



Research Data Networks: Privacy-Preserving Sharing of Protected Health Information

Lucila Ohno-Machado, MD, PhD

Division of Biomedical Informatics

University of California San Diego

21st Century Healthcare



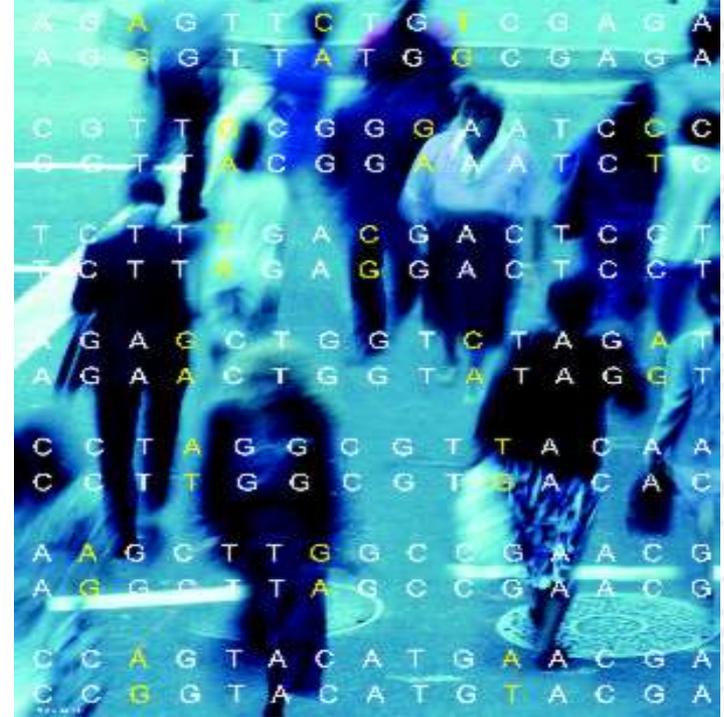
What is the influence of genetics, environment?

What therapies work best for individual patients?



Patient-Centered Outcomes Research

- Genome
 - Arrays, sequencing
- Phenome
 - Personal monitoring
 - Blood pressure, glucose
 - Personal Health Records
 - Behavior monitoring
 - Adherence to medication, exercise
- Environment
 - Air sensors, food quality
 - Location



Source: DOE

Personalized Medicine

Prevention, Diagnosis and Therapy

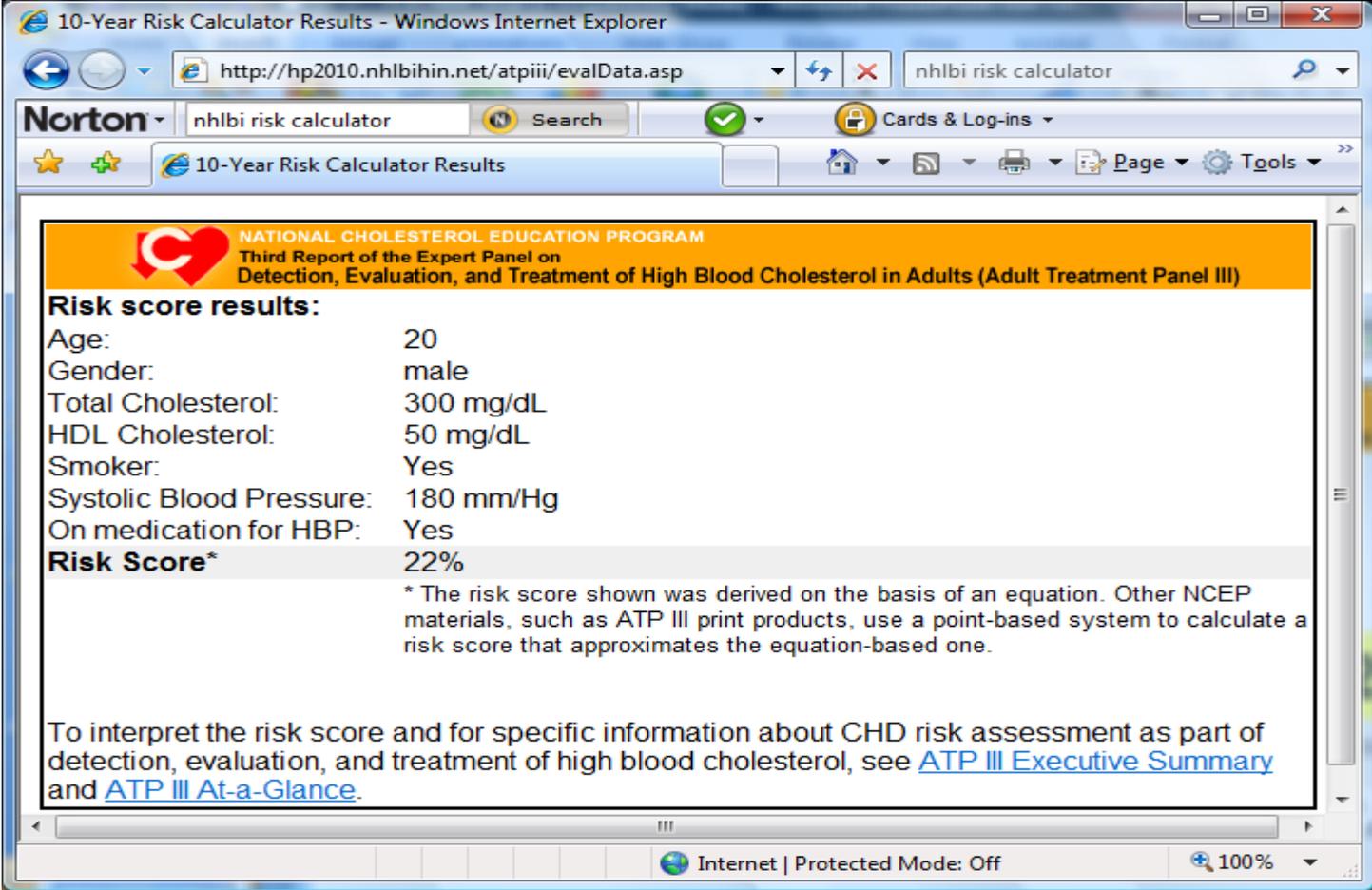
- Genetic predisposition
- Biomarkers
- Pharmacogenomics



Requirement for Handling Big PHI Data

- Secure Electronic Environment
 - Electronic Health Records
 - Genetic Data

Practical Risk Assessment by Clinicians



10-Year Risk Calculator Results - Windows Internet Explorer

http://hp2010.nhlbi.nih.net/atpiii/evalData.asp

nhlbi risk calculator

NATIONAL CHOLESTEROL EDUCATION PROGRAM
Third Report of the Expert Panel on
Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)

Risk score results:

Age:	20
Gender:	male
Total Cholesterol:	300 mg/dL
HDL Cholesterol:	50 mg/dL
Smoker:	Yes
Systolic Blood Pressure:	180 mm/Hg
On medication for HBP:	Yes
Risk Score*	22%

* The risk score shown was derived on the basis of an equation. Other NCEP materials, such as ATP III print products, use a point-based system to calculate a risk score that approximates the equation-based one.

To interpret the risk score and for specific information about CHD risk assessment as part of detection, evaluation, and treatment of high blood cholesterol, see [ATP III Executive Summary](#) and [ATP III At-a-Glance](#).

Internet | Protected Mode: Off 100%

Examples of Drugs with Genetic Information in Their Labels

Table 2. Examples of Drugs with Genetic Information in Their Labels.*

Drug	Sponsor	Indication	Gene or Genotype	Effect of Genotype	Clinical Directive on Label
Abacavir (Ziagen)	GlaxoSmithKline	HIV-1	<i>HLA-B*5701</i>	Hypersensitivity	Black-box warning: "Prior to initiating therapy with abacavir, screening for the <i>HLA-B*5701</i> allele is recommended." "Your doctor can determine with a blood test if you have this gene variation."
Azathioprine (Imuran)	Prometheus	Renal allograft transplantation, rheumatoid arthritis	<i>TPMT*2</i> , <i>TPMT*3A</i> , and <i>TPMT*3C</i>	Severe myelotoxicity	" <i>TPMT</i> genotyping or phenotyping can help identify patients who are at an increased risk for developing Imuran toxicity." "Phenotyping and genotyping methods are commercially available."
Carbamazepine (Tegretol)	Novartis	Epilepsy, trigeminal neuralgia	<i>HLA-B*1502</i>	Stevens–Johnson syndrome or toxic epidermal necrolysis	Black-box warning: "Patients with ancestry in genetically at-risk populations should be screened for the presence of <i>HLA-B*1502</i> prior to initiating treatment with Tegretol. Patients testing positive for the allele should not be treated with Tegretol." "For genetically at-risk patients, high-resolution <i>HLA-B*1502</i> typing is recommended."
Cetuximab (Erbix)	Imclone	Metastatic colorectal cancer	<i>KRAS</i> mutations	Efficacy	"Retrospective subset analyses of metastatic or advanced colorectal cancer trials have not shown a treatment benefit for Erbitux in patients whose tumors had <i>KRAS</i> mutations in codon 12 or 13. Use of Erbitux is not recommended for the treatment of colorectal cancer with these mutations."
Clopidogrel (Plavix)	Bristol-Myers Squibb	Anticoagulation	<i>CYP2C19*2*3</i>	Efficacy	"Tests are available to identify a patient's <i>CYP2C19</i> genotype; these tests can be used as an aid in determining therapeutic strategy. Consider alternative treatment or treatment strategies in patients identified as <i>CYP2C19</i> poor metabolizers."
Irinotecan (Camptosar)	Pfizer	Metastatic colorectal cancer	<i>UGT1A1*28</i>	Diarrhea, neutropenia	"A reduction in the starting dose by at least one level of Camptosar should be considered for patients known to be homozygous for the <i>UGT1A1*28</i> allele." "A laboratory test is available to determine the <i>UGT1A1</i> status of patients."
Panitumumab (Vectibix)	Amgen	Metastatic colorectal cancer	<i>KRAS</i> mutations	Efficacy	"Retrospective subset analyses of metastatic colorectal cancer trials have not shown a treatment benefit for Vectibix in patients whose tumors had <i>KRAS</i> mutations in codon 12 or 13. Use of Vectibix is not recommended for the treatment of colorectal cancer with these mutations."
Trastuzumab (Herceptin)	Genentech	HER2-positive breast cancer	HER2 expression	Efficacy	"Detection of HER2 protein overexpression is necessary for selection of patients appropriate for Herceptin therapy because these are the only patients studied and for whom benefit has been shown." "Several FDA-approved commercial assays are available to aid in the selection of breast cancer and metastatic gastric cancer patients for Herceptin therapy."
Warfarin (Coumadin)	Bristol-Myers Squibb	Venous thrombosis	<i>CYP2C9*2*3</i> and <i>VKORC1</i> variants	Bleeding complications	Includes the following table: Range of Expected Therapeutic Warfarin Doses Based on <i>CYP2C9</i> and <i>VKORC1</i> Genotypes.

* All drug labels were accessed through Drugs@FDA at www.accessdata.fda.gov/scripts/cder/drugsatfda. HIV-1 denotes human immunodeficiency virus type 1, *TPMT* thiopurine methyltransferase, *UGT1A1* UDP glucuronosyltransferase 1 family polypeptide A1, and *VKORC1* vitamin K epoxide reductase complex subunit 1.

Needed Decision Support for Clinicians

The screenshot shows the OpenEMR web interface in a Mozilla Firefox browser window. The interface includes a menu bar (File, Edit, View, History, Bookmarks, Tools, Help) and a status bar indicating the user is logged in as Thomas. The main content area is titled "Prescriptions" and contains a form for adding a new prescription. The form fields are as follows:

- Currently Active:
- Starting Date: August 26, 2009
- Provider: Thomas
- Drug: Lipitor
- Buttons: Drug Lookup, Save
- Notes: (empty text area)
- Add to Medication List: No Yes substitution allowed

A yellow callout box is overlaid on the form, containing the following text:

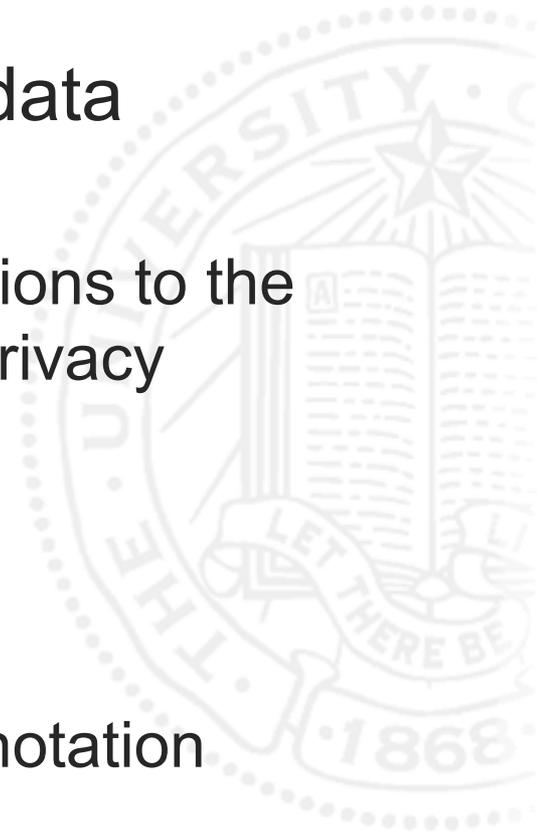
This patient has genotype
VKORC1 GG and CYP2C9 *1*1

Start Warfarin at 5 -7 mg

The left sidebar of the interface shows a navigation menu with categories such as Patient/Client, Medical Record, Fees, Administration, Reports, and Miscellaneous.

How can we accelerate research?

- Build infrastructure to access large data repositories
 - Enhance policy and technological solutions to the problem of individual and institutional privacy
 - Lower the barriers to share data
- Share tools to analyze the data
 - Meta-data: data harmonization and annotation
 - Algorithms and computational facilities



Best Practices and Minimal Standards

Systematic Reviews
(3,057 documents)

- Architectures
- Data harmonization
- Governance
- Privacy protection



Technical Report

Standards in the Use of Collaborative or Distributed Data Networks in Patient Centered Outcomes Research

Principal Investigator: Lucila Ohno-Machado, UCSD

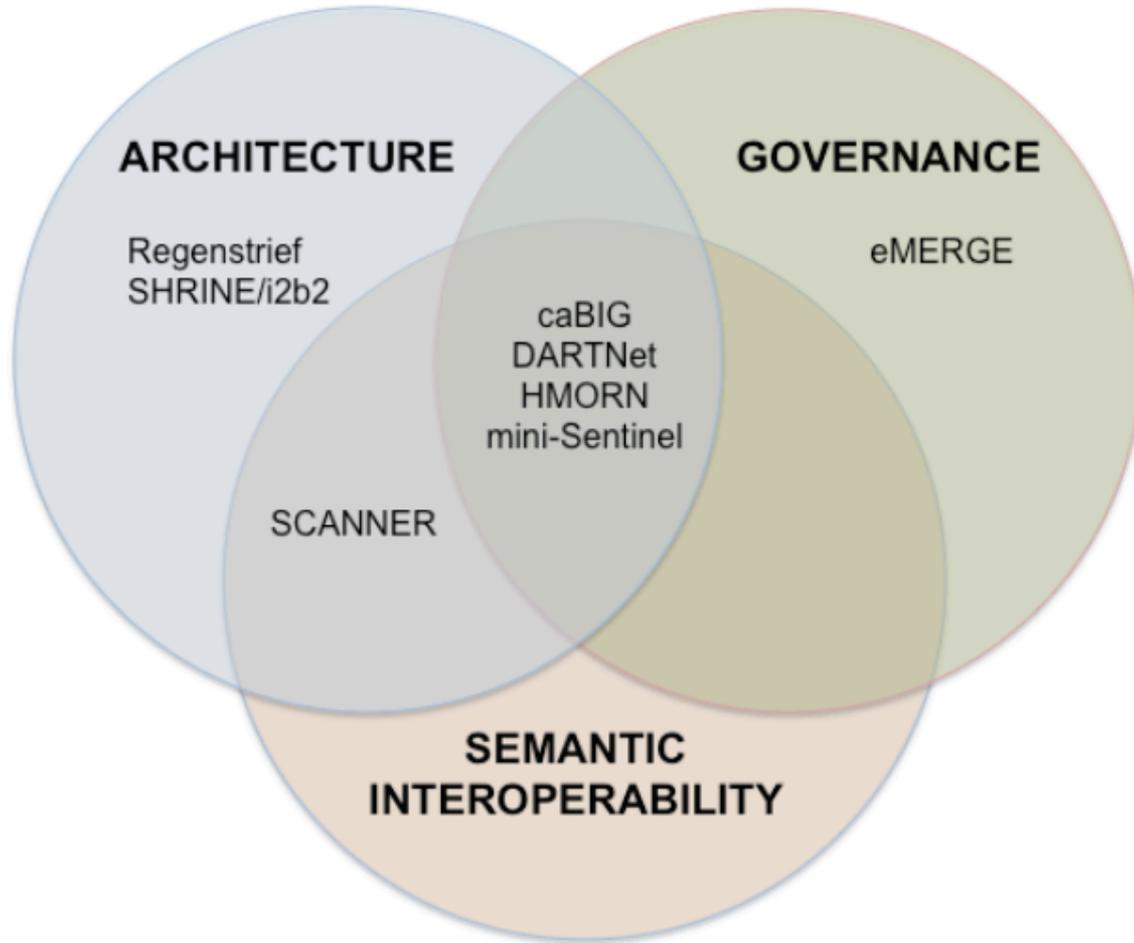
<u>UCSD</u> Michele E. Day Robert El-Kareh Xiaoqian Jiang Hyeoneui Kim	<u>RAND Corporation</u> Daniella Meeker	<u>SFSU</u> Dennis Browe Katherine Kim
--	--	--

Consultants
Jason Doctor (USC)
Aziz Boxwala (UCSD)
Claudiu Farcas (UCSD)
Deven McGraw (Center for Democracy & Technology)
Jaiideep Vaidya (Rutgers University)

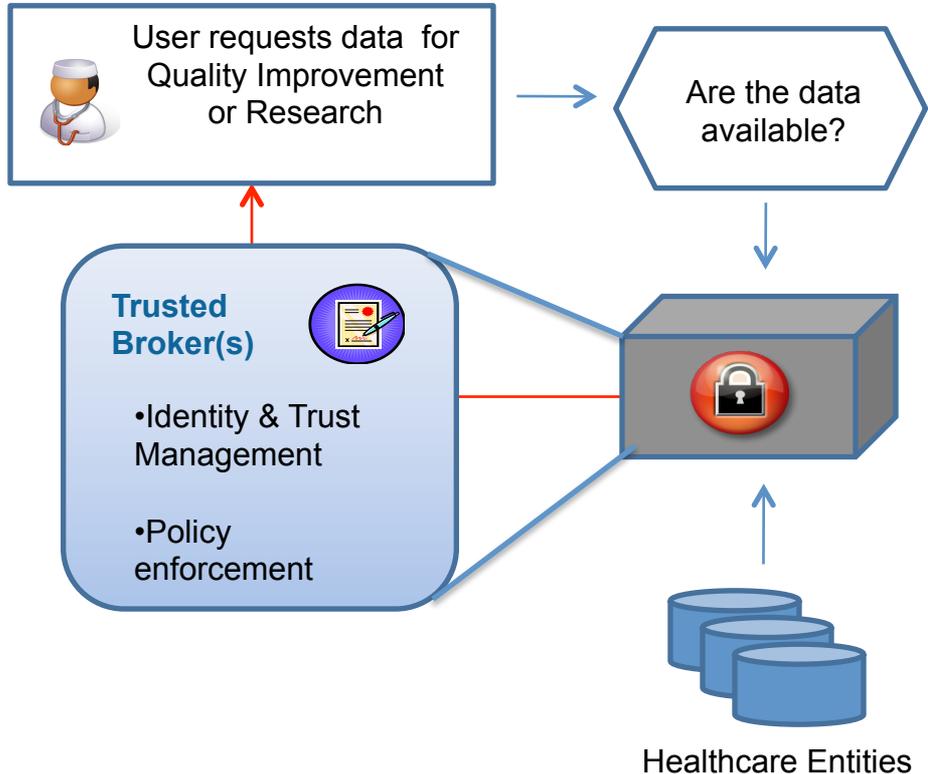
March 15, 2012

UC San Diego
RAND CORPORATION
SF STATE

Some examples



QI and Clinical Research Data Networks



- Scalable networks for comparative effectiveness research
- Re-usable infrastructures to lower barriers to add
 - Policies
 - Studies
 - Institutions

Example: UC ReX - Research eXchange



Funded by the UC Office of the President to the CTSA

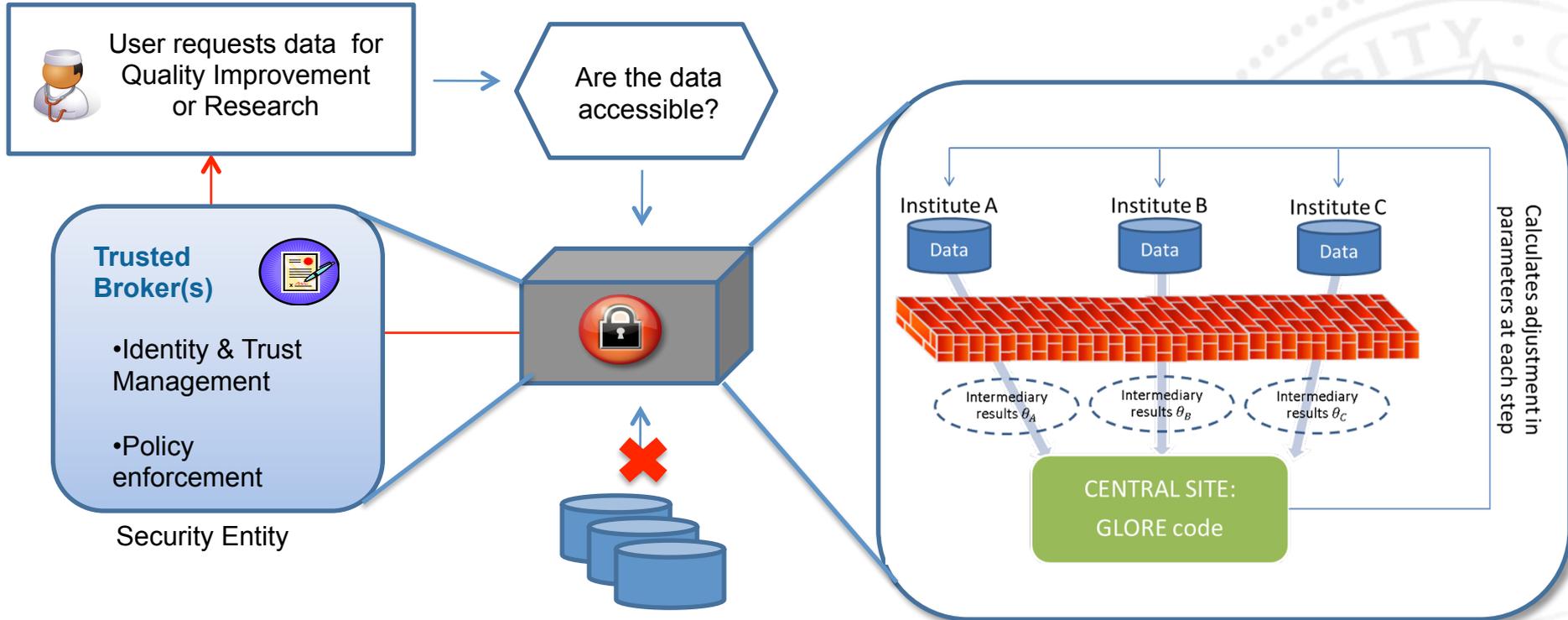
- Current plans: Integration of Clinical Data Warehouses from 5 Medical Centers and affiliated institutions (>10 million patients)
 - Aggregate and individual-level patient data will be accessible according to data use agreements and IRB approval
- Future plans: Integration with clinical trial management systems, biorepositories

Privacy Protection

- Use of clinical, experimental, and genetic data for **research**
 - not primarily for clinical practice (i.e., not for health care)
 - not primarily for quality improvement (i.e., not for IRB exempt activities – regulatory ethics committee)
- Data networks must host and disseminate data according to
 - Federal and state rules and regulations
 - Data owner (e.g., institutional) requirements
 - Consents from individuals



QI and Clinical Research Data Networks



AHRQ R01HS19913 / EDM forum

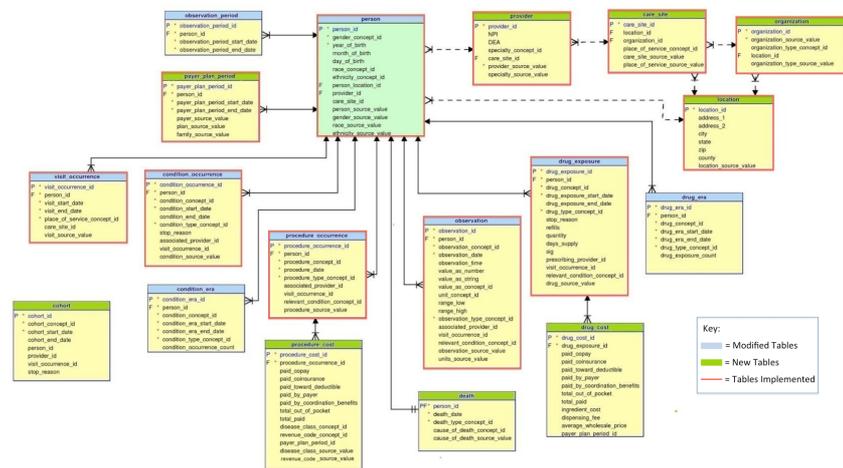
SCANNER

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

Wu Y et al. Grid Binary Logistic Regression (GLORE): Building Shared Models Without Sharing Data. *JAMIA* 2012

Summary of recommendations

- Data Harmonization
 - Common data model
 - Meta-data
- Privacy
 - Access controls, audits
 - Encryption
 - Assess risk of re-identification



- Architectures
 - Distributed
 - Centralized

Models for Data Sharing

- **Cloud Storage:** data exported for computation elsewhere
 - Users download data from the cloud
- **Cloud Compute and Virtualization:** computation goes to the data
 - Users query data in the cloud
 - Users upload algorithms to the cloud

National Centers for Biomedical Computing

Home

NCBC Summary

Calendar

All Hands Meetings

Biological Projects

Biositemaps Projects

Working Group Archive

[Search for NCBC resources in the new Resource Discovery System \(RDS\)](#)



Informatics for Integrating Biology and the Bedside
Brigham and Women's Hospital

National Alliance for Medical
Imaging Computing

National Center for
Integrative Biomedical Informatics
University of Michigan

National Center for Biomedical Ontology
Physics-based Simulation of Biological Structures
Stanford University

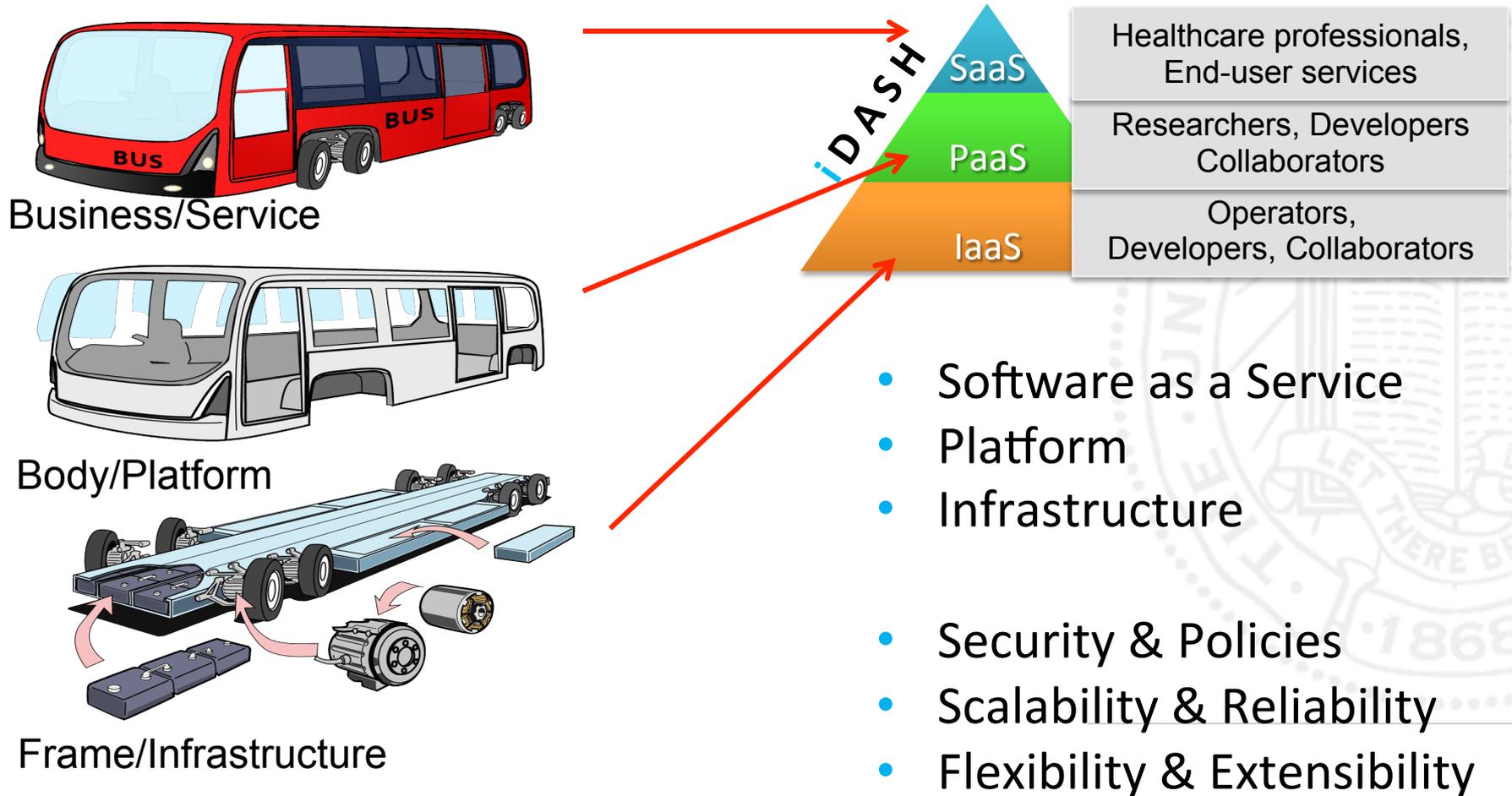
Center for Computational Biology
University of California at Los Angeles

National Center for Multi-Scale
Study of Cellular Networks
Columbia University

iDASH

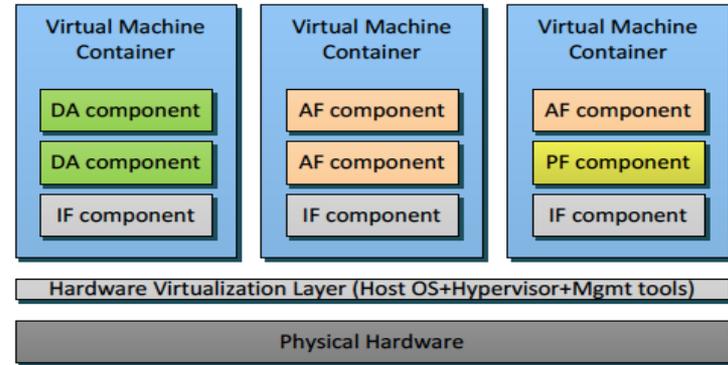
Integrating Data for Analysis, Anonymization and SHaring (iDASH)
University of California at San Diego

Shared Services and Infrastructure



Shared Infrastructure

- 315TB Cloud and project storage for 100s of virtual servers
- 54TB high-speed database and system storage; high-performance parallel databases
- 10Gb redundant network environment; firewall and IDS to address HIPAA requirements
- Multiple-site encrypted storage of critical data



Research data from several institutions:
Clinical & genomic data hosting in a HIPAA compliant facility



Summary of recommendations

- Data Harmonization
 - Common data model
 - Meta-data
- Privacy
 - Access controls, audits
 - Encryption
 - Assess risk of re-identification
- Architectures
 - Distributed
 - Centralized
- Governing body
 - Data use agreements
 - Policy for IP
 - Consent
 - **Include stakeholders**

Patient-Centered Data Sharing

