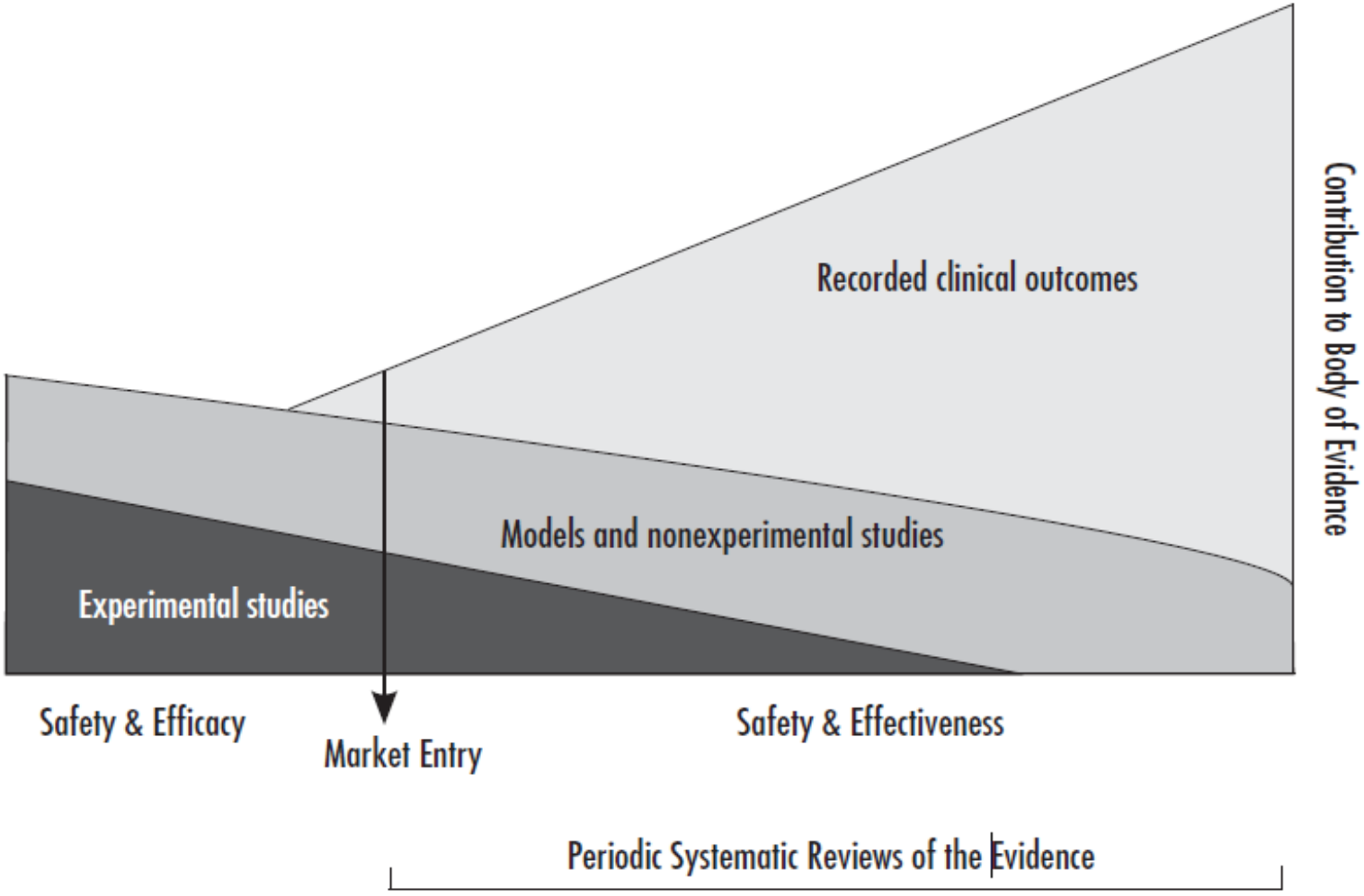
Data Integration in the SCILHS PCORNet Clinical Data Research Network

Shawn Murphy, Ken Mandl, Sebastian Schneeweiss
Principle Investigators

SCILHS scalable collaborative infrastructure for a learning health system

Learning Health System: Evidence Generated from a Well-Instrumented Health System



Patient Centered Outcomes Research Institute

- An **independent, non-profit health research organization** authorized by the Patient Protection and Affordable Care Act of 2010.
- PCORI funds **patient-centered research** to assist patients, caregivers, and other stakeholders in making informed health decisions.

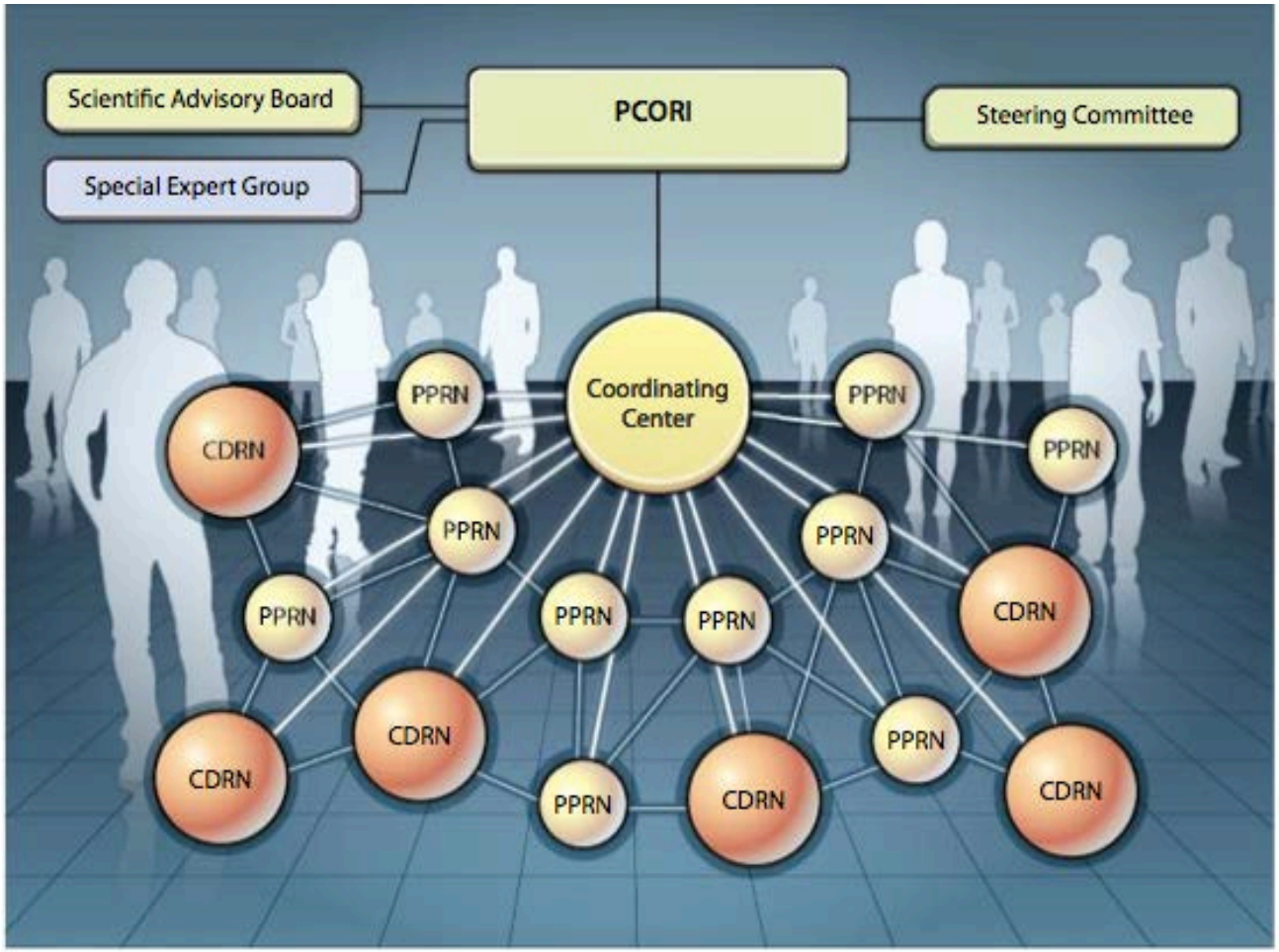
Mission

PCORI helps people make informed healthcare decisions and improves healthcare delivery and outcomes by producing and promoting high integrity, evidence-based information that comes from research **guided by patients, caregivers, and the broader healthcare community.**

Vision

Patients and the public have the information they need to **make decisions** that reflect their **desired health outcomes**.

The National Network of Networks



Clinical Data Research Networks

The Clinical Data Research Networks (CDRNs) will develop the capacity to conduct randomized comparative effectiveness studies using data from clinical practice in a large, defined population. These established or newly developed networks involve two or more healthcare systems, with plans to function as integrated research network.

- C1_Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS)
  Harvard University
- C2_Mid-South CDRN
  Vanderbilt University
- C3_Patient-oriented SCAlable National Network for Effectiveness Research (pSCANNER)
  University of California, San Diego (UCSD)
- C4_Great Plains Collaborative (GPC)
  University of Kansas Medical Center
- C5_Kaiser Permanente & Strategic Partners Patient Outcomes Research To Advance Learning (PORTAL) Network
  Kaiser Foundation Research Institute
- C6_Louisiana CDRN (LACDRN)
  Louisiana Public Health Institute (LPHI)
- C7_National Pediatric Learning Health System (PEDSNet)
  The Children's Hospital of Philadelphia
- C8_New York City Clinical Data Research Network (NYC-CDRN)
  Weill Medical College of Cornell University
- C9_Chicago Area Patient Centered Outcomes Research Network (CAPriCORN)
  The Chicago Community Trust
- C10_Accelerating Data Value Across a National Community Health Center Network (ADVANCE)
  Oregon Community Health Information Network (OCHIN)
- C11_P2aTH: Towards a Learning Health System in the Mid-Atlantic Region
  University of Pittsburgh

PPRNs represent a number of conditions…

Organization	PI	Condition	Population Size
Accelerated Cure Project for Multiple Sclerosis	Robert McBurney	Multiple Sclerosis	20,000
Cincinnati Children's Hospital Medical Center	Peter Margolis	Pediatric Crohn's Disease and Ulcerative Colitis i2b2	15,000
COPD Foundation	Richard Mularski	Chronic Obstructive Pulmonary Disease	50,000
Crohn's and Colitis Foundation of America	R. Balfour Sartor	Inflammatory Bowel Disease (Crohn's disease and ulcerative colitis)	30,000
Global Healthy Living Foundation	Seth Ginsberg	Arthritis (rheumatoid arthritis, spondyloarthritis), musculoskeletal disorders (osteoporosis), and inflammatory conditions (psoriasis)	50,000
Massachusetts General Hospital	Andrew Nierenberg	Major Depressive Disorder and Bipolar Disorder i2b2	50,000
Univ of California, San Francisco	Mark Pletcher	Cardiovascular health	100,000
University of South Florida	Rebecca Sutphen	Hereditary Breast & Ovarian Cancer	17,000

….including rare diseases

Organization	PI	Condition	Population Size
ALD Connect, Inc	Florian Eichler	Adrenoleukodystrophy	3,000
Arbor Research Collaborative for Health	Bruce Robinson	Primary Nephrotic Syndrome, Focal Segmental Glomerulosclerosis, Minimal Change Disease, and Membranous Nephropathy Multiple Sclerosis	1,250
Duke University	Laura Schanberg	Juvenile Rheumatic Disease i2b2	9,000
Epilepsy Foundation	Janice Beulow	Aicardi Syndrome, Lennox-Gastaut Syndrome, Phelan-McDermid Syndrome, Hypothalamic Hamartoma, Dravet Syndrome, Tuberous Sclerosis	1,500
Genetic Alliance, Inc	Sharon Terry	Alström syndrome , Dyskeratosis congenital, Gaucher disease, Hepatitis, Inflammatory breast cancer, Joubert syndrome, Klinefelter syndrome & associated conditions, Psoriasis, Metachromatic leukodystrophy, Pseudoxanthoma elasticum,	50- 50,000
Immune Deficiency Foundation	Kathleen Sullivan	Primary Immunodeficiency Diseases	1,250
Parent Project Muscular Dystrophy	Holly Peay	Duchenne and Becker muscular dystrophy	4,000
Phelan-McDermid Syndrome Foundation	Megan O'Boyle	Phelan-McDermid Syndrome i2b2	737
University of Pennsylvania	Peter Merkel	Vasculitis	500

Key Challenges

- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

The Plan

Common data platform **(i2b2)**

+

Federated queries across sites **(SHRINE)**

+

Point of care apps **(SMART)**

+

Patient-facing technologies **(RedCap, SMART, +)**

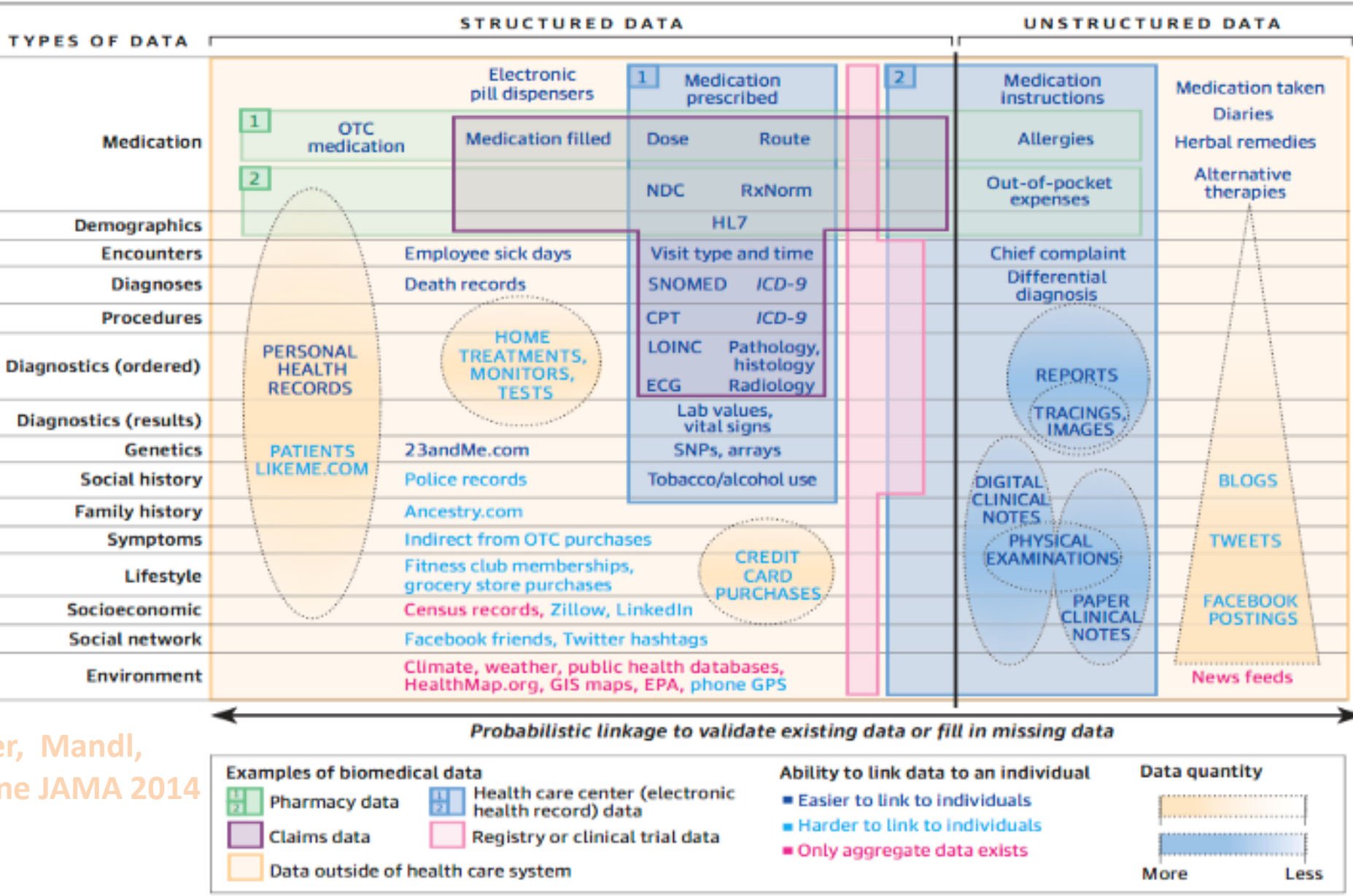
Key Challenges

- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

Health Data

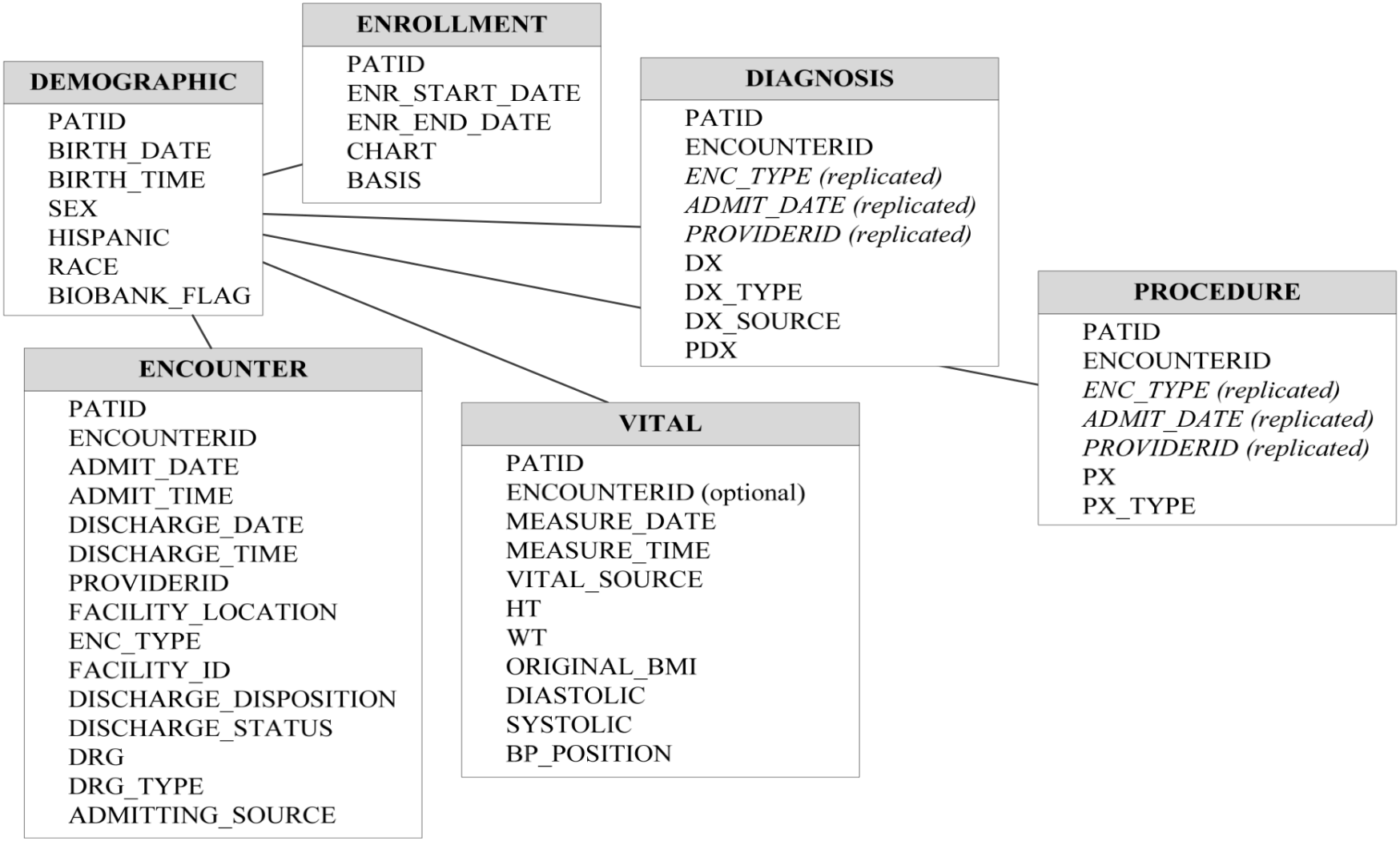
- Myriad types
  - Labs, Meds, Diagnoses are dominant in EHRs
  - Payors and some health systems have claims
  - Patient Reported Outcomes are increasingly used in clinical settings
  - Text and notes
  - Genomic, proteomic, medical device device, smart phone social media data critical to some studies

. The Tapestry of Potentially High-Value Information Sources That May be Linked to an Individual for Use in Health Care



er, Mandl,
ne JAMA 2014

PCORnet CDM v1



PCORnet Common Data Model v3.0

New to v3.0

DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS

DISPENSING
DISPENSINGID
PATID
PRESCRIBINGID (optional)
DISPENSE_DATE
NDC
DISPENSE_SUP
DISPENSE_AMT

DEATH
PATID
DEATH_DATE
DEATH_DATE_IMPUTE
DEATH_SOURCE
DEATH_MATCH_CONFIDENCE

DEATH_CONDITION
PATID
DEATH_CAUSE
DEATH_CAUSE_CODE
DEATH_CAUSE_TYPE
DEATH_CAUSE_SOURCE
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

VITAL
VITALID
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE

CONDITION
CONDITIONID
PATID
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
CONDITION
CONDITION_TYPE
CONDITION_SOURCE

PRO_CM
PRO_CM_ID
PATID
ENCOUNTERID (optional)
PRO_ITEM
PRO_LOINC
PRO_DATE
PRO_TIME
PRO_RESPONSE
PRO_METHOD
PRO_MODE
PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

ENCOUNTER
ENCOUNTERID
PATID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

DIAGNOSIS
DIAGNOSISID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

PROCEDURES
PROCEDURESID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX_DATE
PX
PX_TYPE
PX_SOURCE

LAB_RESULT_CM
LAB_RESULT_CM_ID
PATID
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
RESULT_DATE
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND

PRESCRIBING
PRESCRIBINGID
PATID
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

Data captured from healthcare delivery, direct encounter basis

PCORNET_TRIAL
PATID
TRIALID
PARTICIPANTID
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

Associations with PCORnet clinical trials

HARVEST
NETWORKID
NETWORK_NAME
DATAMARTID
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

Process-related data

Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.

It is a rapidly evolving picture!

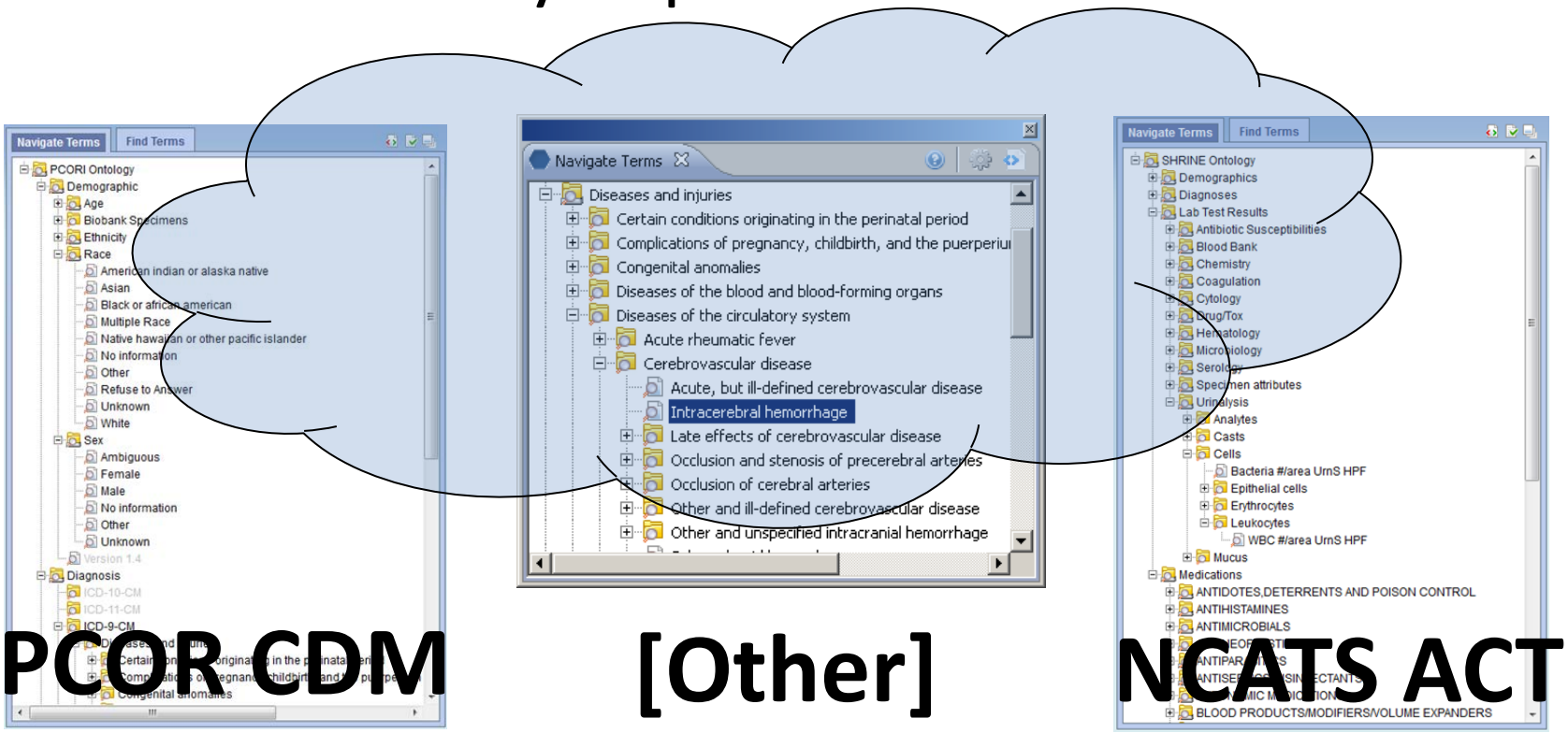
- I2b2 is designed to support health-system-wide data harmonization
- BD2K and beyond

i2b2

- Explicitly organizes and transforms person-oriented clinical data for clinical and translational research
  - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
  - Software is built in a modular pattern that allows additions without disturbing core parts
  - Available as open source at https://www.i2b2.org

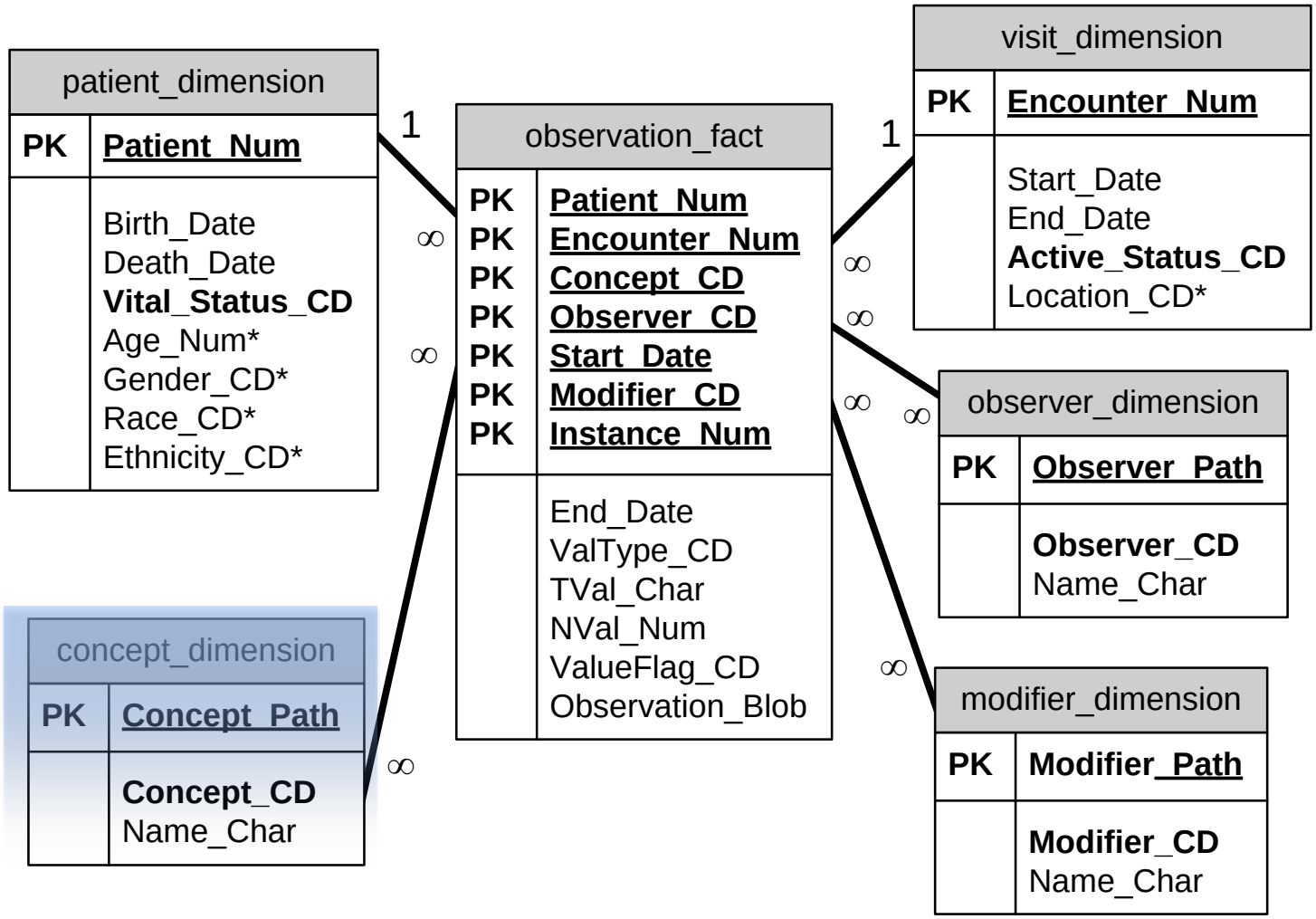
I2b2 is a Giant Data Sponge

Generally Represents Patient Data



PCOR CDM **[Other]** **NCATS ACT**

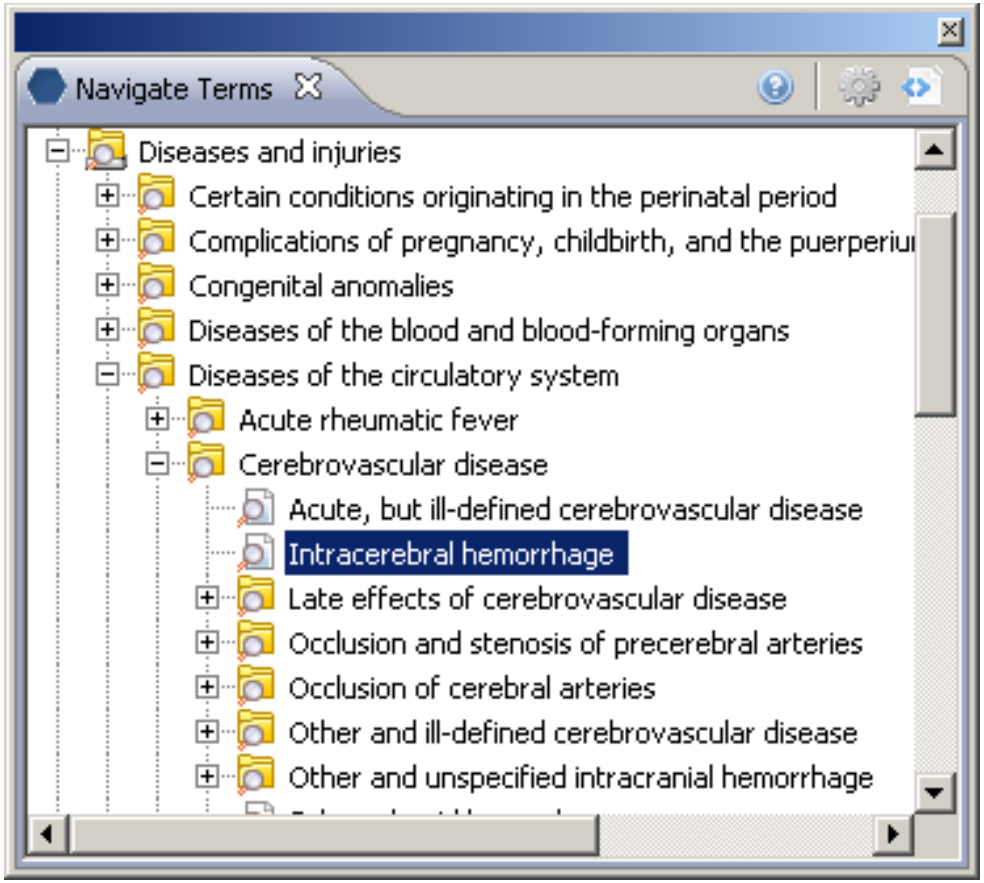
i2b2 Star Schema



Concept_dimension “look” …

CONCEPT_PATH	CONCEPT_CD	NAME_CHAR
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.4	Toxic myopathy
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.5	Myopathy in endocrine diseases …
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.6	Symptomatic inflammatory myop…
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.8	Other myopathies
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.81	Critical illness myopathy
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.89	Other myopathies
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde…	ICD9:359.9	Myopathy, unspecified
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(260) Kwashiorkor\	ICD9:260	Hypoproteinosis
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(261) Nutritional maras…	ICD9:261	Marasmus
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(262) Other severe pr…	ICD9:262	Other severe protein-calorie mal…
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263	Other and unspecified protein-c…
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263.0	Malnutrition of moderate degree
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263.1	Malnutrition of mild degree
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263.2	Dwarfism, nutritional
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263.8	Other protein-calorie malnutrition
\i2b2\Diagnoses\Nutritional deficiences (260-269)\(263) Other and unspe…	ICD9:263.9	Protein-calorie undernutrition

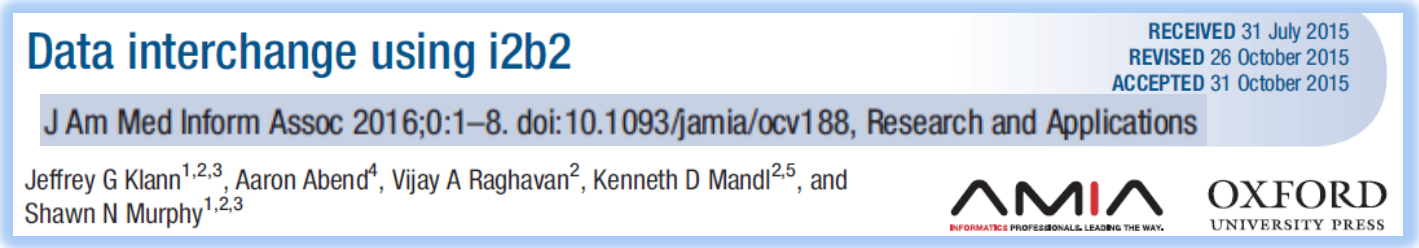
Metadata Presentation



Diseases and injuries \ Diseases of the circulatory system \
Cerebrovascular disease \ Intracerebral hemorrhage

Adapting i2b2 to PCORNet Data Model

1. Ontology-Driven Physical Transformation into PCORNet Common Data Model with a Generalizable approach to adapt to other Common Data Models



Data interchange using i2b2

RECEIVED 31 July 2015
REVISED 26 October 2015
ACCEPTED 31 October 2015

J Am Med Inform Assoc 2016;0:1–8. doi:10.1093/jamia/ocv188, Research and Applications

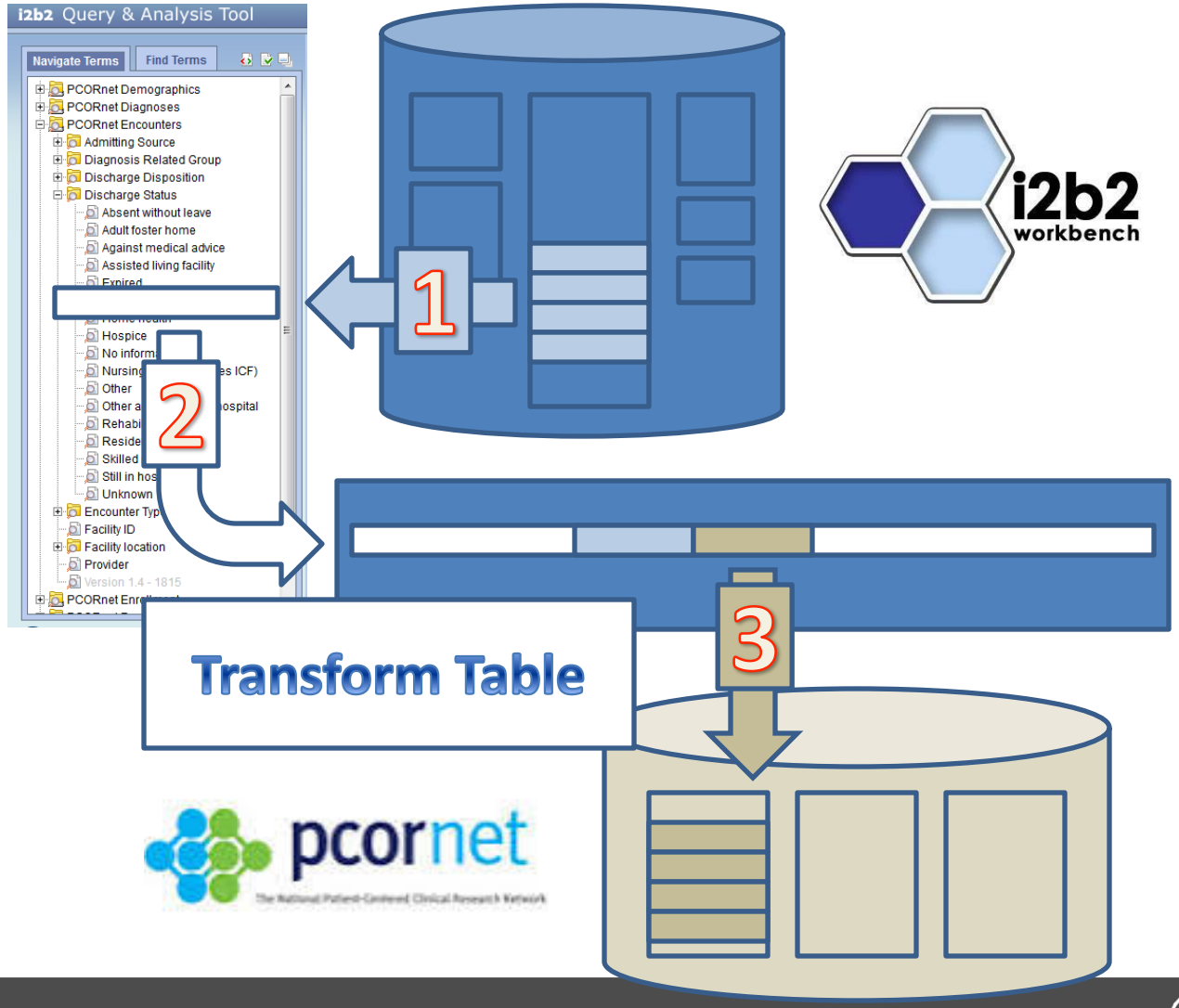
Jeffrey G Klann[1,2,3], Aaron Abend[4], Vijay A Raghavan[2], Kenneth D Mandl[2,5], and Shawn N Murphy[1,2,3]

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD UNIVERSITY PRESS

2. Implement a SHRINE adapter enabling participation in a peer-to-peer network

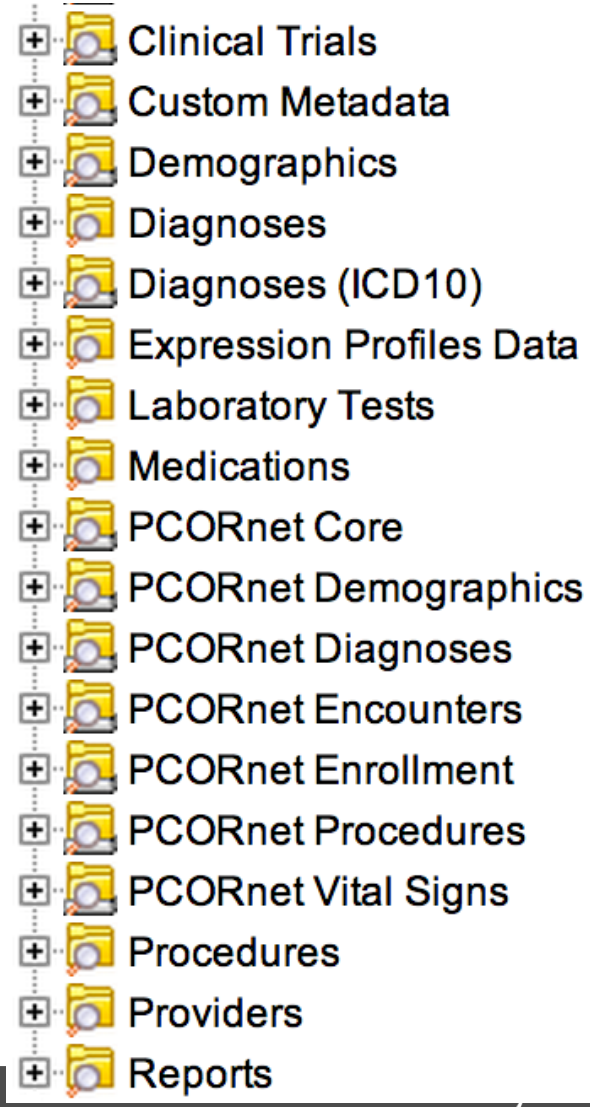
Physical data transformations to non-i2b2 formats ontology driven



New Information Model Ontology

Consensus Ontology can live alongside other ontologies

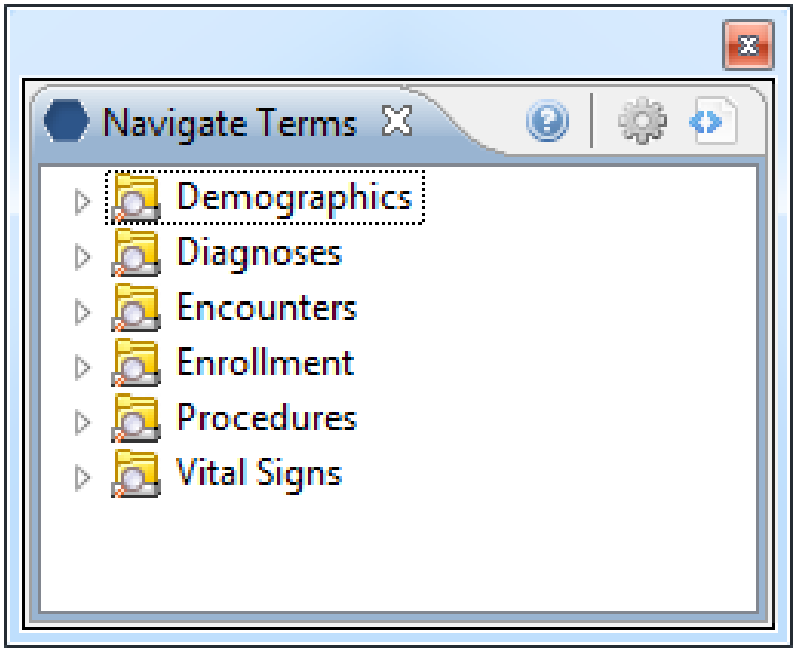
(For example: PCORNet CDM ontology and the i2b2 demo ontology in this case)



PCORI-centric Use Case

A requirement in SCILHS is that all i2b2 instances have to map their ontologies to a central PCORI CDM ontology.

- Local codes need to be integrated into the PCORI hierarchy
- Tools and strategies to assist with this effort



ICD-9 plain_code example

- If your institution uses ICD-9 but in a plain_code format…
  - Change the c_basecode to match your code format.

PCORI_BASECODE	C_BASECODE
ICD9:250.1	2501

```
update pcornet
set c_basecode = substring(pcori_basecode, 6, 25)
where pcori_basecode like 'ICD9:%'
and c_fullname like ‘\PCORI\DIAGNOSIS\09\%‘;

update pcornet
set c_basecode = replace(c_basecode, '.', '')
where pcori_basecode like 'ICD9:%'
and c_fullname like ‘\PCORI\DIAGNOSIS\09\%‘;
```

Why does SCILHS not just only use PCORNet Common Data Model for All Queries??

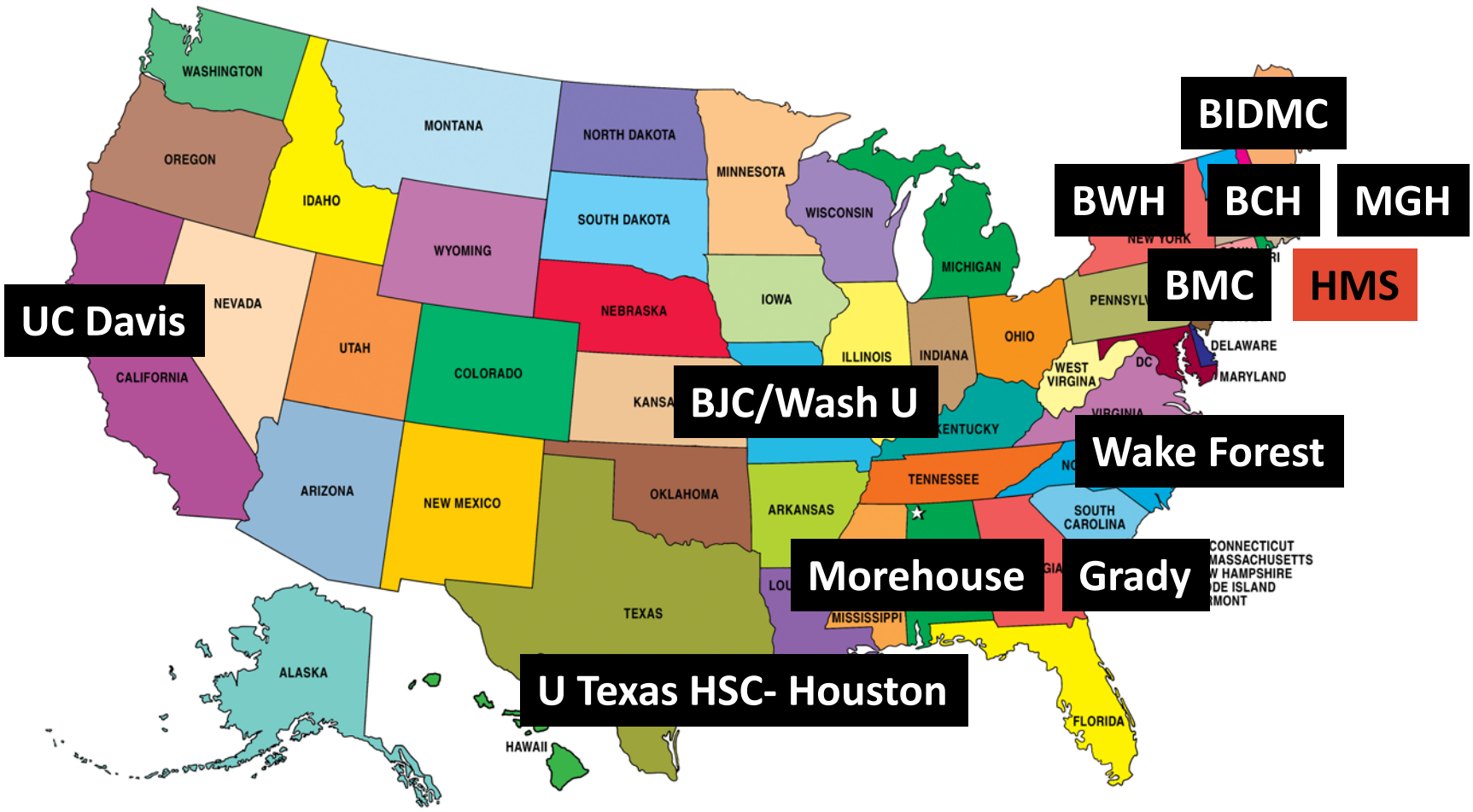
- PCORNet CDM is a measured approach to create a core dataset queryable by a central authority
- But i2b2 is being used as a complete workflow to support clinical trials in a peer-to-peer network

Key Challenges

- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

SCILHS 11 Participating Sites

Phase II: 10/1/15 – 9/30/18



Agreements/Policy

- Governance Agreement
  - Executed March, 2016
- Master Reliance Agreement
  - Implemented June, 2015
- Contracting & BAAs (MySCILHS IVR)
- PCORNet Data Sharing Agreement v1
  - 6/11 signed
- SCILHS DSA & DUA template (in progress)
- Engagement Policy (in progress)

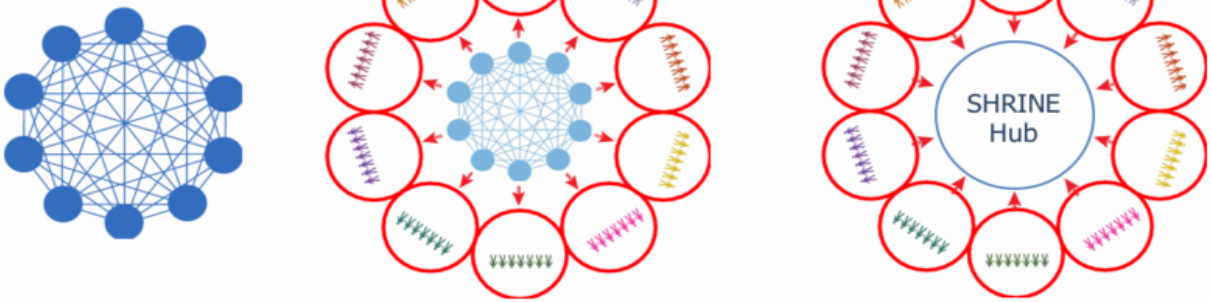
What is SCILHS?

Jun 2, 2016

About the Network

Leveraging an informatics infrastructure that our investigators have developed over the past 15 years, the **Department of Biomedical Informatics at Harvard Medical School** is working across a growing number of **healthcare centers throughout the U.S.** to develop the Scalable Collaborative Infrastructure for a Learning Health System (SCILHS, pronounced "skills"). An **open-source** platform, the system covers more than **10 million patients** and enables clinician and patient participation in research. As of this writing, **a dozen clinical studies are underway** that involve SCILHS.

SCILHS is also a Clinical Data Research Network (CDRN) in the **Patient-Centered Outcomes Research Institute's PCORnet**, a national effort to instantiate a 'network of networks' that supports large-scale comparative effectiveness research. SCILHS uses Informatics for Integrating Biology and the Bedside (i2b2) as its technical backbone and has adopted the PCORnet Common Data Model (CDM) as its foundation for interoperable data exchange. SCILHS uses the Shared Health Research Information Network (SHRINE) platform to enable live distributed querying across sites.



10 EHRs

+ 10 i2b2 Sidecars

+ 1 SHRINE Hub to connect them all

Feasibility / Clinical Study with no contact Information

1

Researcher requests access to the SCILHS Network and define patient cohorts using Shrine Web Client

2

Researcher requests data from SCILHS sites using Shrine Web Client

3

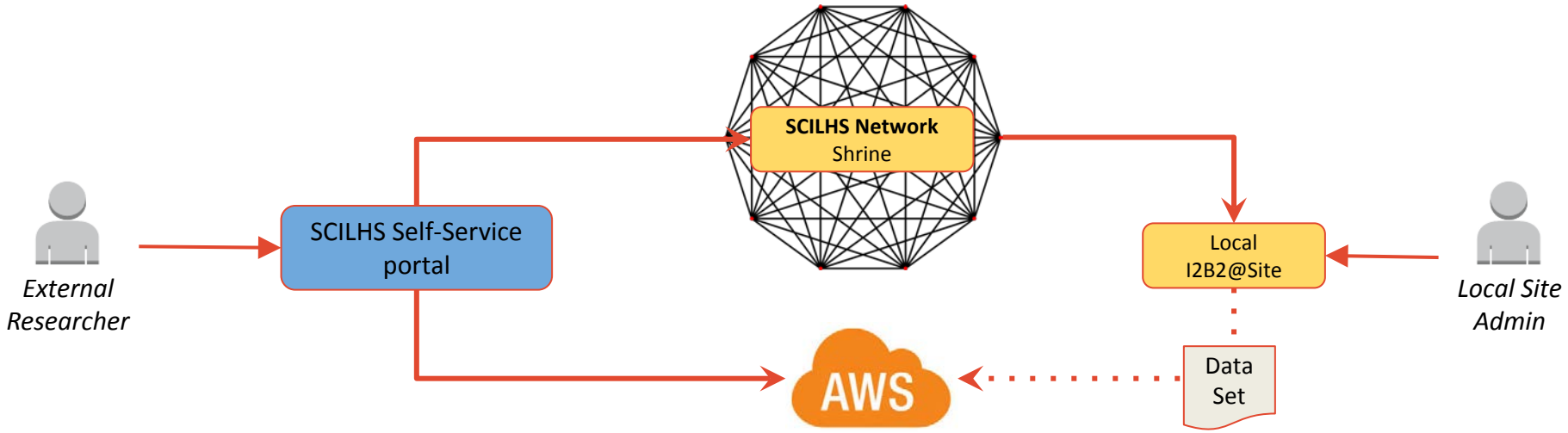
Local Site Admin reviews and approves the query

4

Local Site Admin generates data set file and uploads it into an AWS space

5

Researcher accesses data sets in AWS and performs phenotype computation



SCILHS scalable collaborative infrastructure for a learning health system

SCHILS SSP Panel

ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/index

Isaac

Guest

Request Access

Navigation

- Home
- The SCILHS Network
- Tutorials
- Shrine Demo
- Analytical Resources Demo
- Request Access / Login

What is SCILHS?

Jun 2, 2016

About the Network

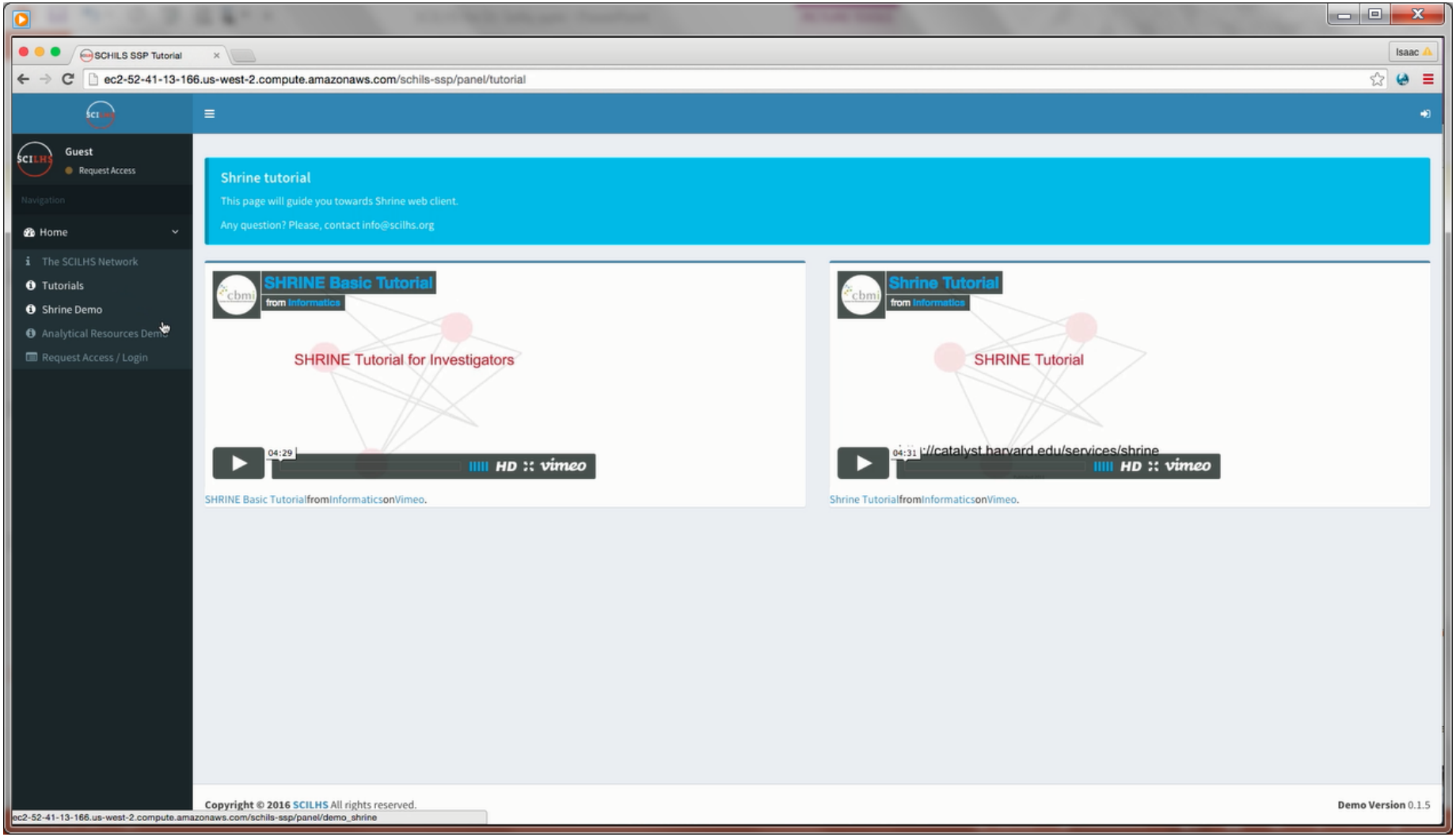
Leveraging an informatics infrastructure that our investigators have developed over the past 15 years, the **Department of Biomedical Informatics at Harvard Medical School** is working across a growing number of **healthcare centers throughout the U.S.** to develop the Scalable Collaborative Infrastructure for a Learning Health System (SCILHS, pronounced "skills"). An **open-source** platform, the system covers more than **10 million patients** and enables clinician and patient participation in research. As of this writing, **a dozen clinical studies are underway** that involve SCILHS.

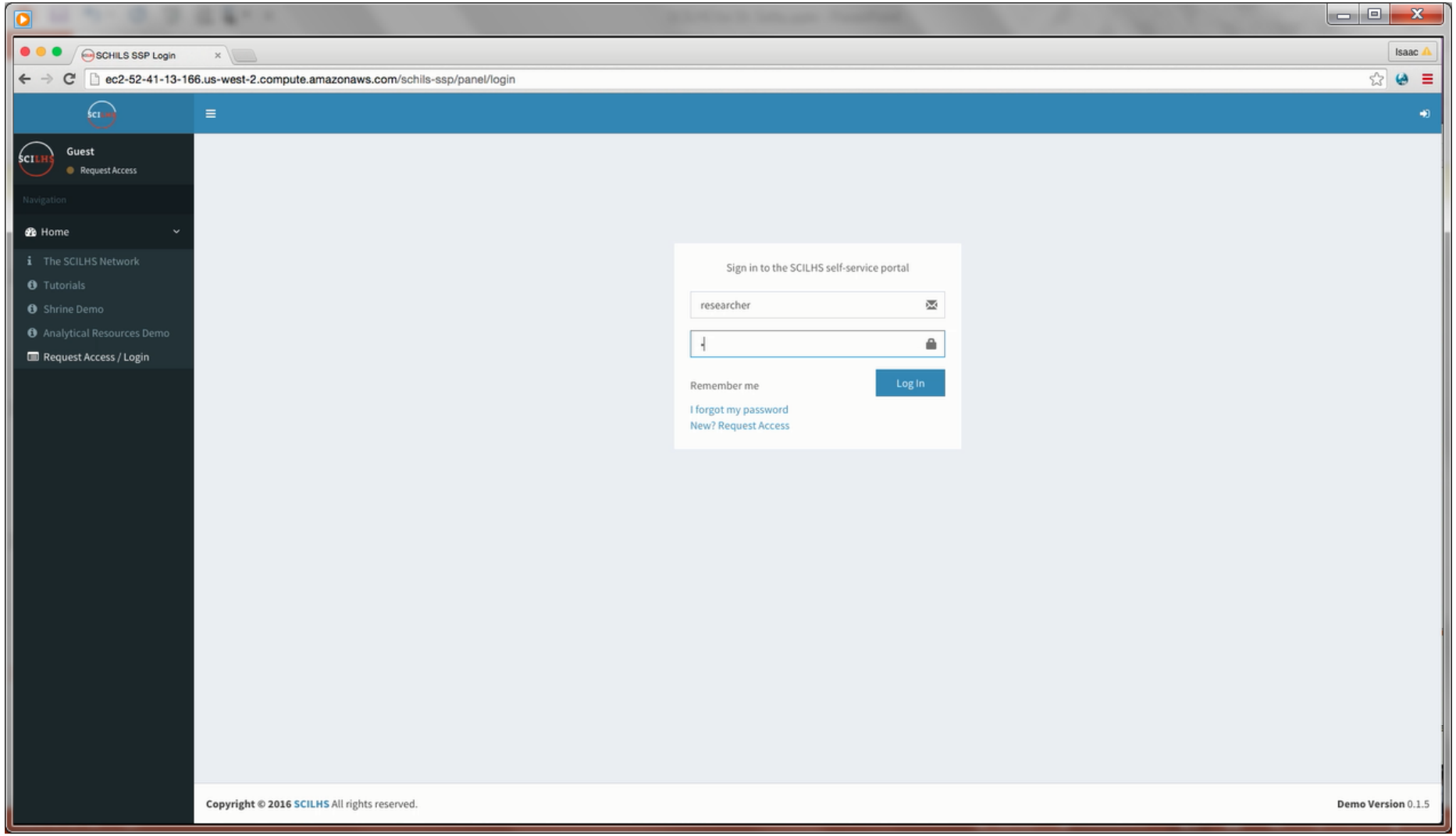
SCILHS is also a Clinical Data Research Network (CDRN) in the **Patient-Centered Outcomes Research Institute's PCORnet**, a national effort to instantiate a

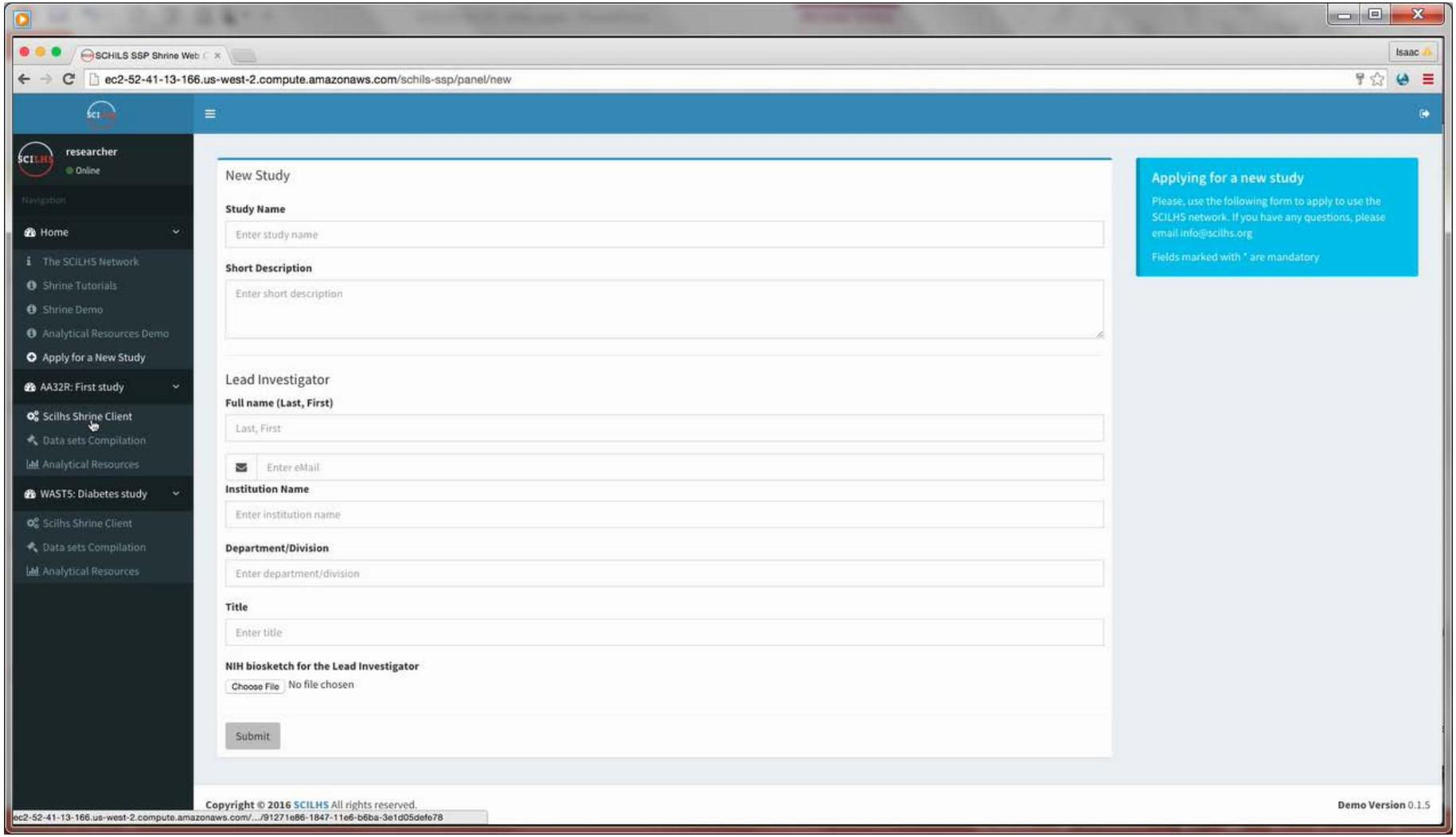
Demo Version 0.1.5

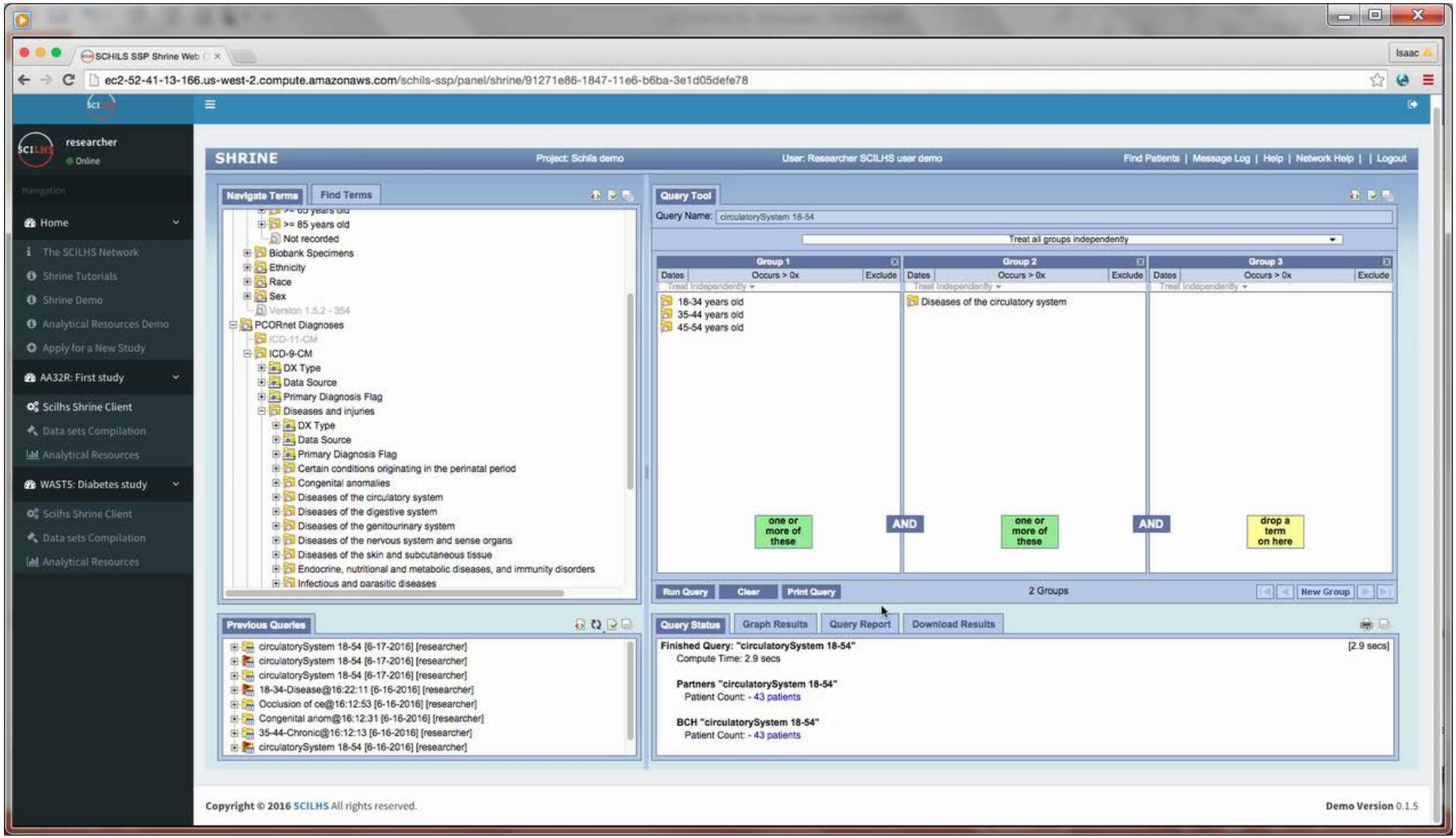
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/index

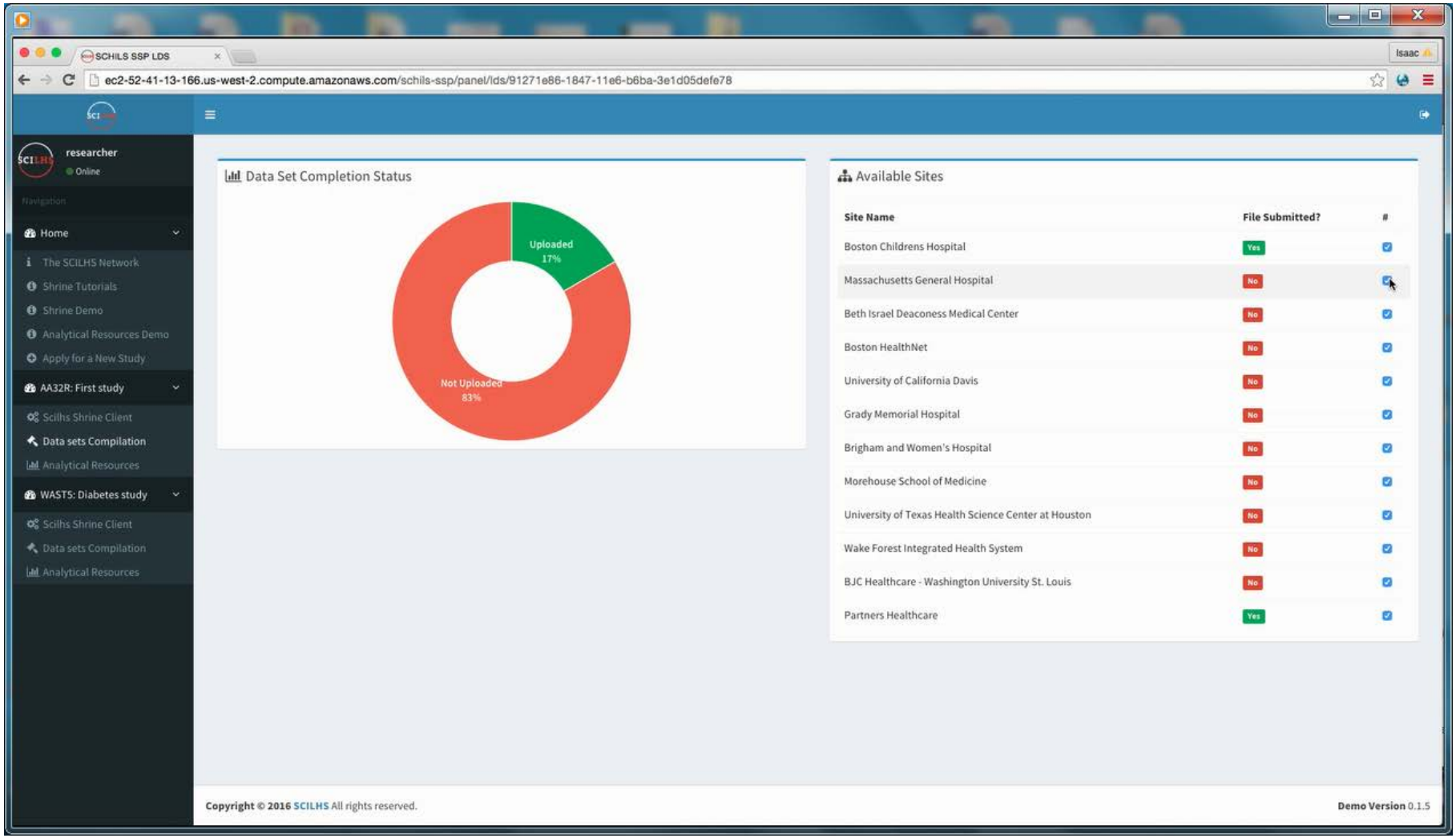
SCILHS scalable collaborative infrastructure for a learning health system

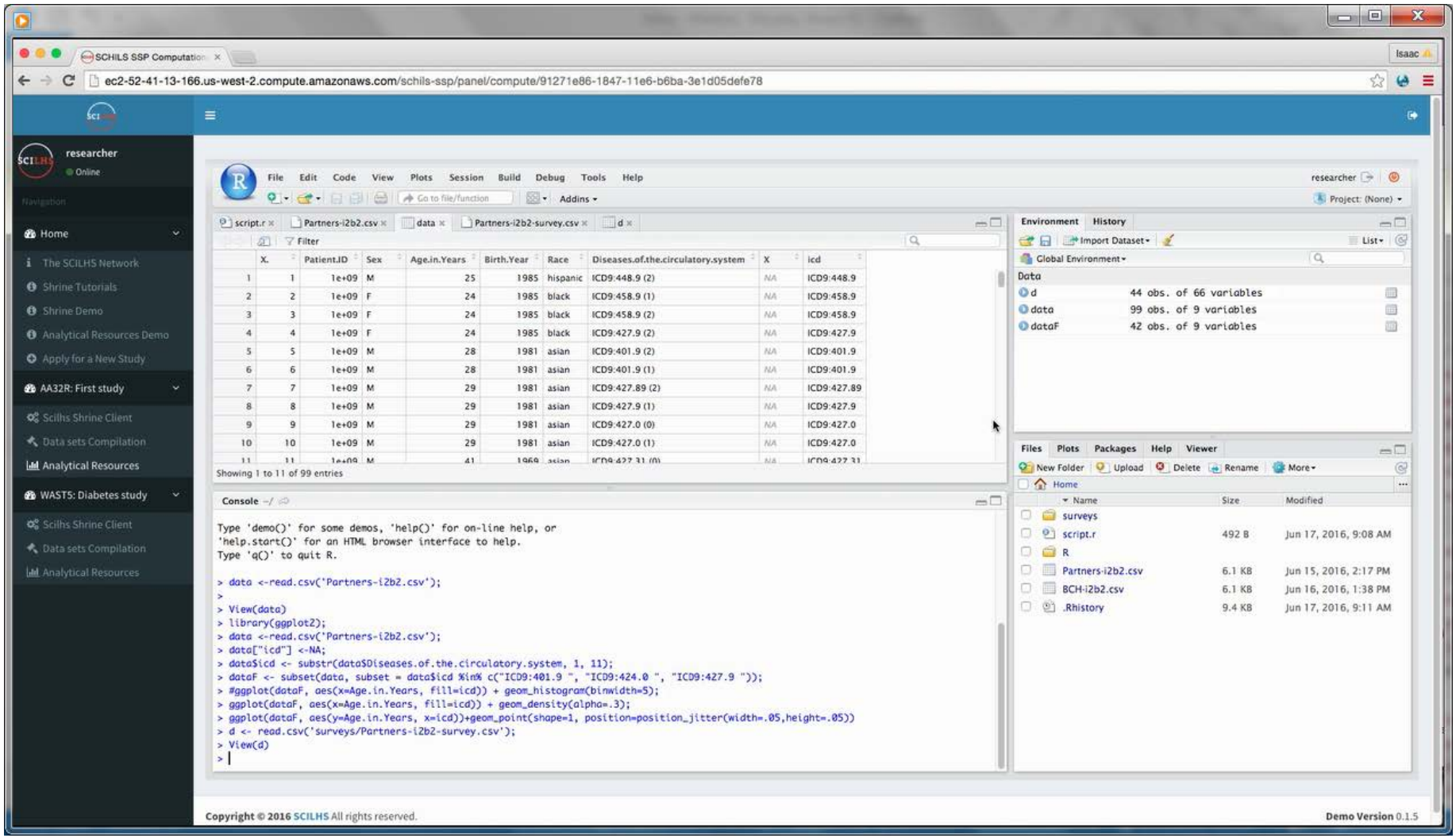

SCHILS SSP Tutorial
Isaac
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/tutorial
Guest
Request Access
Navigation
Home
The SCILHS Network
Tutorials
Shrine Demo
Analytical Resources Demo
Request Access / Login
Shrine tutorial
This page will guide you towards Shrine web client.
Any question? Please, contact info@scilhs.org
SHRINE Basic Tutorial
from Informatics
SHRINE Tutorial for Investigators
04:29
HD :: vimeo
SHRINE Basic TutorialfromInformaticsonVimeo.
Shrine Tutorial
from Informatics
SHRINE Tutorial
04:31
HD :: vimeo
Shrine TutorialfromInformaticsonVimeo.
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/demo_shrine

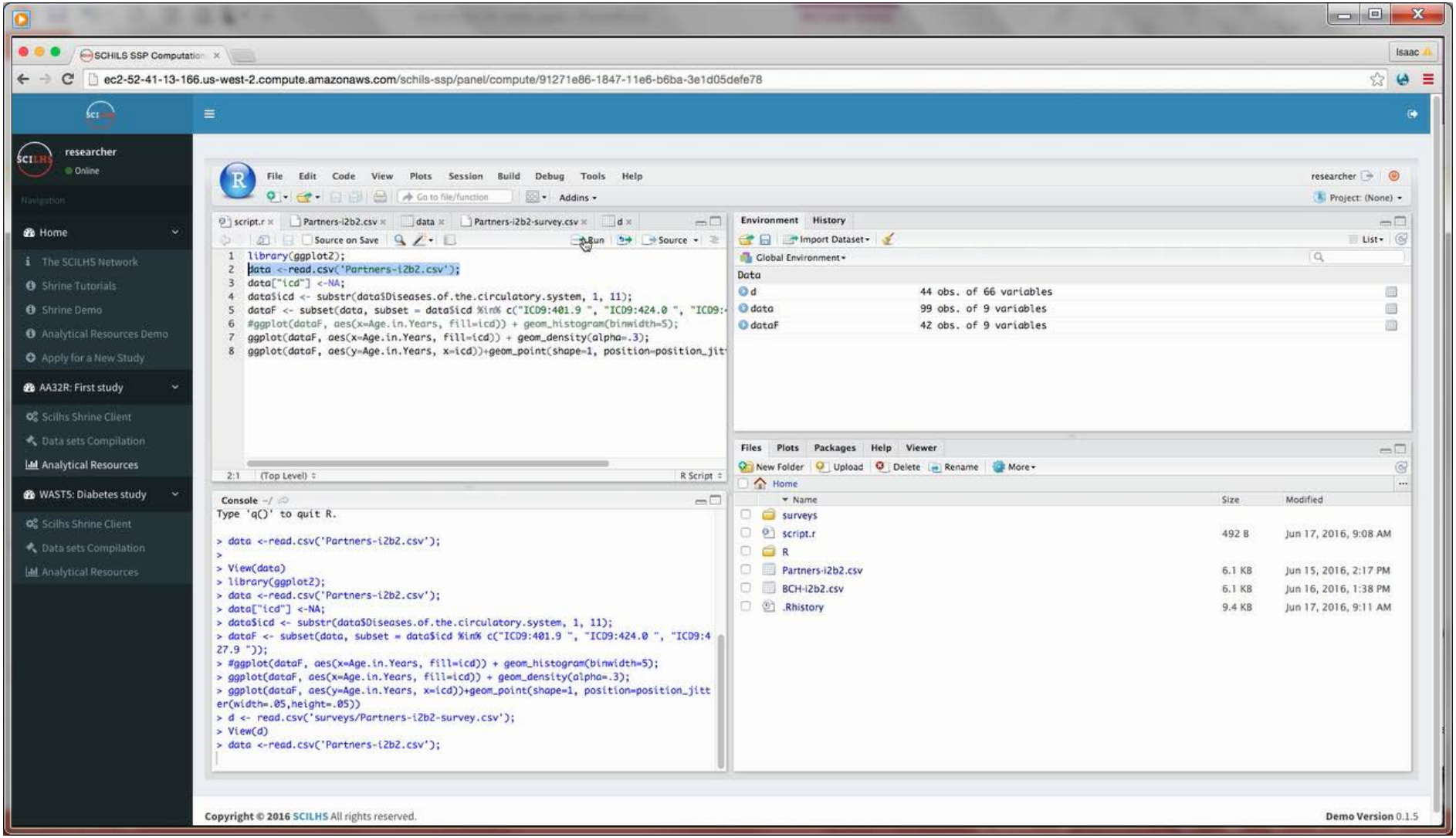

SCHILS SSP Login
Isaac
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/login
Guest
Request Access
Navigation
Home
The SCILHS Network
Tutorials
Shrine Demo
Analytical Resources Demo
Request Access / Login
Sign in to the SCILHS self-service portal
researcher
Remember me
Log In
I forgot my password
New? Request Access
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5

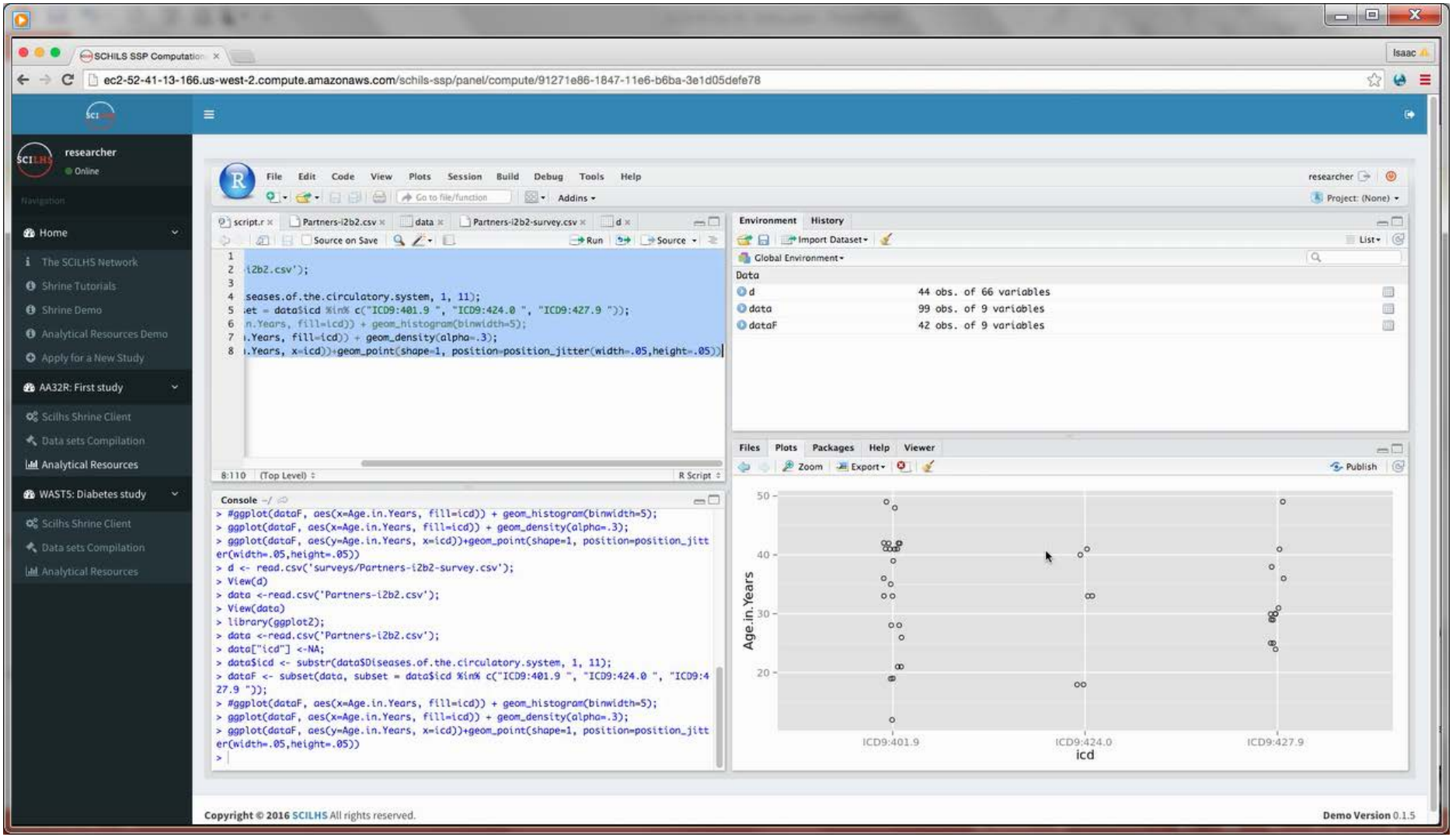

SCHILS SSP Shrine Web
Isaac
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/new
researcher
Online
Navigation
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
New Study
Study Name
Enter study name
Short Description
Enter short description
Lead Investigator
Full name (Last, First)
Last, First
Enter eMail
Institution Name
Enter institution name
Department/Division
Enter department/division
Title
Enter title
NIH biosketch for the Lead Investigator
Choose File
No file chosen
Submit
Applying for a new study
Please, use the following form to apply to use the SCILHS network. If you have any questions, please email info@scilhs.org
Fields marked with * are mandatory
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/.../91271e86-1847-11e6-b6ba-3e1d05defe78

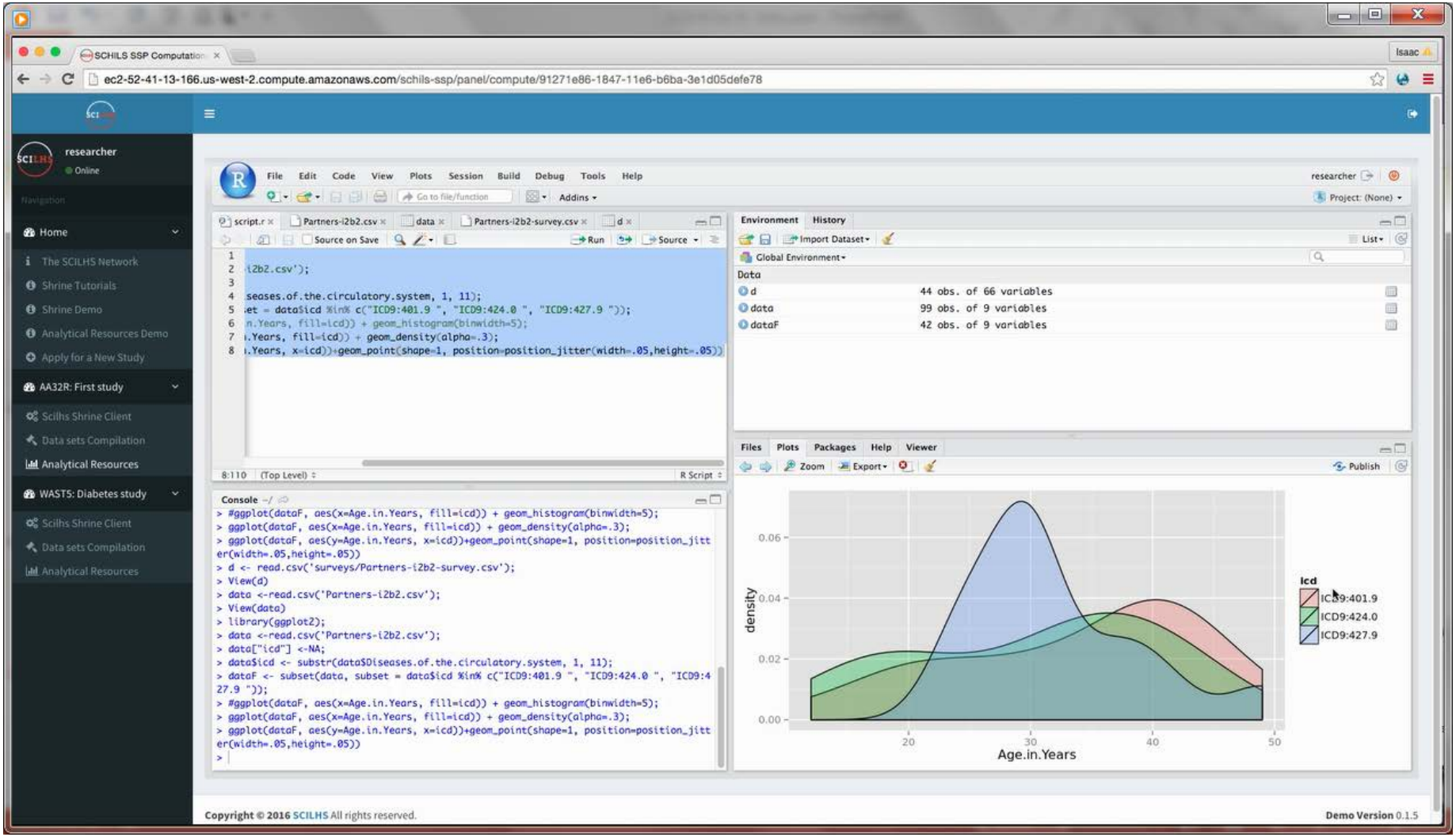

SCHILS SSP Shrine Web
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/shrine/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Navigation
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
SHRINE
Project: Schils demo
User: Researcher SCILHS user demo
Find Patients | Message Log | Help | Network Help | | Logout
Navigate Terms
Find Terms
>= 85 years old
Not recorded
Biobank Specimens
Ethnicity
Race
Sex
Version 1.5.2 - 354
PCORnet Diagnoses
ICD-11-CM
ICD-9-CM
DX Type
Data Source
Primary Diagnosis Flag
Diseases and injuries
DX Type
Data Source
Primary Diagnosis Flag
Certain conditions originating in the perinatal period
Congenital anomalies
Diseases of the circulatory system
Diseases of the digestive system
Diseases of the genitourinary system
Diseases of the nervous system and sense organs
Diseases of the skin and subcutaneous tissue
Endocrine, nutritional and metabolic diseases, and immunity disorders
Infectious and parasitic diseases
Query Tool
Query Name: circulatorySystem 18-54
Treat all groups independently
Group 1
Group 2
Group 3
Dates
Occurs > 0x
Exclude
Treat Independently
18-34 years old
35-44 years old
45-54 years old
Diseases of the circulatory system
one or more of these
AND
one or more of these
AND
drop a term on here
Run Query
Clear
Print Query
2 Groups
New Group
Previous Queries
circulatorySystem 18-54 [6-17-2016] [researcher]
circulatorySystem 18-54 [6-17-2016] [researcher]
circulatorySystem 18-54 [6-17-2016] [researcher]
18-34-Disease@16:22:11 [6-16-2016] [researcher]
Occlusion of ce@16:12:53 [6-16-2016] [researcher]
Congenital anom@16:12:31 [6-16-2016] [researcher]
35-44-Chronic@16:12:13 [6-16-2016] [researcher]
circulatorySystem 18-54 [6-16-2016] [researcher]
Query Status
Graph Results
Query Report
Download Results
Finished Query: "circulatorySystem 18-54"
[2.9 secs]
Compute Time: 2.9 secs
Partners "circulatorySystem 18-54"
Patient Count: - 43 patients
BCH "circulatorySystem 18-54"
Patient Count: - 43 patients
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5


SCHILS SSP LDS
Isaac
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/lds/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Navigation
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
Data Set Completion Status
Uploaded
17%
Not Uploaded
83%
Available Sites
Site Name
File Submitted?
#
Boston Childrens Hospital
Yes
Massachusetts General Hospital
No
Beth Israel Deaconess Medical Center
No
Boston HealthNet
No
University of California Davis
No
Grady Memorial Hospital
No
Brigham and Women's Hospital
No
Morehouse School of Medicine
No
University of Texas Health Science Center at Houston
No
Wake Forest Integrated Health System
No
BJC Healthcare - Washington University St. Louis
No
Partners Healthcare
Yes
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5


SCHILS SSP Computation
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/compute/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
File Edit Code View Plots Session Build Debug Tools Help
script.r
Partners-i2b2.csv
data
Partners-i2b2-survey.csv
d
Showing 1 to 11 of 99 entries
Environment History
Import Dataset
Global Environment
Data
d 44 obs. of 66 variables
data 99 obs. of 9 variables
dataF 42 obs. of 9 variables
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home
surveys
script.r 492 B Jun 17, 2016, 9:08 AM
R
Partners-i2b2.csv 6.1 KB Jun 15, 2016, 2:17 PM
BCH-i2b2.csv 6.1 KB Jun 16, 2016, 1:38 PM
.Rhistory 9.4 KB Jun 17, 2016, 9:11 AM
Console
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> data <-read.csv('Partners-i2b2.csv');
> View(data)
> library(ggplot2);
> data <-read.csv('Partners-i2b2.csv');
> data["icd"] <-NA;
> data$icd <- substr(data$Diseases.of.the.circulatory.system, 1, 11);
> dataF <- subset(data, subset = data$icd %in% c("ICD9:401.9 ", "ICD9:424.0 ", "ICD9:427.9 "));
> #ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_histogram(binwidth=5);
> ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_density(alpha=.3);
> ggplot(dataF, aes(y=Age.in.Years, x=icd))+geom_point(shape=1, position=position_jitter(width=.05,height=.05))
> d <- read.csv('surveys/Partners-i2b2-survey.csv');
> View(d)
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5


SCHILS SSP Computation
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/compute/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
File Edit Code View Plots Session Build Debug Tools Help
script.r
Partners-i2b2.csv
data
Partners-i2b2-survey.csv
d
1 library(ggplot2);
2 data <-read.csv('Partners-i2b2.csv');
3 data["icd"] <-NA;
4 data$icd <- substr(data$Diseases.of.the.circulatory.system, 1, 11);
5 dataF <- subset(data, subset = data$icd %in% c("ICD9:401.9 ", "ICD9:424.0 ", "ICD9:
6 #ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_histogram(binwidth=5);
7 ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_density(alpha=.3);
8 ggplot(dataF, aes(y=Age.in.Years, x=icd))+geom_point(shape=1, position=position_jit
Environment History
Global Environment
Data
d 44 obs. of 66 variables
data 99 obs. of 9 variables
dataF 42 obs. of 9 variables
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home
Name Size Modified
surveys
script.r 492 B Jun 17, 2016, 9:08 AM
R
Partners-i2b2.csv 6.1 KB Jun 15, 2016, 2:17 PM
BCH-i2b2.csv 6.1 KB Jun 16, 2016, 1:38 PM
.Rhistory 9.4 KB Jun 17, 2016, 9:11 AM
Console
Type 'q()' to quit R.
> data <-read.csv('Partners-i2b2.csv');
> View(data)
> library(ggplot2);
> data <-read.csv('Partners-i2b2.csv');
> data["icd"] <-NA;
> data$icd <- substr(data$Diseases.of.the.circulatory.system, 1, 11);
> dataF <- subset(data, subset = data$icd %in% c("ICD9:401.9 ", "ICD9:424.0 ", "ICD9:427.9 "));
> #ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_histogram(binwidth=5);
> ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_density(alpha=.3);
> ggplot(dataF, aes(y=Age.in.Years, x=icd))+geom_point(shape=1, position=position_jitter(width=.05,height=.05))
> d <- read.csv('surveys/Partners-i2b2-survey.csv');
> View(d)
> data <-read.csv('Partners-i2b2.csv');
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5


SCHILS SSP Computation
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/compute/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
File Edit Code View Plots Session Build Debug Tools Help
script.r
Partners-i2b2.csv
data
Partners-i2b2-survey.csv
Environment History
Global Environment
Data
d 44 obs. of 66 variables
data 99 obs. of 9 variables
dataF 42 obs. of 9 variables
Files Plots Packages Help Viewer
Zoom Export Publish
Age.in.Years
icd
ICD9:401.9
ICD9:424.0
ICD9:427.9
Console
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5


SCHILS SSP Computation
ec2-52-41-13-166.us-west-2.compute.amazonaws.com/schils-ssp/panel/compute/91271e86-1847-11e6-b6ba-3e1d05defe78
researcher
Online
Home
The SCILHS Network
Shrine Tutorials
Shrine Demo
Analytical Resources Demo
Apply for a New Study
AA32R: First study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
WAST5: Diabetes study
Scilhs Shrine Client
Data sets Compilation
Analytical Resources
File Edit Code View Plots Session Build Debug Tools Help
script.r
Partners-i2b2.csv
data
Partners-i2b2-survey.csv
d
Source on Save
Run
Source
Environment
History
Import Dataset
Global Environment
Data
d 44 obs. of 66 variables
data 99 obs. of 9 variables
dataF 42 obs. of 9 variables
Files
Plots
Packages
Help
Viewer
Zoom
Export
Publish
density
Age.in.Years
icd
ICD9:401.9
ICD9:424.0
ICD9:427.9
Console
> #ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_histogram(binwidth=5);
> ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_density(alpha=.3);
> ggplot(dataF, aes(y=Age.in.Years, x=icd))+geom_point(shape=1, position=position_jitter(width=.05,height=.05))
> d <- read.csv('surveys/Partners-i2b2-survey.csv');
> View(d)
> data <-read.csv('Partners-i2b2.csv');
> View(data)
> library(ggplot2);
> data <-read.csv('Partners-i2b2.csv');
> data["icd"] <-NA;
> data$icd <- substr(data$Diseases.of.the.circulatory.system, 1, 11);
> dataF <- subset(data, subset = data$icd %in% c("ICD9:401.9 ", "ICD9:424.0 ", "ICD9:427.9 "));
> #ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_histogram(binwidth=5);
> ggplot(dataF, aes(x=Age.in.Years, fill=icd)) + geom_density(alpha=.3);
> ggplot(dataF, aes(y=Age.in.Years, x=icd))+geom_point(shape=1, position=position_jitter(width=.05,height=.05))
Copyright © 2016 SCILHS All rights reserved.
Demo Version 0.1.5

Active SCILHS Studies

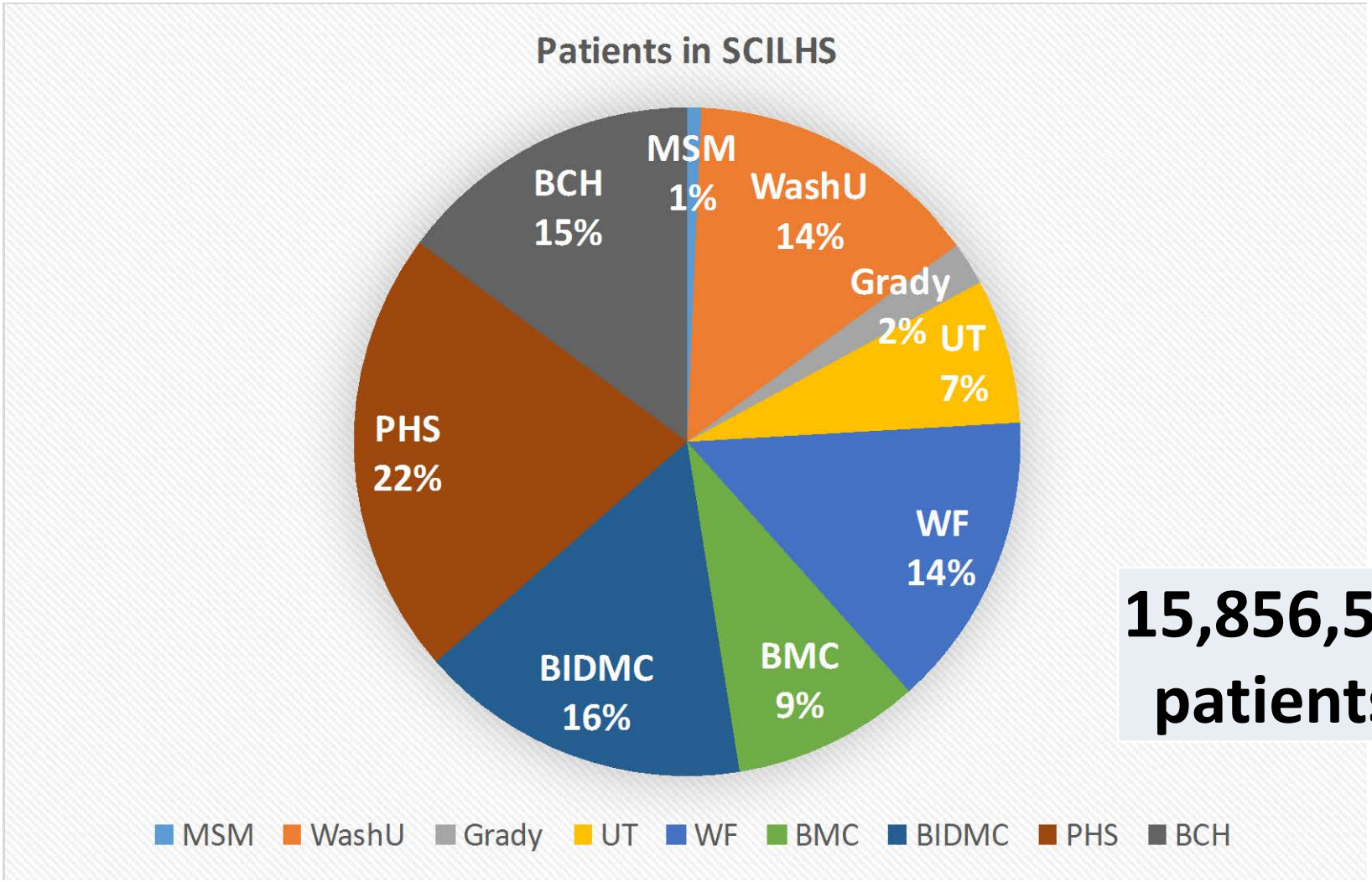
- Bariatric Surgery Obesity Demonstration Project (Tavakkoli)
- Pediatric Antibiotic Obesity Demonstration Project (Taveras)
- Suicide Prediction – Fuss Proposal (Smoller and Reis)
- Parkinson's Clinical Trial – Data Recruitment (Schwarzschild)
- INVESTED Clinical Trial (Solomon)
- RELIANCE Clinical Trial (Pending SSRC Review)

Represents all studies from Phase I and II

SCILHS Studies Pending Funding Decision

- PCORnet Health System Demonstration Project, Effectiveness of Case Management Programs in ACOs (Metlay) - FUNDED
- Methods to collect longer-term PROs for Trauma Registries (Haider)
- Population-Scale Predictive Modeling for Suicide Risk (Reis)

Characteristics of Network



15,856,539 patients!

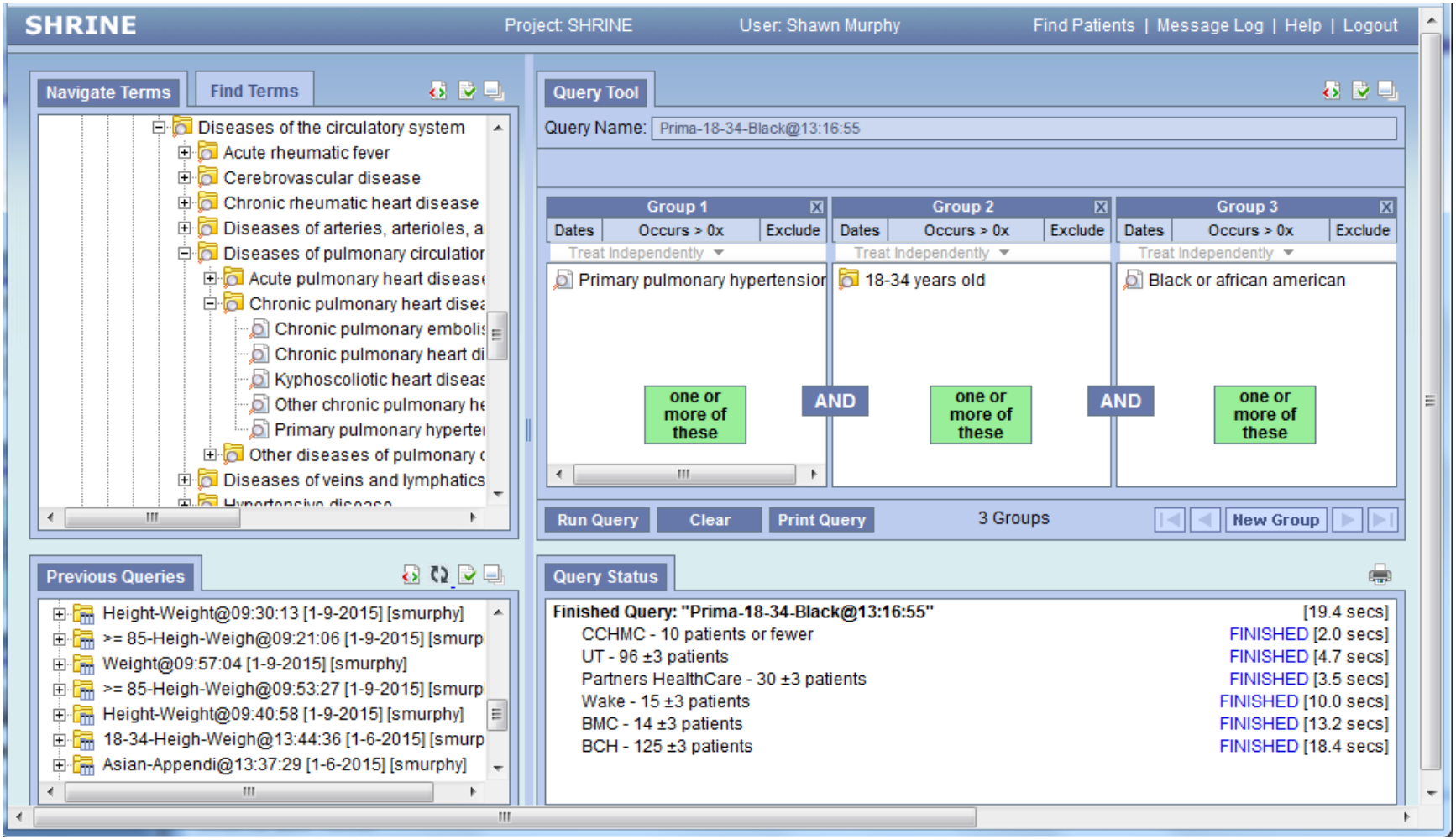
Key Challenges

- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

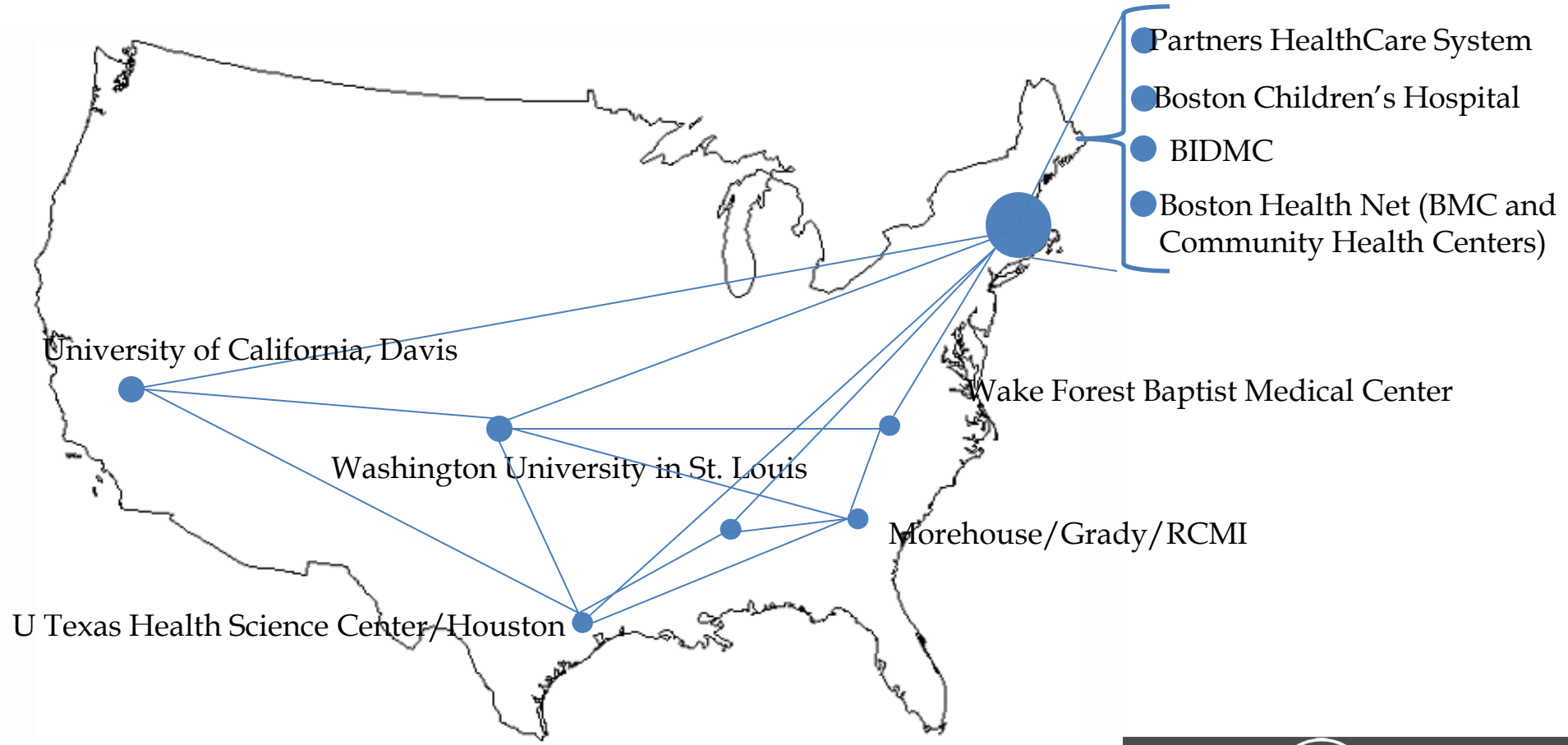
SHRINE and i2b2

- Rapid, real time query
- Increasingly complex queries possible
- PHI completely protected

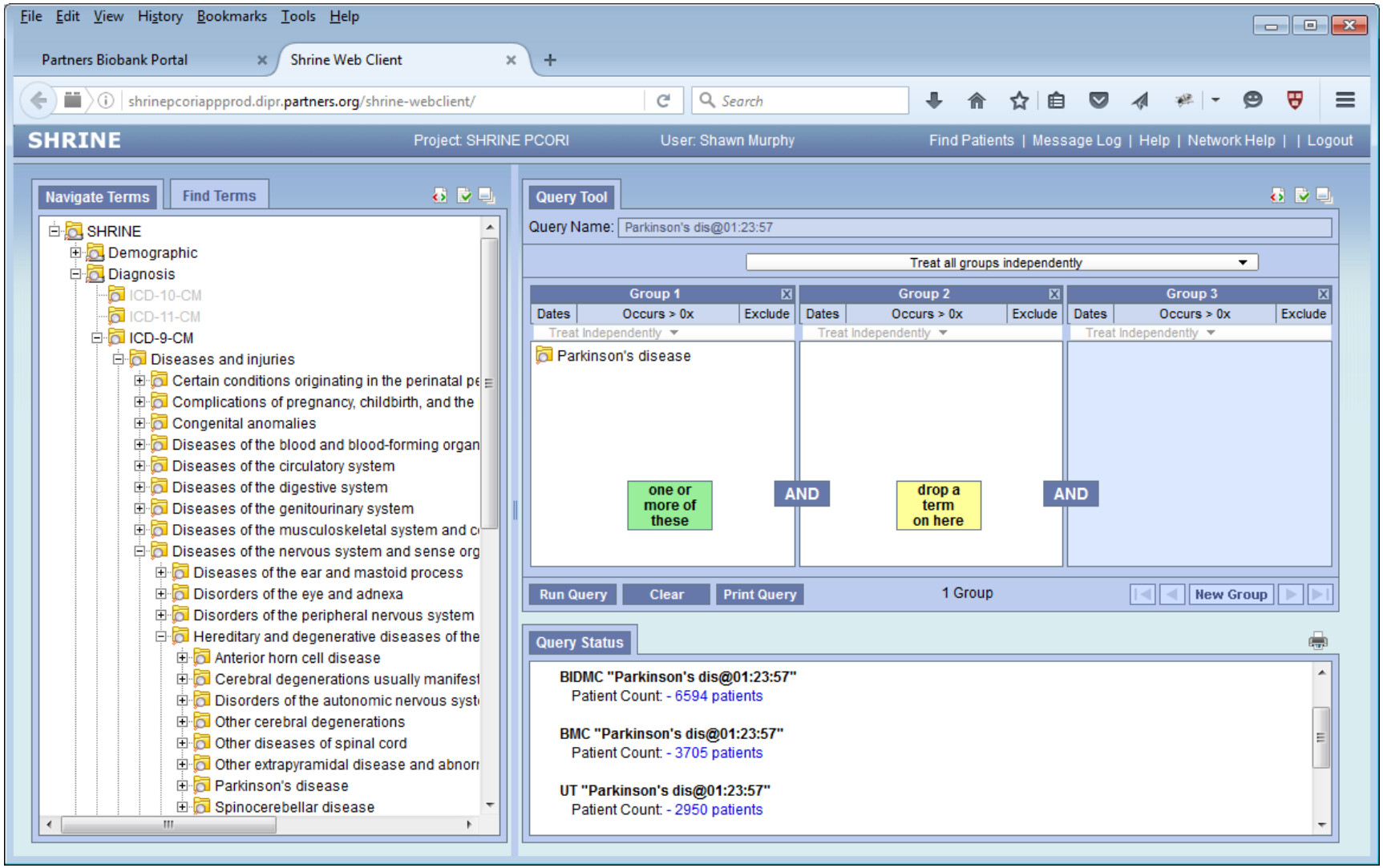
Researchers formulate queries across hospitals



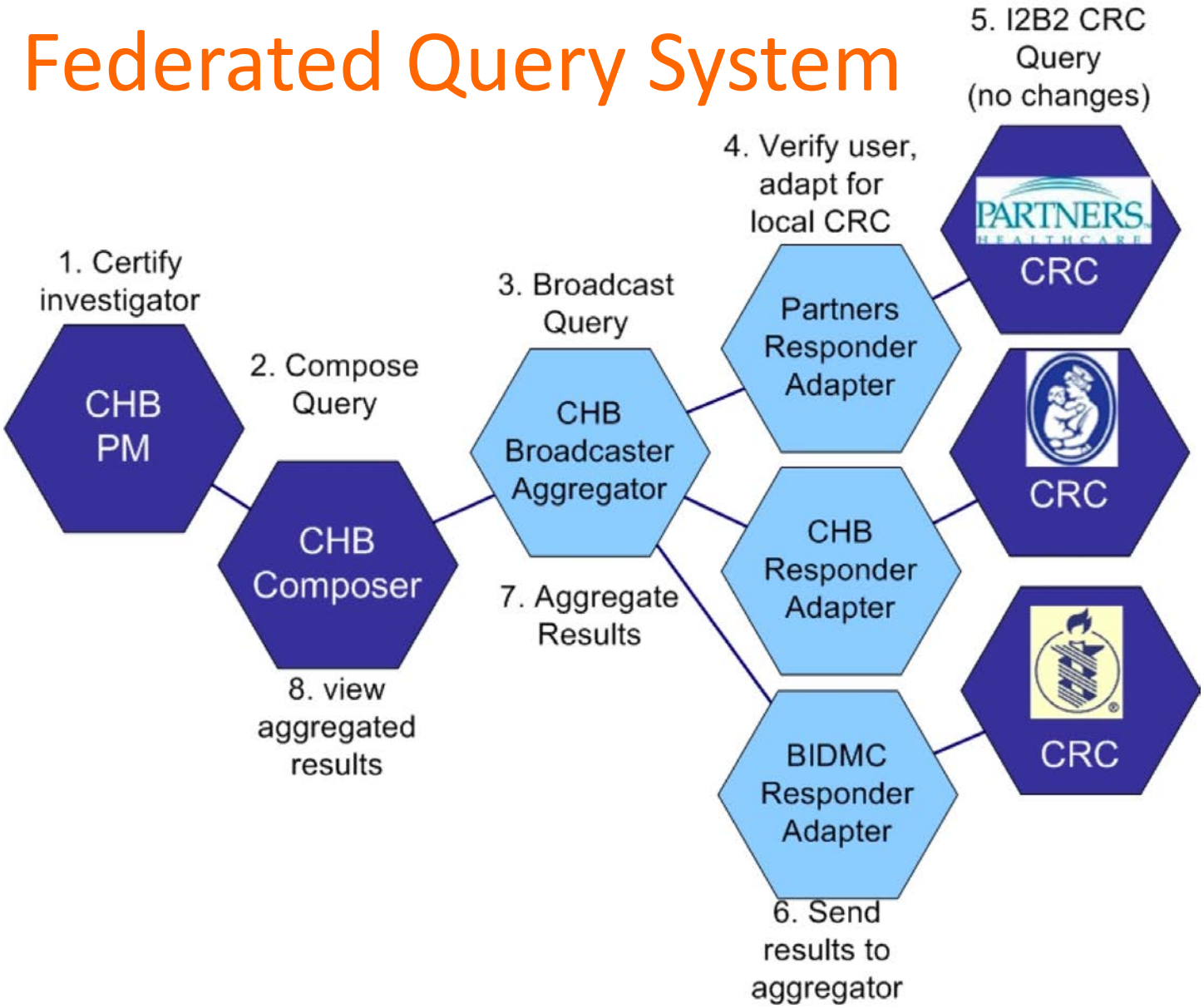
SCILHS Clinical Data Research Network



Network is Peer-to-Peer



Federated Query System



- **Measuring Quality of Decisions About Treatment of Depression**
  ClinicalTrials.gov Identifier: NCT01152307

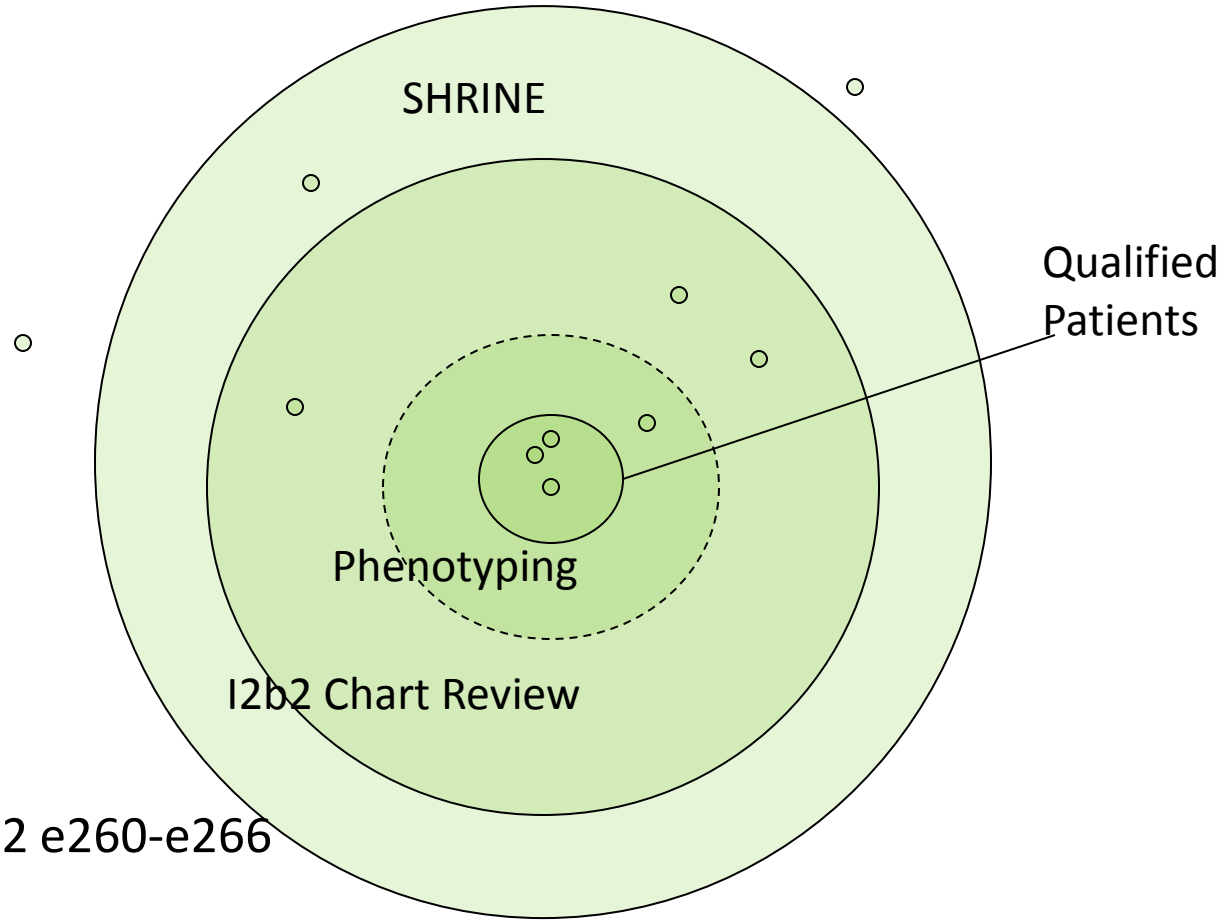
- Eligibility

  Ages Eligible for Study: 18 Years and older Genders Eligible for Study: Both Accepts Healthy Volunteers: No

- Inclusion Criteria:
- Aged 18+
- In the last 12 months, talked to a health care provider about starting or stopping a treatment (prescription medicine for depression or counseling)

- Exclusion Criteria:
- none

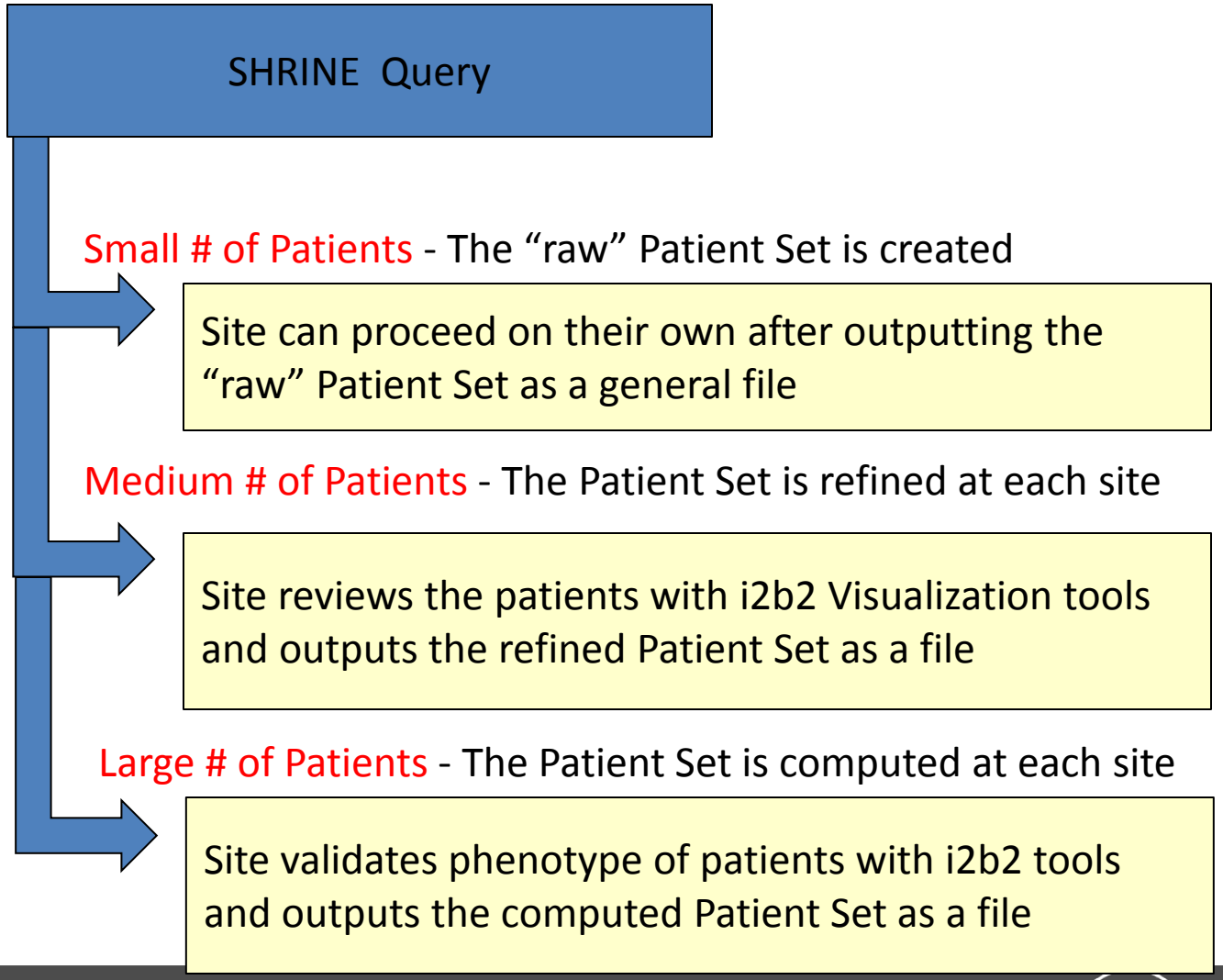
Strategy to Find Qualified Patients for Clinical Trials



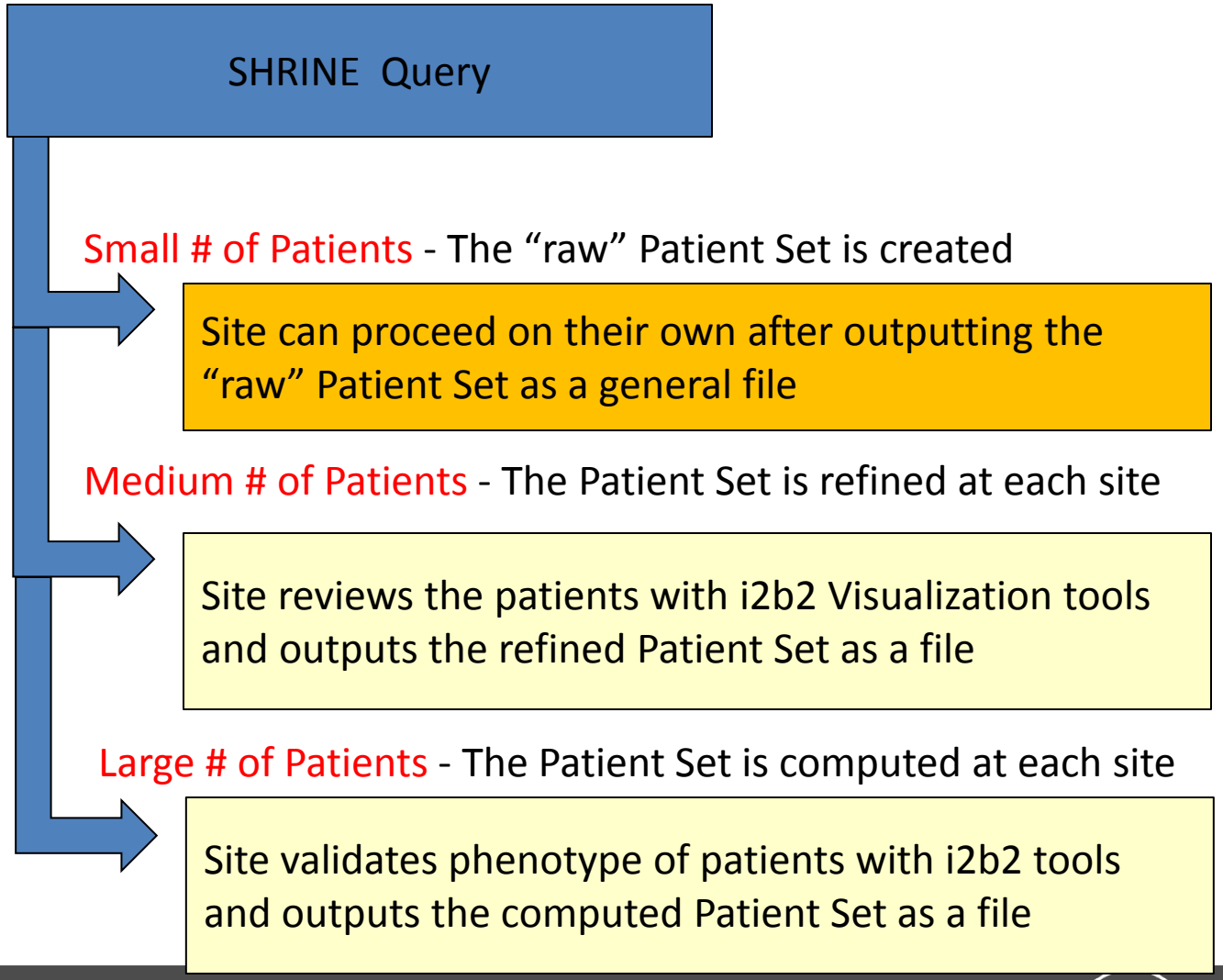
Jack London
JAMIA 2013;20:e2 e260-e266

Design-phase prediction of potential cancer clinical trial accrual success using a research data mart

Handling Clinical Trials in an i2b2-SHRINE Network depends on Number of Patients found



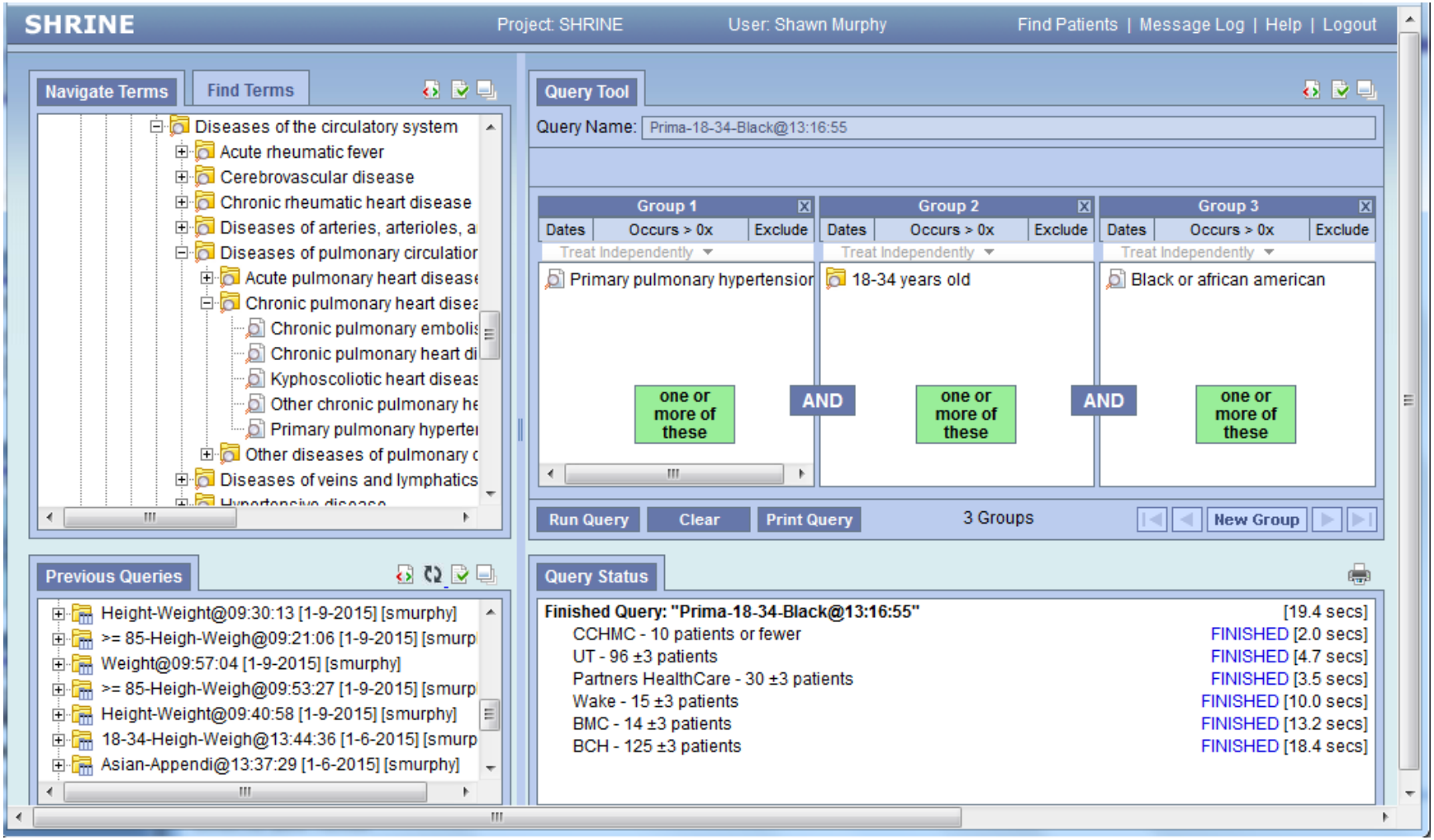
If a Small Number of Patients are Identified in SHRINE Query



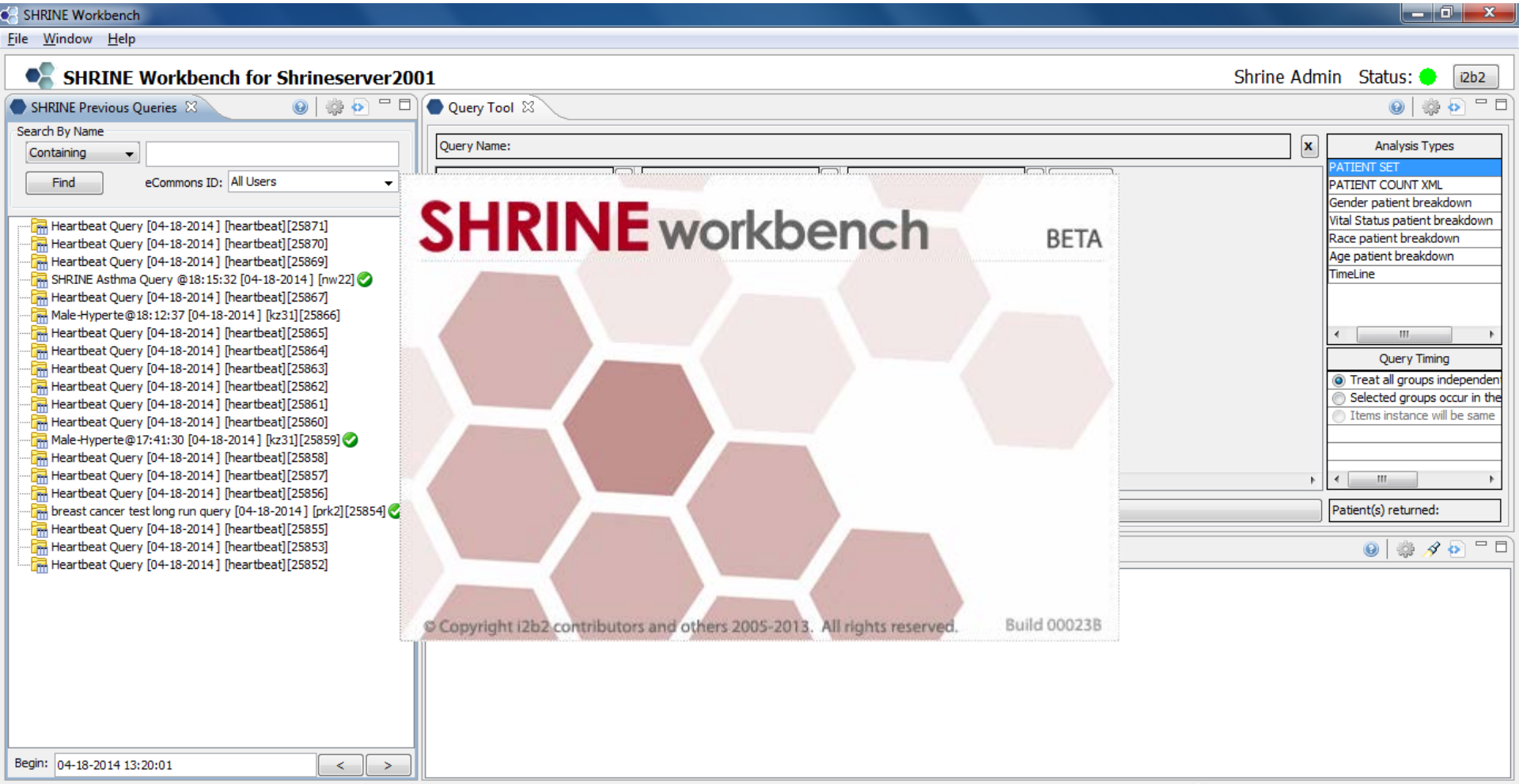
Clinician Sensitivities in the Recruitment Step – why the detailed data stays at the sites

- Control of Protected Health Information
  - PHI remains at the hospitals when collected under a waiver of consent
- Control of recruitment policies
- Patients as a scarce and valuable resource
  - Revenue and scientific recognition are tied to recruitment for clinical trails

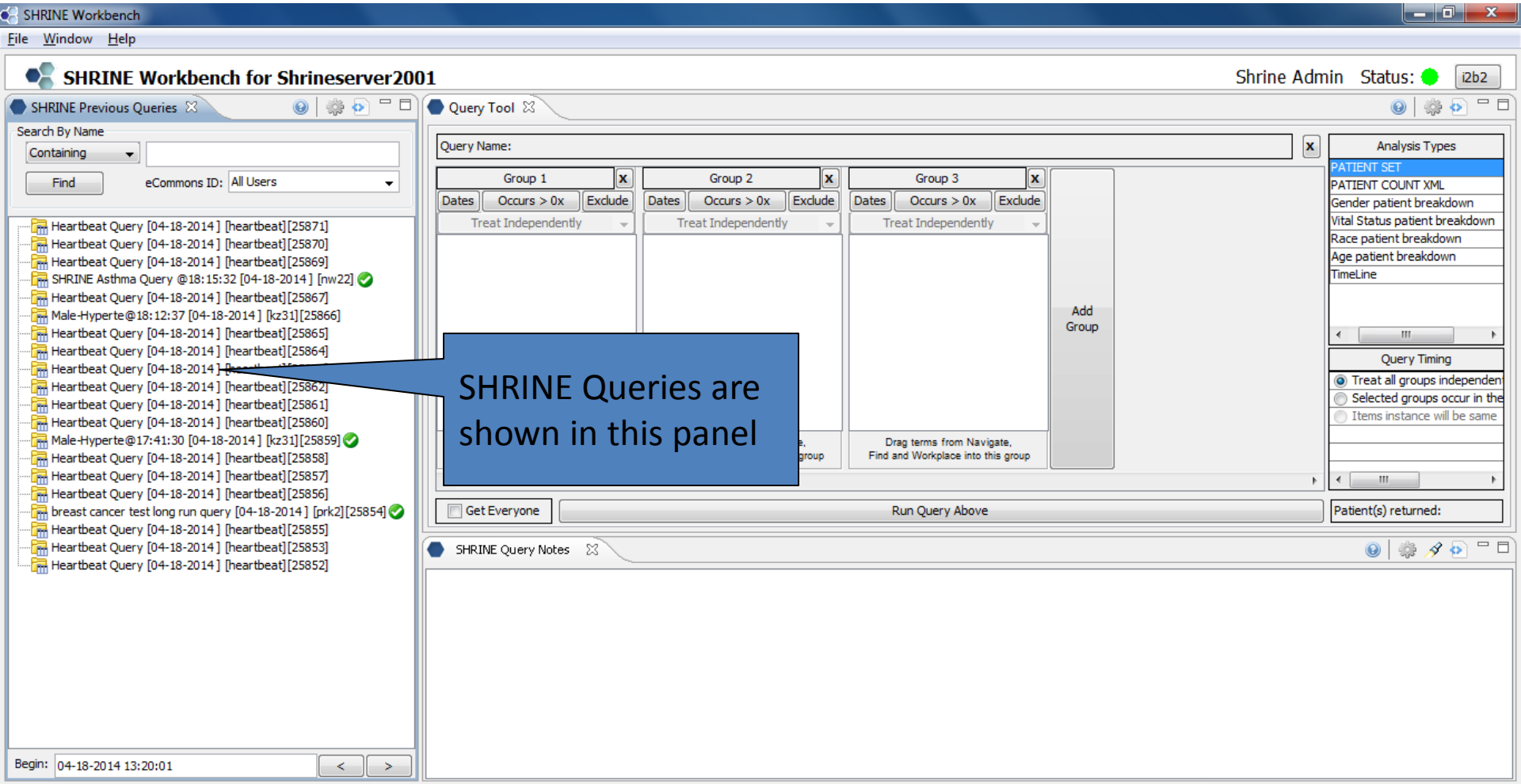
Run Query Using SHRINE



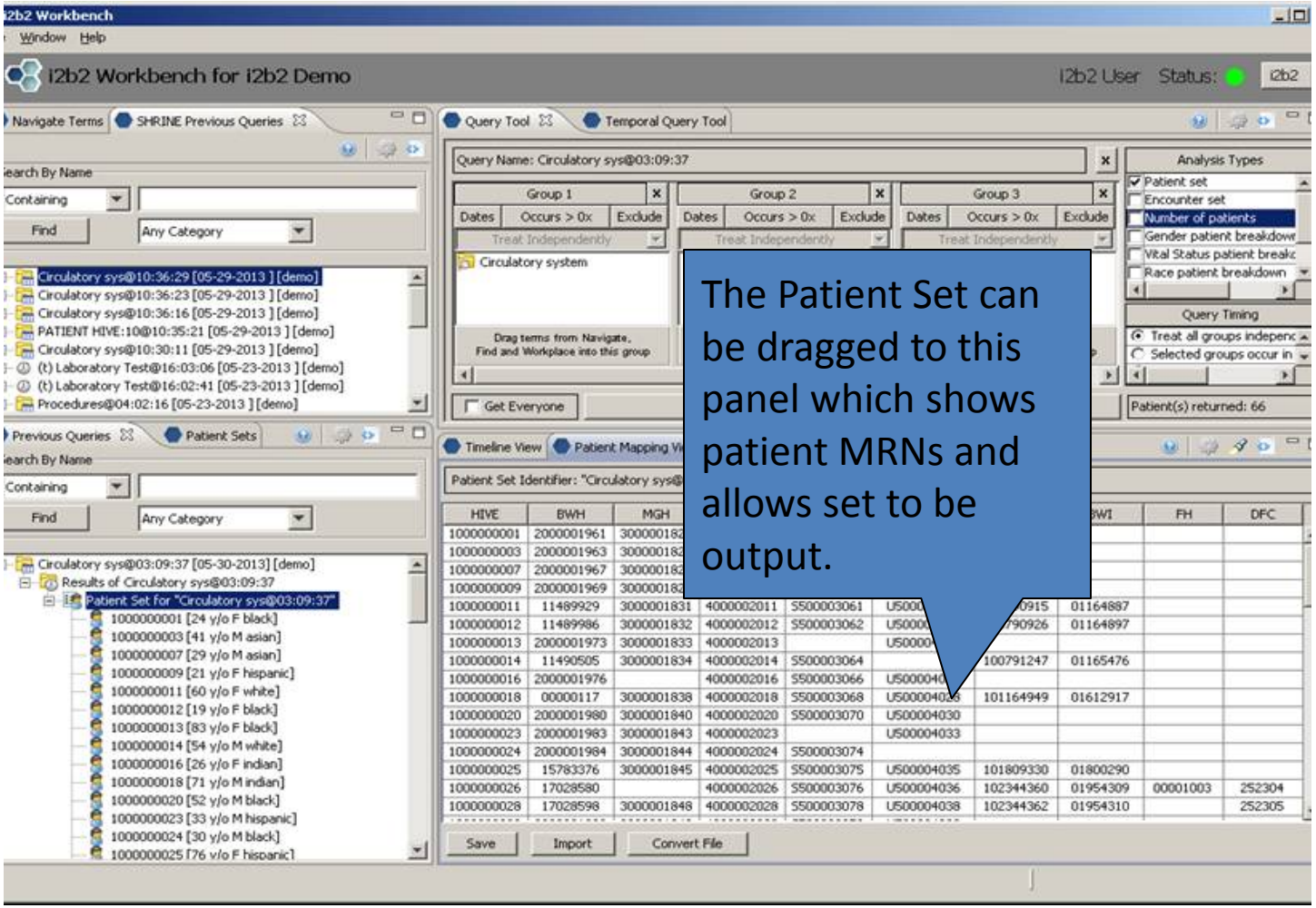
Local Site Admin
starts up i2b2 workbench



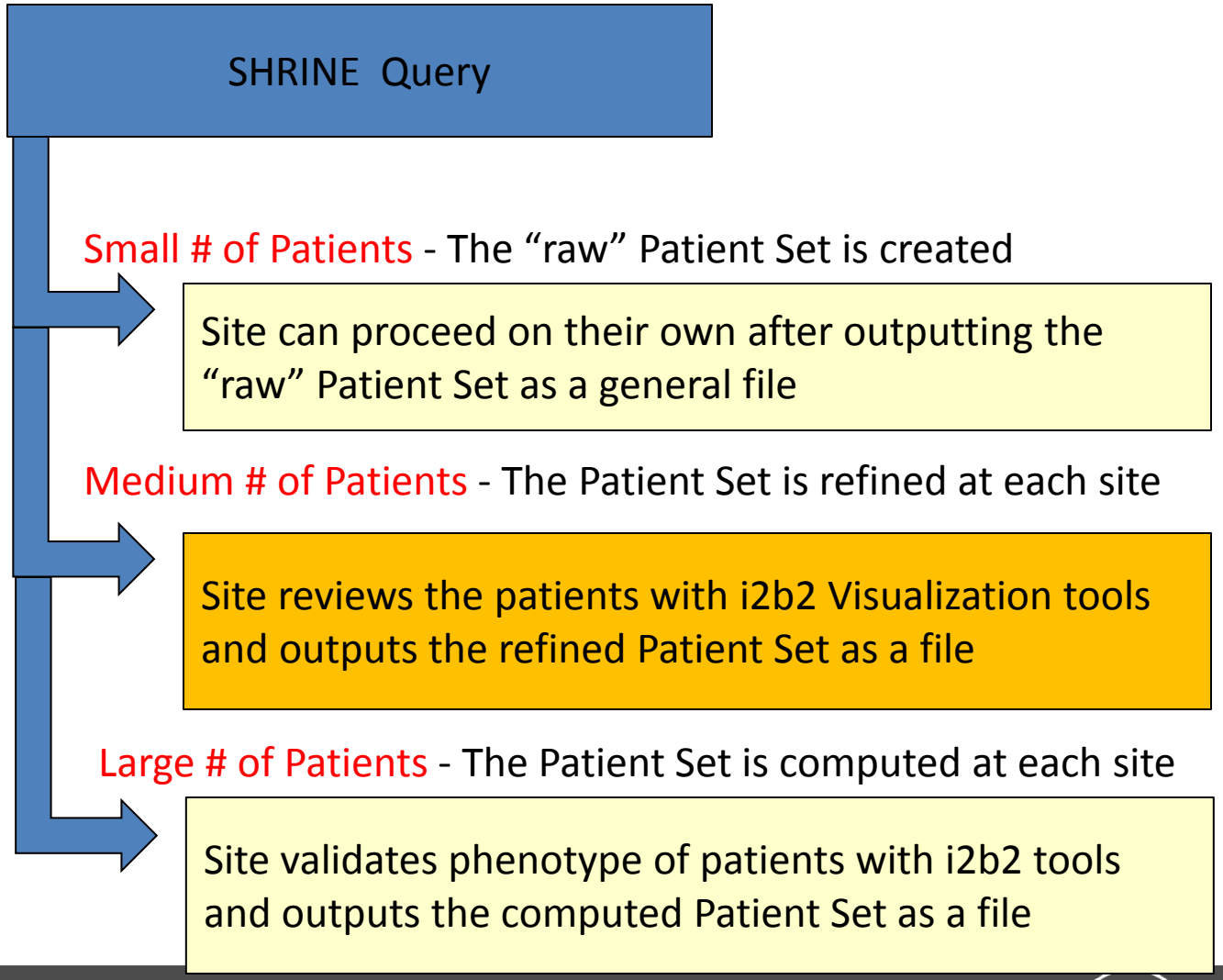
Local Site admin reviews query



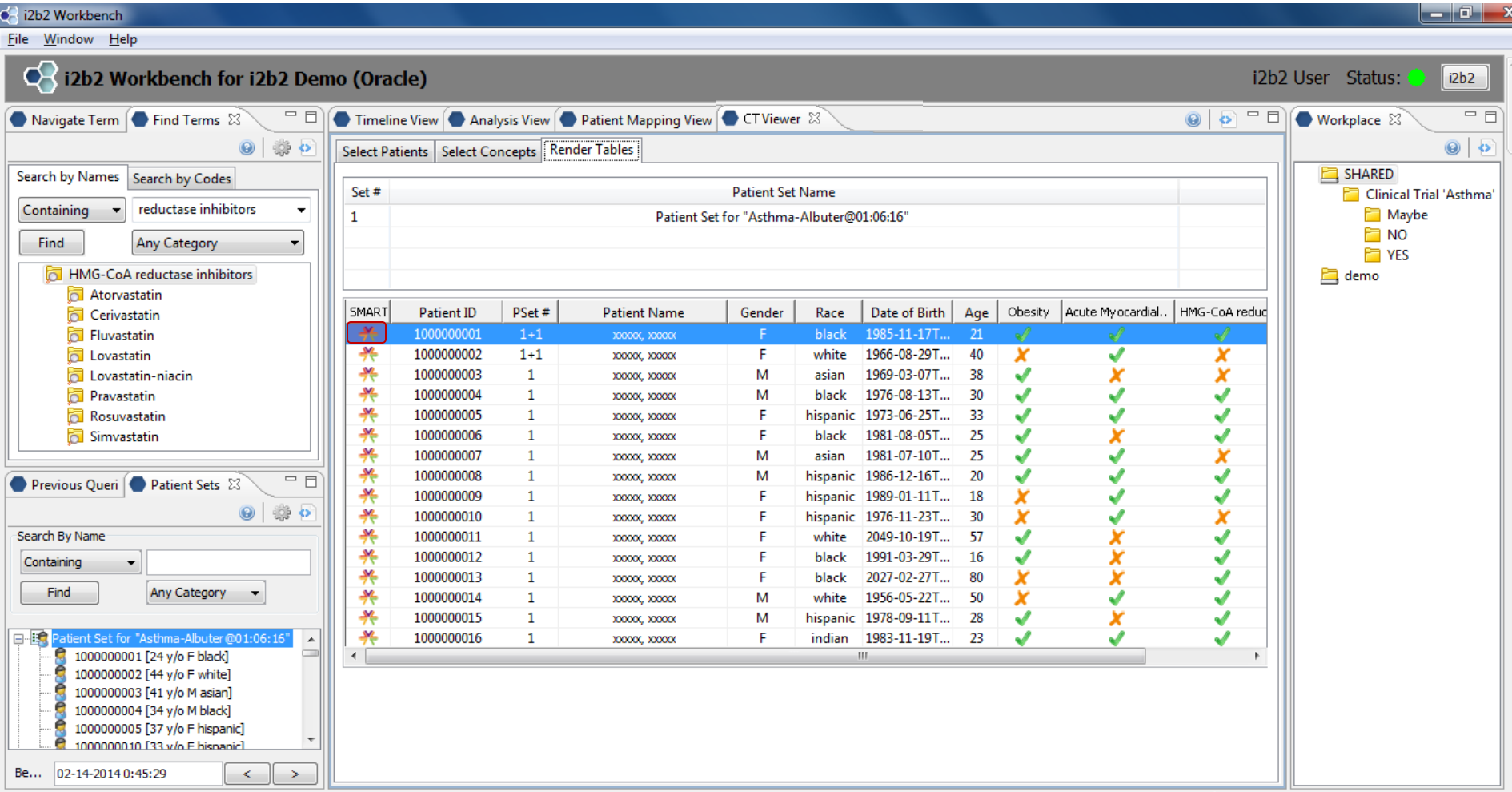
Can output “raw” Patient Set



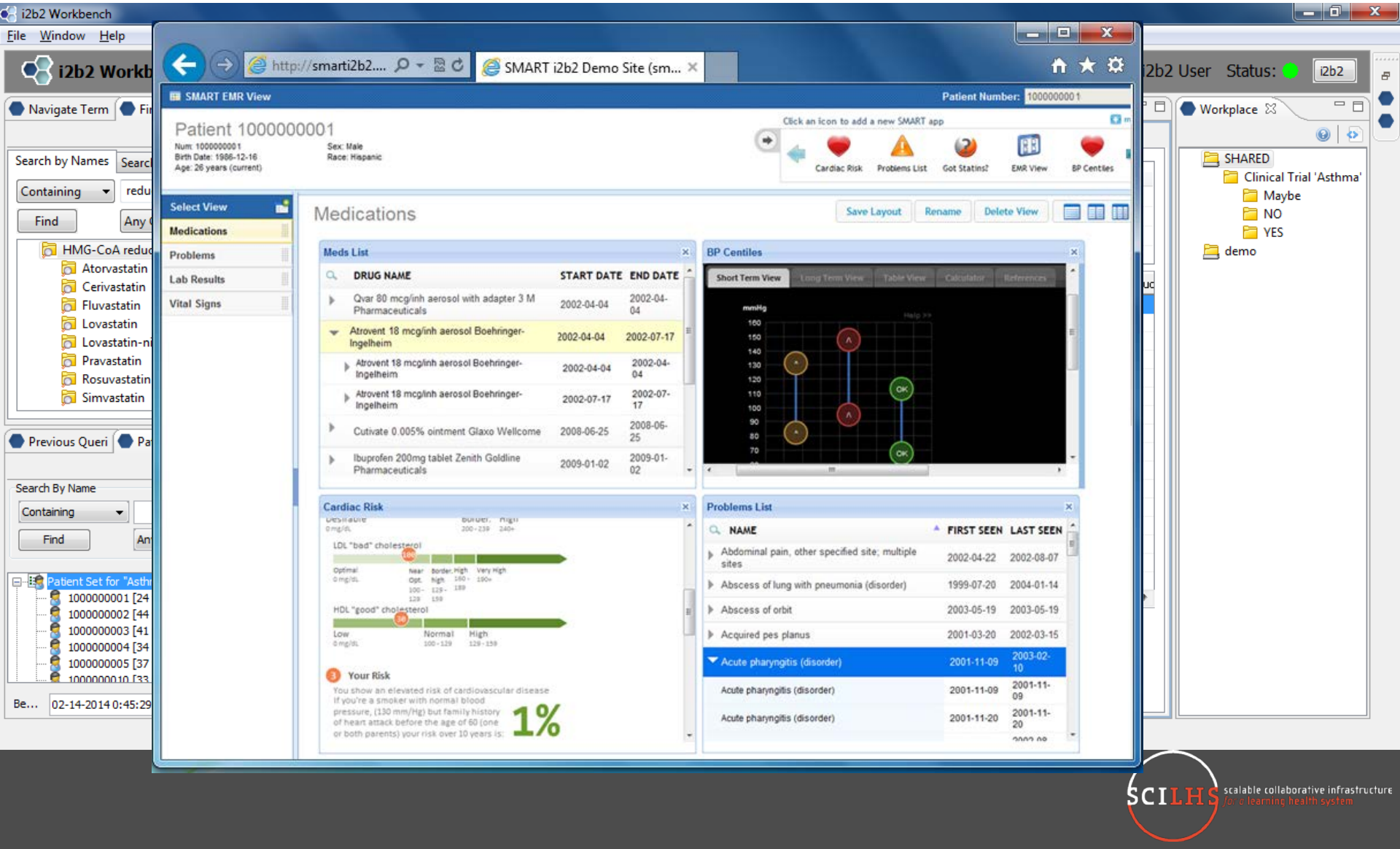
Handling Clinical Trials in an i2b2-SHRINE Network



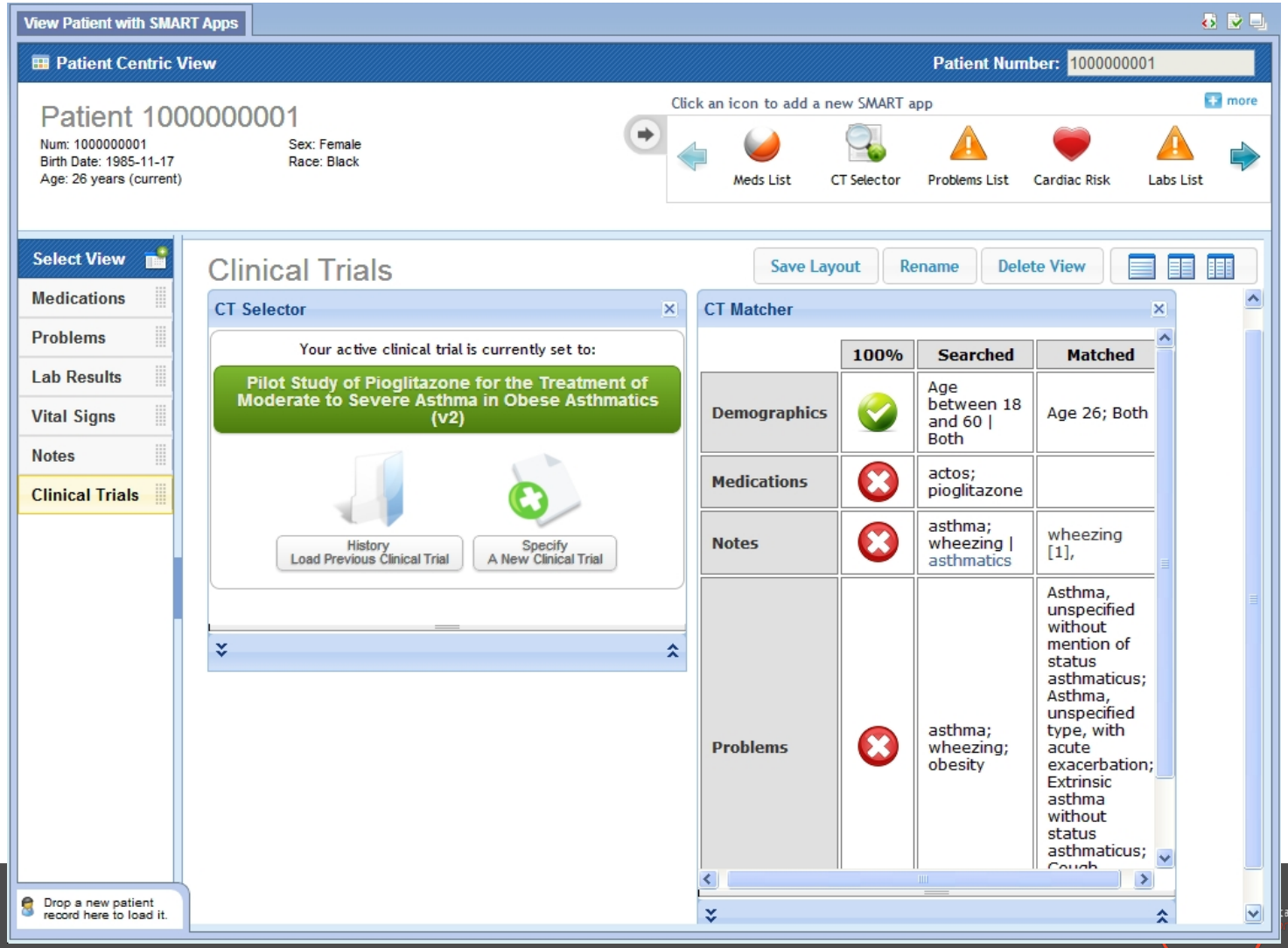
Or can go on to Review Patients in i2b2



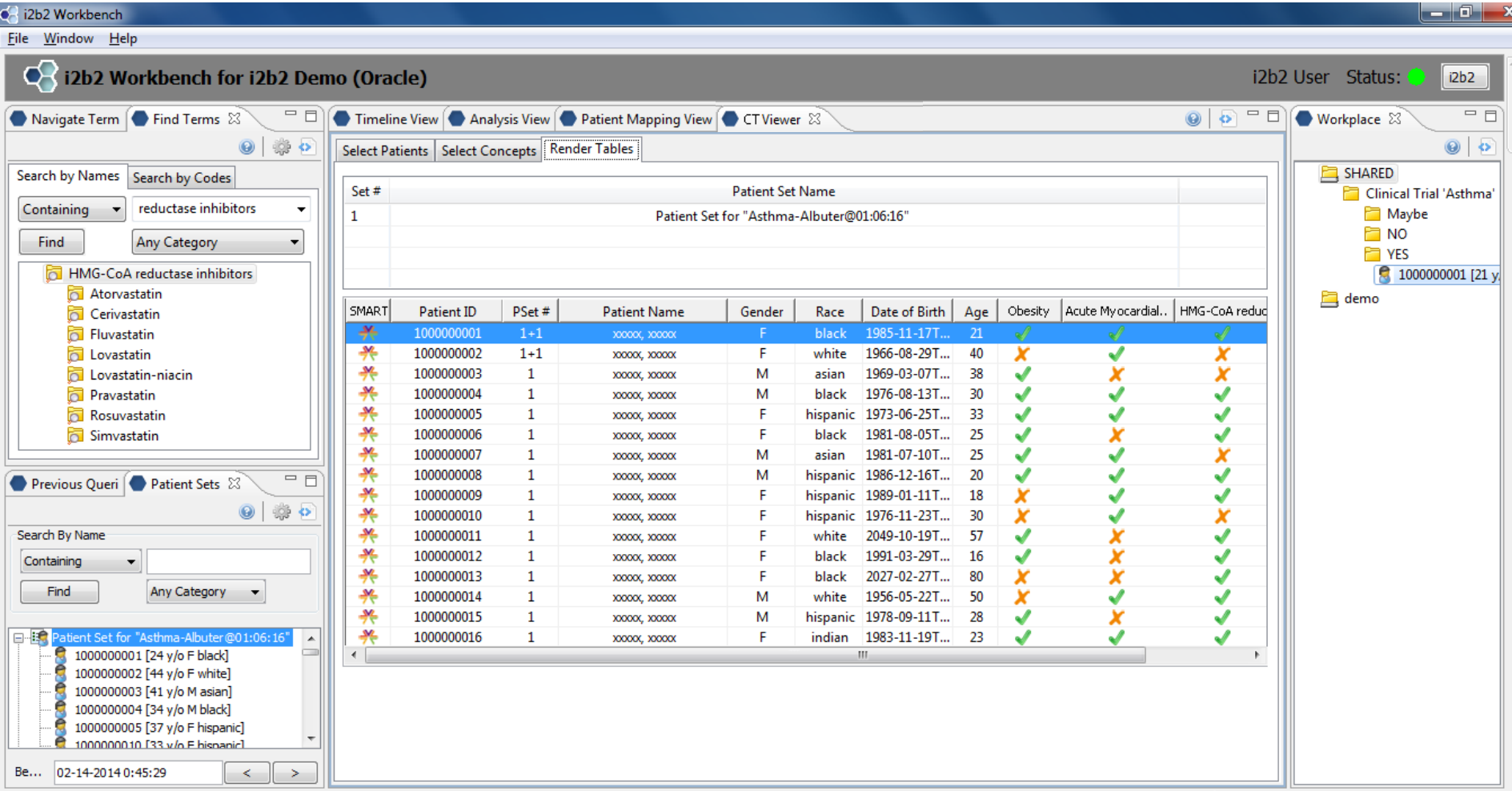
Review Patients with SMART tools



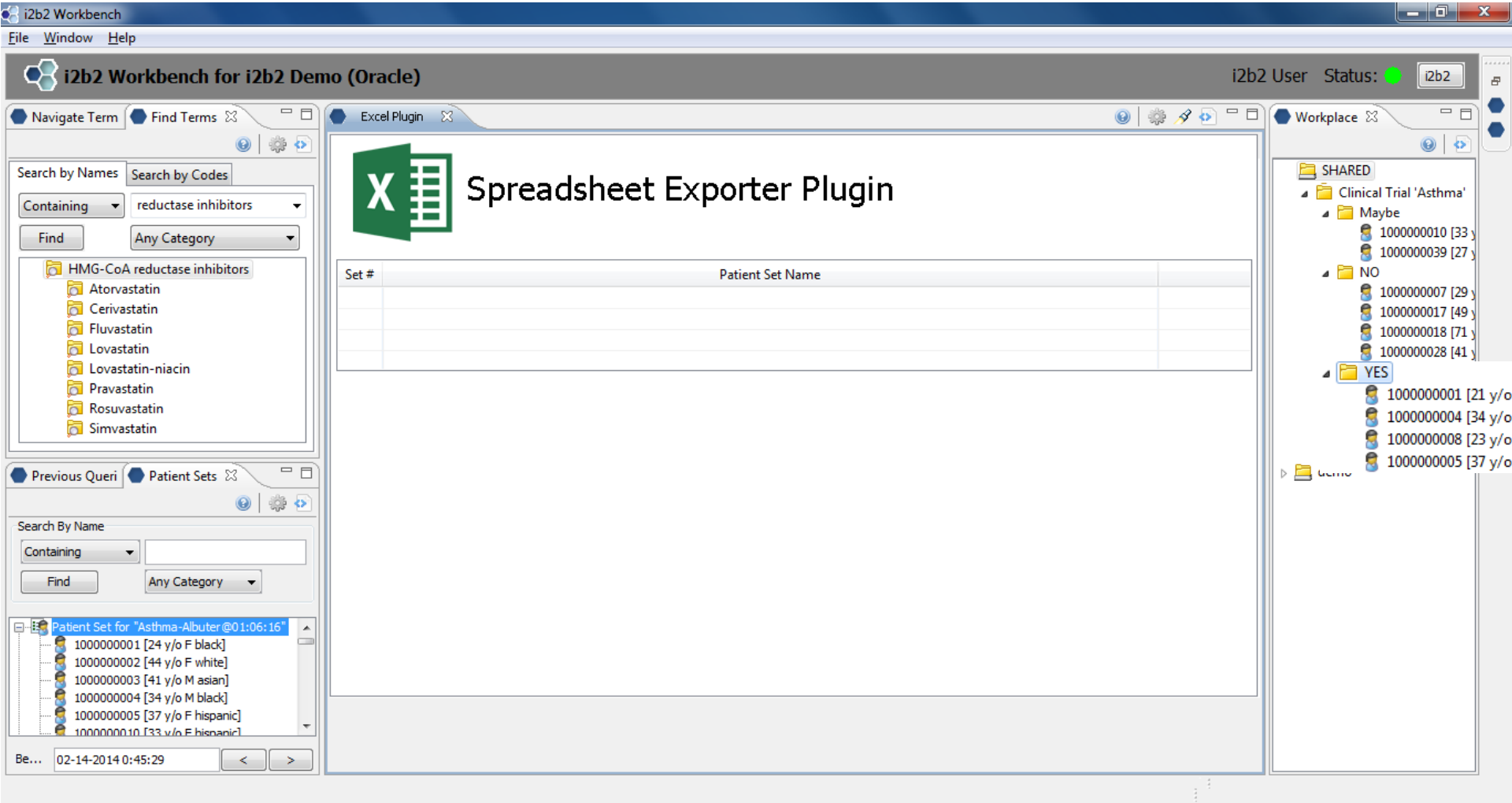
A “Criteria Matcher” App to review the “list”



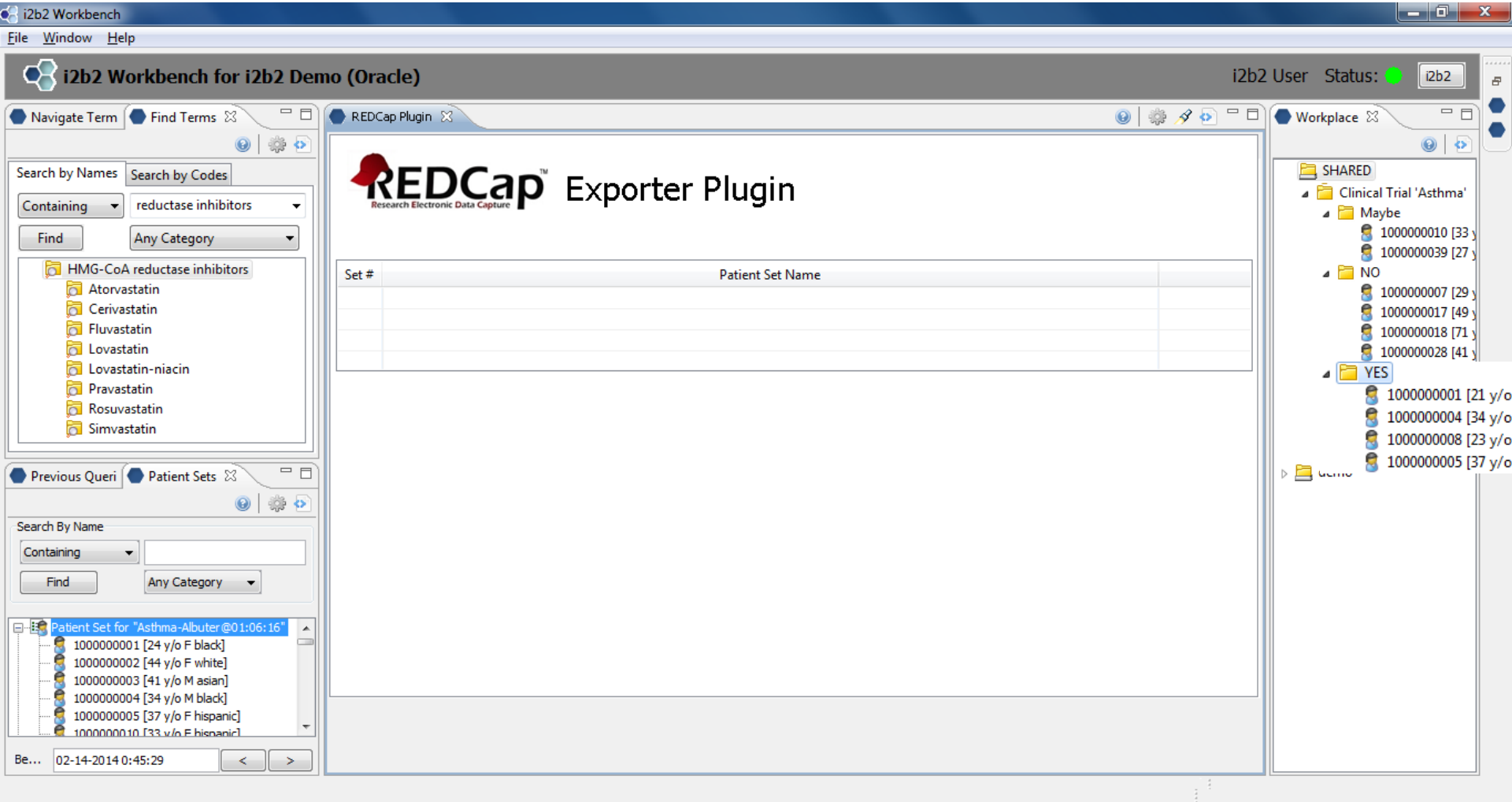
Classify *Patients*



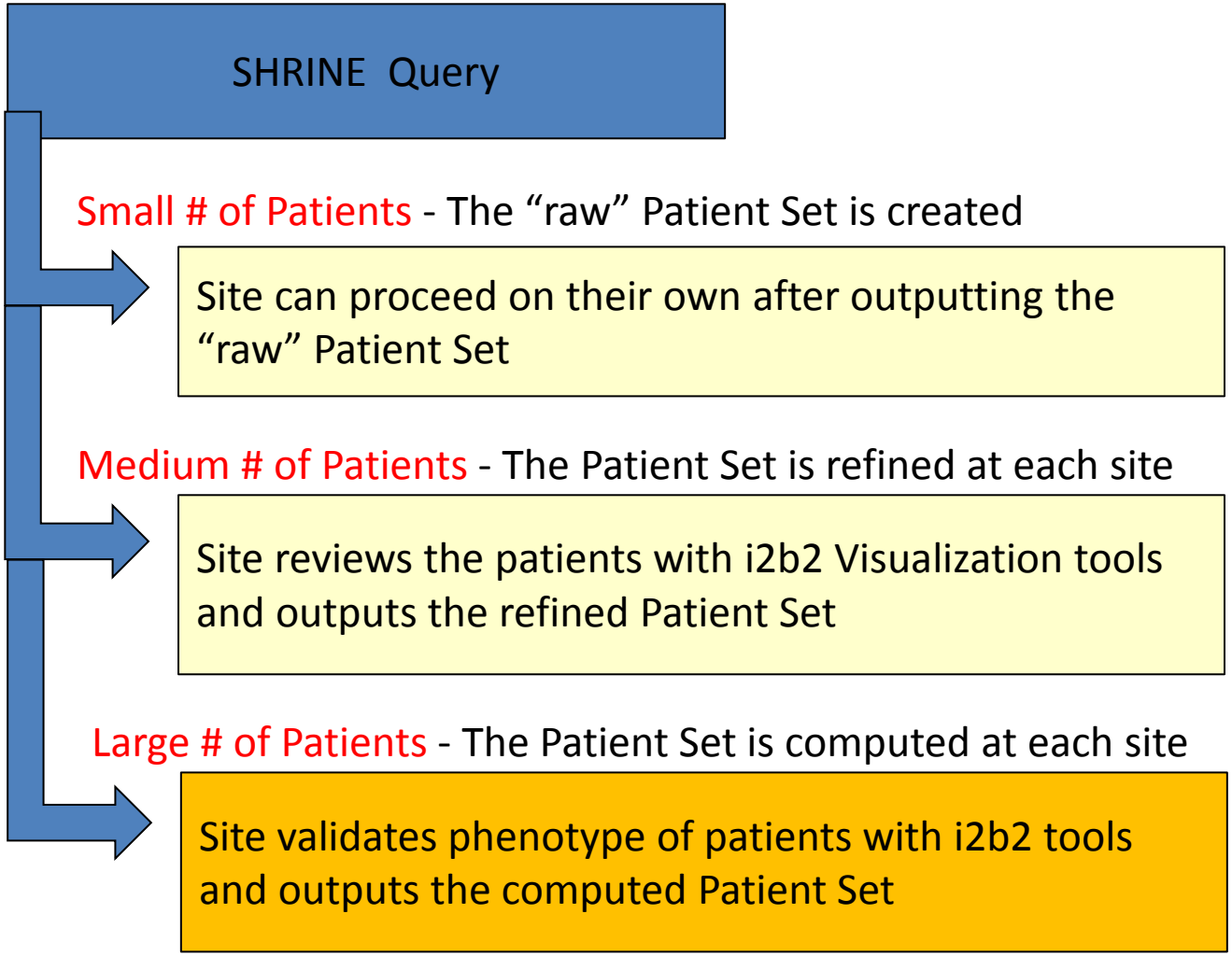
Output the “refined” Patient Set

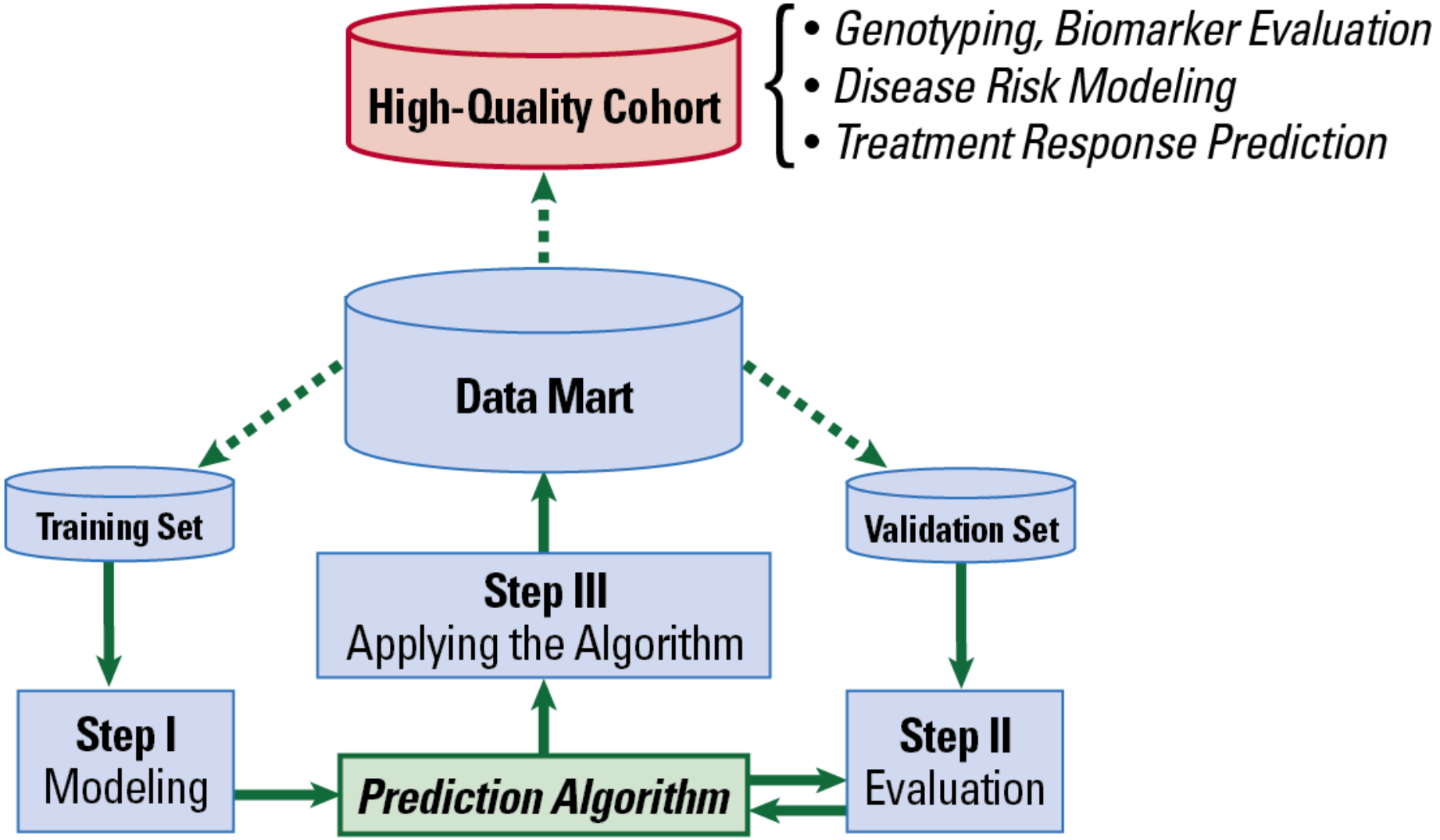


Often Output to REDCap



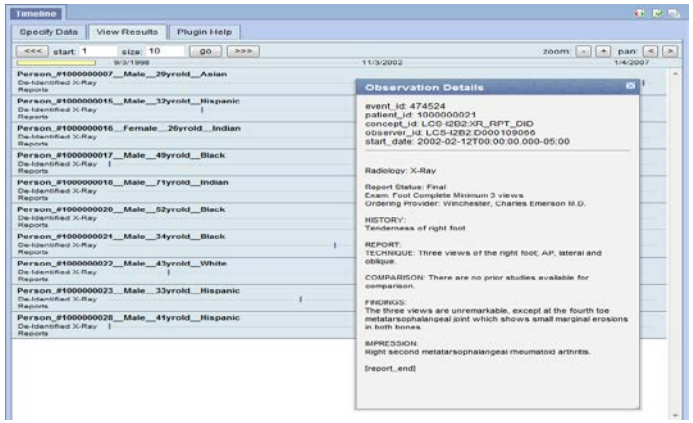
When Large numbers of patients found use Machine Learning Approach



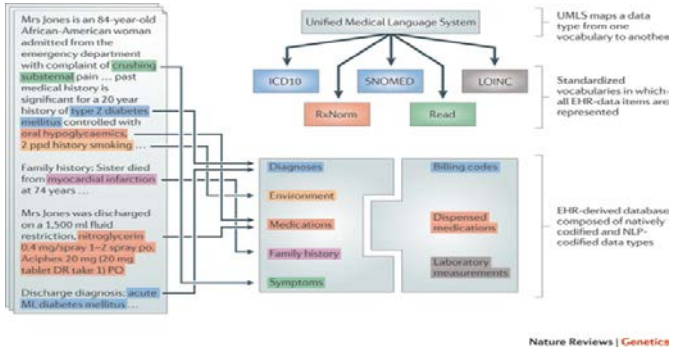

High-Quality Cohort
• Genotyping, Biomarker Evaluation
• Disease Risk Modeling
• Treatment Response Prediction
Data Mart
Training Set
Validation Set
Step III
Applying the Algorithm
Step I
Modeling
Prediction Algorithm
Step II
Evaluation

Creating Quality Data with Supervised Machine Learning

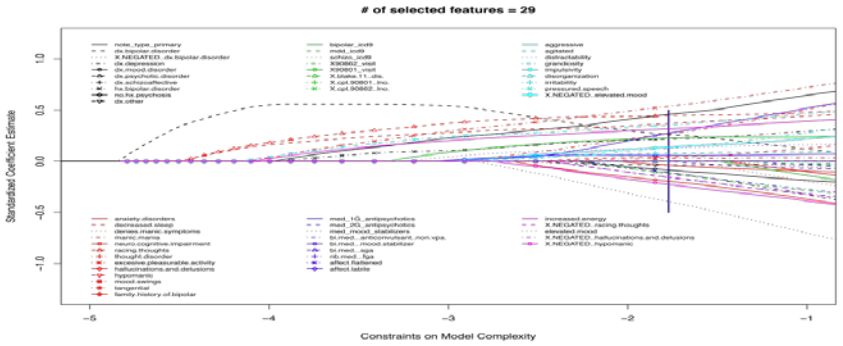
1. **Create a gold standard training set**.



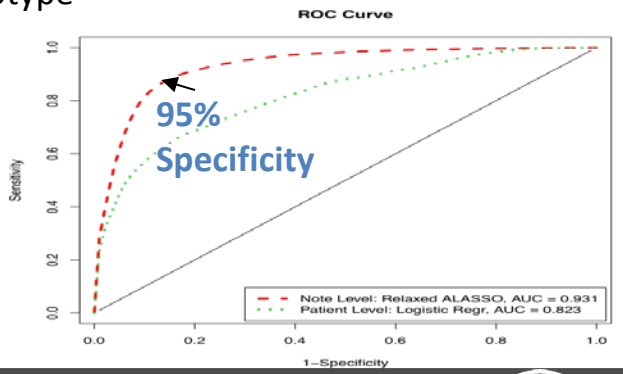
2. **Create a comprehensive list of features** (concepts/variables) that describe the phenotype of interest



3. **Develop the classification algorithm**. Using the data analysis file and the training set from step 1, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm



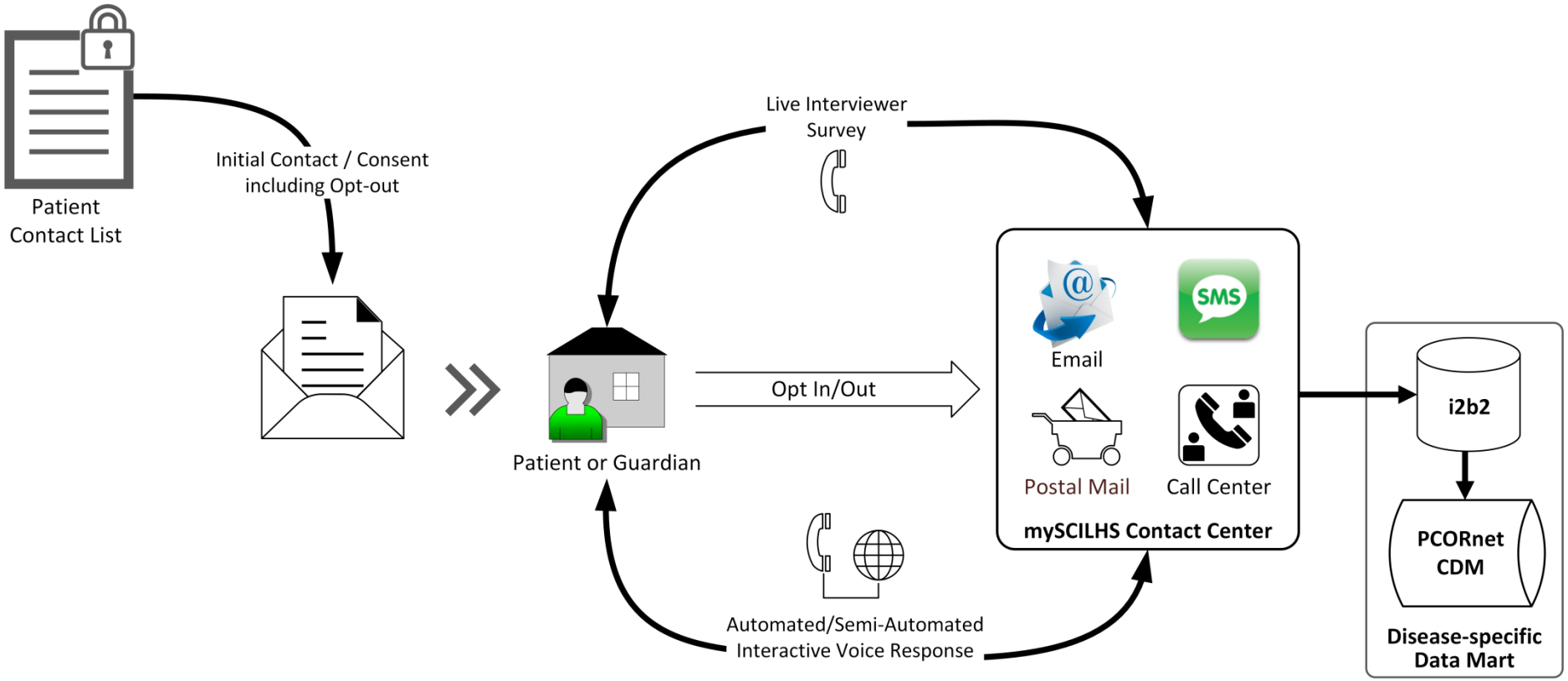
4. **Apply the algorithm to all subjects** in the superset and assign each subject a probability of having the phenotype

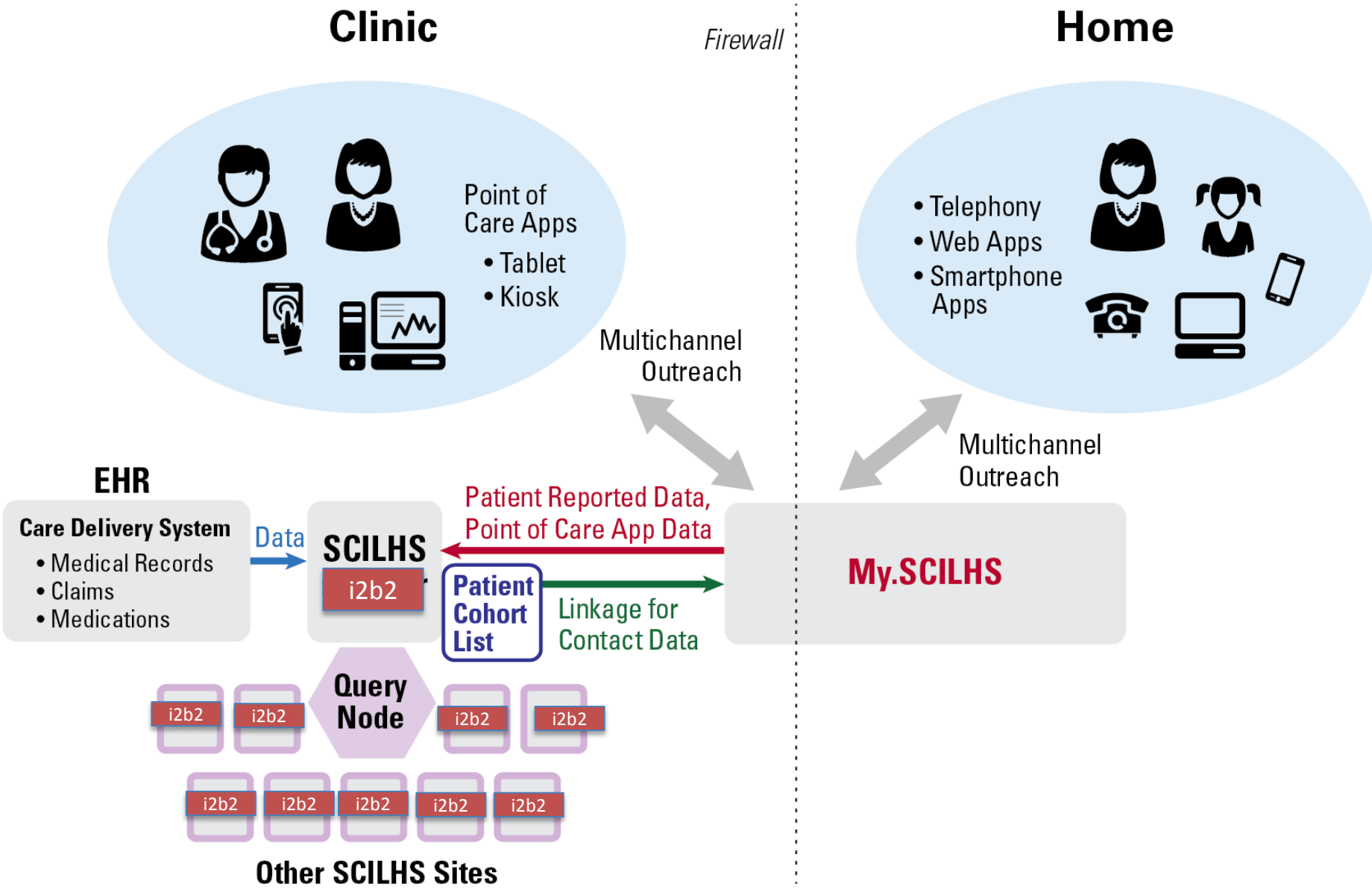


Key Challenges

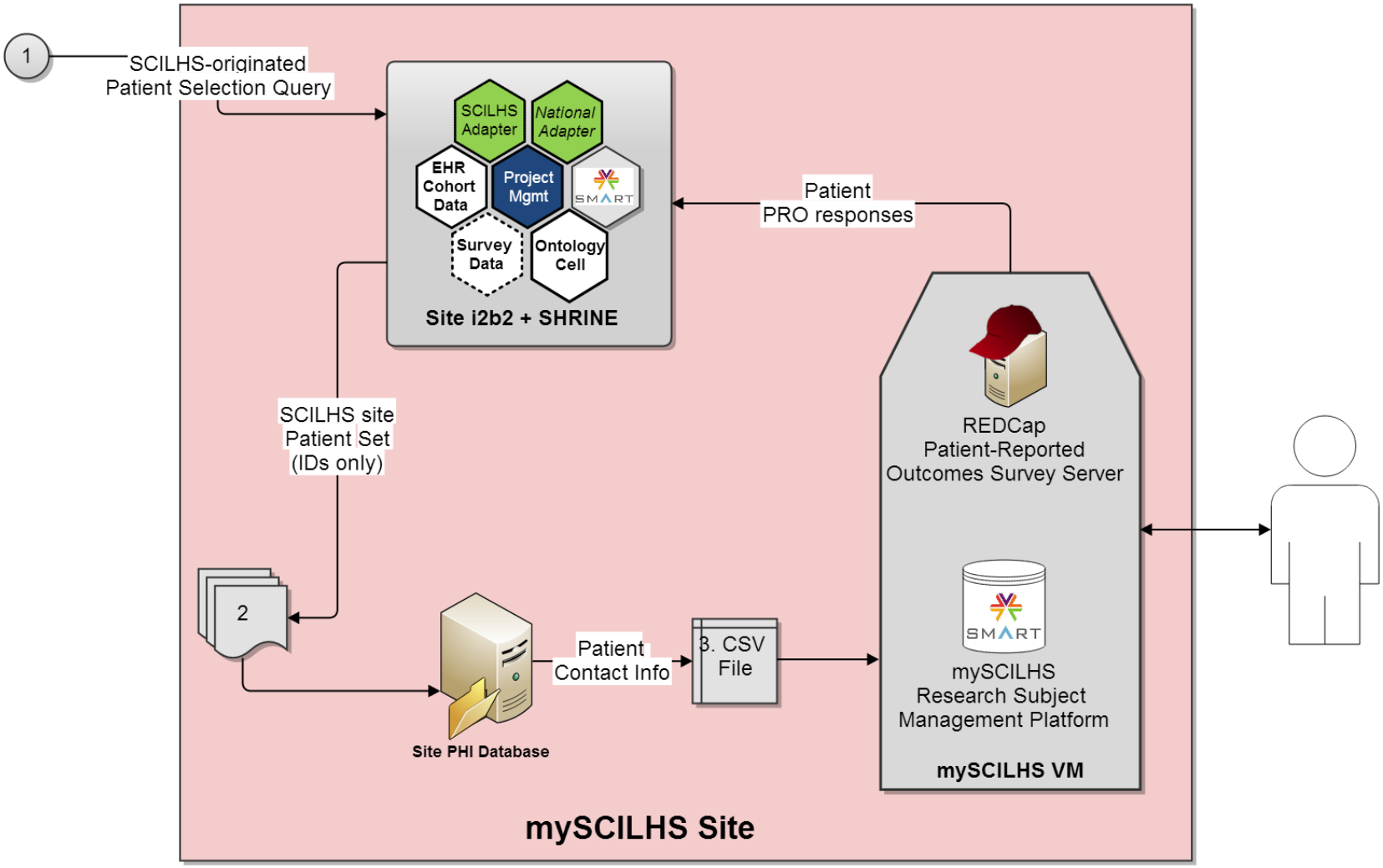
- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

A full workflow from query to contact



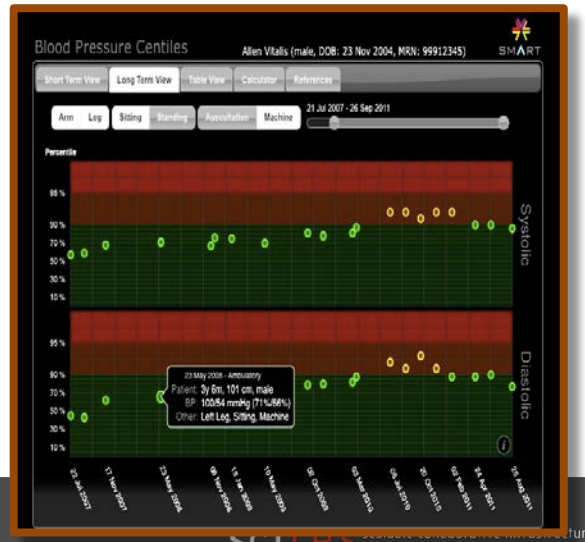

Clinic
Firewall
Home
Point of Care Apps
• Tablet
• Kiosk
• Telephony
• Web Apps
• Smartphone Apps
Multichannel Outreach
Multichannel Outreach
EHR
Care Delivery System
• Medical Records
• Claims
• Medications
Data
SCILHS
i2b2
Patient Reported Data, Point of Care App Data
Patient Cohort List
Linkage for Contact Data
My.SCILHS
Query Node
i2b2
Other SCILHS Sites

mySCILHS


1
SCILHS-originated
Patient Selection Query
SCILHS
Adapter
National
Adapter
EHR
Cohort
Data
Project
Mgmt
SMART
Survey
Data
Ontology
Cell
Site i2b2 + SHRINE
Patient
PRO responses
SCILHS site
Patient Set
(IDs only)
REDCap
Patient-Reported
Outcomes Survey Server
2
Site PHI Database
Patient
Contact Info
3. CSV
File
SMART
mySCILHS
Research Subject
Management Platform
mySCILHS VM
mySCILHS Site

Designing the App Store for Health

SMART

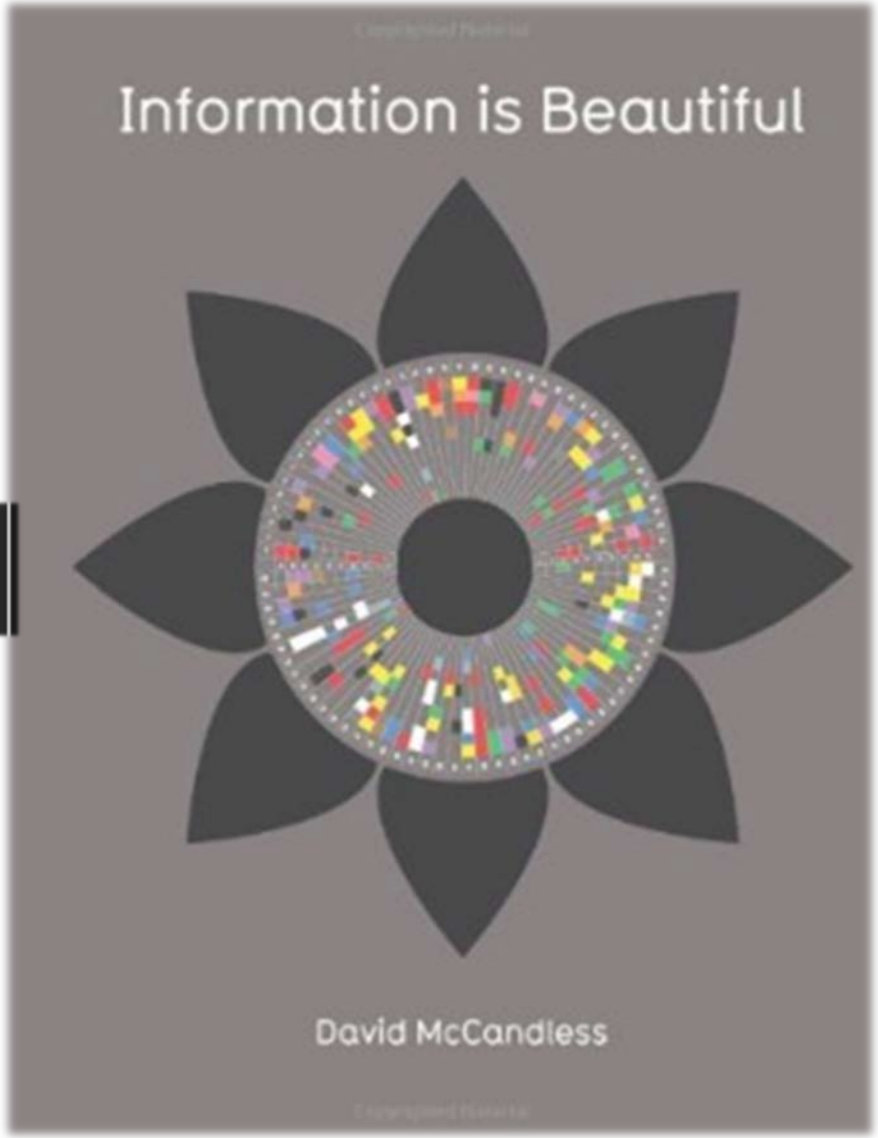


SCILHS

Inspired by a

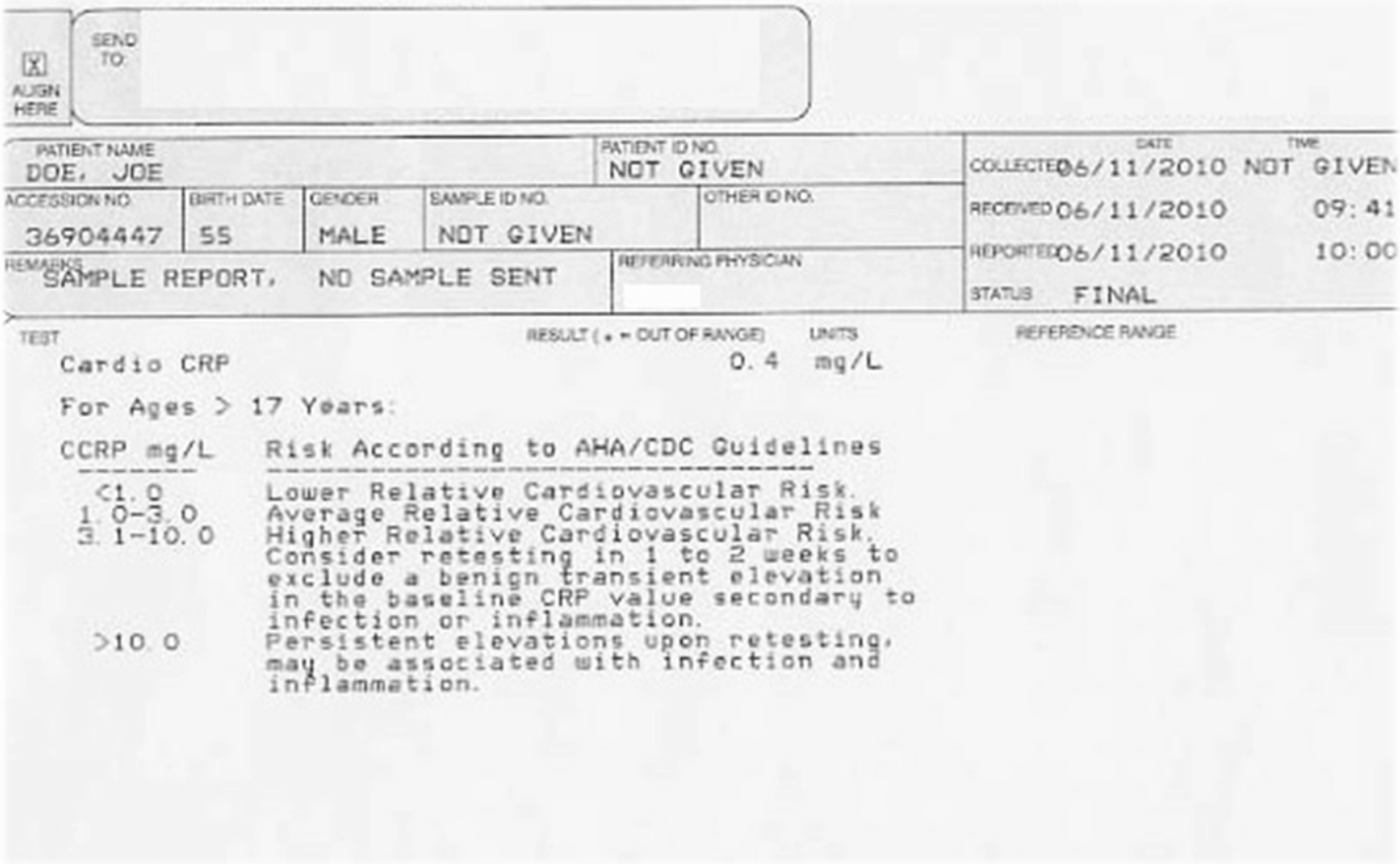
WIRED

18.12 Issue



Design Challenge

State-of-the-Art ???



SIGN HERE

SEND TO:

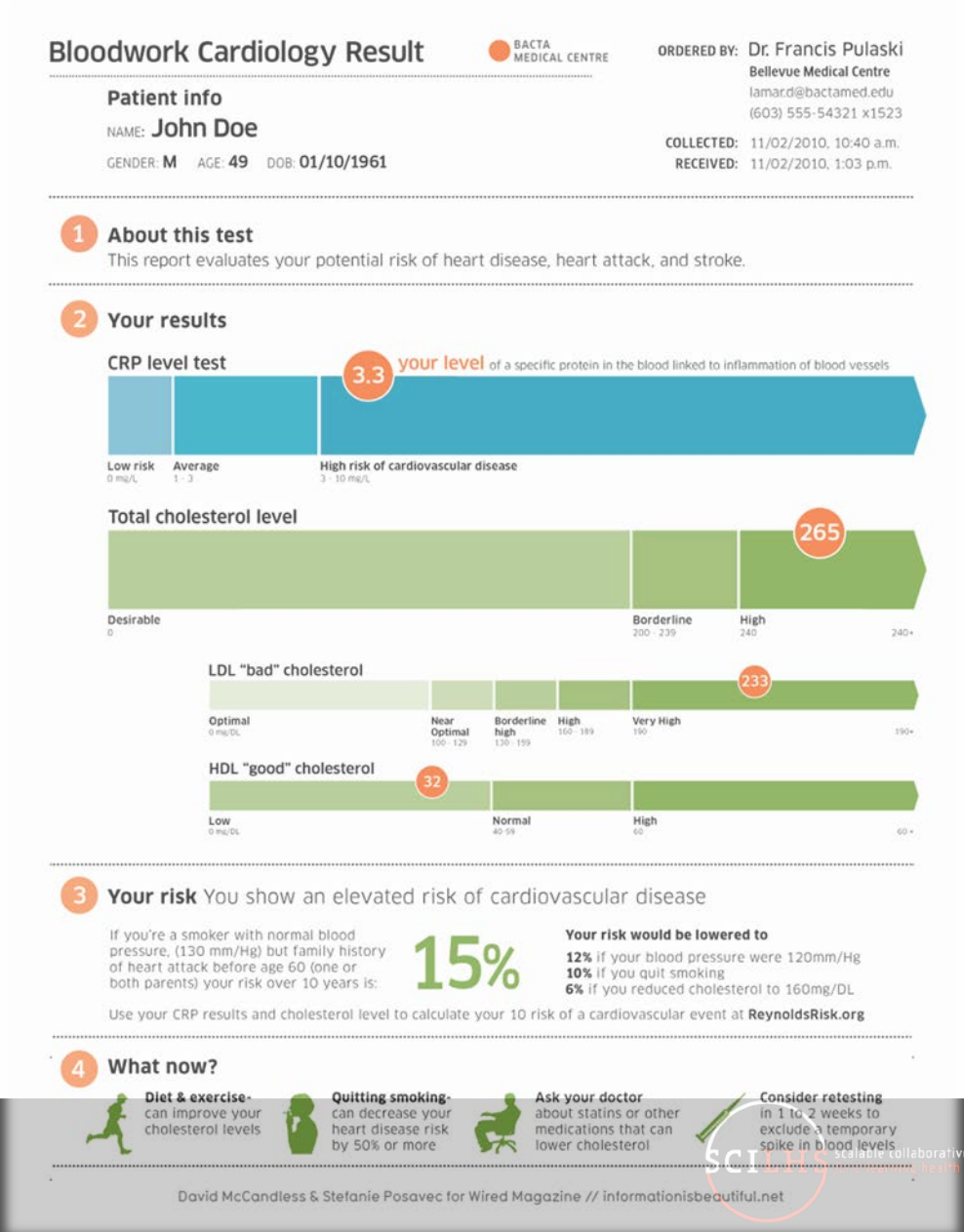
PATIENT NAME				PATIENT ID NO.		DATE	TIME
DOE, JOE				NOT GIVEN		COLLECTED 06/11/2010	NOT GIVEN
ACCESSION NO.	BIRTH DATE	GENDER	SAMPLE ID NO.	OTHER ID NO.		RECEIVED 06/11/2010	09:41
36904447	55	MALE	NOT GIVEN			REPORTED 06/11/2010	10:00
REMARKS: SAMPLE REPORT, NO SAMPLE SENT				REFERRING PHYSICIAN		STATUS FINAL	

TEST	RESULT (* = OUT OF RANGE)	UNITS	REFERENCE RANGE
Cardio CRP	0.4	mg/L	

For Ages > 17 Years:

CCRP mg/L	Risk According to AHA/CDC Guidelines
<1.0	Lower Relative Cardiovascular Risk.
1.0-3.0	Average Relative Cardiovascular Risk
3.1-10.0	Higher Relative Cardiovascular Risk. Consider retesting in 1 to 2 weeks to exclude a benign transient elevation in the baseline CRP value secondary to infection or inflammation.
>10.0	Persistent elevations upon retesting, may be associated with infection and inflammation.

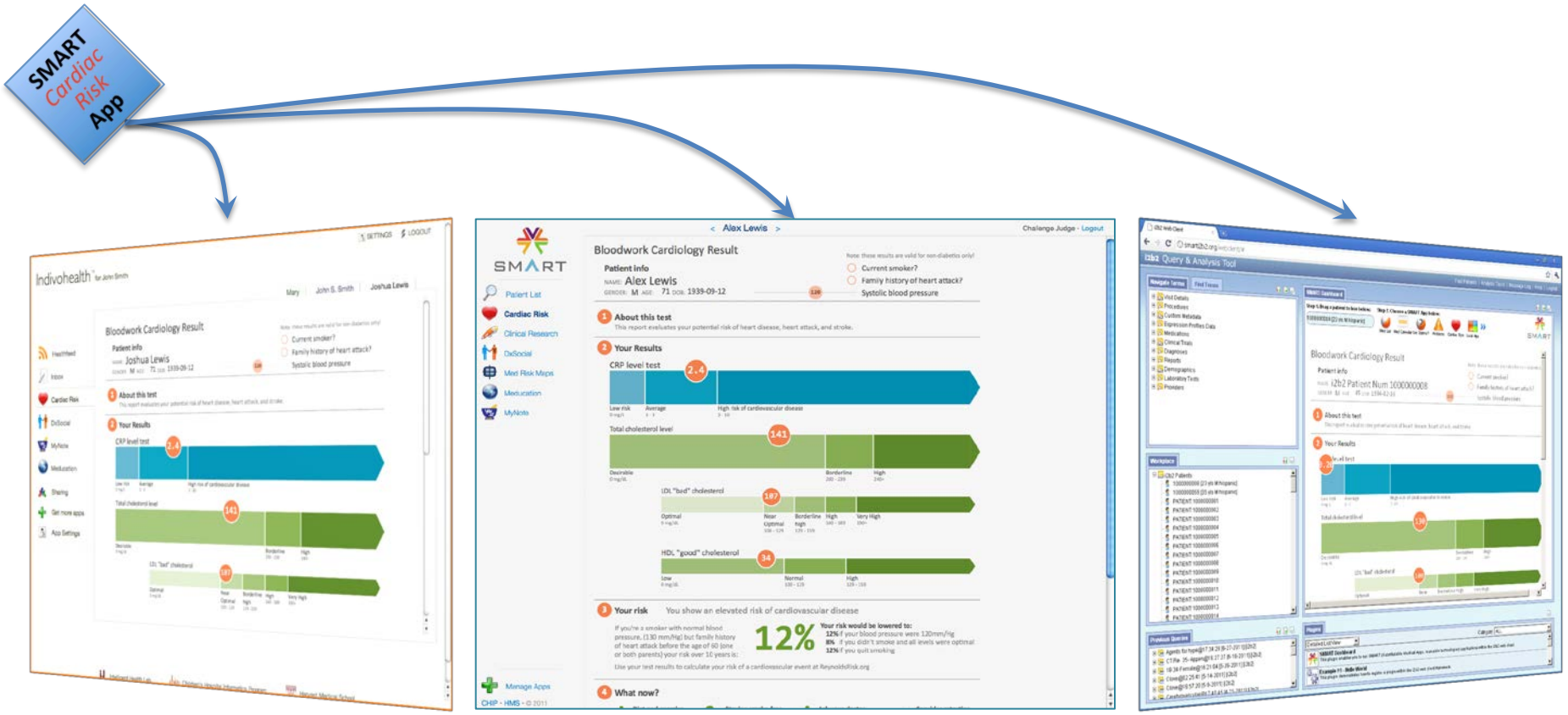
An
Inspired Design
from Dave McCandless
(cc license)



1 Design + 1 Developer + 1 Week

SMART *Cardiac Risk* App

1 SMART App in 3 SMART Systems



Trial Eligibility

Trial Eligibility

Female | Male

43 y

Boston, MA

rheumatoid arthritis +methotrexate

Cancel

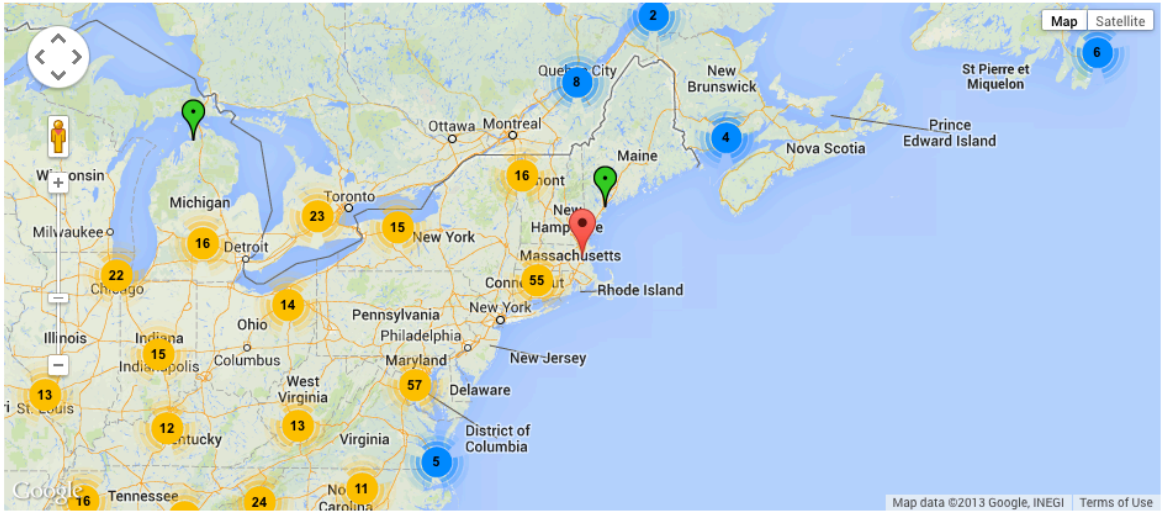
Refresh

Intervention / Observation

- [x] **Biological (14)**
- [ ] Device (1)
- [ ] Dietary Supplement (1)
- [x] **Drug (69)**
- [ ] Observational (14)
- [ ] Other (4)
- [ ] Procedure (1)

Trial phase

- [x] **N/A (8)**
- [x] **Phase 1 (8)**
- [x] **Phase 2 (23)**
- [x] **Phase 3 (17)**
- [x] **Phase 4 (26)**



Hide Map (2333 trial locations)

Key Challenges

- Governance of Data Types
- Sustainability
- Patient Privacy
- Patient Engagement

The Plan

Common data platform **(i2b2)**

+

Federated queries across sites **(SHRINE)**

+

Point of care apps **(SMART)**

+

Patient-facing technologies **(RedCap, SMART, +)**

Tribute to…

- I2b2 Core Team
  - Christopher Herrick
  - Isaac Kohane
  - Susanne Churchill
  - Griffin Weber
  - Paul Avillach
  - Michael Mendis
  - Lori Phillips
  - Janice Donahoe
  - Nich Wattanasin
  - Wayne Chan
  - David Wang
  - Bill Wang
  - Vivian Gainer
  - Andrew Cagan
- SMART i2b2 Team
  - Nich Wattanasin
  - Kenneth Mandl
  - Joshua Mandel
- i2b2 – SHRINE Team
  - William Simons
  - Douglas MacFadden
- tranSMART Team
  - Paul Avillach
  - Michael McDuffy
- mySCILHS Team
  - Marc Natter
  - Isaac Pinol Catadau
  - Stanley Boykin

I2b2, SHRINE, and SMART Information and Software on the Web

i2b2 Homepage (https://www.i2b2.org)
i2b2 Software (https://www.i2b2.org/software)
i2b2 Community Site (https://community.i2b2.org)
SHRINE at Harvard (http://shrine.catalyst.harvard.edu)
SHRINE Software:
(https://open.med.harvard.edu/display/SHRINE/Software)
SMART i2b2 Homepage (http://smarti2b2.org)
SMART Platforms Homepage (http://smartplatforms.org)

THANK YOU