iDASH HIPAA-Cloud
Sharing and Protecting Human Subjects Data

May 2015
IDI presentation

Lucila Ohno-Machado, MD, MBA, PhD
Biomedical Informatics, University of California San Diego
Models for ‘Sharing’ Data

Clinical Data Network – UC-ReX

- Clinical Data Warehouses from 5 Medical Centers and affiliated institutions exchange (>13 million patients)
  - Translational research
  - Patient safety surveillance
  - Quality improvement

UC-ReX was formed in 2010
Currently using i2b2/SHRINE

Request for Results on Cohort

Results on Harmonized Data

- Currently
  - Count queries (i.e., how many >65yo with afib taking dabigatran had a stroke?)
  - Demographics, meds, labs, diagnoses. Data quality checks

- Future steps
  - Individual-level patient data delivery (recruitment)
  - Cross-institutional analytics

Data Matching Function: Map onto data dictionaries, use Data Warehouses derived from EHRs
Predictive modeling and adjustment for cofounders require lots of data.
Some institutions cannot move data outside their firewalls, we can bring computation to the data.

User requests data for Quality Improvement or Research

Diverse Healthcare Entities in 3 different states (federal, state, private)

Trusted Broker(s)
- Identity & Trust Management
- Policy enforcement

Security Entity

Phase 1

21 Million people

Standardized data

Data governance

9 health systems

funded by pcori
Phase 2

31 Million people

Same standards

Data governance

23 health systems

University of Washington CTSA
Critical Access Hospital Network 20k
WWAMI Region Practice & Research Network
UC ReX
UC Davis 2.3M
UCSF 3.2M
UCLA 4.3M
UC Irvine 300k
UCSD 2.3M
San Mateo Medical Center 77k
Cedars-Sinai 2M

Family Medicine Residency of Idaho 39k
Idaho State University Pocatello Family Medicine 16k
Intermountain Healthcare 2M
VA VINCI 11M
SAFTiNet
University of Colorado Health System 672k

Bi-State Primary Care Association 122k
Cherokee Health Systems 89k
Denver Health Hospital Authority 150k
Colorado Community Managed Care Network 350k

USC Keck 2M
LA Children’s 200k
LA DHS 600K
Altamed 200k
Children’s Clinic 24k
Queenscare 19k

Bi-­‐State	
Primary	Care
AssociaBon	
122k
Cherokee	Health Systems	89k
Denver	Health Hospital	Authority	150k
Colorado	Community Managed	Care Network 350k

USC

SCANNER

University of Texas Houston (NLP)
Breaches Affecting 500 or More Individuals

As required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals. These breaches are now posted in a new, more accessible format that allows users to search and sort the posted breaches. Additionally, this new format includes brief summaries of the breach cases that OCR has investigated and closed, as well as the names of private practice providers who have reported breaches of unsecured protected health information to the Secretary. The following breaches have been reported to the Secretary:

**Full Dataset** [CSV format (18 KB) XML format (57 KB)]

Select a column head to sort by that column. Select again to reverse the sort order. Select an individual record to display it in full below the table.

<table>
<thead>
<tr>
<th>Name of Covered Entity</th>
<th>State</th>
<th>Individuals Affected</th>
<th>Date of Breach</th>
<th>Type of Breach</th>
<th>Location of Breach Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st response Medical Transport Corp.</td>
<td>MD</td>
<td>552</td>
<td>06/15/2012-10/01/2012</td>
<td>Unauthorized Access/Disclosure</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>ABQ HealthPartners</td>
<td>NM</td>
<td>778</td>
<td>2012-12-20</td>
<td>Theft</td>
<td>Laptop</td>
</tr>
<tr>
<td>Accendo</td>
<td>AZ</td>
<td>175,350</td>
<td>2011-01-01</td>
<td>Unauthorized Access/Disclosure</td>
<td>Paper</td>
</tr>
<tr>
<td>Access Medical Group</td>
<td>PR</td>
<td>7,606</td>
<td>2012-01-11</td>
<td>Theft</td>
<td>Laptop</td>
</tr>
</tbody>
</table>

Supported by the NIH Grant U54 HL108460 to the University of California, San Diego
Whole Exome Pipeline Workflow

**NORMAL DNA**
- .fastq
- .bam
- .refined.bam
- .realigned.bam
- Variant calling (VarScan)
- .vcf
- .annotated.vcf

**TUMOR DNA**
- .fastq
- .bam
- .refined.bam
- .realigned.bam
- Copy Number calling (VarScan)
- .copy_number

Alignment (BWA)
Duplicate removal (Picard)
Quality Recalibration (GATK)
Indel Realignment (GATK)

**Databases**
- dbNSFP
- ExAC
- dbSNP
- COSMIC

**Variant Annotation**
- (Oncotator, VariantTools)

**Pipeline Infrastructure**
- OmicsPipe
- Seq-Ware

**QC and Metadata**
- Variant counts
- Overlap with databases
- Substitution Profiles

Supported by the NIH Grant U54 HL108460 to the University of California, San Diego
Genomes are Biometrics

- Biometrics are Protected Health Information (PHI)
- PHI requires HIPAA
  - Biometrics require HIPAA
iDASH On-Demand Resources

On-demand
Virtualized
Elastic
Resilient
Compute
And
Storage
Technology

AUTOMATED
✓ Compute nodes
✓ Memory
✓ Disk storage
✓ Networking

compute request,
direct upload & download of proprietary data, tool, recipe

upload & download data

safe HIPAA-compliant Annotated Data deposit box Environment

Safe HIPAA-compliant Annotated
Data deposit box Environment

middleware and HIPAA security developed by iDASH

HIPAA and non-public data

public data, tools, recipes

Powered by VMware

Powered by MIDAS

Supported by the NIH Grant U54 HL108460 to the University of California, San Diego
Repeatable Results

Workflow
- Short reads
- Index reference
- Align to reference
- Call variants
- Annotate variants
- Pick high impact
- Deleterious SNPs

Blueprint
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Instance
- Reference DB
- Configuration
- Helper tools
- OS

External Data

Context
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Workflow
- Short reads
- Index reference
- Align to reference
- Call variants
- Annotate variants
- Pick high impact
- Deleterious SNPs

Blueprint
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Instance
- Reference DB
- Configuration
- Helper tools
- OS

External Data

Context
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Workflow
- Short reads
- Index reference
- Align to reference
- Call variants
- Annotate variants
- Pick high impact
- Deleterious SNPs

Blueprint
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Instance
- Reference DB
- Configuration
- Helper tools
- OS

External Data

Context
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Workflow
- Short reads
- Index reference
- Align to reference
- Call variants
- Annotate variants
- Pick high impact
- Deleterious SNPs

Blueprint
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Instance
- Reference DB
- Configuration
- Helper tools
- OS

External Data

Context
- Reference DB
- Test data
- Configuration
- Helper tools
- OS

Supported by the NIH Grant U54 HL108460 to the University of California, San Diego
Plans

- Scale up the infrastructure (cores, disk, network)
  - the plan was shared with IDI and there are many ways to collaborate and mutually benefit from a common vision

- Serve HIPAA-storage and compute needs of a larger UCSD community
  - Expand library of workflows/pipelines
  - FISMA ATO

- Hub infrastructure for large HIPAA-data networks
  - Continue data extraction services and privacy technology research