation
**Architecture Design Considerations**

<table>
<thead>
<tr>
<th>Centralized</th>
<th>Federated</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Loss of control of local data</td>
<td>+ Local control of data</td>
</tr>
<tr>
<td>- Centralized resource needs own $</td>
<td>- Local sites need own $</td>
</tr>
<tr>
<td>- Local sites need some $</td>
<td>+ No central resource $</td>
</tr>
<tr>
<td>- Unbalanced relationships “what do I get/what must I give” –</td>
<td>+ Be a seed or a leech</td>
</tr>
<tr>
<td>+ up front expectation setting</td>
<td>+ Ad hoc participation</td>
</tr>
<tr>
<td>+ Centralized access roles</td>
<td>- Decentralized access roles (this is shifting)</td>
</tr>
<tr>
<td>+ Single main point for security</td>
<td>- Multiple points of security</td>
</tr>
<tr>
<td></td>
<td>+ Local security locally controlled</td>
</tr>
<tr>
<td>- Query speed and data requests a function of central schema/concurrent</td>
<td>+ Query speed and data requests remain</td>
</tr>
<tr>
<td>uses, size of resource</td>
<td>constant, or as fast as the slowest active</td>
</tr>
<tr>
<td></td>
<td>node</td>
</tr>
<tr>
<td>- Scalability: resource can vanish if loss of $ or commitment</td>
<td>+ Scalability: network can grow beyond</td>
</tr>
<tr>
<td></td>
<td>local funding/commitment</td>
</tr>
</tbody>
</table>
Constellation of Policy Issues

• Privacy, confidentiality, identifiability
• Data use, data access
• Returning results, incidental findings, consent, notification, benefit sharing
• Stewardship, governance
• Intellectual property, public private partnerships
Inter-institutional Examples

• Multi-institution PACS of radiological studies for secondary readings in clinical trials (caGRID)
• “Prep to study” discovery of potential patient #’s to support clinical trials or aid in hypothesis generating (i2b2)
• Discovery of human studies descriptive data to support re-use in future trials (caGRID/i2b2)
• Multi-institutional query capabilities to support pharmacovigilance (i2b2)
• Institutionally based Comparative Effectiveness networks (i2b2/caGRID)
Data Sharing via Shared Expectations
Federated Policy Approaches

• Assume all control begins with local IRB approval
• Assume limited ability to control secondary downstream uses
• Focus on adherence to law (HIPAA), human subjects protection, reduction of risk,
• Assessments: 1. “repository” ownership -> 2. uses -> then 3. users
Challenges

• Dependence on and varying interpretations of anonymization vs. de-identification
• Uneven local implementations – difficult to harmonize across sites
• Coordinating Data Use Agreements vs Individual IRB approvals vs Project IRB’s
• Data Use Agreements still nascent, limited scope (see caBIG DSIC KC for some examples), untested
Assessing Risk

• Significant institutional sensitivity to the use of such systems
  - potential disruptive technologies
  - common desire to use, less to lead

• Levels of risk inherent in extracting, delivering and analyzing different modalities of data
  – HIPAA-2 disclosure, competitive business intelligence, public relations
Assessing Utility

- Semantic and syntactic transformation to support de-identification has implications on utility
- Coordinating development of policy requires common use/semantic models
- Limited policy or systems available to normalize complex resource to researcher relationships
- Hypothesis: best researcher use is a governed approach that puts them in the query-seat
  - How to facilitate this? Currently data is technically “not human subjects”, yet sensitivity and emphasis on secure control remains
Intersecting policy, ethics and data utility

- Regulations are just the floor of data sharing policy
- Tie policy to standardized terminologies, where possible (see HITSP IS 158 Clinical research)
- Tie standards and policy to interpretation of HIPAA
- Tie all above to secure and auditable systems
- Evaluate impact of above:
  - Strict interpretation of HIPAA may remove at-risk patients from data sets
  - Obfuscating zip code renders patients in rural communities “invisible”
- Develop policy that plans for responding to the worst case, hopes for the best case. Revisit regularly.
CICTR/HOMERUN Project
- Data sharing by example

- 4.2 million patients
- 2 states/3 institutions – moving to 4 states/7 institutions Q1/11
- I2b2/SHRINE
- Local IRB/ad-hoc DUA
- One-way de-identified, obfuscated aggregate, blurred
- Redact or “zero-out” classes of sensitive populations as to protect them
- Common terminology mappings/user interfaces
Multi-Institutional Use-cases and Users

• Anonymized cohort discovery for clinical trial recruitment
  – Current: aggregate counts and institutional source
  – Future:
    » Descriptive metadata
    » Local HIPAA de-identified Limited Data Sets

• Intended users:
  – Clinical translational investigators/study teams
  – Informaticians
  – Terminologists
  – Comparative Effectiveness Research
i2b2CICTR: Iterative Tech, Policy, Use, Users
(funded) Initiatives where informatics, data sharing and policy are intersecting

- eMERGE (ELSI)
- CTSA (KFCs on ethics/Informatics, groups on standards, *omics, data sharing)
- CEER (Center for Excellence in Ethical, Legal and Social Research)
- caHUB (Cancer Human Biobank), DSIC (caBIG)
- SHARP (ONC/NIH)
Lessons in Developing Policy from Health Care Reform in Bimodal House

• You might not be able to achieve consensus
  ... But you can at least consider standards
• You might not be able to accommodate everyone’s preferences and values
  .. But you will need advocate/stakeholders to test this
• You sometimes just need to make policy decisions that will based on core values, best evidence, and best available options
• Your solution for today may not be your solution tomorrow
Questions

• nicka@uw.edu