

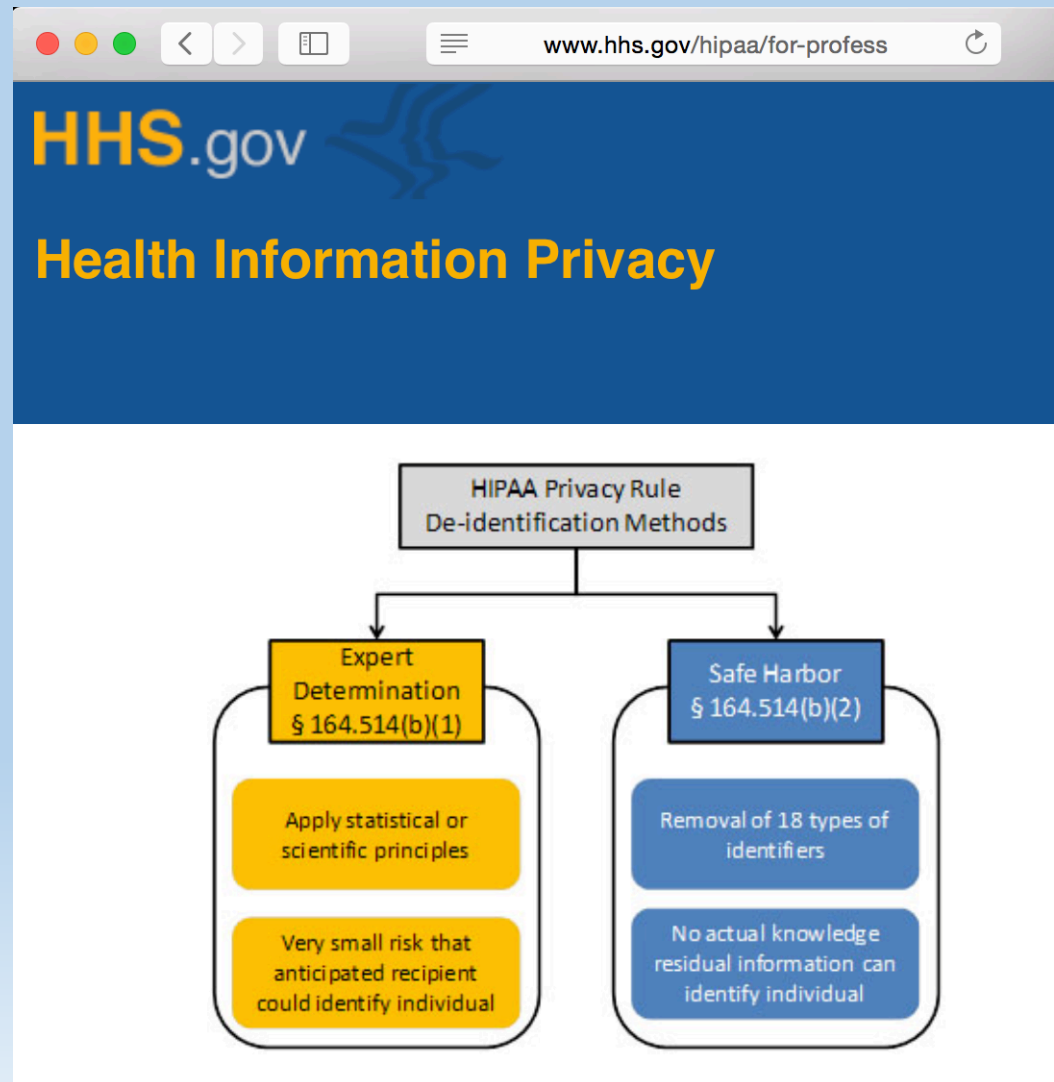


# **Privacy Preserving Technology and Policy: Institution and Patient- Centered Approaches**

Lucila Ohno-Machado, MD, PhD  
University of California San Diego  
San Diego VA

## HIPAA 'De-identified' data

- removal of 18 identifiers, such as dates, **biometrics**, names, etc.
- expert certification of low risk of re-identification
- 'Limited' data sets have 'de-identified data' plus dates



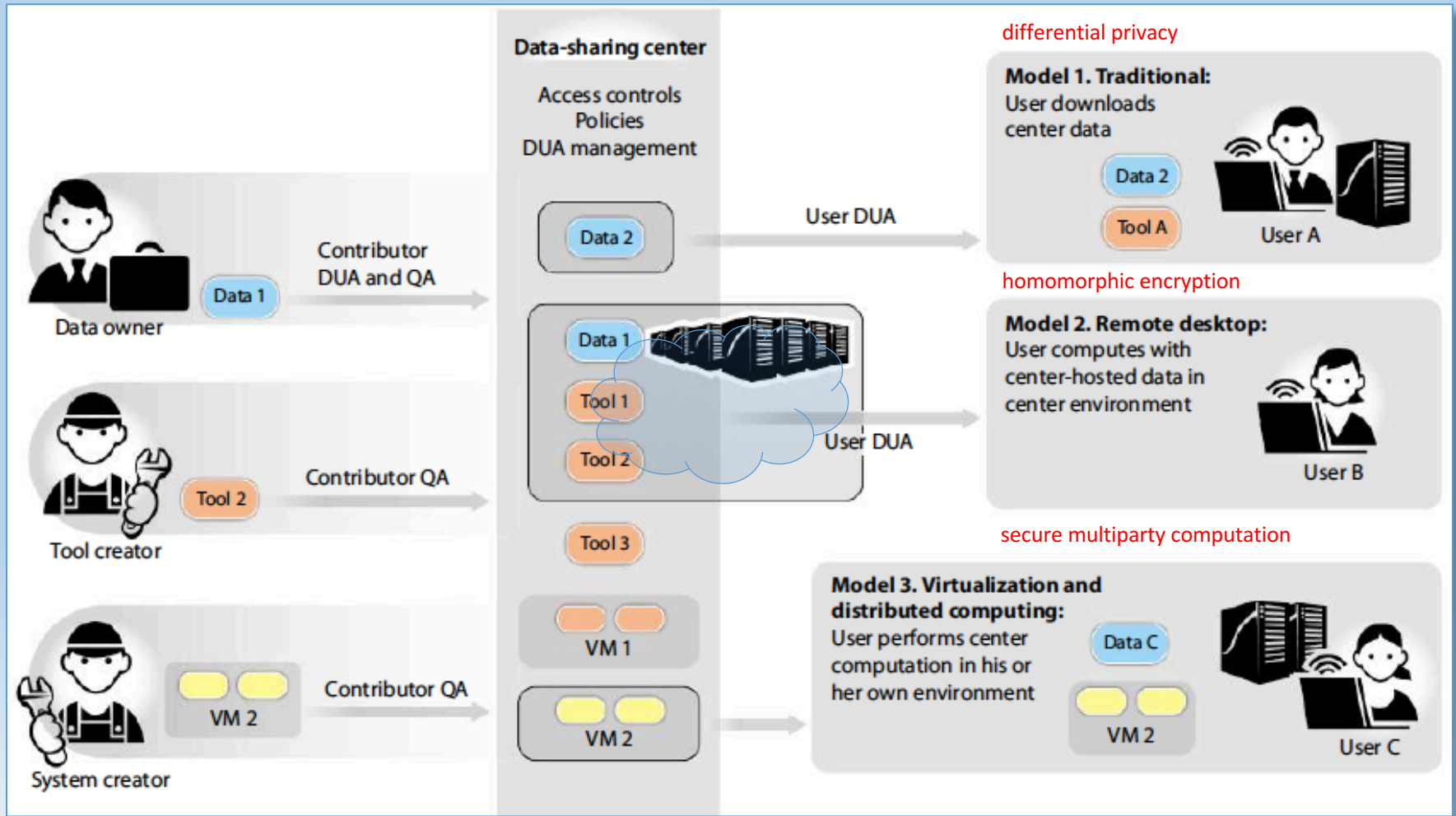
# Sharing Human Subjects Data

- Some data typically can be shared without patient's consent
  - 'de-identified' data
  - Limited data sets ('de-identified' except for dates)
- Even 'de-identified' data can be re-identified
- Patient privacy is a major concern when sharing data

# A National Center for Biomedical Computing focused on Data Sharing and Analysis



# A Center for Sharing Human Data



**DATA CONTRIBUTOR AGREEMENT  
BETWEEN  
THE UNIVERSITY OF CALIFORNIA, SAN DIEGO  
AND**

# DATA CONTRIBUTOR AGREEMENT

This **DATA CONTRIBUTOR AGREEMENT** (AGREEMENT) is by and between The Regents of California, on behalf of its San Diego campus and iDASH Data Repository (“iDASH”) and \_\_\_\_\_ (“Contributing Institution”), and is effective as of \_\_\_\_\_ (“Effective Date”).

**WHEREAS**, UCSD was funded by the National Institutes of Health to establish “Integrating Anonymization, and Sharing” (iDASH), a National Center for Biomedical Computing (NC) fundamental challenges to research progress by providing a secure, privacy-preserving env researchers can analyze genomic, transcriptomic, clinical, behavioral, and social data relev

**WHEREAS**, UCSD has established a data registry known as the “iDASH Data Repository”

**WHEREAS**, Collaborators, desire to contribute its data (Data Identifier), participate in the i and obtain access to the iDASH Data Repository for research purposes.

**WHEREAS**, UCSD is the designated iDASH Repository Host, and is responsible for manag collecting, protecting, and providing access to the data) and will receive all data from the C

**WHEREAS**, Terms used but not otherwise defined in this Agreement shall have the same m defined in the Health Insurance Portability and Accountability Act’s Privacy and Security amened from time to time (“HIPAA”).

**NOW THEREFORE**, the Parties agree as follows:

1. **PURPOSE**. To receive subject data from Collaborators, and from other institutions, inclu the iDASH Data Repository for research and publication purposes.
2. **REPOSITORY HOST**. The UCSD iDASH Data Repository will host, manage, and be resp including data integrity, registration process, receiving and signing of forms, granting u maintaining a log of authorized users, as well as allocation of costs for repository maint If services ever need to be discontinued for any reason, users will be notified 8 weeks i
3. **COLLABORATOR**. Collaborator understands and agrees that the iDASH Data Repository Collaborator’s data sets. Each Collaborator, and any other institution that wishes to sul Data Repository, is required to execute a copy of this Agreement.

4. **DESCRIPTION OF DATA**. The data set titled “\_\_\_\_\_” contains  
Collaborator agrees (*select only one option*):

- Protected Health Information or other individually identifiable health information
- A Limited Data Set
- A De-identified Data Set
- No Human Data

*(Include a brief description of data, type of data, source of data, original name of study, and supporting agency, source format)*

\_\_\_\_\_

5. **ACCEPTABLE USE OF DATA**. Except as otherwise specified herein, iDASH Data Repository Authorized Users may use and disclosure the Collaborative Data Set necessary to conduct research.  
\_\_\_\_\_ (“Identify, if any, research qualifications or authorized user access requirements.”)

Collaborator agrees (*select only one option*):

- The Collaborative Data Set will be limited to the following individuals, groups, or communities  
\_\_\_\_\_
- The Collaborative Data Set will be made publicly available.

6. **DATA USE BY RESEARCHERS**. iDASH agrees to require that any iDASH Data Repository participants (“Authorized Users”) to whom it provides access to the data contained within the Collaborator’s Data Set execute a Data User Agreement.

7. **CONFIDENTIALITY**. Collaborator and UCSD agree to abide by applicable United States state and federal laws, including the Security and Privacy Rules of HIPAA, in contributing data to the iDASH Data Repository and in maintaining the repository and granting access to researchers. iDASH is not a part of UCSD’s covered component and is not a business associate under HIPAA. Nevertheless, UCSD expressly agrees that it shall maintain the iDASH Data Repository consistent with the standards set forth in the HIPAA Security and Privacy Rules. Collaborator represents and warrants that it has the legal authority to deposit at UCSD and data derived or otherwise originating outside the United States of America: and that no foreign law applies to such data once deposited. UCSD MAKES NO REPRESENTATION OR WARRANTY AS TO COMPLIANCE WITH ANY FOREIGN LAW, REGULATION, OR STANDARD.

8. **NON-TRANSFERABILITY**. UCSD shall not transfer any information contributed by the Collaborator under this “Agreement” to individuals or organizations who are not active participants in the iDASH Data Repository pursuant to a signed Data Contributor Agreement.

9. **NON-ASSIGNABILITY**. Neither party may assign this Data Contributor Agreement or any right under this Data Contributor Agreement without the prior written consent of the other party.

10. **NOTIFICATION OF DATA BREACH**. The iDASH Data Repository will notify the Collaborator of any instances of which it is aware that the confidentiality of the Collaborative Data Set has been breached. All other uses and disclosures not authorized by this Agreement are prohibited.

**DATA USE AGREEMENT  
BETWEEN  
THE UNIVERSITY OF CALIFORNIA, SAN DIEGO  
AND**

[INSERT CONSUMER/COLLABORATOR'S ORGANIZATION]

# DATA USE AGREEMENT

This data use agreement (the "Agreement") is by and between The Regents of the University of California on behalf of its San Diego campus and iDASH Data Repository ("iDASH") and [REDACTED] ("Receiving Institution"), and is effective as of [REDACTED] (the "DUA Effective Date").

**WHEREAS**, UCSD was funded by the National Institutes of Health to establish "Integrated Data Anonymization, and Sharing" (iDASH), a National Center for Biomedical Computing to address fundamental challenges to research progress by providing a secure, privacy preserving environment in which researchers can analyze genomic, transcriptomic, clinical, behavioral, and social data.

**WHEREAS**, this Agreement sets forth the terms and conditions pursuant to which UCSD hereby disclose certain Data to the Data Recipient.

**WHEREAS**, the Recipient, desires to participate in the iDASH Data Repository as well as the iDASH Data Repository for research purposes.

**WHEREAS**, the terms used but not otherwise defined in this Agreement shall have the same meanings as those same terms in Title 45 of the Code of Federal Regulations, Part 164.

**NOW THEREFORE**, the Parties agree as follows:

1. **PURPOSE.** Receiving Institution's researchers wish to access specified data set [REDACTED] (Data Collection Title) maintained for [REDACTED] research purposes (the "Repository Data").
2. **AUTHORIZED USERS.** Receiving Institution shall assure that only Authorized Users (those individuals and who have signed this Agreement) are provided access to the Repository Data.
3. **PERMITTED USE AND DISCLOSURES.** The Recipient may use and disclose the limited data described herein:  
[REDACTED]  
[REDACTED]  
[REDACTED]

5. **DATA.** The Repository Data to which Receiving Institution **MAY** have access under this Agreement include (select one):

- Protected health information or other individually identifiable private information ("Identifiable Data");
- A Limited Data Set, as defined at 45 C.F.R. § 164.514(e) ("LDS"); or
- A De-Identified Data Set, as defined at 45 C.F.R. § 164.514(b) ("DDS").

(a) If the Repository Data include Identifiable Data, the Receiving Institution: (i) certifies that it has and shall maintain throughout the term of this Agreement a Federal wide Assurance with the U.S. Department of Health and Human Services, Office for Human Research Protections ("OHRP"); (ii) shall ensure that the Authorized Users secure approval for its use from an Institutional Review Board registered with OHRP; (iii) shall ensure that the Authorized Users perform their research, regardless of funding source, consistent with the substantive requirements of the Common Rule (45 CFR part 46), as well as with other applicable law; and (iv) shall maintain the Repository Data securely, consistent with the standards set forth at 45 C.F.R. part 164, subpart C, regardless of whether the Receiving Institution is a "Covered Entity" under the Health Insurance Portability and Accountability Act of 1996 and its Privacy and Security Rules ("HIPAA").

(b) If the data to be accessed include only LDS or DDS, then Institution and each Authorized User agrees as follows: (i) not to use or further disclose the LDS or DDS or any information contained therein other than as permitted by this Agreement or required by applicable law; (ii) to use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this Agreement; (iii) to promptly report to iDASH any use or disclosure of the LDS or DDS not provided for by this Agreement to which Receiving Institution or any Authorized User becomes aware; (iv) to ensure that any agents, including subcontractors, to whom Receiving Institution or any Authorized User provides the LDS or DDS or any part of it agrees to the same restrictions and conditions that apply to Receiving Institution and the Authorized Users under this Agreement; and (v) not to use the information contained in the LDS or DDS to identify any individual whose information is contained in the LDS or DDS, nor to contact them under any circumstances. In the event any Authorized User makes a discovery that it believes requires contact with such an individual, the Authorized User shall make a request in writing that iDASH facilitate recontact if possible and iDASH will make reasonable efforts to do so.

(c) The Receiving Institution and Authorized Users shall comply with any additional requirements or restrictions placed on the Repository Data by data contributors of which they are informed by iDASH in writing or through any electronic means.

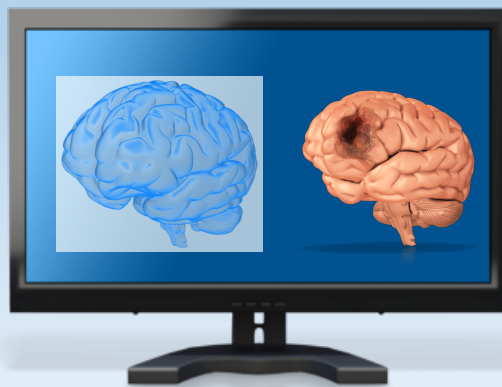
6. **NON-TRANSFERABILITY.** Recipient shall not transfer any information obtained or accessed under this DUA to individuals or organizations not specified in this DUA. There are no intended third party beneficiaries to this DUA. Without in any way limiting the foregoing, it is the Parties' specific intent that nothing contained in this DUA gives rise to any right or cause of action, contractual or otherwise, in or on behalf of the individuals whose data is Used or Disclosed pursuant to this DUA. Each Data Recipient, and any other institution that wishes to receive data from the iDASH Data Repository, is required to execute a copy of this DUA.

7. **BREACH.** In the event the Receiving Institution or any Authorized User becomes aware of any use of the Repository Data that is not expressly authorized under this Agreement, iDASH may: (i) terminate this Agreement upon notice; (ii) disqualify the Receiving Institution or any Authorized User from participation in iDASH; and (iii) report the inappropriate use or disclosure to the collaborators who contributed the data to the iDASH Repository; any individuals whose privacy may have been compromised; any regulatory agencies with jurisdiction; and others as may be required by applicable law.

# Who Contributed Human Subjects Data to iDASH

- PIs from studies in which patient consent permitted sharing
- PIs who did not have funds to host the data and their distribution
- PIs who did not want to deal with individual requests and execute DUAs every time someone requested a dataset
- PIs who did not know how to set up a HIPAA-compliant environment for sharing

find all data about low grade astrocytomas with IDH mutations



astrocytoma grade 2 idh

Search for data set  Search for repository

[Advanced Search](#) [help](#)

astrocytoma grade 2 idh

Data Types

- Unspecified (119)
- Gene Expression (85)
- Omics Data (75)
- Clinical Trials (15)
- Data from Papers (8)
- Imaging Data (6)
- Phenotype (4)

Repositories

- BioProject (91)
- ArrayExpress (77)
- OmicsDI (75)
- Figshare (27)
- PeerJ (8)
- CIL (6)
- GEO (4)
- GEMMA (4)
- dbGaP (3)
- Dryad (1)

More...

Accessibility

- Download (296)

Authorization

- None (296)
- Restricted (0)

Displaying 20 of 296 results for "astrocytoma grade 2 idh"

20 Per Page

First « 1 2 3 4 5 6 7 8 9 10 » Last

Send To i

Genome-Wide Profiling of **Astrocytic Gliomas** BioProject

**ID:** PRJNA119515

**Keywords:** Other

**Access Type:** download

Genome-Wide Profiling of **Astrocytic Gliomas** ArrayExpress

**ID:** E-GEOD-18166

**Description:** comparative genomic hybridization in 131 diffuse **astrocytic gliomas**, including 87 primary glioblastomas (pGBIV), 13 secondary glioblastomas (sGBIV), 19 **anaplastic astrocytomas**

Expression profiling of lower-**grade** diffuse **astrocytic glioma** BioProject

**ID:** PRJNA150783

**Keywords:** Transcriptome or Gene expression

**Access Type:** download

Low **grade gliomas subtype** analysis BioProject

**ID:** PRJEB9786

**Keywords:** Transcriptome or Gene expression

**Access Type:** download

Low **grade gliomas subtype** analysis ArrayExpress

**ID:** E-MTAB-3708

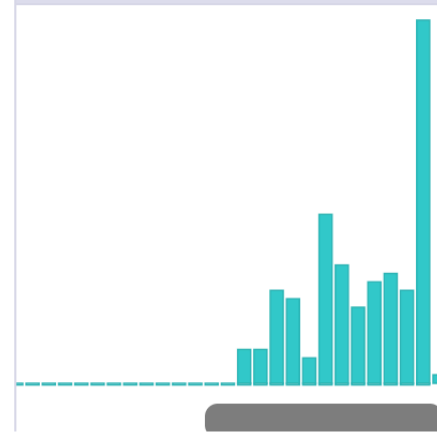
**Description:** Low **grade gliomas** (LGG; WHO **grade 2 astrocytomas**, oligodend...

Expression profiling of lower-**grade** diffuse **astrocytic glioma** ArrayExpress

**ID:** E-GEOD-35158

**Description:** Diffuse **gliomas** represent the most prevalent class of primary brain tumor. Despite significant recent advances in the understanding of gli...

Results by year



Recent Activity

Q astrocytoma grade 2 idh(data)

[See more and save search query](#)

Synonyms

- Childhood Cerebral Astrocytoma
- Grade I Astrocytoma
- Astrocytoma, Anaplastic
- Subependymal Giant Cell Astrocytoma

## Engaging The Community Toward a Data Discovery Index (DataMed v3.0)

Search for data through bioCADDIE



Search for data set  Search for repository

[Advanced Search](#) [help](#)

### Statistics



**74 REPOSITORIES**



**15 DATA TYPES**

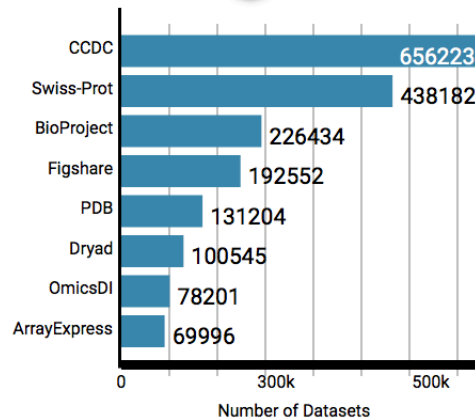


**2,336,403 DATASETS**



**4 PILOT PROJECTS**

### Top 8 Repositories



### New Features



- **July 31, 2017. v3.0**
  - » API for DataMed
  - » Autocompleting of words
  - » Reporting of broken links
  - » Context for search terms
  - » Increase in number of repositories that are indexed
- **Feb 28, 2017. v2.0**
  - » [Increase coverage to more repositories](#)
  - » Duplicate datasets display feature
  - » Usability enhancements based on user feedback and user interviews
  - » [User-reported issues resolved](#)

NIH U24AI117966

### Pilot Projects



#### GWAS Finder

Search literatures for "Genome-Wide Association Studies". [Read More](#)



#### iSEE-DELVE

Search Visualization project for Big Data. [Read More](#)



#### DataRank

Find most suitable datasets for you. [Read More](#)



#### Data Citation Discovery

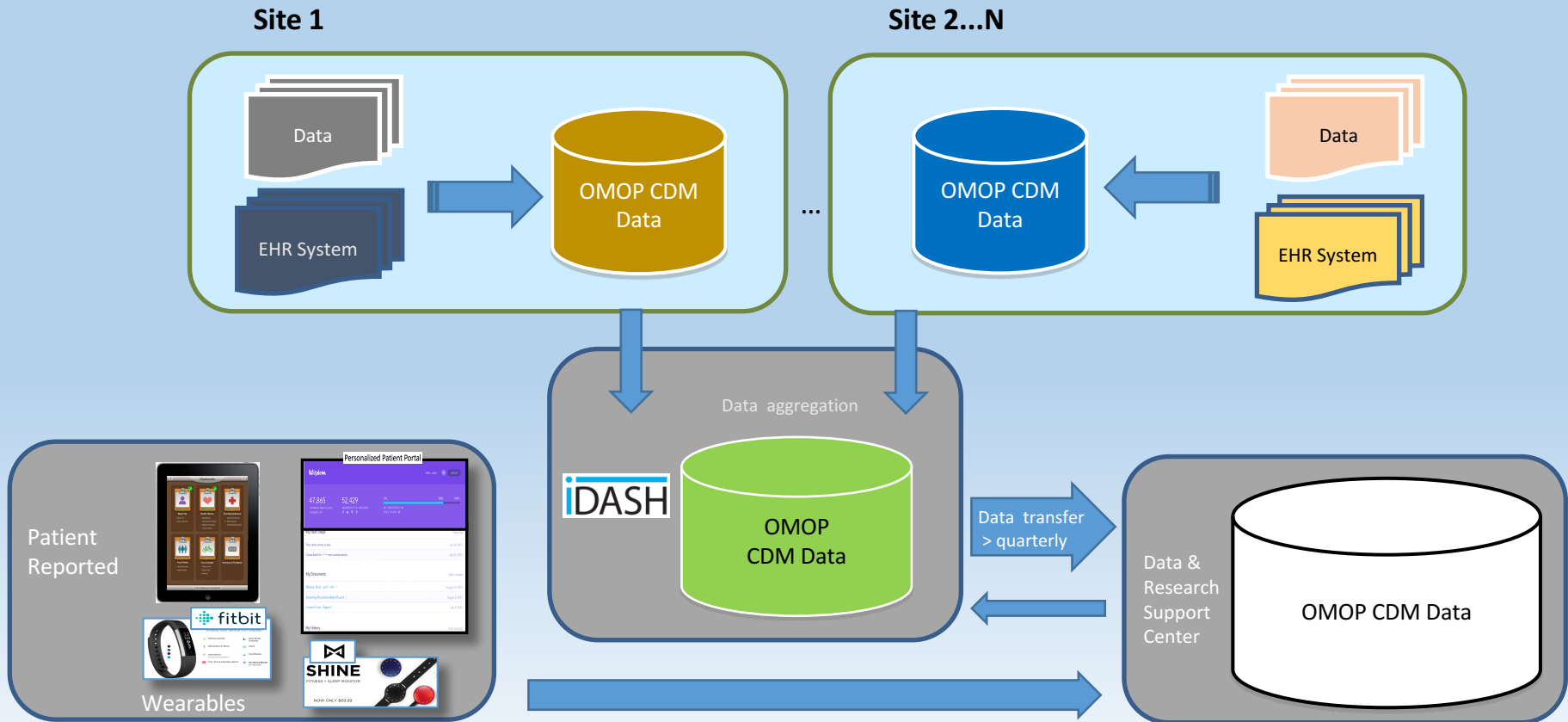
Citation and Data Access Metrics Development applied to RCSB Protein Data Bank. [Read More](#)

# California Data for

# All of Us<sup>SM</sup>

THE FUTURE OF HEALTH BEGINS WITH YOU

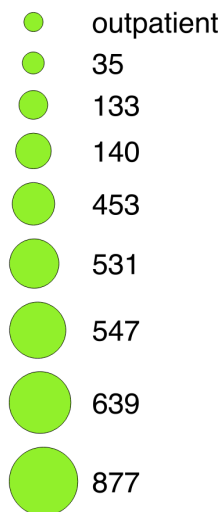
The  
Precision  
Medicine  
Initiative<sup>®</sup>



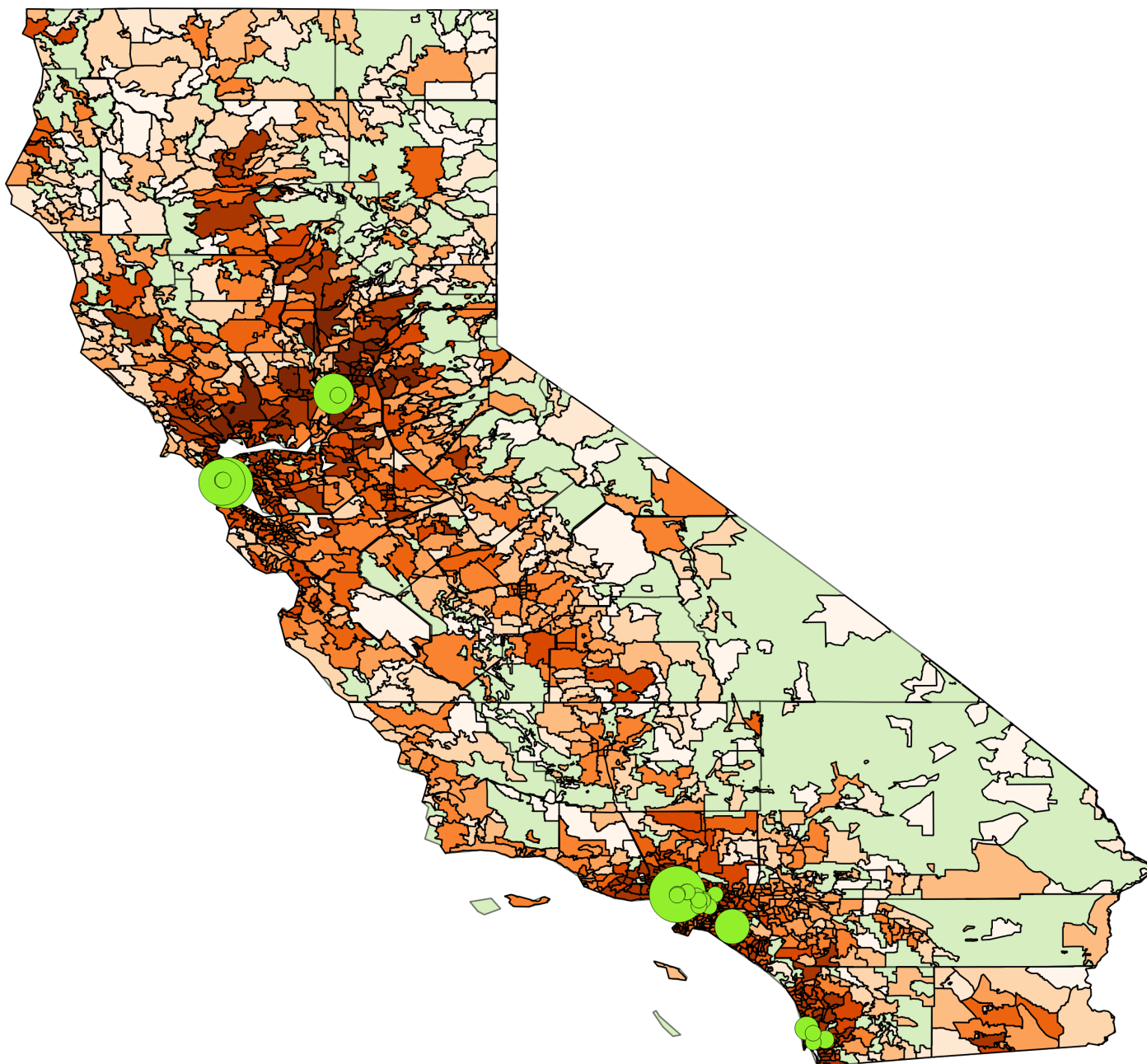
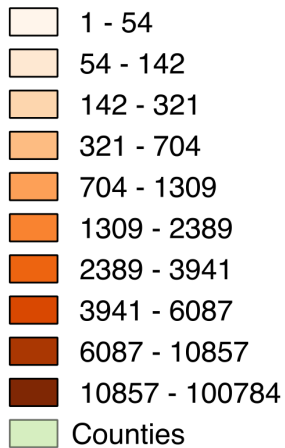
# California Precision Medicine Consortium for the All of Us Research Program

## CAPMC

### Clinical Sites Licensed Beds



### Patient Count by Zip



# Patient Consent is not always broad

## Examples of conditions in some consent forms

- Data should only be used for research in “X” (e.g., cancer)
  - Indexing data so datasets can be found: does this count as “research”?
- Only studies in which institution “Y” participates can be conducted
- No study of ancestry
- etc.

# Privacy technology methods

## Data-centric

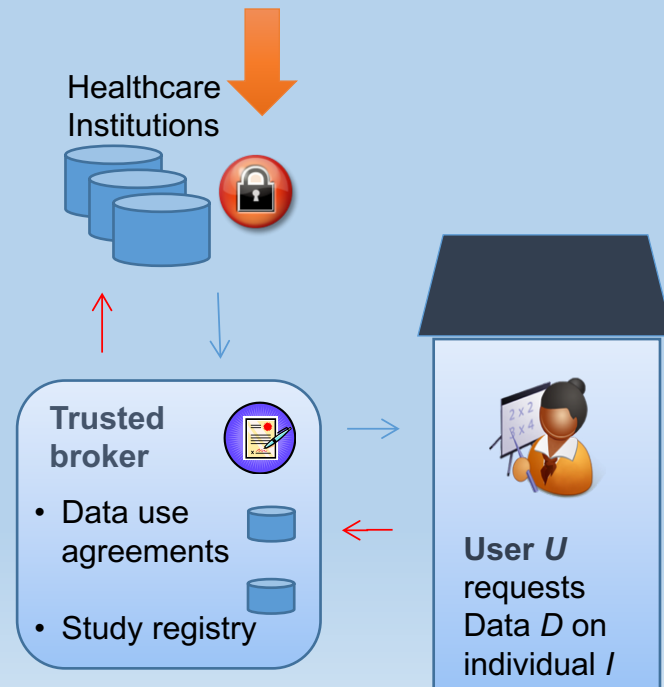
- Add noise to data  
(e.g., differential privacy)
- Operate on encrypted data  
(e.g., homomorphic encryption)
- Multiparty computation  
(e.g., distributed analytics)



## Institution/People-centric

- Data broker  
(e.g., clinical data research networks)
- Patient-defined data sharing permissions  
(e.g., consent management)

# Institution-Centered Strategies



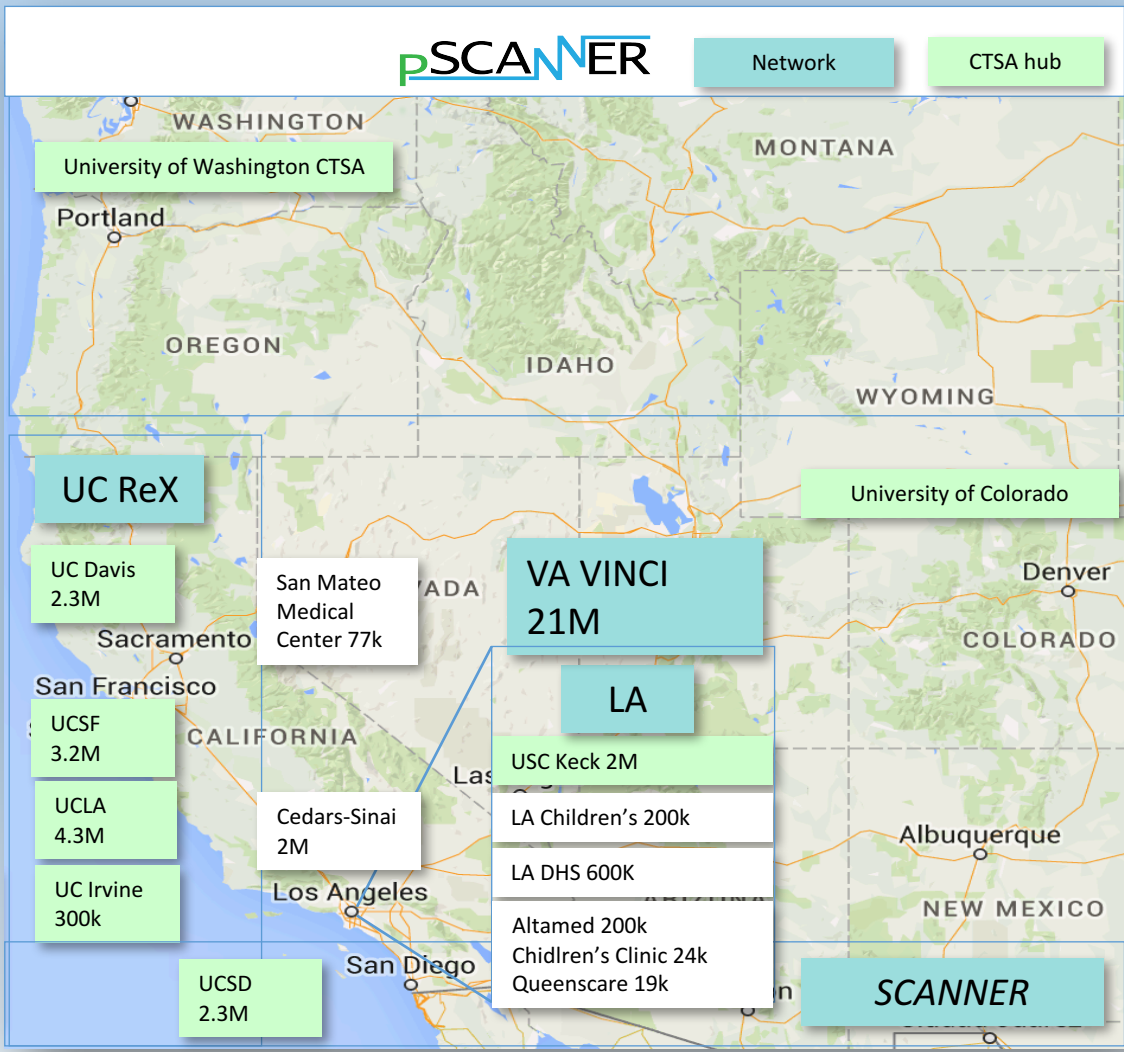
# Clinical Data Research Network

patient-centered  
Scalable National  
Network for  
Effectiveness  
Research

30 million  
people

14 health  
systems

member of  
network of  
networks



# How this works

- Federated data for distributed analytics
- Common data model
- Minimal computational infrastructure
- Compatible institutional policies
- Agreement on rules of engagement
- Shared ethics principles
- Research studies conducted on portions of the data





**Is it ethical to share?**

People have not been explicitly asked and don't know who is sharing what



**Could people choose?**

Is it practical?  
What if massive number of people withdraw their data?

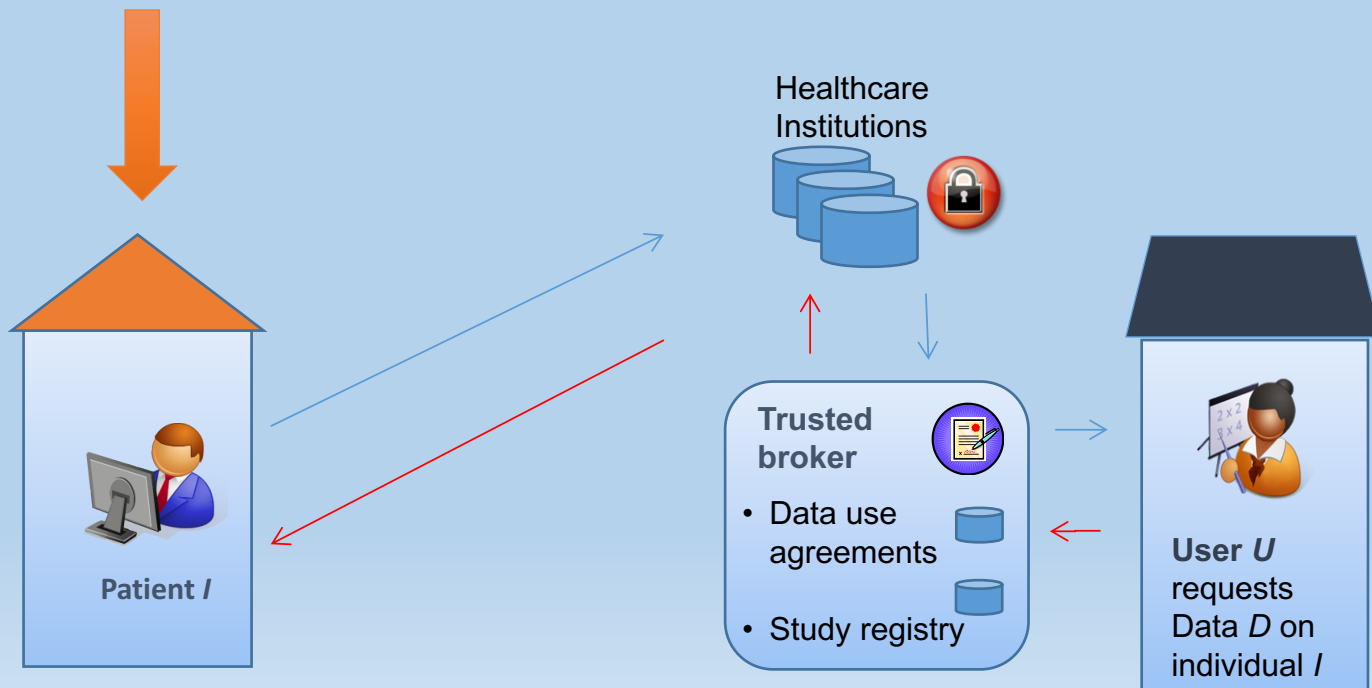


**Is it ethical not to share?**

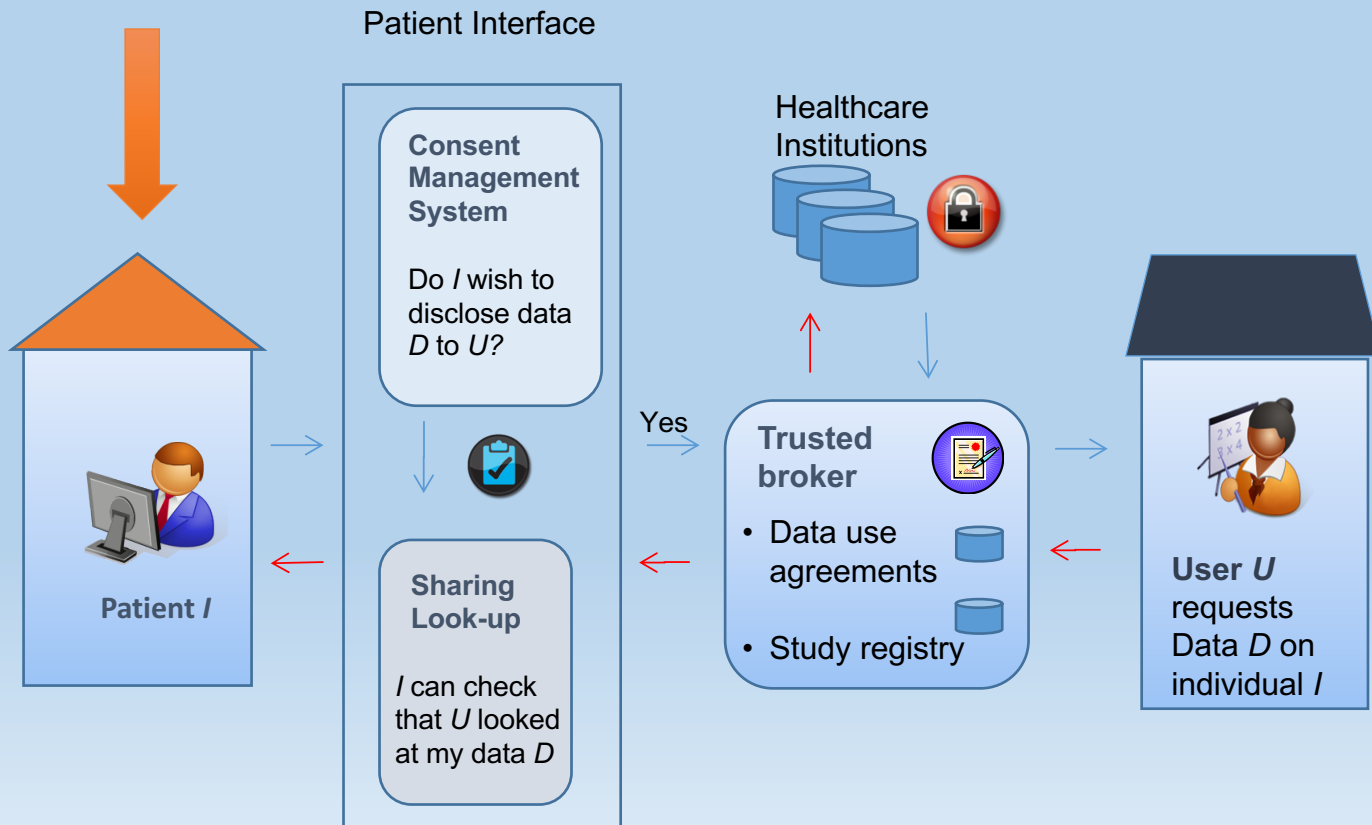
New discoveries and acceleration of science depend on sharing



# People-Centered Strategies



# People-Centered Strategies



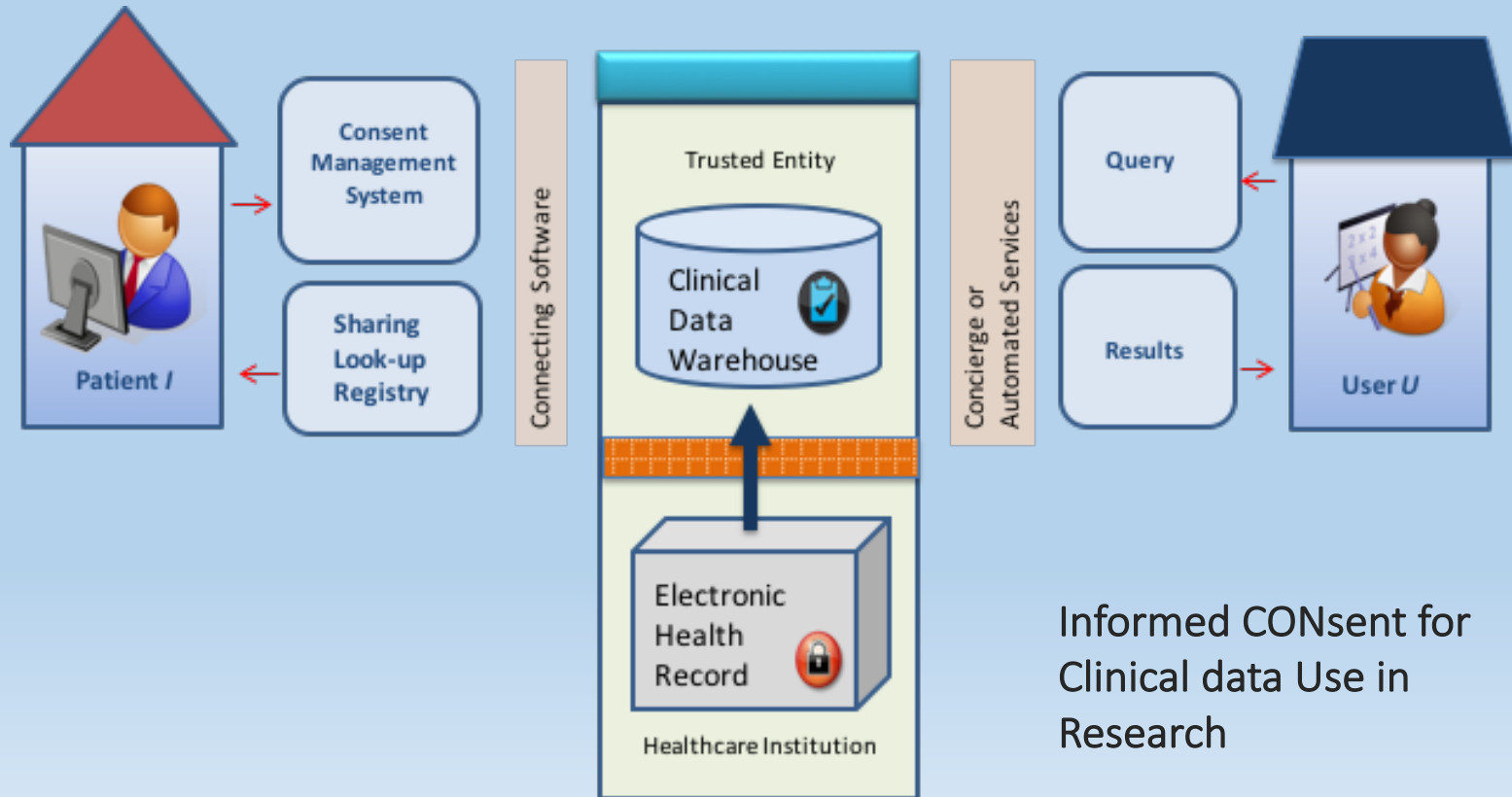
# What if many patients decline use of data?

- Some clinics refused to participate
- Research could be affected
  - Fewer sample sizes
  - Wrong inferences

Someone's opinion (not ours):

“If you can't guarantee that a given non-disclosed item cannot be inferred from disclosed data, then it is best to do nothing”

# iCONCUR



Informed CONsent for  
Clinical data Use in  
Research

Browser: ctri-iconcur.ucsd.edu

Navigation: Google, PubMed, UCSD Library, Blink, GoogleScholar, eRA, YouTube, Bank, Save to Mendeley, ICONCUR, UCRP, MeetingRoom, Epic PLAY

Header: **ICONCUR** Informed Consent for Clinical Data and Sample Use for Research | My UCSD Chart | Logout

Left Sidebar: Home, My Sharing Choices, Admin

## My Sharing Choices

What clinical data am I sharing?	Who can access the clinical data I share?		
	UCSD and VA Hospital	Non-profit Organizations	For-profit Organizations
<b>Demographics</b>			
<b>Age</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Ethnicity</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Gender</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Race</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Socioeconomic Information</b>			
<b>Education Level</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Insurance Status</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Marital Status</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Occupation</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Income</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Sexuality</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>Past Pregnancy</b>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

*Courtesy of H Kim*

<ul style="list-style-type: none"> <li>✔ Demographics <ul style="list-style-type: none"> <li>ⓘ Age</li> <li>ⓘ Ethnicity</li> <li>ⓘ Gender</li> <li>ⓘ Race</li> </ul> </li> <li>✔ Socioeconomic Information <ul style="list-style-type: none"> <li>ⓘ Education Level</li> <li>ⓘ Insurance Status</li> <li>ⓘ Marital Status</li> <li>ⓘ Occupation</li> <li>ⓘ Income</li> <li>ⓘ Sexuality</li> <li>ⓘ Past Pregnancy</li> <li>ⓘ Anthropometrics</li> <li>ⓘ Vital Signs</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>✔ Current or Previous Disease or Condition <ul style="list-style-type: none"> <li>ⓘ Substance abuse related Disease or Condition</li> <li>ⓘ Mental Health Disease or Condition</li> <li>ⓘ Sexual or Reproductive Disease or Condition</li> <li>ⓘ Other</li> </ul> </li> <li>✔ Family's Current or Previous Disease or Condition <ul style="list-style-type: none"> <li>ⓘ Substance abuse related Disease or Condition</li> <li>ⓘ Mental Health Disease or Condition</li> <li>ⓘ Sexual or Reproductive Disease or Condition</li> <li>ⓘ Other</li> </ul> </li> <li>✔ Laboratory and Test Results <ul style="list-style-type: none"> <li>ⓘ Genetic Test</li> <li>ⓘ Sexually Transmitted Disease Test</li> <li>ⓘ Test on Drug Screening</li> <li>ⓘ Other</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>ⓘ Therapy or Treatment Procedures</li> <li>ⓘ Medications</li> <li>✔ Social History <ul style="list-style-type: none"> <li>ⓘ Alcohol Consumption Status</li> <li>ⓘ Recreational Drug Use</li> <li>ⓘ Smoking Status</li> </ul> </li> <li>✔ Health Care Encounter <ul style="list-style-type: none"> <li>ⓘ Location of the hospital or clinic</li> <li>ⓘ Physician's Name</li> <li>ⓘ Speciality of the hospital or clinic</li> <li>ⓘ Visit Dates</li> <li>ⓘ Charges and billing related to encounters</li> </ul> </li> <li>✔ Tissue and Blood Sample Usage <ul style="list-style-type: none"> <li>ⓘ Tissue Sample</li> <li>ⓘ Blood Sample</li> </ul> </li> </ul>
---	--	--

+ Sharing for stem cell research

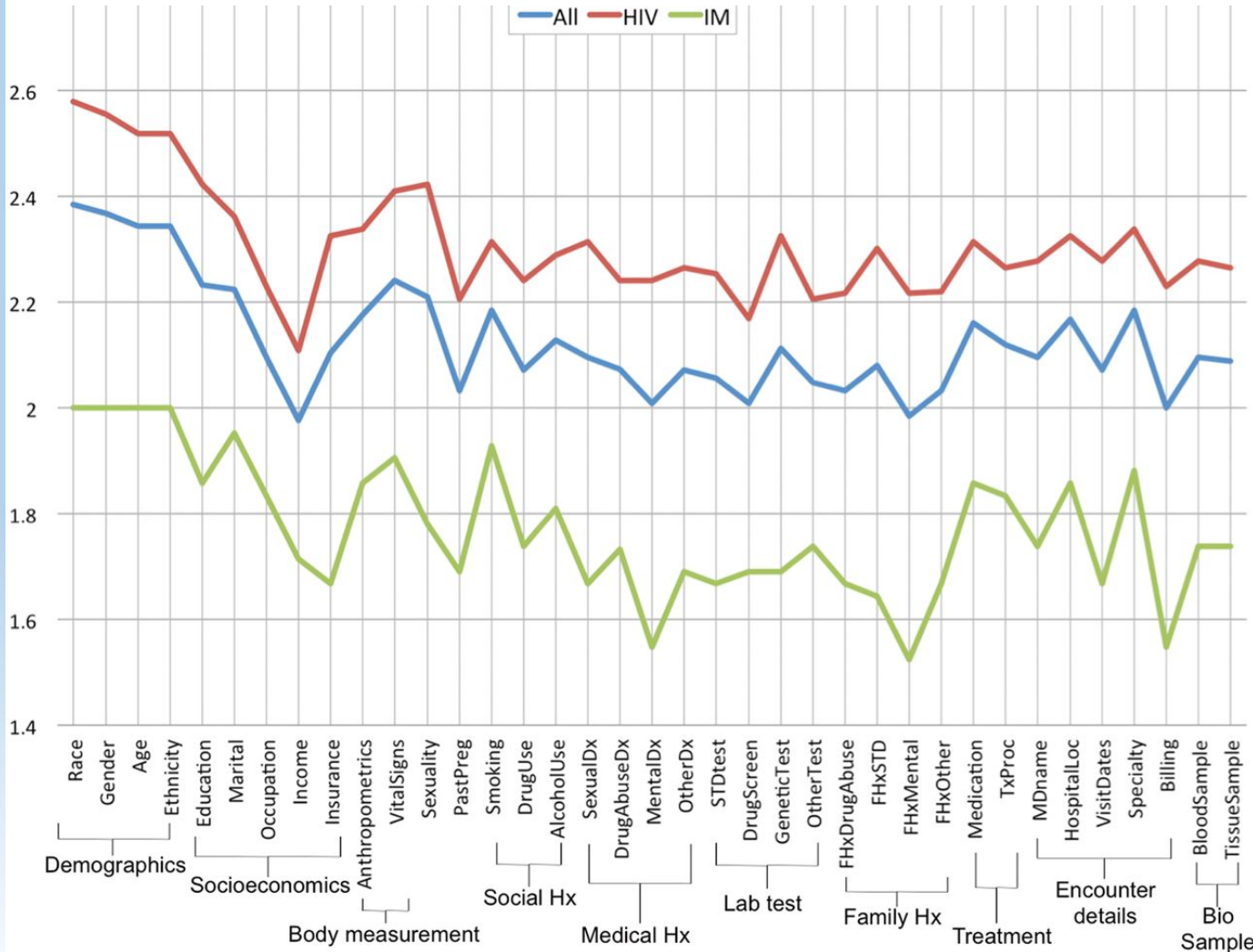
# iCONCUR: informed consent for clinical data and bio-sample use for research

RECEIVED 18 February 2016  
 REVISED 27 May 2016  
 ACCEPTED 11 July 2016

Hyeoneui Kim,<sup>1</sup> Elizabeth Bell,<sup>1</sup> Jihoon Kim,<sup>1</sup> Amy Sitapati,<sup>1,2</sup> Joe Ramsdell,<sup>2</sup>  
 Claudiu Farcas,<sup>1</sup> Dexter Friedman,<sup>3</sup> Stephanie Feudjio Feupe,<sup>1</sup> Lucila Ohno-Machado,<sup>1,4</sup>



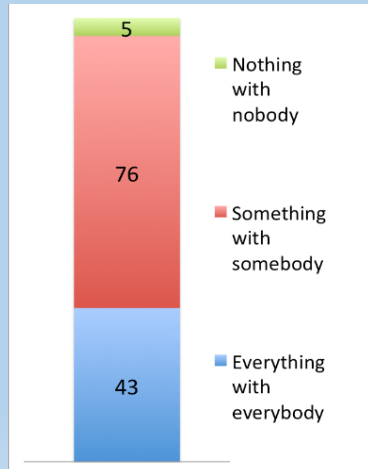
## ABSTRACT



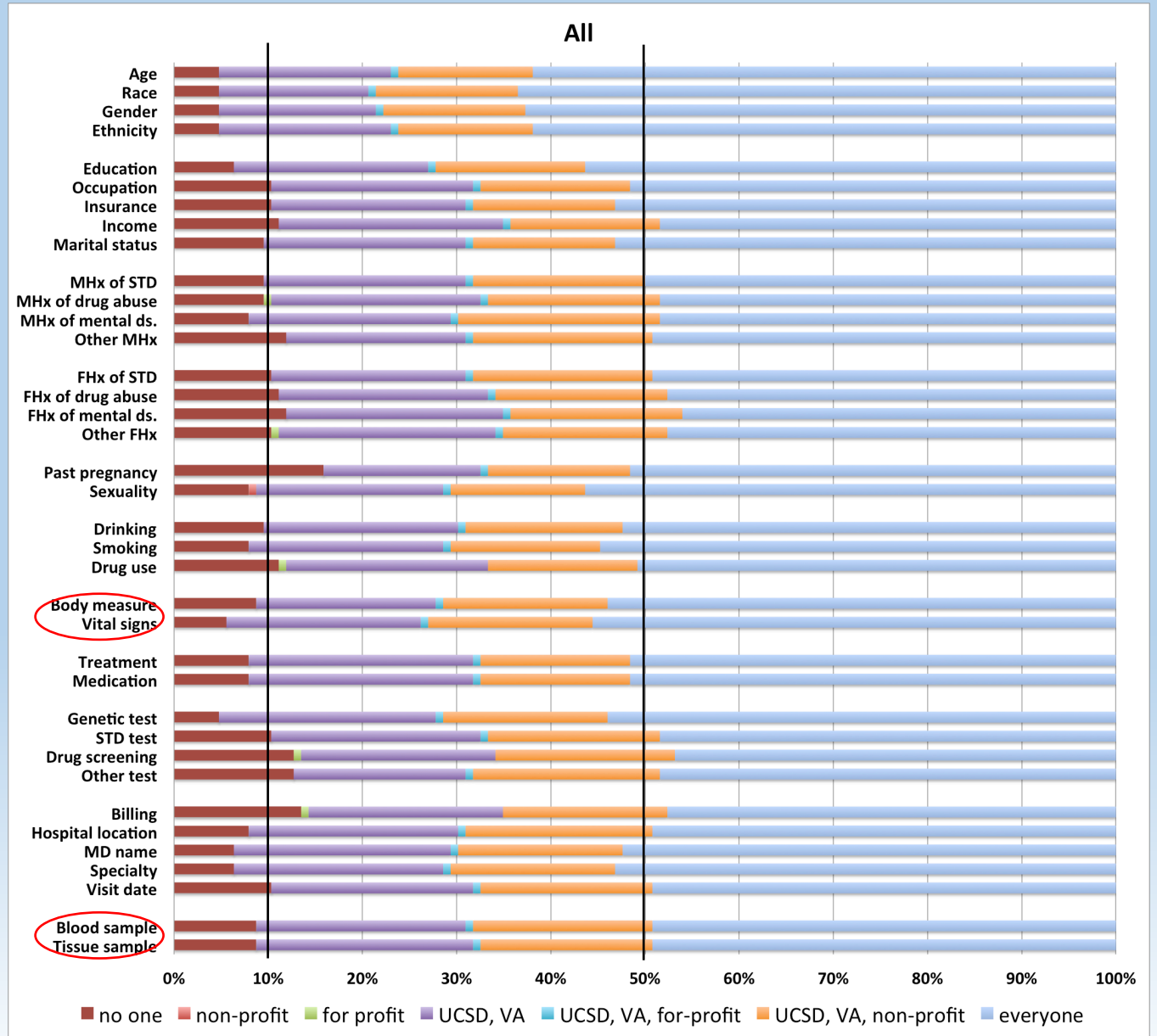
Mean willingness-to-share score by individual data item

(HIV: HIV clinic, IM: IM clinic, Hx: History, FHx: Family History, Dx: Diagnosis, STD: Sexually Transmitted Disease, MD: Medical Doctor)

# iCONCUR



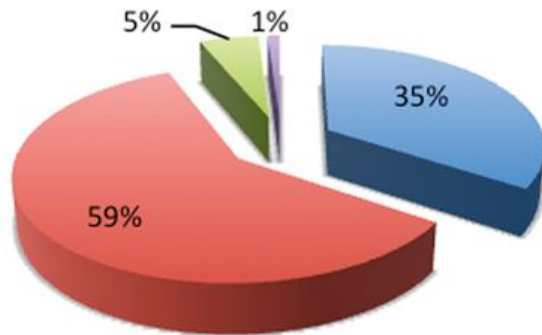
*Preliminary data on >1k patients*



# Post-selection survey

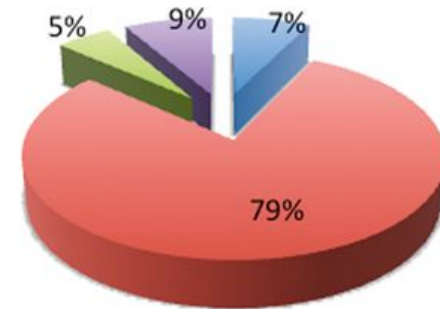
**(a) Does having these choices make you feel differently about sharing your medical data?**

■ More willing to share ■ No change ■ Less willing to share ■ other



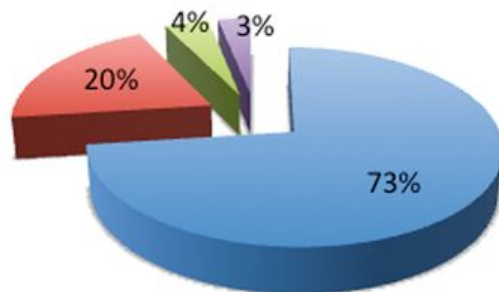
**(b) How does your willingness to share your medical data for healthcare compare to your willingness to share for research?**

■ Don't know ■ Equally  
■ More willing for healthcare ■ More willing for research



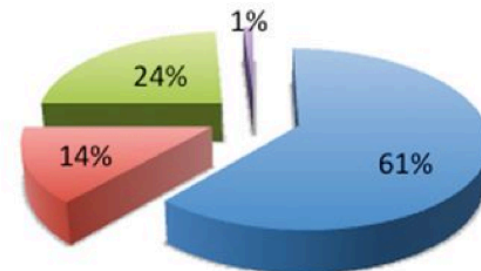
**(c) Would you feel more comfortable sharing your medical information if you know who is using it for research?**

■ More comfortable ■ Indifferent ■ Less comfortable ■ Other



**(e) If you could receive an update each time someone uses your medical data for research, would you like to be notified?**

■ Yes, each time ■ Indifferent  
■ Don't need to know ■ Quarterly or yearly



# Big Data are a Big Deal



**App data**  
no regulation



**EHRs**  
can be linked



**Databases**  
can be purchased



**Social network**  
can be used

# Funding Sources

NIH R01HG008802

Patient-Centered Outcomes Research  
Institute  
CDRN-1306-04819

NIH T15LM011271

NIH U24AI117966

NIH OT2OD024611

NIH U54HL1084600