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

Lucila Ohno-Machado, MD, PhD
Division of Biomedical Informatics
University of California San Diego

PCORI Workshop 7/2/12

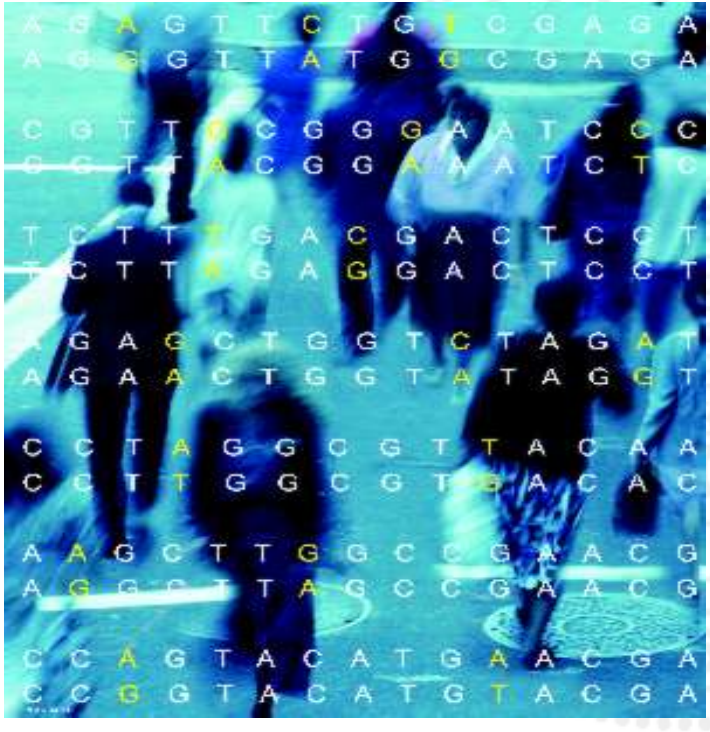
21st Century Healthcare

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

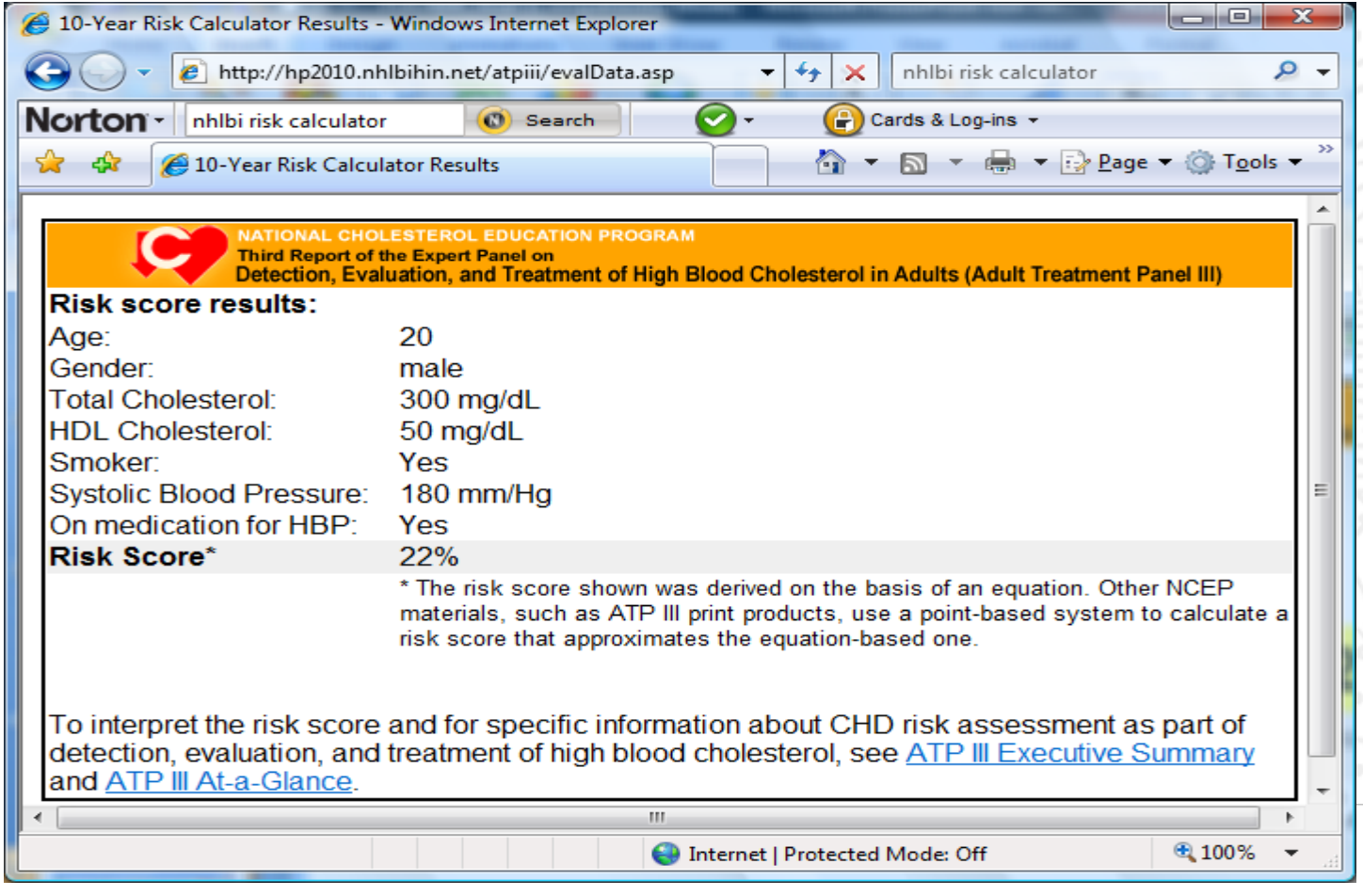
Prevention, Diagnosis and Therapy

- Genetic predisposition
- Biomarkers
- Pharmacogenomics

Requirement for Handling Big PHI Data

- Secure Electronic Environment
  - Electronic Health Records
  - Genetic Data

Practical Risk Assessment by Clinicians



Examples of Drugs with Genetic Information in Their Labels

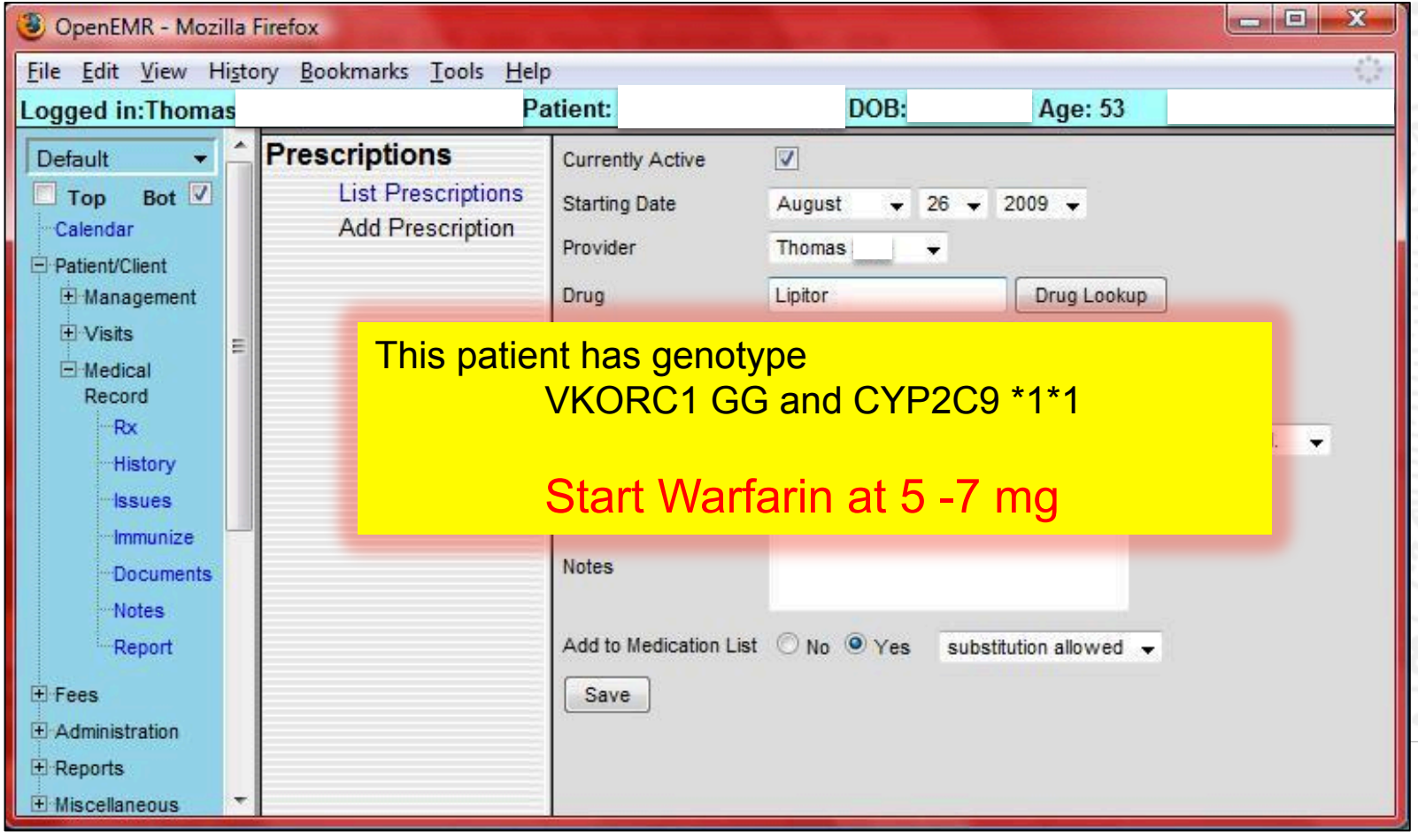
Table 2. Examples of Drugs with Genetic Information in Their Labels.*

Drug	Sponsor	Indication	Gene or Genotype	Effect of Genotype	Clinical Directive on Label
Abacavir (Ziagen)	GlaxoSmithKline	HIV-1	*HLA-B*5701*	Hypersensitivity	Black-box warning: "Prior to initiating therapy with abacavir, screening for the *HLA-B*5701* 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	*TPMT*2*, *TPMT*3A*, and *TPMT*3C*	Severe myelotoxicity	"*TPMT* 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	*HLA-B*1502*	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 *HLA-B*1502* 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 *HLA-B*1502* typing is recommended."
Cetuximab (Erbitux)	Imclone	Metastatic colorectal cancer	*KRAS* 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 *KRAS* 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	*CYP2C19*2*3*	Efficacy	"Tests are available to identify a patient's *CYP2C19* genotype; these tests can be used as an aid in determining therapeutic strategy. Consider alternative treatment or treatment strategies in patients identified as CYP2C19 poor metabolizers."
Irinotecan (Camptosar)	Pfizer	Metastatic colorectal cancer	*UGT1A1*28*	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 *UGT1A1*28* allele." "A laboratory test is available to determine the *UGT1A1* status of patients."
Panitumumab (Vectibix)	Amgen	Metastatic colorectal cancer	*KRAS* mutations	Efficacy	"Retrospective subset analyses of metastatic colorectal cancer trials have not shown a treatment benefit for Vectibix in patients whose tumors had *KRAS* 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	*CYP2C9*2*3* and *VKORC1* variants	Bleeding complications	Includes the following table: Range of Expected Therapeutic Warfarin Doses Based on *CYP2C9* and *VKORC1* 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.

Hudson KL. *N Engl J Med* 2011

Needed Decision Support for Clinicians



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

commissioned by pcori

Technical Report

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

Principal Investigator: Lucila Ohno-Machado, UCSD

UCSD
Michele E. Day
Robert El-Kareh
Xiaoqian Jiang
Hyeoneui Kim

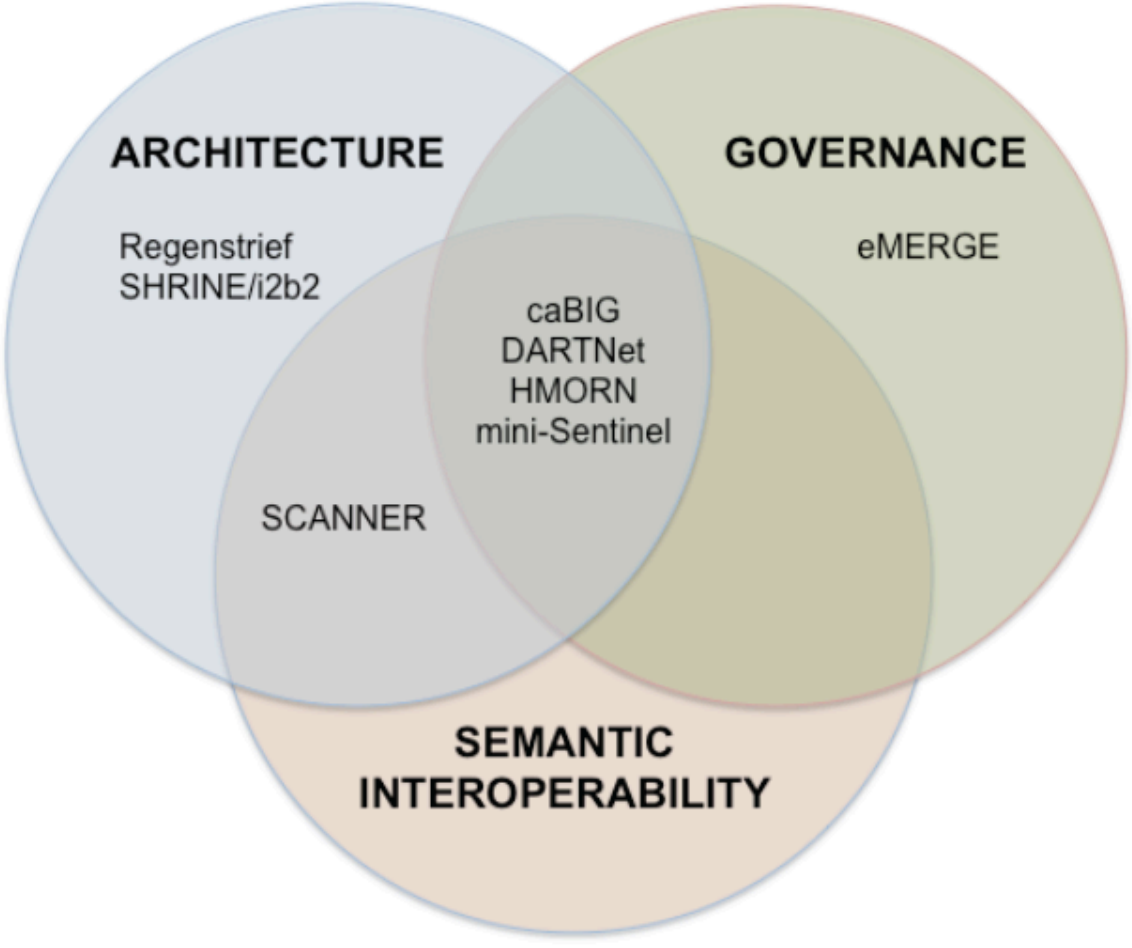
RAND Corporation
Daniella Meeker

SFSU
Dennis Browe
Katherine Kim

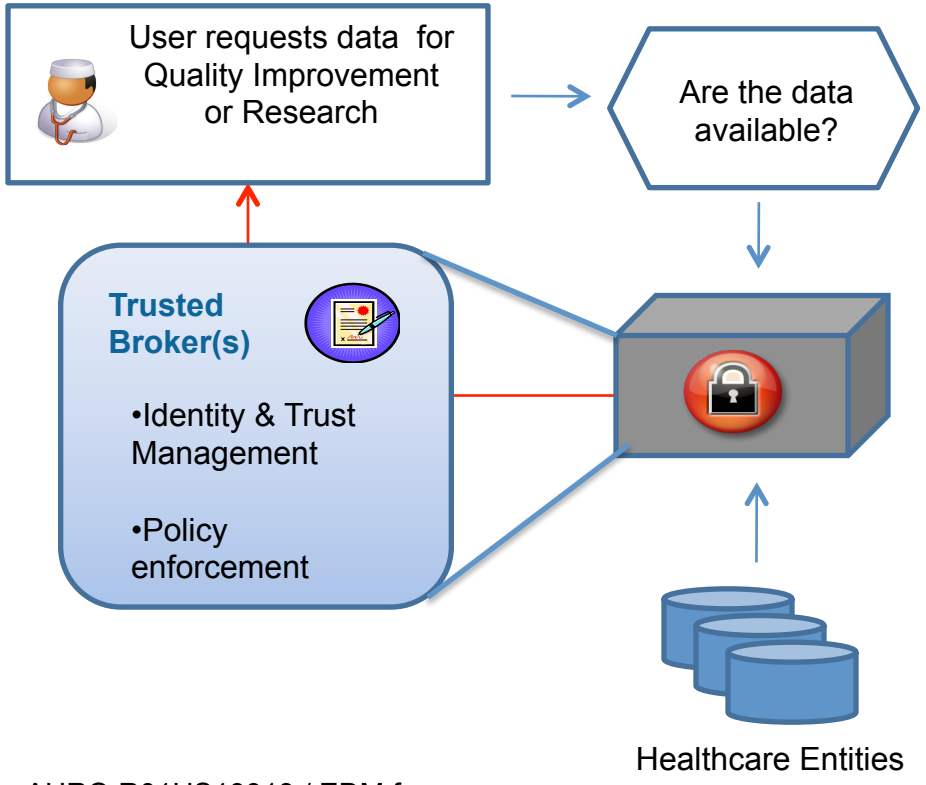
Consultants
Jason Doctor (USC)
Aziz Boxwala (UCSD)
Claudiu Farcas (UCSD)
Deven McGraw (Center for Democracy & Technology)
Jaideep Vaidya (Rutgers University)

March 15, 2012

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

AHRQ R01HS19913 / EDM forum

SCANNER

Example: UC ReX - Research eXchange

- 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

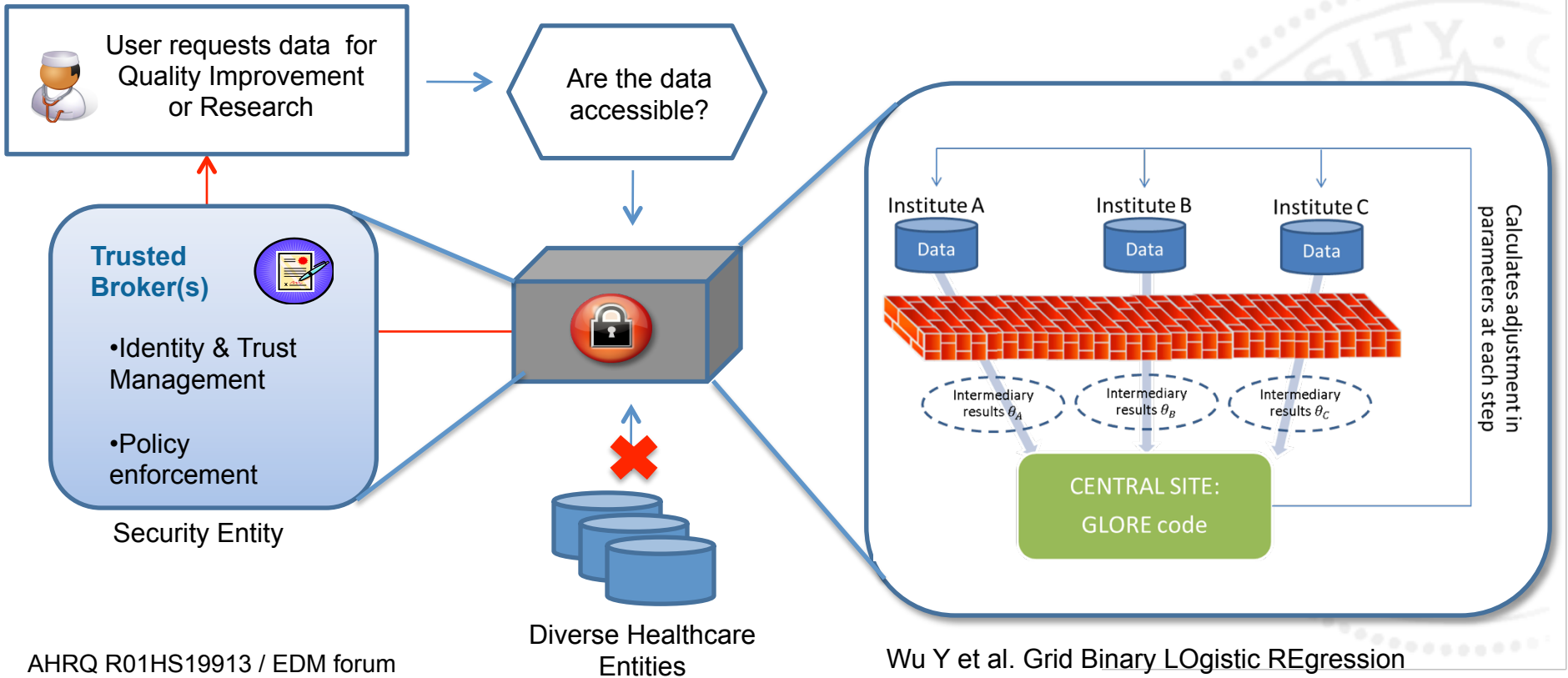
Funded by the UC Office of the President to the CTSAs

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



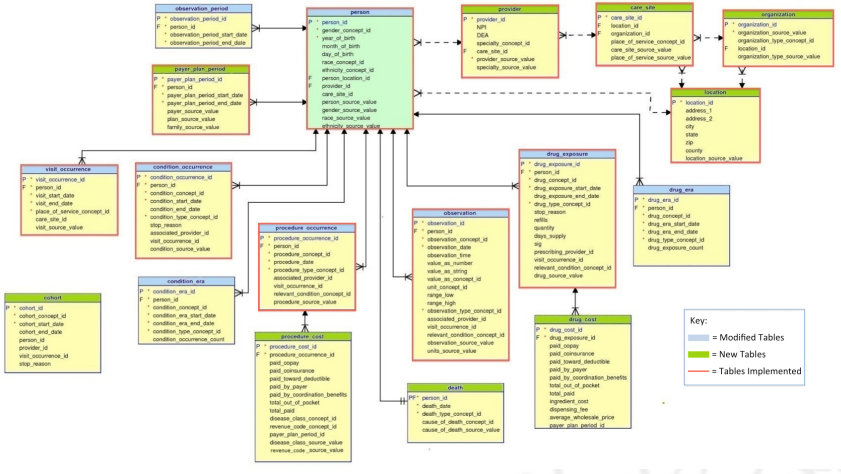
AHRQ R01HS19913 / EDM forum

SCANNER

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

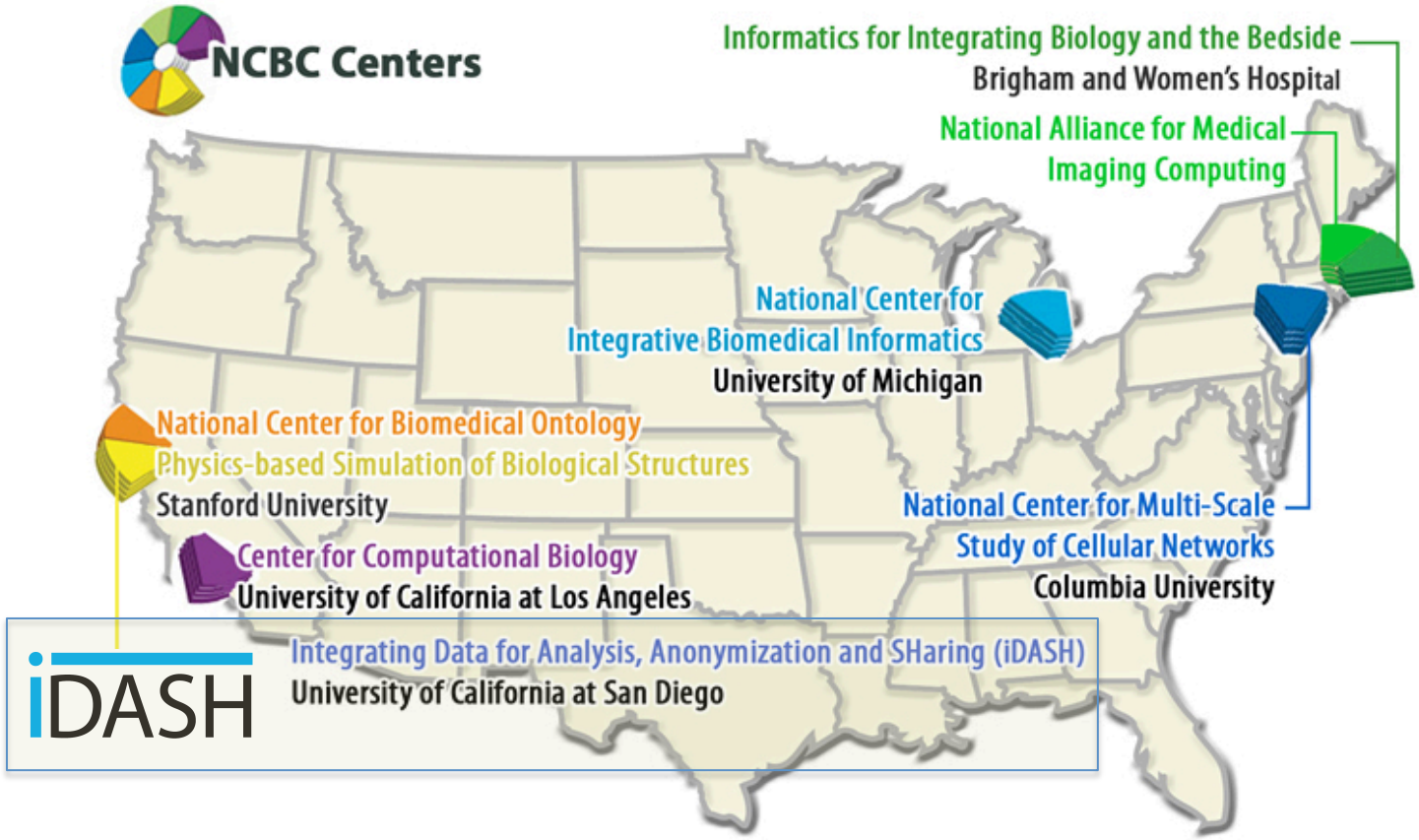
Models for Data Sharing

- 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

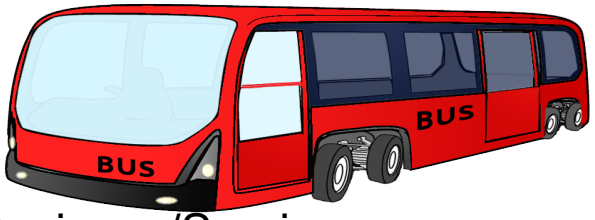
NIH Roadmap

National Centers for Biomedical Computing

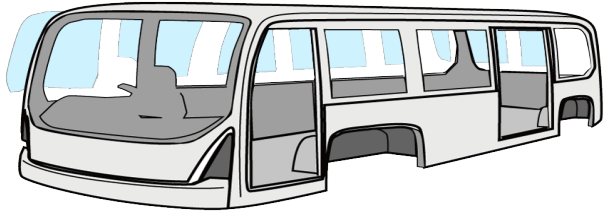
Search for NCBC resources in the new Resource Discovery System (RDS)



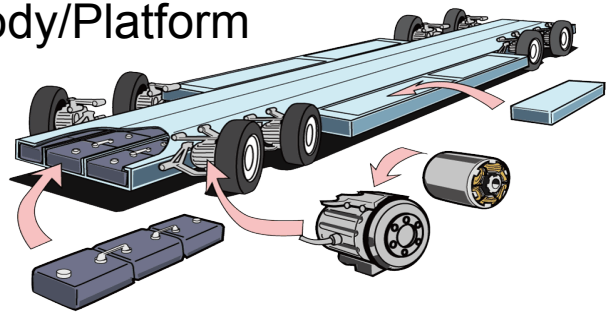
Shared Services and Infrastructure



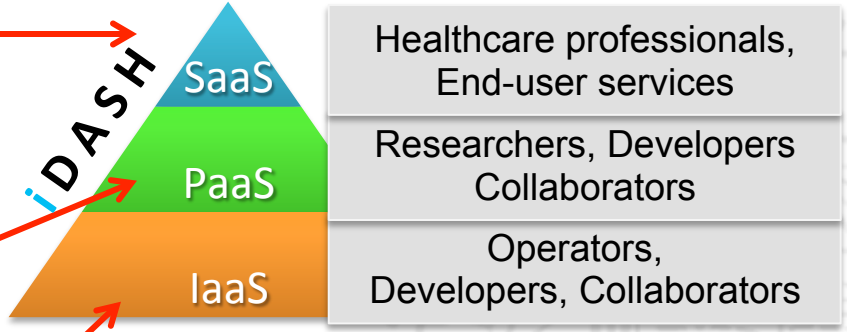
Business/Service



Body/Platform



Frame/Infrastructure

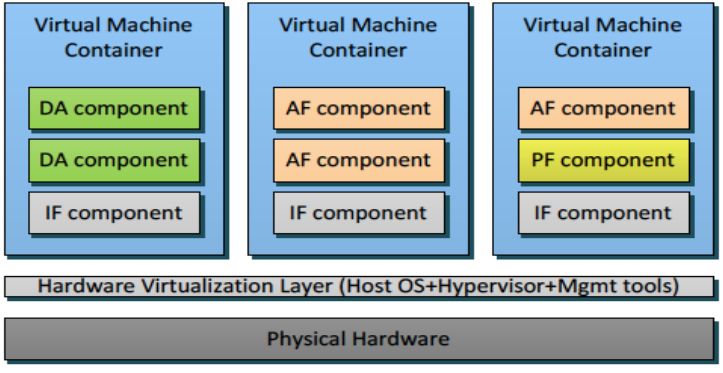


- Software as a Service
- Platform
- Infrastructure

- Security & Policies
- Scalability & Reliability
- Flexibility & Extensibility

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

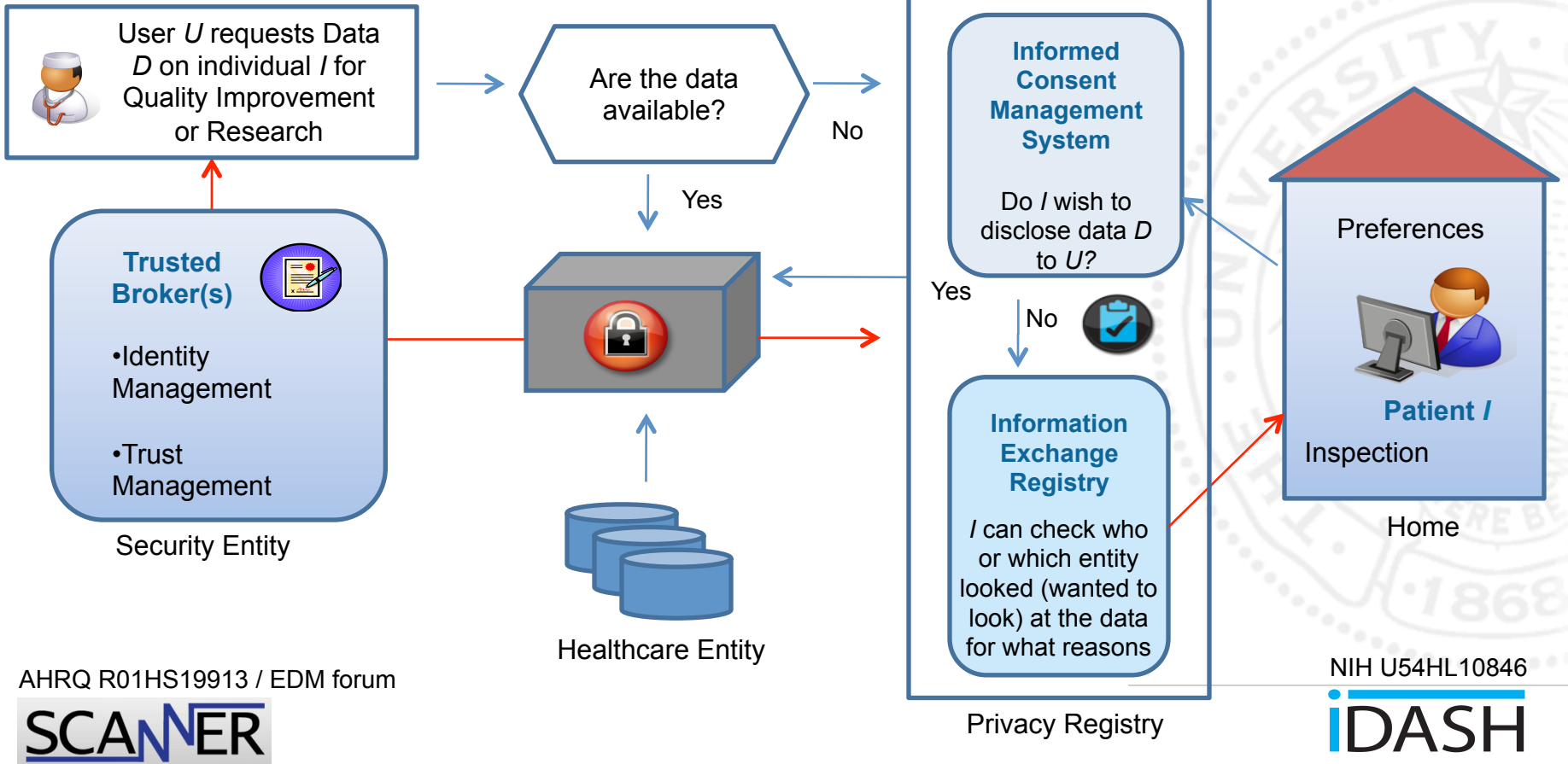


Research data from several institutions:
Clinical & genomic data hosting in a HIPAA compliant facility

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



AHRQ R01HS19913 / EDM forum
SCANNER
NIH U54HL10846
iDASH