

# Enabling the Healthcare Enterprise for Discovery Research with i2b2

*Shawn Murphy MD, Ph.D.*



## High Throughput Method for supporting Research

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research



# High Throughput Method for supporting Research

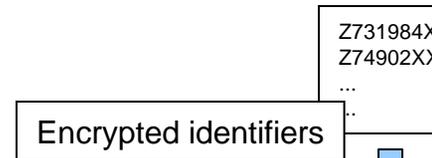
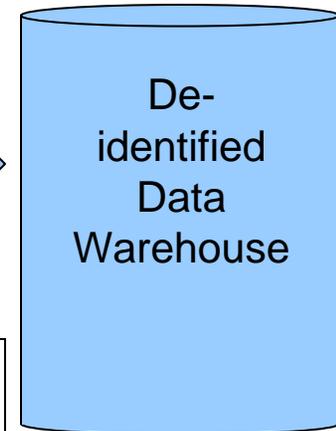
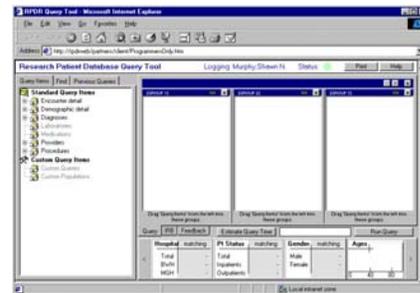
- Set of patients is selected from medical record data in a high throughput fashion
  - Eugene Braunwald
  - Henry Chueh
  - John Glaser
  - Diane Keogh

# Research Patient Data Registry exists at Partners Healthcare

## 1) Queries for aggregate patient numbers

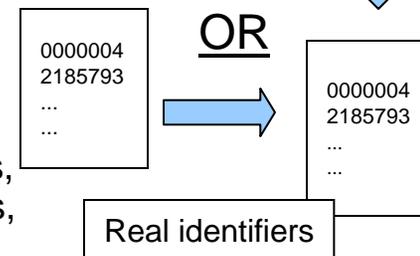
- Warehouse of in & outpatient clinical data
- 4.6 million Partners Healthcare patients
- 1.2 billion diagnoses, medications, procedures, genomics, laboratories, & physical findings coupled to demographic & visit data
- Authorized use by faculty status
- Clinicians can construct complex queries
- Queries cannot identify individuals, internally can produce identifiers for (2)

Query construction in web tool



## 2) Returns identified patient data

- Start with list of specific patients, usually from (1)
- Authorized use by IRB Protocol
- Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary, endoscopy), and images into a Microsoft Access database and text files.



Test ID	Test Description	Result	Result Text	Abnormal Flag	Reference	Unit Reference	Range
SOPTT	Supernat APPT	21.8					22.0-36.0
SOPTT	APTT	32.8					22.0-36.0
SOPTT	APTT	37.8					22.0-36.0
SOPTT	APTT	46.4					22.0-36.0
SOPTT	APTT	43.7			MODERATELY H		22.0-36.0
SOPTT	APTT	26.7					22.0-36.0
SOPTT	APTT	23.7					22.0-36.0
SOPTT	APTT	38.4					22.0-36.0
SOPTT	APTT	34.7					22.0-36.0
SOPTT	APTT	34.6					22.0-36.0
SOPTT	APTT	24.7					22.0-36.0
SOPTT	Supernat APPT	37.7					22.0-36.0
SOPTT	APTT	34.5					22.0-36.0
SOPTT	APTT	40.0			H		22.0-36.0
SOPTT	APTT	40.0			H		22.0-36.0
SOPTT	Supernat APPT	35.2			High Note H		22.0-36.0
SOPTT	APTT	33.8					22.0-36.0
SOPTT	Supernat APPT	34.3					22.0-36.0
SOPTT	APTT	37.8					22.0-36.0
SOPTT	APTT	22.8					22.0-36.0
SOPTT	APTT	37.4					22.0-36.0
SOPTT	APTT	37.2			ULT HEMOLYSIS H		22.0-36.0
SOPTT	APTT	36.1					22.0-36.0
SOPTT	APTT	36.4			MODERATE HE H		22.0-36.0

# Security and Patient Confidentiality of Step 1

- All patients at Partners are added
  - HIPAA notification that their data may be used for research upon registration.
- RPDR data is anonymized at the Query Tool.
  - Aggregated numbers are obfuscated to prevent identification of individuals; automatic lock out occurs if pattern suggests identification of an individual is being attempted.

## A Security Architecture for Query Tools used to Access Large Biomedical Databases

Shawn N. Murphy, MD, Ph.D. and Henry C. Chueh, MD, M.S.

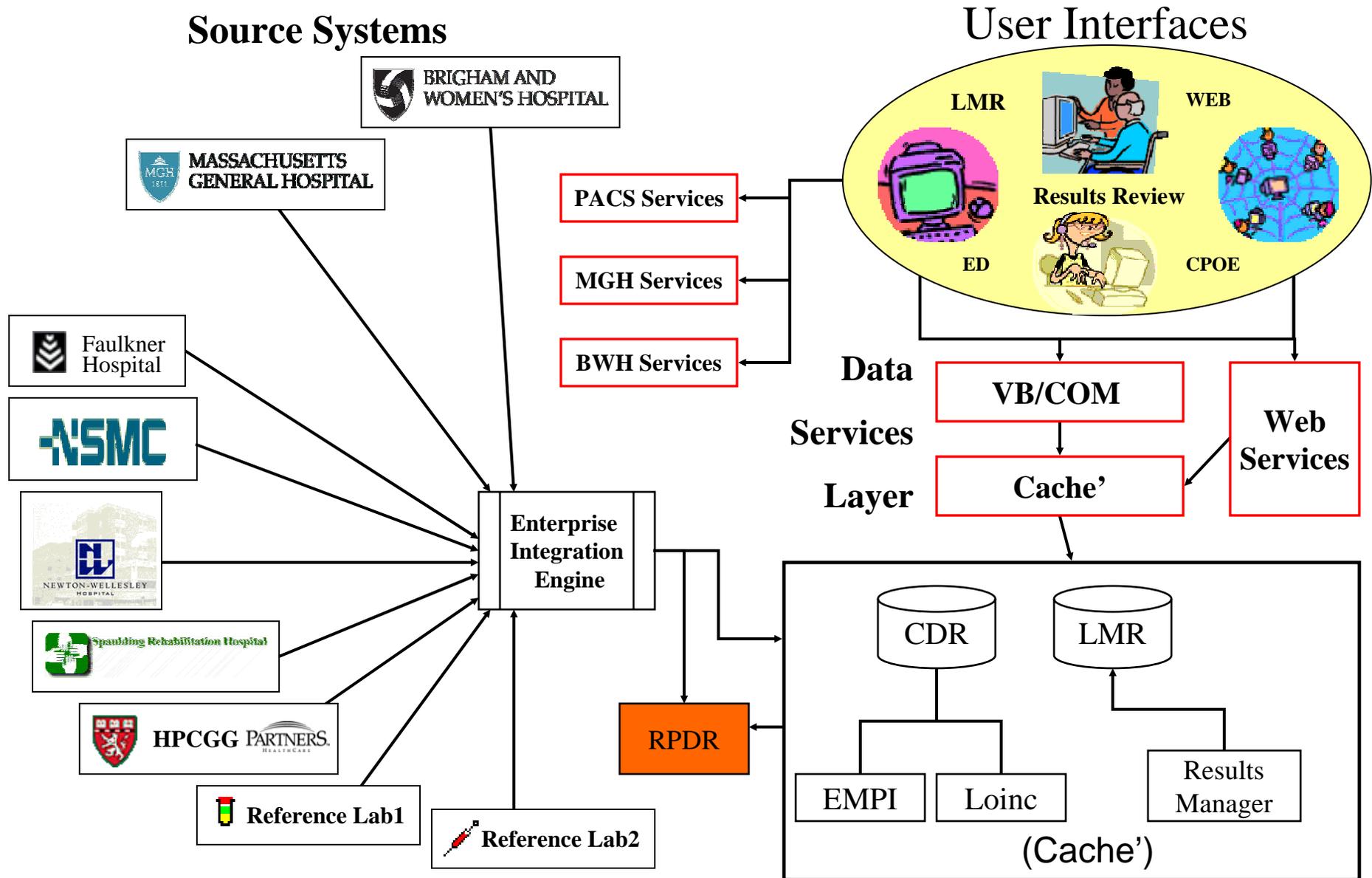
Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA.

- Queries done in Query Tool available for review by RPDR team, a user lock out will specifically direct a review.
- De-identified data warehouse is a “Limited Data Set” by HIPAA
  - Medical record numbers are encrypted and obvious identifiers are removed from data.
- Concept of “established medical investigator” is promoted by classification as a faculty sponsor.

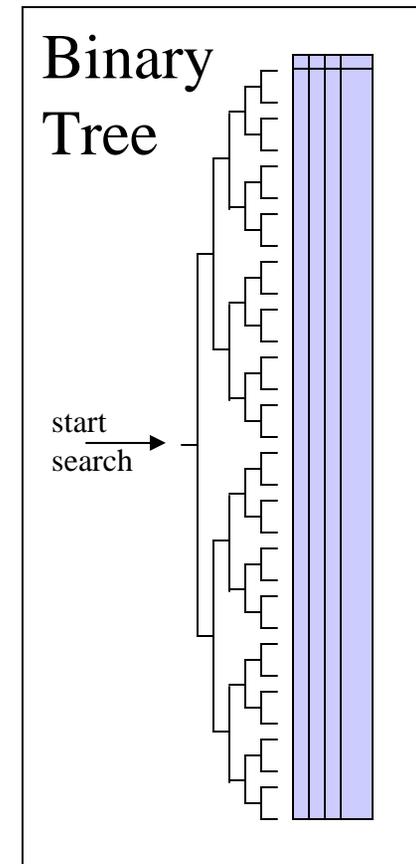
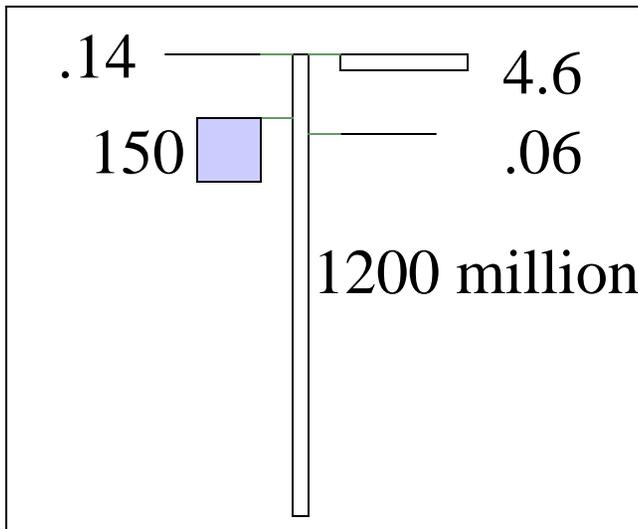
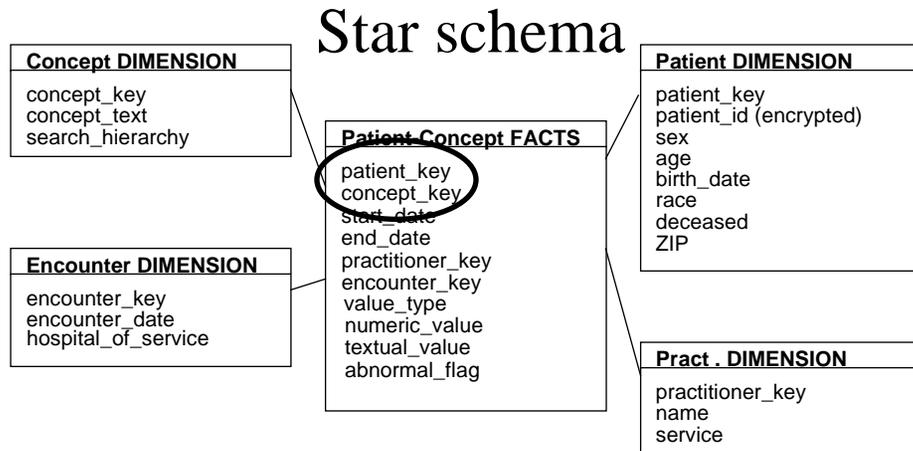
# Security and Patient Confidentiality of Step 2

- Only studies approved by the Institutional Review Board (IRB) are allowed to receive identified data.
- Queries may be set up by workgroup member, but faculty sponsor on IRB protocol must directly approve all queries that return identified data.
- Special controls exist when distributing data regarding HIV antibody and antigen test results, substance abuse rehab programs, and genetic data, due to specific state and federal laws.
- Queries that return identified data are reviewed (retrospectively) by the IRB.

# Getting data into the RPDR from clinical systems



# Organizing data in the Clinical Data Warehouse



Query items

Person who is using tool

Query construction

Results - broken down by number distinct of patients

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- [-] Encounter detail
  - [+] Clinics
  - [-] Hospital of service
    - Brigham and Womens Hospi
    - Massachusetts General Hospi
  - [+] Inpatient or outpatient
  - [+] Special Inpatients
- [+] Demographic detail
- [+] Diagnoses
- [+] Laboratory tests
- [+] Medications
- [+] Providers
- [+] Procedures

**Custom Query Items**

- [+] Query Library
- [+] Patient Sets

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]

[GROUP 2]

[GROUP 3]

Drag 'Query Items' from the left into these groups.

Create Query Export Patients Import Patients  Run Query

Gender	patients
Male	-
Female	-

Age	patients
0	
40	
80	

Race	patients
IABHWOU	

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

- Demographic detail
  - Age
  - Country
  - Gender
  - Language
  - Marital Status
  - Race
  - Religion
  - Veteran
  - Vital Status
    - Deceased
    - Living
    - not recorded
  - Zip codes
- Diagnoses
- Laboratory tests
- Medications
- Providers
- Procedures
- Custom Query Items
  - Query Library
  - Patient Sets

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1] [GROUP 2] [GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients Run Query

Gender	patients
Male	-
Female	-

Age

patients

0 40 80

Race	patients
IABHWOU	-

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

- Encounter detail
- Demographic detail
- Diagnoses**
  - Circulatory system
    - Acute Rheumatic fever
    - Arterial vascular disease
    - Cerebrovascular disease
    - Chronic Rheumatic heart diseases
    - Disease of capillaries
    - Diseases of pulmonary circulatory system
    - Hypertensive disease
    - Ischemic heart disease
      - Acute myocardial infarction
      - Angina pectoris
      - Old myocardial infarction
      - Other acute and subacute forms of myocardial infarction
      - Other forms of chronic ischemic heart disease
    - Other forms of heart disease
    - Venous and lymphatic disease
  - Conditions in the perinatal period
  - Congenital anomalies

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1] [GROUP 2] [GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients Run Query

Gender	patients
Male	-
Female	-

Age

patients

0 40 80

Race	patients
IABHWOU	-

Vital

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit

Address http://rpdrweb/partners/client/ Go Links >>

**Research Patient Database Query Tool** Logging: Shawn Murphy, MD Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
- Laboratory tests
  - Cell Counts
  - Cell Markers
  - Chemistry
    - Anemia Related Studies
    - Blood Gases
    - Cardiac Tests
      - CK-BB (CK1)
      - CK-MB (CK2)
      - CK-MB Index**
      - Cpk isoenzymes
      - Cpk isoenzymes
      - Cpk isoenzymes
    - CK-MB QUANT
    - CK-MM (CK3)
    - CPK
    - LDH

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]

The items of [1] are joined, and then intersected with other groups.

[GROUP 2]

Drag 'Query Items' from the left into these groups.

[GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients  Run Query

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th>Gender</th><th>patients</th></tr> <tr><td>Male</td><td style="text-align: center;">-</td></tr> <tr><td>Female</td><td style="text-align: center;">-</td></tr> </table>	Gender	patients	Male	-	Female	-	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th>Age</th></tr> <tr><td style="text-align: center;">0 40 80</td></tr> </table>	Age	0 40 80	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th>Race</th></tr> <tr><td style="text-align: center;">IABHWOU</td></tr> </table>	Race	IABHWOU	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th>Vital</th><th>patients</th></tr> <tr><td>Alive</td><td style="text-align: center;">-</td></tr> <tr><td>Dead</td><td style="text-align: center;">-</td></tr> </table>	Vital	patients	Alive	-	Dead	-
Gender	patients																		
Male	-																		
Female	-																		
Age																			
0 40 80																			
Race																			
IABHWOU																			
Vital	patients																		
Alive	-																		
Dead	-																		

Contains commands for working with the selected items.

RPDR Query Tool - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites

Address http://rpdrweb/partners/client/ Go Links

**Research Patient Database Query Tool** Logging: Shawn Murphy, MD Status: ● Data Status Home/Help

Query Items Find Previous Queries

**Standard Query Items**

- [-] Encounter detail
- [-] Demographic detail
- [-] Diagnoses
- [-] Laboratory tests
- [-] Medications
  - [-] Alternative medicines
  - [-] Anti-infectives
    - [-] Amebicides
    - [-] Aminoglycosides
    - [-] Anthelmintics
    - [-] Antifungals
      - [-] Amphotericins
        - [-] Amphotericin B
        - [-] Amphotericin B cl
        - [-] Amphotericin B lip
        - [-] Amphotericin B lip
        - [-] Liposomal amphotericin (ambisome) - lmr 2908
        - [-] Azole antifungals

Print

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]

[GROUP 2]

[GROUP 3]

The items of [1] are joined, and then intersected with other groups. Drag items from the 'Query Items' and 'Find' tabs on the left into these groups. Drag items from the 'Query' and 'Find' tabs on the left into these groups.

NEW GROUP

Create Query Request Data Manage Results Run Query

medications \ anti-infectives \ antifungals \ amphotericins \ liposomal amphotericin (ambisome) - lmr 2908

	Race	Vital	patients
Male	IABHWOU	Alive	-
Female		Dead	-

0 40 80

Done Local intranet

RPDR Query Tool - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Recycle Bin Mail Print Camera Taskbar Bluetooth

Address http://rpdweb/partners/client/ Go Links

**Research Patient Database Query Tool** Logging: Shawn Murphy, MD Status: ● Data Status Home/Help

Query Items Find Previous Queries

- Laboratory tests
- Medications
- Microbiology
- Molecular Medicine
  - Genomics
    - Hearing Loss
    - Hypertrophic Cardiomyope
    - Marfan Syndrome, TYPE I
    - Non-Small Cell Lung Canc
      - EGFR
        - 2125G>A (Respor
        - 2126A>T (Unclas
        - 2155G>A (Unclas
        - 2155G>A (Unclas
        - 2155G>T (Respor
        - 2156G>C (Respor

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1] [GROUP 2] [GROUP 3]

Print

Drag items from the 'Query Items' and 'Find' tabs on the left into these groups.

NEW GROUP

Molecular Medicine \ Genomics \ Non-Small Cell Lung Cancer \ EGFR \ 2155G>T (Responsive)

2232C>T (Novel F  
2235\_2243del (Re  
2235\_2249del (Re  
2236\_2250del (Pa  
2237\_2248delinsC  
2237\_2251del (L

Create Query Request Data Manage Results Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done Local intranet

RPDR Query Tool - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print Mail Stop Taskbar Bluetooth

Address http://rpdrweb/partners/client/ Go Links

**Research Patient Database Query Tool** Logging: Shawn Murphy, MD Status: Data Status Home/Help

Query Items Find Previous Queries

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
- Laboratory tests
- Medications
- Microbiology
- Molecular Medicine
- Health History
  - Vital Signs
    - Blood pressure
    - Heart Rate
    - Height
    - Respiratory rate
    - Temperature
    - Weight
- Providers
- Procedures
- Transfusion services

**Custom Query Items**

- Query Library
- Patient Sets

**Choose value of Heart Rate**

Searches using health history can be constrained to be within certain values through the dialog below.

No value ?    Please select operator:   
 By high/low flag ?    GREATER THAN (>)   
 By numeric value ?    Please enter value: 100

Click on a bar segment to help you specify a value or range:

units = beats/minute    OK    Clear    Cancel

Create Query    Request Data    Manage Results    Run Query

Gender	patients	Age	Vital
Male	-	0    40    80	Alive
Female	-	IABHWOU	Dead

Done Local intranet

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
- Laboratory tests
- Medications
- Providers
  - BWH
    - Administration
    - Anesthesiology, Perioper...
    - Cntr-Biochem And Biophy...
    - Data Processing
    - Dermatology
    - Emergency Medicine
    - Fellows, Residents and Ir...
    - Medicine
    - Midwife
    - Neonatology
    - Neurology
    - Neurosurgery
    - Newborn Medicine

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1] [GROUP 2] [GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients  Run Query

Gender	patients
Male	-
Female	-

Age

patients

0 40 80

Race	patients
IABHWOU	-

Vital

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

- Medications
- Providers
- Procedures
  - Diagnostic and nonsurgical pro
  - Obstetrical procedures
  - Operations on cardiovascular system
    - Incision, excision, and occl
    - Operations on valves and s
    - Operations on vessels of he
      - Bypass anastomosis for
      - Heart revascularization
  - Other operations on ve
  - Removal of coronary ar
  - Other operations on heart a
  - Other operations on vessek
  - Operations on digestive system
  - Operations on ear
  - Operations on endocrine system
  - Operations on eye
  - Operations on female genital sy

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < Specificity >

[GROUP 1] [GROUP 2] [GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients Run Query

Gender	patients
Male	-
Female	-

Age

patients

0 40 80

Race	patients
IABHWOU	-

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
  - Circulatory system
    - Acute Rheumatic fever
    - Arterial vascular disease
    - Cerebrovascular disease
    - Chronic Rheumatic heart
    - Disease of capillaries
    - Diseases of pulmonary ci
    - Hypertensive disease
    - Ischemic heart disease
      - Acute myocardial infarction
      - Angina pectoris
      - Old myocardial infarct
      - Other acute and sub
      - Other forms of chroni
      - Other forms of heart disea
      - Venous and lymphatic dis
    - Conditions in the perinatal per

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]

- Acute myocardial infarction

The items of [1] are joined, and then intersected with other groups.

[GROUP 2]

Drag 'Query Items' from the left into these groups.

[GROUP 3]

Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients  Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Qu

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
  - Circulatory system
    - Acute Rheumatic
    - Arterial vascular c
    - Cerebrovascular
    - Chronic Rheumat
    - Disease of capille
    - Diseases of pulm
    - Hypertensive dise
    - Ischemic heart di
      - Acute myoca
      - Angina pecto
      - Old myocardi
      - Other acute e
      - Other forms o
      - Other forms of he
      - Venous and lymph
    - Conditions in the perit

**Please complete...**

Please supply a name for this query:

Just Diagnoses AMI

Create a patient set from this query [What's this?](#)

Please choose one of these two timing models ...

Either (1) Some item from each group must have occurred in the same visit; or  
 (2) Some item from each group may have occurred at any visit. For example, if one wanted to find patients whose diabetes was treated at the MGH while an inpatient, choose the timing model where items occurred at the same visit.

Same Visit

Patient	Date	Groups		
		1	2	3
0000004	12/15/1999	MGH	Inpatient	Diabetes

Any Visit

Patient	Date	Groups		
		1	2	3
0000004	2/28/1999	B'WH	Inpatient	Seizure
0000004	10/05/1999	MGH	Outpatient	Afib
0000004	12/15/1999	B'WH	Outpatient	Diabetes

Run the query as background job

Run Query Cancel Query

Drag 'Query Items' from the... these groups.

NEW GROUP

Run Query

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

Research Patient Database Query Tool Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- [-] Encounter detail
- [-] Demographic detail
- [-] Diagnoses
  - [-] Circulatory system
    - [-] Acute Rheumatic fever
    - [-] Arterial vascular disease
    - [-] Cerebrovascular disease
    - [-] Chronic Rheumatic heart
    - [-] Disease of capillaries
    - [-] Diseases of pulmonary ci
    - [-] Hypertensive disease
    - [-] Ischemic heart disease
      - [-] Acute myocardial info
      - [-] Angina pectoris
      - [-] Old myocardial infarct
      - [-] Other acute and sub
      - [-] Other forms of chroni
      - [-] Other forms of heart disea
      - [-] Venous and lymphatic dis
    - [-] Conditions in the perinatal per

**Just Diagnoses AMI**

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]	[GROUP 2]	[GROUP 3]
Acute myocardial infarction		

**Please wait ...**

Querying the RPDR...

Cancel Query

The items of [1] are joined, and then intersected with other groups. Drag 'Query Items' from the left into these groups. Drag 'Query Items' from the left into these groups. NEW GROUP

Create Query Export Patients Import Patients Running query ... Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

**Standard Query Items**

- Encounter detail
- Demographic detail
- Diagnoses
  - Circulatory system
    - Acute Rheumatic fever
    - Arterial vascular disease
    - Cerebrovascular disease
    - Chronic Rheumatic heart
    - Disease of capillaries
    - Diseases of pulmonary ci
    - Hypertensive disease
    - Ischemic heart disease
      - Acute myocardial info
      - Angina pectoris
      - Old myocardial infarct
      - Other acute and sub
      - Other forms of chroni
      - Other forms of heart disea
      - Venous and lymphatic dis
    - Conditions in the perinatal per

**Just Diagnoses AMI**

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1] [GROUP 2] [GROUP 3]

Acute myocardial infarction

The items of [1] are joined, and then intersected with other groups. Drag 'Query Items' from the left into these groups. Drag 'Query Items' from the left into these groups.

NEW GROUP

Create Query Export Patients Import Patients 46388 patients. Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool**    Logging: Jeanne M. Guerin, RN    Status: ●    Print    Help

Query Items    Find    Previous Queries

Search for: mb index

Laboratory tests    Find

CK-MB Index

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT    Sensitivity <    > Specificity

[GROUP 1]	[GROUP 2]	[GROUP 3]
Acute myocardial infarction		

The items of [1] are joined, and then intersected with other groups.    Drag 'Query Items' from the left into these groups.    Drag 'Query Items' from the left into these groups.    NEW GROUP

Create Query    Export Patients    Import Patients    Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done    Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

Search for: mb index

Laboratory tests Find

CK-MB Index

GROUPS DO NOT HAVE

[GROUP 1]

Acute myoc...

**Choose value of CK-MB Index**

Searches by Lab values can be constrained by the high/low flag set by the performing laboratory, or by the values themselves.

No value ?
  By high/low flag ?
  **By numeric value:** ?

Please select operator: GREATER THAN (>)

Please enter value: 3.50

Click on a bar segment to help you specify a value or range:

units = %

OK Clear Cancel

The items of [1] are intersected with

Create Query

Gender	patients
Male	-
Female	-

Age
0 40 80

Race
IABHWOU

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool**    Logging: Jeanne M. Guerin, RN    Status: ●    Print    Help

Query Items    Find    Previous Queries

Search for: mb index

Laboratory tests    Find

CK-MB Index

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT    Sensitivity <    > Specificity

[GROUP 1]	[GROUP 2]	[GROUP 3]
Acute myocardial infarction	CK-MB Index > 3.50	

The items of [1] are joined, and then intersected with other groups.    The items of [2] are joined, and then intersected with other groups.    Drag 'Query Items' from the NEW GROUP to these groups.

Create Query    Export Patients    Import Patients    Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done    Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Qu

Search for: mb index  
Laboratory tests

CK-MB Index

**Please complete...**

Please supply a name for this query:  
AMI and CK-MB>3.5

Create a patient set from this query [What's this?](#)

Please choose one of these two timing models ...

Either (1) Some item from each group must have occurred in the same visit; or  
(2) Some item from each group may have occurred at any visit. For example, if one wanted to find patients whose diabetes was treated at the MGH while an inpatient, choose the timing model where items occurred at the same visit.

Same Visit

Patient	Date	Groups		
		1	2	3
0000004	12/15/1999	MGH	Inpatient	Diabetes

Any Visit

Patient	Date	Groups		
		1	2	3
0000004	2/28/1999	B'WH	Inpatient	Seizure
0000004	10/05/1999	MGH	Outpatient	Afib
0000004	12/15/1999	B'WH	Outpatient	Diabetes

Run the query as background job

Run Query Cancel Query

Specificity

[GROUP 3]

NEW GROUP

Drag 'Query Items' from the these groups.

Run Query

Vital	patients
Alive	-
Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool** Logging: Jeanne M. Guerin, RN Status: ● Print Help

Query Items Find Previous Queries

Search for: mb index

Laboratory tests Find

CK-MB Index

**AMI and CK-MB > 3.5**

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < > Specificity

[GROUP 1]	[GROUP 2]	[GROUP 3]
Acute myocardial infarction	CK-MB Index > 3.50	

The items of [1] are joined, and then intersected with other groups.    The items of [2] are joined, and then intersected with other groups.    Drag 'Query Items' from the NEW GROUP to these groups.

Create Query Export Patients Import Patients 11430 patients. Run Query

Gender	patients	Age	Race	Vital	patients
Male	-	0 40 80	IABHWOU	Alive	-
Female	-			Dead	-

Done Trusted sites

RPDR Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address http://rpdweb/partners/client/

**Research Patient Database Query Tool**    Logging: Jeanne M. Guerin, RN    Status: ●    Print    Help

Query Items    Find    Previous Queries

Search for: mb index

Laboratory tests    Find

CK-MB Index

**AMI and CK-MB > 3.5 analysis**    [Navigation icons]

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT    Sensitivity <    > Specificity

[GROUP 1] Acute myocardial infarction

The items of [1] are joined, and then intersected with other groups.

[GROUP 2] CK-MB Index > 3.50

The items of [2] are joined, and then intersected with other groups.

[GROUP 3]

Drag 'Query Items' from the NEW GROUP these groups.

Create Query    Export Patients    Import Patients    11430 patients.    Run Query

Gender	patients
Male	7342
Female	4088

Age

max: 3484

Race

max: 10078

Vital	patients
Alive	-
Dead	-

Ages 40-49, 551 patients

Done    Trusted sites

### ***Welcome to the RPDR Data Request Wizards***

The RPDR is a HIPAA compliant system, which returns aggregate patient information via a Query Tool, based on user-defined criteria. With proper IRB approval, RPDR users can:

- use their previously queried patient set
- or import their own approved set of Medical Record Numbers

to request detailed or identified patient clinical data. These wizards are included in the RPDR for human research investigators to request identified patient data from their respective Partners sites.

#### **You are now launching a wizard in order to request identified patient data.**

Your request must conform and comply with the allowances of your Partners sponsored IRB human studies protocol. This responsibility rests entirely on the faculty sponsor who is requesting the identified data or who is approving the request of identified data from a workgroup member. It is very important that the correct IRB protocol number be chosen for each request of protected health information.

This information is protected under the Partners Privacy and Confidentiality Policy and provided with approval by the Human Research Committee only for the use specified in your protocol number. It may not be used for any other purpose without specific approval by the Human Research Committee. It may not be distributed to any individual not specifically authorized under that approval. The data must be managed in a manner that complies with HIPAA Security Regulations.

I accept responsibility for the data returned by this query.

[Accept](#) [Cancel](#)

#### **Partners Healthcare System HIPAA Compliance**

Additional HIPAA information for the research community is available from these links, sponsored by Partners and the Human Research Council (PHRC).

[HIPAA and the Privacy Rule](#)  
[HIPAA Central](#)

RPDR Detailed Data Request Wizard -- Web Page Dialog

**RPDR DETAILED DATA REQUEST WIZARD**

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**Select the sites from which you would like to receive data**  
(Please note that different sites and Institutional Review Boards (IRBs) may have different policies regarding obtaining patient data. Detail of the policies can be read [here](#))

- Massachusetts General Hospital (MGH)
- Brigham and Women's Hospital (BWH)
- Newton Wellesley Hospital (NWH)
- Spaulding Rehabilitation Hospital (SRH)
- Faulkner Hospital (FH)
- North Shore Medical Center (NSMC)

Help   < Back   **STEP 2**   Next >   Cancel

RPDR Detailed Data Request Wizard -- Web Page Dialog

**RPDR DETAILED DATA REQUEST WIZARD**

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**Select protocol number(s)**

Partners IRB (required):

Title: RPDR protocol - Demonstration IRB number for Dr. Murphy  
Status: Active

Newton Wellesley Hospital IRB:

Title: test  
Status: Active

Spaulding Rehabilitation Hospital IRB:

Options for returned set of patients:

- Create a static set of patients from this query that can be used in other RPDR queries
- Rerun the base query shown above to obtain a fresh set of patients

Help < Back **STEP 3** Next > Cancel

RPDR Detailed Data Request Wizard -- Web Page Dialog

**RPDR DETAILED DATA REQUEST WIZARD**

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**Select others who require access to this data**

Selected faculty sponsor(s):  
Shawn Murphy, MD

**Available**

Peng, Zhaoping Z.  
Nalichowski, Ruth

**Chosen**

Gainer, Vivian S

Add ->  
Add All ->>  
<- Remove  
<<- Remove All

Help < Back **STEP 5** Next > Cancel

RPDR Detailed Data Request Wizard -- Web Page Dialog

**RPDR DETAILED DATA REQUEST WIZARD**

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**Please select if you would like a HIPAA-defined (deidentified) limited data set or an identified data set**

What's a limited data set?

**Limited Data Set**

- The files that result from this request will be available in a protected file share with no special encryption.

**Identified Data Set**

- The text files that result from this request will be encrypted and the Microsoft Access file will be password protected. In order to access the data, a password will be provided.

Help < Back **STEP 8** Next > Cancel

**RPDR Detailed Data Request Wizard -- Web Page Dialog**

**RPDR DETAILED DATA REQUEST WIZARD**  
Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**Select the types of data that should be returned from the RPDR**  
**Only data allowed by your protocol should be chosen**  
(Identified data sets will always return a set of identified patient medical numbers)

**Detail Data Items**

- Demographic Data
- Identifying Patient Information - not available for Limited Data Sets
- LMR (Longitudinal Medical Record)
- Medications, Diagnoses and Procedures
- Medications, Diagnoses and Procedures from Billing Data - only visits where query criteria occur all in the same visit
- Patient Clinical Reports- not available for Limited Data Sets
  - Cardiology Reports
  - Discharge Summaries
  - Endoscopy Reports
  - Microbiology Data
  - Operative Notes
  - Pathology Reports
  - Pulmonary Reports
  - Radiology Reports
  - Transfusion Data, Blood Bank Data

Help   < Back   **STEP 9**   Next >   Cancel

**RPDR DETAILED DATA REQUEST WIZARD -- Web Page Dialog**

**RPDR DETAILED DATA REQUEST WIZARD**  
Using IRB#2000P000000 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

Search for [Back to top view](#) **Find**

**Laboratory Data Items**

- Laboratory tests
  - Cell Counts
  - Cell Markers
  - Chemistry
    - Anemia Related Studies
    - Blood Gases
    - Blood Gases/Oximetry
    - Cardiac Tests
      - BNP (Group:BNP)
      - CK Isos (Group:CK-MMT)
      - CK-BB (CK1) (Group:CK1/BB)
      - CK-BB% (Group:%CK1BB)
      - CK-MB (CK2) (Group:CK2/ME)
      - CK-MB Index (Group:CKMBR)**
      - CK-MB QUANT (Group:CPK-M)
      - CK-MB% (Group:%CKMB)

**Selected Data Items**

- CK-MB Index (Group:CKMBRI) (Loinc:2158-4)

**Please select the laboratory tests that should be returned**  
(Drag the lab items from the left panel and drop into the panel below)

**STEP 8**

Help < Back Next > Cancel

RPDR DETAILED DATA REQUEST WIZARD

Using IRB#2000P000000 (found in the RPDR Identified database) to obtain data from the RPDR  
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

**RPDR DISCLAIMER: All requests, such as this one, are reviewed by the IRB to assure compliance with the written protocol. If the data requested is not authorized by the protocol, those involved with its retrieval face disciplinary action including loss of federal funding, loss of employment, and/or criminal prosecution.**

[Read Full IRB Policy](#)

Enter IRB Password:

Help

< Back

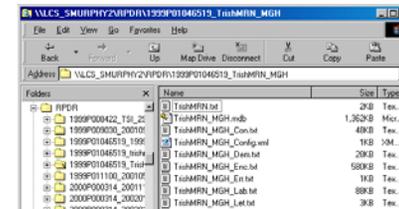
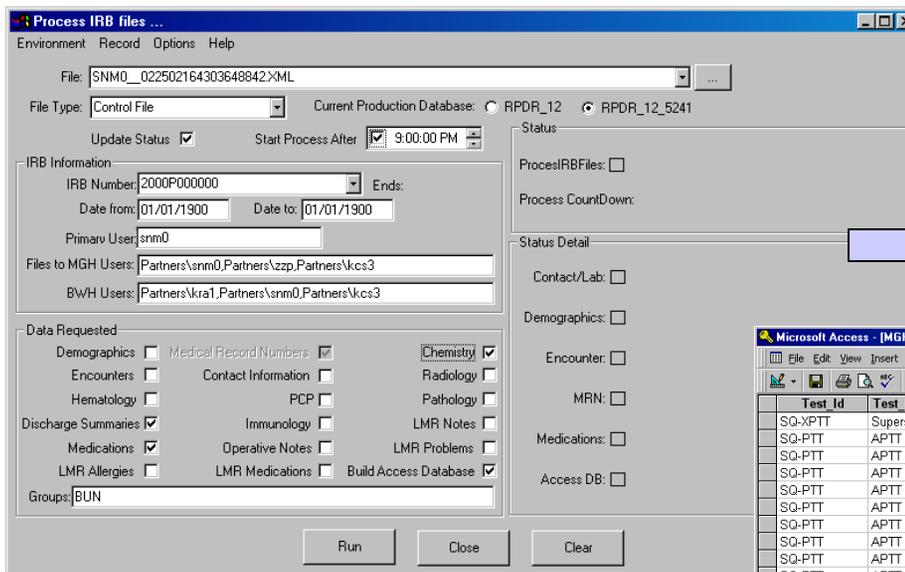
STEP 11

Next >

Cancel

# Identified data is gathered

Output files placed in special directory



The Microsoft Access window displays a table with the following columns: Test Id, Test Description, Result, Result Text, Abnormal Flag, Reference, and Range. The table contains multiple rows of test data, including Superstat APTT, APTT, and Hemolysis tests.

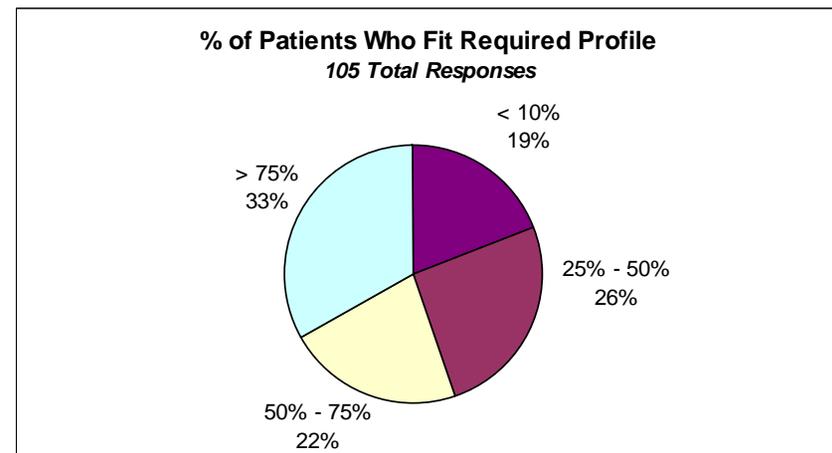
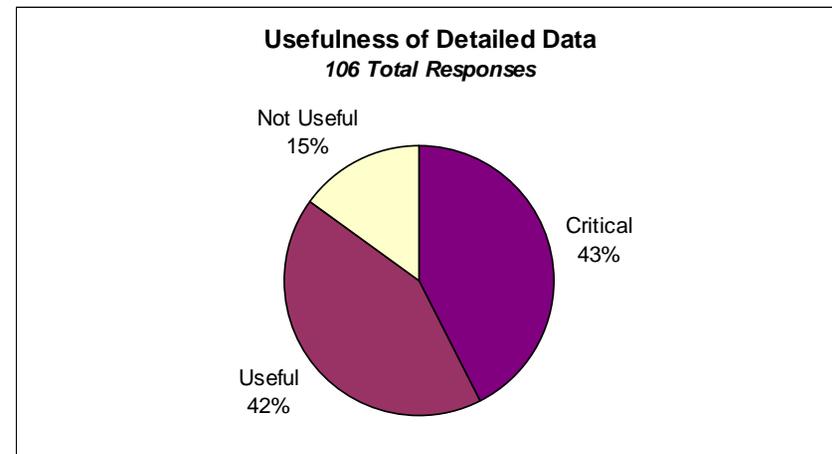
Test Id	Test Description	Result	Result Text	Abnormal Flag	Reference	Range
SQ-XPTT	Superstat APTT	29.8			sec	22.1-35.1
SQ-PTT	APTT	32.8			sec	22.1-35.1
SQ-PTT	APTT	37.8		H	sec	22.1-35.1
SQ-PTT	APTT	46.4		H	sec	22.1-35.1
SQ-PTT	APTT	43.1	MODERATELY	H	sec	22.1-35.1
SQ-PTT	APTT	25.7			sec	22.1-35.1
SQ-PTT	APTT	23.7			sec	22.1-35.1
SQ-PTT	APTT	25.4			sec	22.1-35.1
SQ-PTT	APTT	24.7			sec	22.1-35.1
SQ-PTT	APTT	24.0			sec	22.1-35.1
SQ-PTT	APTT	24.7			sec	22.1-35.1
SQ-XPTT	Superstat APTT	31.3			sec	22.1-35.1
SQ-PTT	APTT	34.5			sec	22.1-35.1
SQ-PTT	APTT	40.0		H	sec	22.1-35.1
SQ-PTT	APTT	45.0		H	sec	22.1-35.1
SQ-XPTT	Superstat APTT	55.2	*** Note: New n	H	sec	22.1-35.1
SQ-PTT	APTT	33.6			sec	22.1-35.1
SQ-XPTT	Superstat APTT	34.3			sec	22.1-35.1
SQ-PTT	APTT	37.9		H	sec	22.1-35.1
SQ-PTT	APTT	22.6			sec	22.1-34.1
SQ-PTT	APTT	37.4		H	sec	22.1-34.1
SQ-PTT	APTT	37.2	SLT HEMOLYS	H	sec	22.1-34.1
SQ-PTT	APTT	35.1		H	sec	22.1-34.1
SQ-PTT	APTT	36.4	MODERATE HE	H	sec	22.1-34.1

Data is gathered from RPDR and other Partners sources

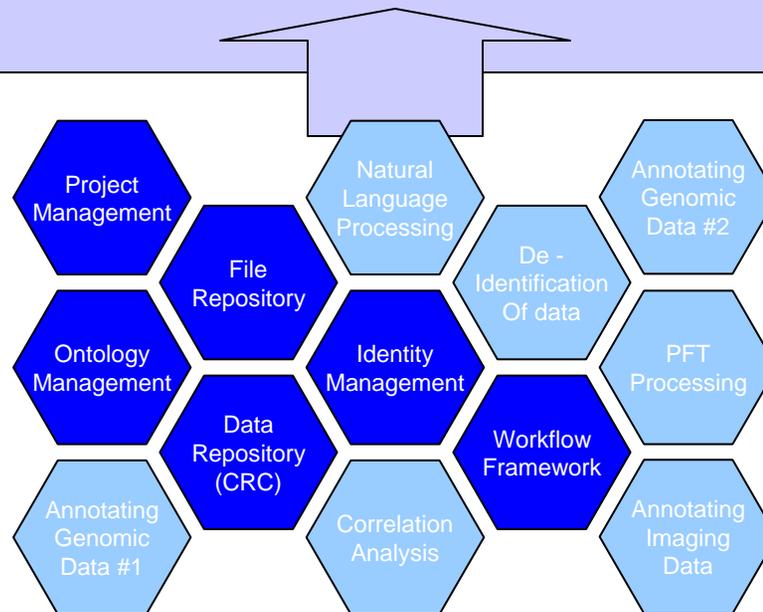
Files include a Microsoft Access Database

# 2007's usage of RPDR

- 1,580 registered users, 332 new in 2007
- 294 teams gathering data for research studies
- 815 identified patient data sets returned to these teams, containing data for of 8.8 million patient records.
- From a survey of 153 teams
  - Importance of the data received from the RPDR was evaluated in relation to the study it was supporting.
  - The adequacy of the match of a patient profile that could be obtained through the RPDR query tool was estimated.
- \$94-136 million total research support critically dependent on RPDR from patient data received throughout life of funding.
- ~300 data marts were created to support hospital operations, representing about 80 million patient records



# Enterprise-wide repurposing and distribution of medical record data for research



- Enable high performance collection of medical record data for querying and distribution
  - Enterprise web client
  - Create patient cohorts for further investigation
- Enable discovery within data on enterprise wide scale
  - Relevance networks
  - Health Surveillance

# Enterprise web client

## <http://services.i2b2.org/webclient/>

The screenshot shows the i2b2 Query & Analysis Tool web client interface. The browser window title is "i2b2 Web Client - Microsoft Internet Explorer" and the address bar shows "http://services.i2b2.org/webclient/#". The interface is divided into several sections:

- Navigate Terms / Find Terms:** A search area with "Search by Names" and "Search by Codes" tabs. The "Search by Names" tab is active, showing a search for "multiple sclerosis" with a "Find" button and a dropdown menu set to "contains". Below the search input, the results list "Multiple sclerosis".
- Workplace:** A file explorer view showing a tree structure with "SHARED" and "demo" folders.
- Query Tool:** A central area for building queries. It shows a "Query Name" field with the value "Moderat-Multipl@02:14:47". Below this are three columns representing query groups: "Group 1", "Group 2", and "Group 3". Each group has a "Dates" field, an "Occurs > 0x" field, and an "Exclude" checkbox. Group 1 contains "Moderate Alzheimers dise:", Group 2 contains "Multiple sclerosis", and Group 3 is empty. Below the groups are three buttons: "none of these" (red), "AND" (blue), "one or more of these" (green), "AND" (blue), and "drop a term on here" (yellow).
- Previous Queries:** A list of recent queries, including "Moderat-Multipl@02:14:47 [2-24-2009] [demo]", "Moderat-Multipl@02:14:24 [2-24-2009] [demo]", "BetaBlockerStudy [2-24-2009] [demo]", "Coronal@21:39:04 [2-24-2009] [demo]", "Age-Circu-WBC (@21:36:01 [2-24-2009] [demo]", and "Age-Circu-WBC (@21:32:35 [2-24-2009] [demo]".
- Query Status:** A box showing the execution status: "Executing query...", "Elapsed time (seconds): 2.8", "Query Finished...", and "Matching patients: 2".

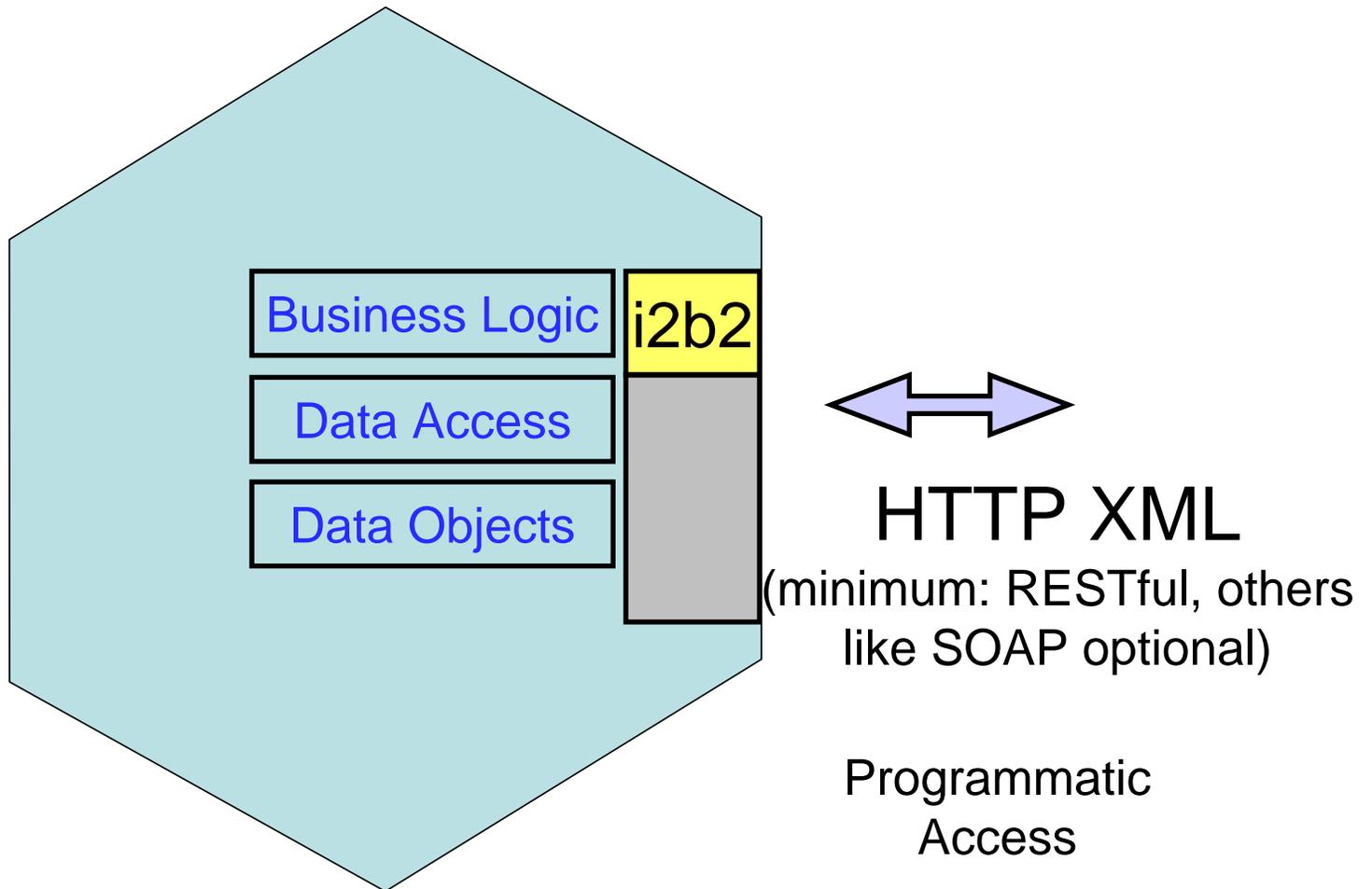
The browser's taskbar at the bottom shows the "Internet" icon.



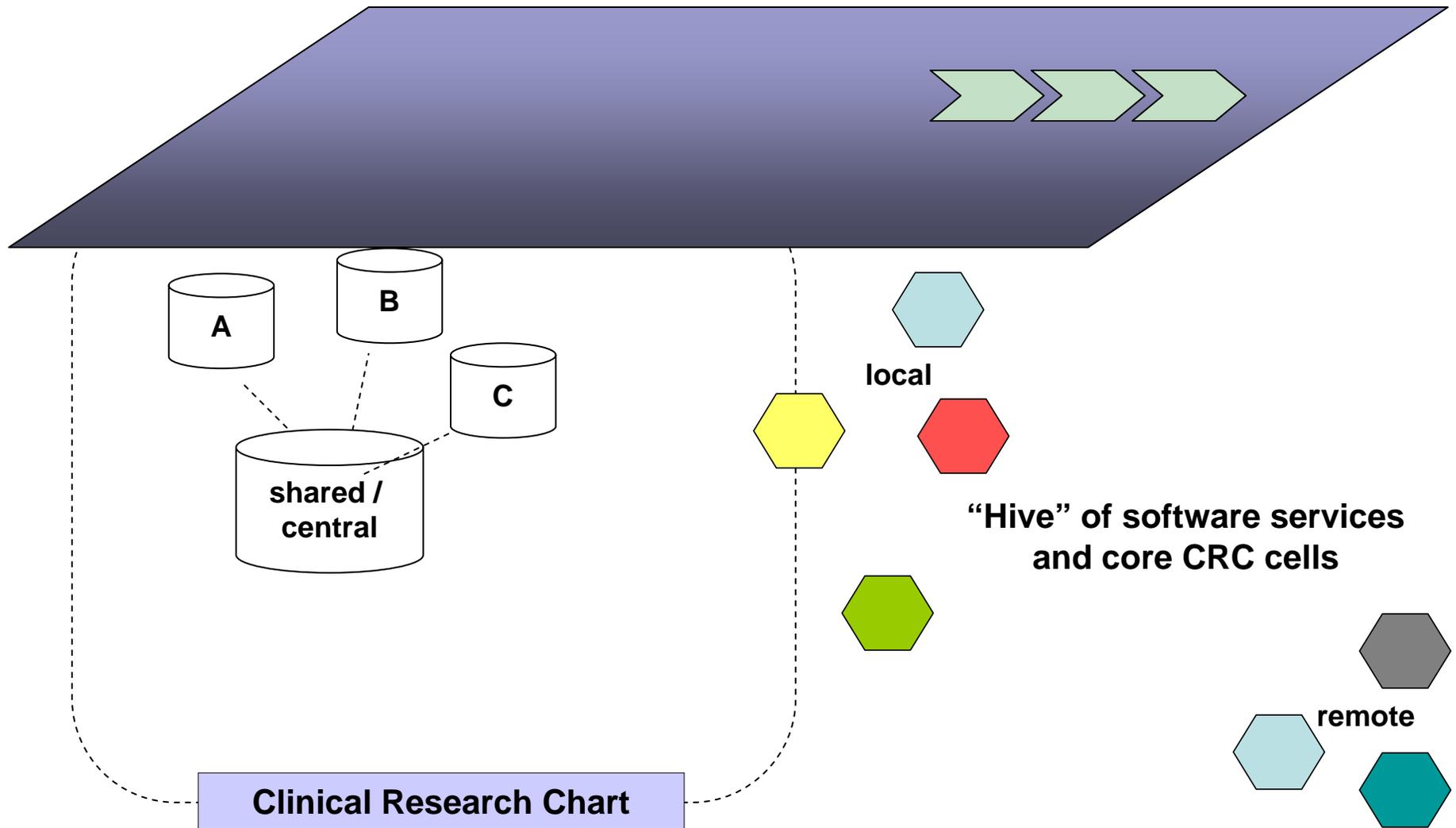
# Technical Overview

- Formed as a collection of interoperable services provided by i2b2 Cells
- Loosely coupled
- Makes no assumptions about proximity
- Connected by Web services
- Activity can be directed manually or automatically

# i2b2 Cell: Canonical Hive Unit



# i2b2 Environment





-  CTSAs\* Adopting i2b2
-  CTSAs\* Evaluating i2b2 platform
-  Academic Medical Centers Adopting i2b2 Platform
-  Foreign Medical Centers Adopting i2b2 Platform

# Community

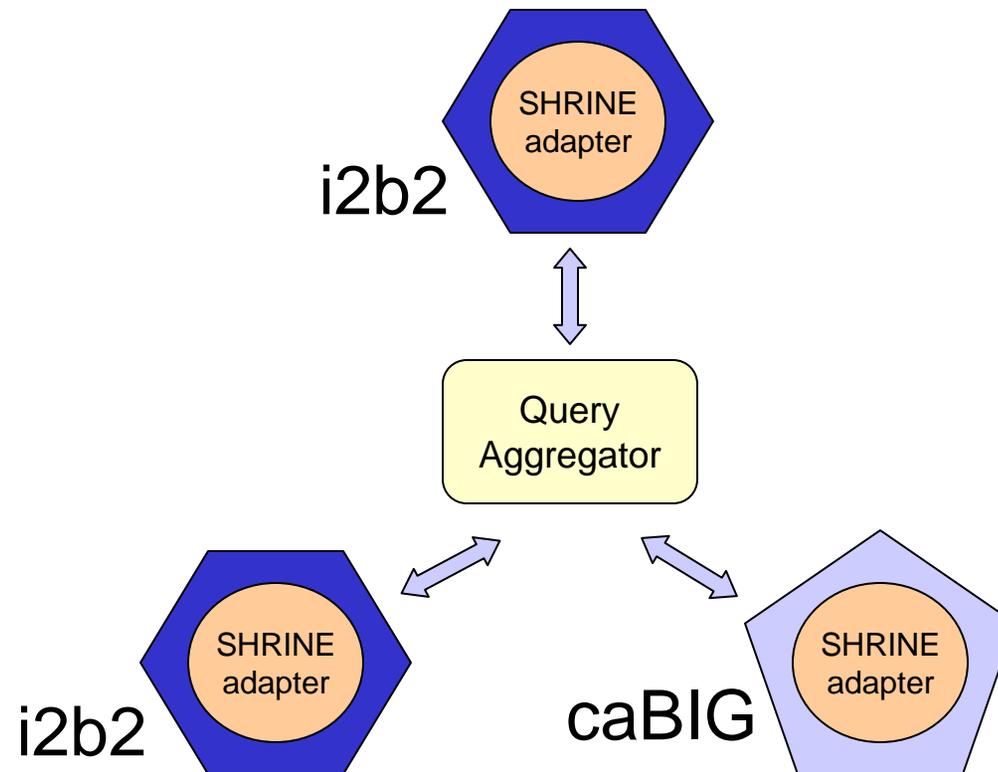
## United States

- Beth Israel Deaconess Hospital, Boston, MA
- Boston University School of Medicine, Boston, MA
- Brigham and Women's Hospital, Boston, MA
- Children's Hospital, Boston, MA
- Denver Children's Hospital, Denver, CO
- Cincinnati Children's Hospital, Cincinnati, OH
- Cleveland Clinic, Cleveland, OH
- Weil Medical College of Cornell, NYC, NY
- Group Health Cooperative
- Harvard Medical School, Boston, MA
- Massachusetts General Hospital, Boston, MA
- Maine Medical Center, Portland, ME
- Marshfield Clinic, Wisconsin
- Morehouse School of Medicine, Atlanta, GA
- Oregon Health & Science University, Portland, OR
- Ohio State University Medical Center, Columbus, OH
- Philadelphia Children's Hospital, Philadelphia, PA
- Renaissance Computing Institute, Chapel Hill, NC
- Tufts New England Medical Center, Boston, MA
- University of California Davis, Davis, CA
- University of California San Francisco, SF, CA
- University of Massachusetts Medical School, Worcester, MA
- University of Michigan Medical Center, Ann Arbor, MI
- University of Pennsylvania School of Medicine, Philadelphia, PA
- University of Rochester Medical Center, Rochester, NY
- University of Texas Health Sciences Center Houston, Houston, TX
- University of Texas Health Sciences Center San Antonio, SA, TX
- University of Texas Health Sciences Center Southwestern,
- Utah Health Science Center, Salt Lake City, UT
- University of Washington, Seattle, WA

## International

- Georges Pompidou Hospital, Paris, France
- University of Goettingen, Goettingen, Germany
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

# SHRINE (Shared Research Informatics Network) = Distributed Queries



Central “aggregator” broadcasts query to local hospital “adaptors”, which return aggregate counts only

# SHRINE

**i2b2 Query & Analysis Tool** Logout

**Navigate Terms** Find Terms

- Demographics
  - Age
  - Gender
    - Female
    - Male
    - Unknown
  - Language
  - Marital-Status
  - Race
  - Vital-Status
  - Diagnoses

drag an item from here

**Query Tool**

Query Name: \_\_\_\_\_

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
			Regional enteritis Ulcerative colitis	Male		Neoplasms		

one or more of these    AND    none of these    AND    one or more of these

Run Query    New Query    3 Groups    New Group

**Previous Queries**

- Noninfe-Female@16:39:03
- Nonin-Femal-Neopl@16:39:11
- Circulatory sys@16:39:18
- Circulatory sys@17:01:03
- Circula-Events @17:04:11
- Demographics@17:04:21
- 10-19 years old@17:04:39
- Albanian@02:31:25

**Query Status**

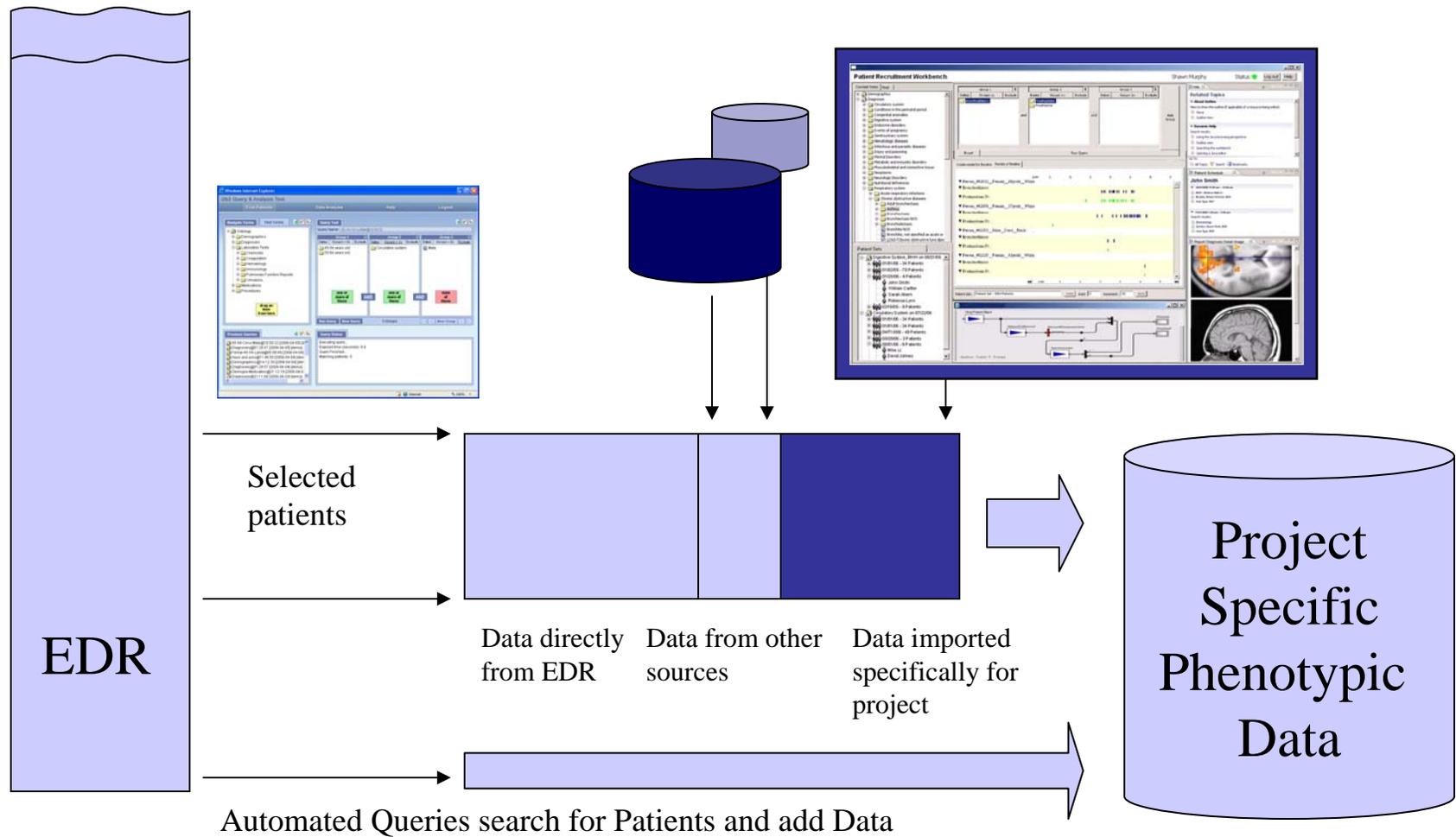
Executing query...  
Elapsed time (seconds): 14.0  
Query Finished...  
Matching patients (hospital 1): 332 (+/-3)  
Matching patients (hospital 2): 16 (+/-3)  
Matching patients (hospital 3): 151 (+/-3)



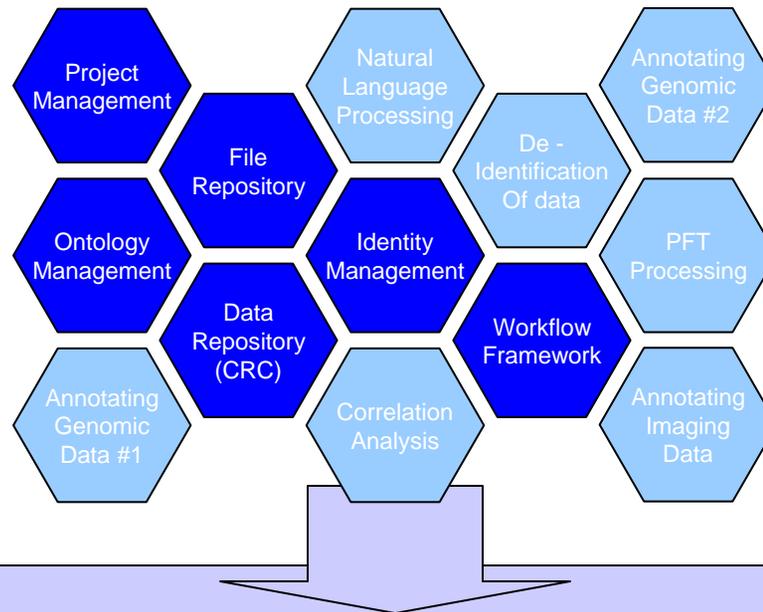
# High Throughput Method for supporting Research

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
  - Isaac Kohane
  - John Glaser
  - Susanne Churchill
  - Henry Chueh
  - Griffin Weber
  - Michael Mendis
  - Andrew McMurry
  - Vivian Gainer
  - Lori Phillips
  - Rajesh Kuttan
  - Wensong Pan

# Set of patients is selected through Enterprise Repository and data is gathered into a data mart



- Repurpose medical record information for research studies
  - I2b2 Workbench
  - Natural language processing
- Enable Genomic Studies and Public Health
  - Tissue/blood selection
  - Health Reporting



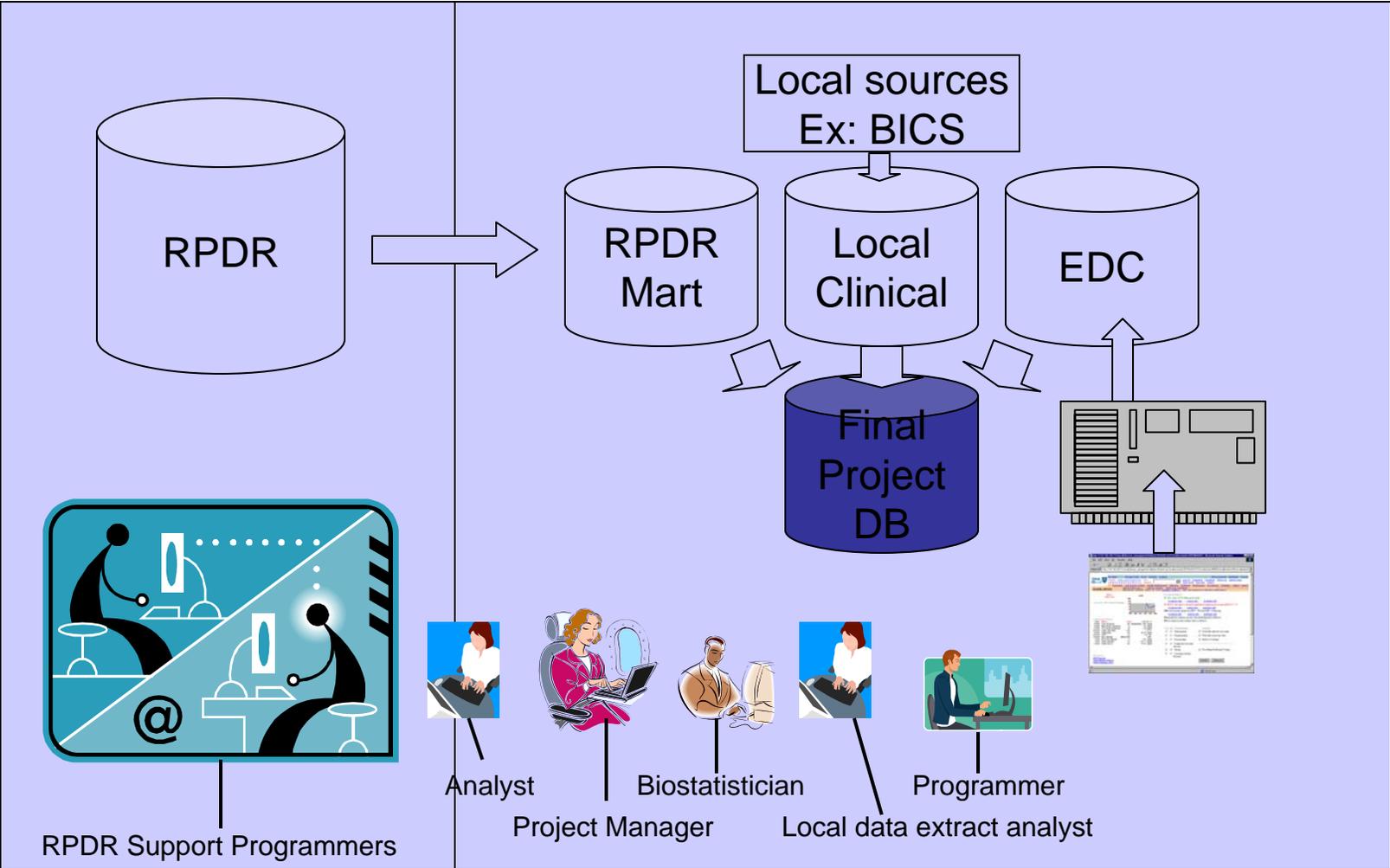
Use of medical record data in clinical studies  
focused upon teamwork and workflow

# Data is available through the i2b2 Workbench

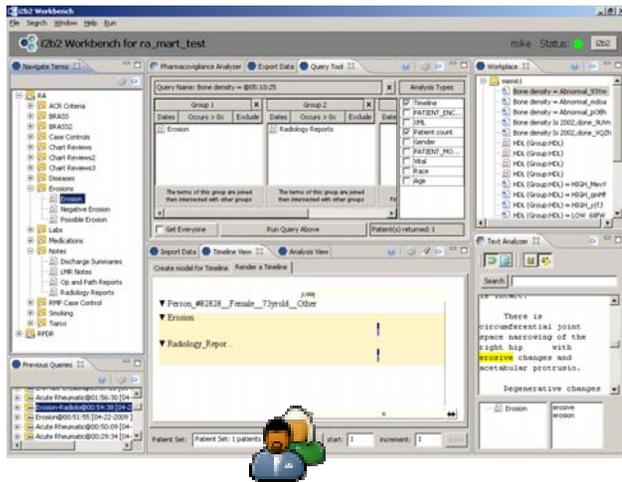
The screenshot displays the 'Patient Recruitment Workbench' interface. At the top, the user is identified as 'Shawn Murphy' with a 'Status' indicator and 'Log out' and 'Help' buttons. The interface is divided into several main sections:

- Concept Trees:** A hierarchical tree on the left lists medical categories such as 'Demographics', 'Diagnoses', 'Respiratory system', and 'Patient Sets'. The 'Respiratory system' is expanded to show 'Chronic obstructive diseases' and 'Asthma'.
- Query Groups:** Three groups are defined: 'Group 1' containing 'Bronchodilators', 'Group 2' containing 'Prednisolone' and 'Prednisone', and 'Group 3' which is currently empty. The groups are connected by 'and' operators.
- Patient Timelines:** A central area shows timelines for several patients, including 'Person\_#82032\_Female\_68yroid\_White', 'Person\_#82098\_Female\_37yroid\_White', 'Person\_#82202\_Male\_Deaf\_Black', and 'Person\_#82229\_Female\_63yroid\_White'. Each patient's record shows vertical bars representing the occurrence of 'Bronchodilators' and 'Prednisolone-Pr...'. A 'Patient Set' of 3894 patients is also indicated.
- Diagnostic Images:** Two windows on the right show medical images. The top one is a 'Report Diagnosis Detail Image' showing a cross-section of a brain with orange highlights. The bottom one is a sagittal MRI scan of a brain.
- Workflow Diagram:** At the bottom, a flowchart shows the data processing pipeline: 'Drop Patient Object' feeds into 'ObjectToRecord', which then feeds into 'RecordDisassembler'. The output of 'RecordDisassembler' goes to 'Synchronizer', which then feeds into two 'Display' components.

# Team support for Projects



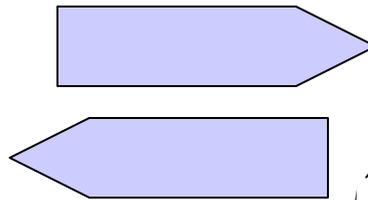
# NLP Workflow



I2b2 Project Investigators

Results Delivery

Communication



NLP Specialists

# NLP (and comedy) is not pretty

Programmer's File Editor - [050210\_1629\MiniDem1.txt]

File Edit Options Template Execute Macro Window Help

SOCIAL HISTORY: The patient is married with four grown daughters, **uses tobacco**, has wine with dinner. **Smoker**

PRINCIPAL DIAGNOSIS: LEFT LOWER LOBE PNEUMONIA

SECONDARY DIAGNOSES:

1. CHI SOCIAL HISTORY: The patient is a **nonsmoker**. No alcohol. **Non-Smoker**

2. HELICOBACTER PYLORI

HISTORICAL SOCIAL HISTORY: **Negative for tobacco**, alcohol, and IV drug abuse.

PAST MEDICAL HISTORY: (1) Hip fracture. (2) Bronchiectasis.

BRIEF RESUME OF HOSPITAL COURSE:

M 63 yo woman with COPD, **50 pack-yr tobacco (quit 3 wks ago)**, **Past Smoker**

ALLERGIES: (1) Aspirin. (2) Ciprofloxacin. (3) Penicillin.

SOCIAL HISTORY: The patient lives alone and denies tobacco or

F SOCIAL HISTORY: The patient lives in rehab, married. **Unclear smoking** history **???**

from the admission note...

PHYSICAL EXAMINATION: Temperature 97.2, pulse 60, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils

LABORATORY DATA: Sodium 148, potassium 3.4, chloride 87, bicarb

HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of **Lactobacillus** acidophilus to populate her gut. **Hard to pick**

Chest x-ray revealed hyperinflated lungs and

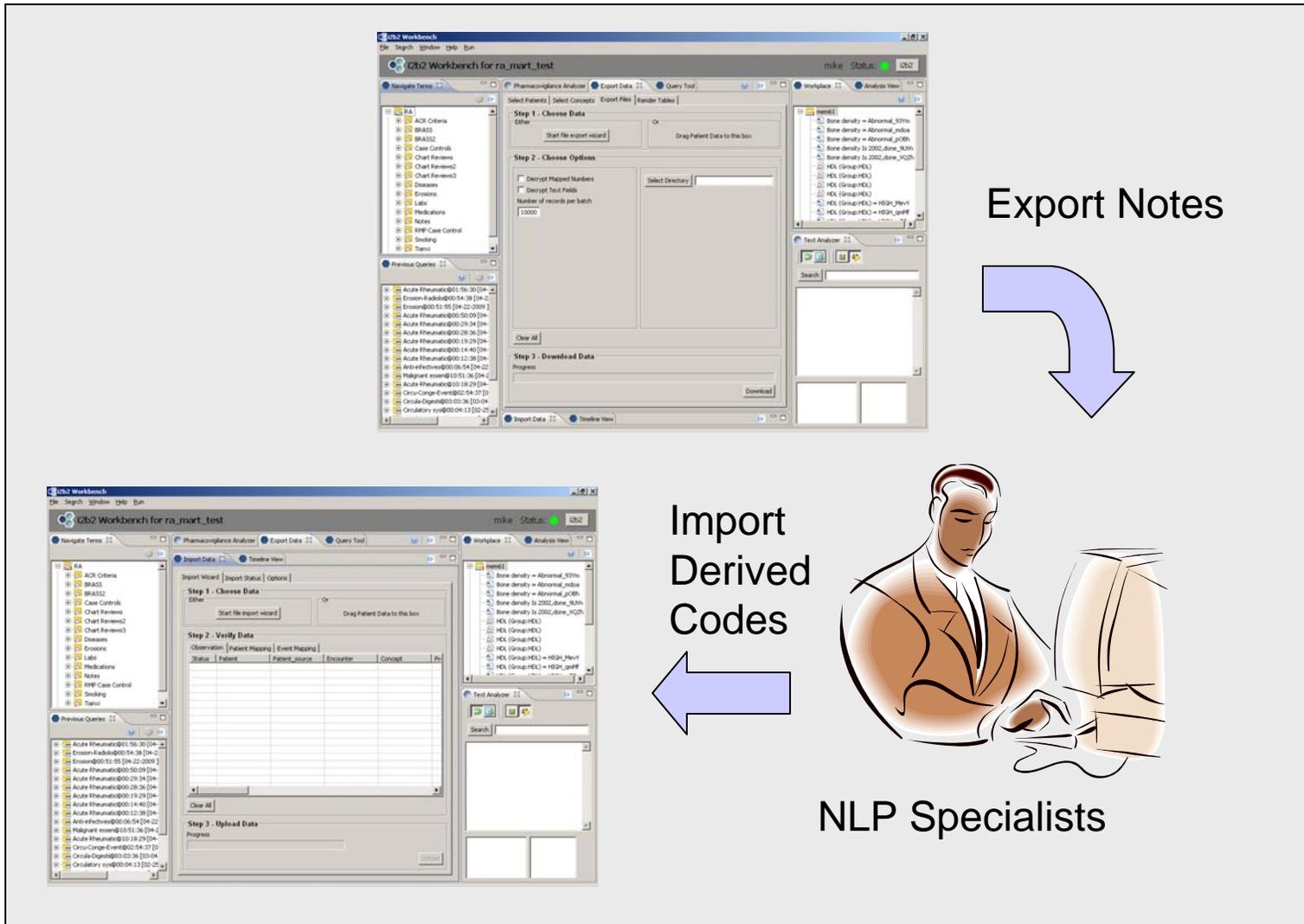
HOSPITAL COURSE: The patient was seen and evaluated by the

The patient was discharged home on 07/19/77 to rehab to have a 14-day course of acetaminophen

SH: widow, lives alone, 2 children, no **tob/alcohol**. **Hard to pick**

Ln 44 Col 1 274 WR Rec Off No Wrap DOS INS NUM

# NLP Specialists Workstation



# Investigator Review

The screenshot displays the i2b2 Workbench interface for a user named 'mike'. The main window is titled 'i2b2 Workbench for ra\_mart\_test'. The interface is divided into several panes:

- Navigate Terms:** A tree view on the left showing a hierarchy of terms under 'RA', including 'ACR Criteria', 'BRASS', 'Erosions', and 'Radiology Reports'. 'Erosion' is currently selected.
- Pharmacovigilance Analyzer:** The central pane shows a query named 'Bone density = @05:10:25'. It features two groups: 'Group 1' containing 'Erosion' and 'Group 2' containing 'Radiology Reports'. Below the groups are buttons for 'Get Everyone', 'Run Query Above', and 'Patient(s) returned: 1'.
- Analysis Types:** A list of analysis options on the right, with 'Timeline' and 'Patient count' checked.
- Workplace:** A list of search results on the right, including 'Bone density = Abnormal\_93Ym', 'HDL (Group:HDL)', and 'HDL (Group:HDL) = HIGH\_MevY'.
- Timeline View:** The bottom pane shows a timeline for 'Person\_#82828\_Female\_73yrold\_Other'. It lists 'Erosion' and 'Radiology\_Report...' with a vertical bar indicating a point in time. The timeline is labeled '2/99'.
- Text Analyzer:** A pane on the bottom right showing a search for 'erosive' in a text document. The search results highlight 'erosive changes and acetabular protrusio.' and 'Degenerative changes'.

# Select patients for clinical trials

The screenshot displays the BIRN Workbench software interface, which is used for managing and viewing medical data. The main window is titled "BIRN Workbench for Demo Group" and shows a "Query Tool" with a "Timeline View" tab selected. The "Select Patients" section displays a table with columns for Row #, Decision, PSet #, Patient ID, Patient Name, Gender, Race, Date of Birth, Age, and MRNs. The table contains 22 rows of data, with the first four rows having a "Yes" decision and the remaining 18 rows having a "No" or "Und" decision.

Row #	Decision	PSet #	Patient ID	Patient Name	Gender	Race	Date of Birth	Age	MRNs
1	Yes	1	100000001	xxxxxx, xxxxxx	F	black		74	
2	Yes	1	100000002	xxxxxx, xxxxxx	F	white		55	
3	Yes	1	100000003	xxxxxx, xxxxxx	M	asian		73	
4	Und	1	100000004	xxxxxx, xxxxxx	M	black		28	
5	No								
6	Und								
7	No								
8	Und								
9	Und								
10	Yes								
11	Und								
12	No								
13	Und								
14	No								
15	Und								
16	Und								
17	Und								
18	Und								
19	Und								
20	Und								
21	Und								
22	Und								

The interface also shows a "Previous Queries" panel on the left, listing various image sets and patient sets. A smaller window in the foreground shows a detailed view of a patient's data, including a "List" of events (Index, Name, Time) and a large image of a brain scan. The image is labeled "OAS1\_0001\_MR1/RAW/OAS1\_0001\_MR1\_mpr-1\_anon.img". The patient ID is 100000001, and the subject is identified as "OAS1\_0001\_MR1/RAW/OAS1\_0001\_MR1\_mpr-1\_anon.img". The interface includes navigation controls for the image, such as "1", "2", "4", and "edit" buttons, and a status bar at the bottom showing "Subject" and "Increment" information.

# Workbench display of Mutual Inform. Correlations

i2b2 Workbench for Demo Data

Vladimir Valtchinov Status: ● Wiki

Correlation Analysis Cell

Concepts Selection >> Calculate Valid Intervals >> Calculate Correlations >> Results | About

Pairs Results | Compare Graphs

Search And Display

Metric: MIC

TYPE 1 or TYPE 2 is: (shows only pairs with this type)

Threshold: (sets threshold value on selected metric)

Top: (most correlated pairs on selected metric)

Calculation method: NORMAL

Show: All results

type 2	Description 2	MIC	Pearson	Vect
97_s_at	29100	0.8636	0.8831	1390
00_s_at	28971	0.8581	0.8891	1390
96_s_at	84060	0.795	0.765	1390
00_s_at	80233	0.7393	0.8119	1390
97_s_at	29100	0.6629	0.7993	1390
96_s_at	84060	0.6348	0.5554	1390
97_s_at	29100	0.6099	0.5387	1390
00_s_at	28971	0.5928	0.5447	1390
00_s_at	28971	0.5615	0.5783	1390
02_s_at	9214	0.5565	0.9631	1390

yED Graph Options

Show metric

Negative metric colour

221597\_s\_at

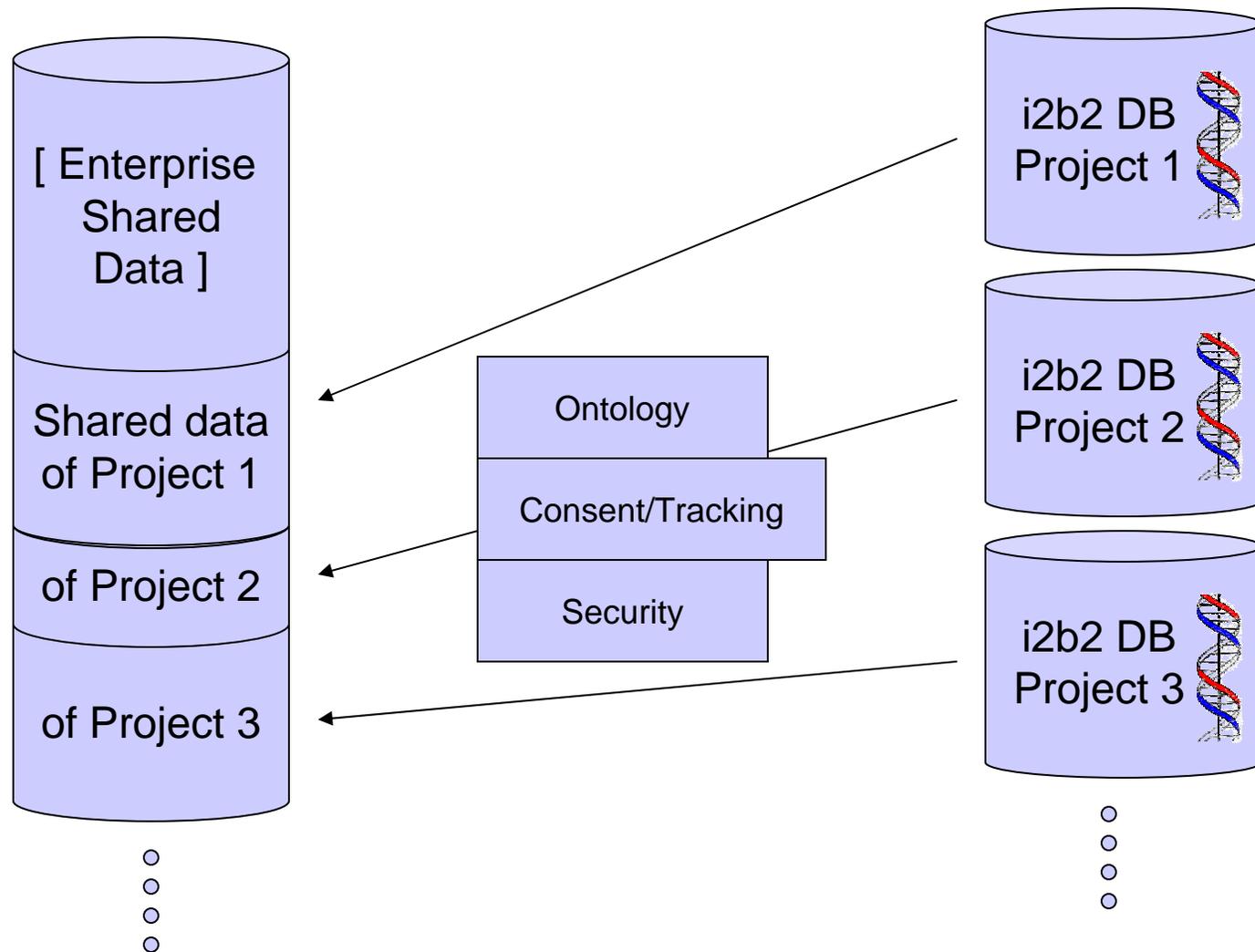
Overlap in time: Scatter plot

Row No.	Vector 1	Vector 2
1	392.10287	1089.9792
2	737.32404	1621.6381
3	586.41644	863.0268
4	589.6359	1308.3345
5	444.2756	1392.4344
6	790.24536	1030.8126
7	525.5525	890.79297
8	572.63	1724.3481
9	545.0431	1450.0469
10	575.0798	1538.5139
11	646.5513	1274.1931
12	301.8552	1118.3007
13	536.53143	1036.4288
14	635.1086	1126.2897
15	898.6878	1378.0665
16	284.69394	1368.532

Patient Set: All

Close

# Project data can be added back to Enterprise Repository

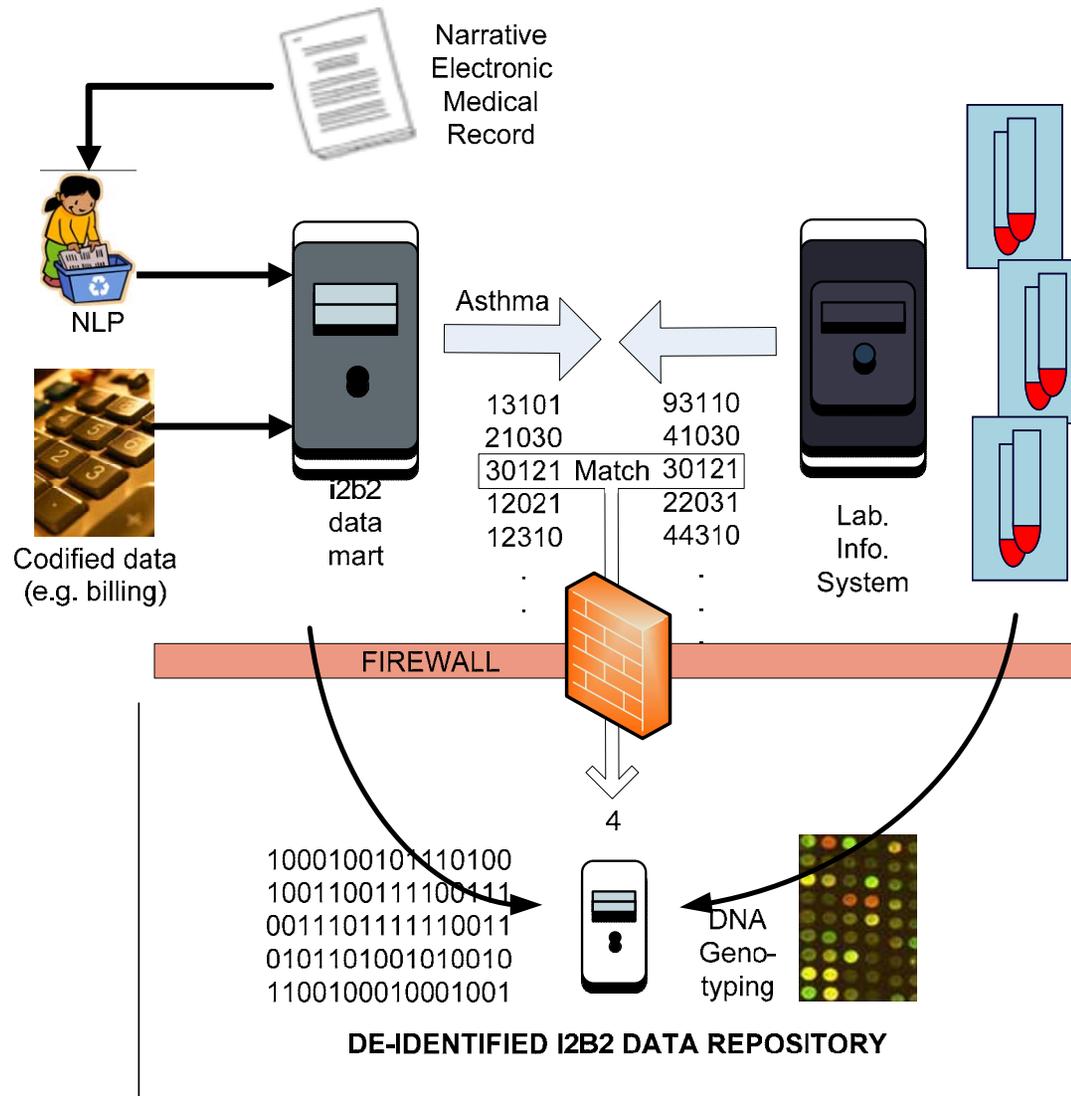




## High Throughput Method for supporting Research

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the BETR/Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
  - Lynn Bry
  - Natalie Boutin

# Genotype samples and compare to controls



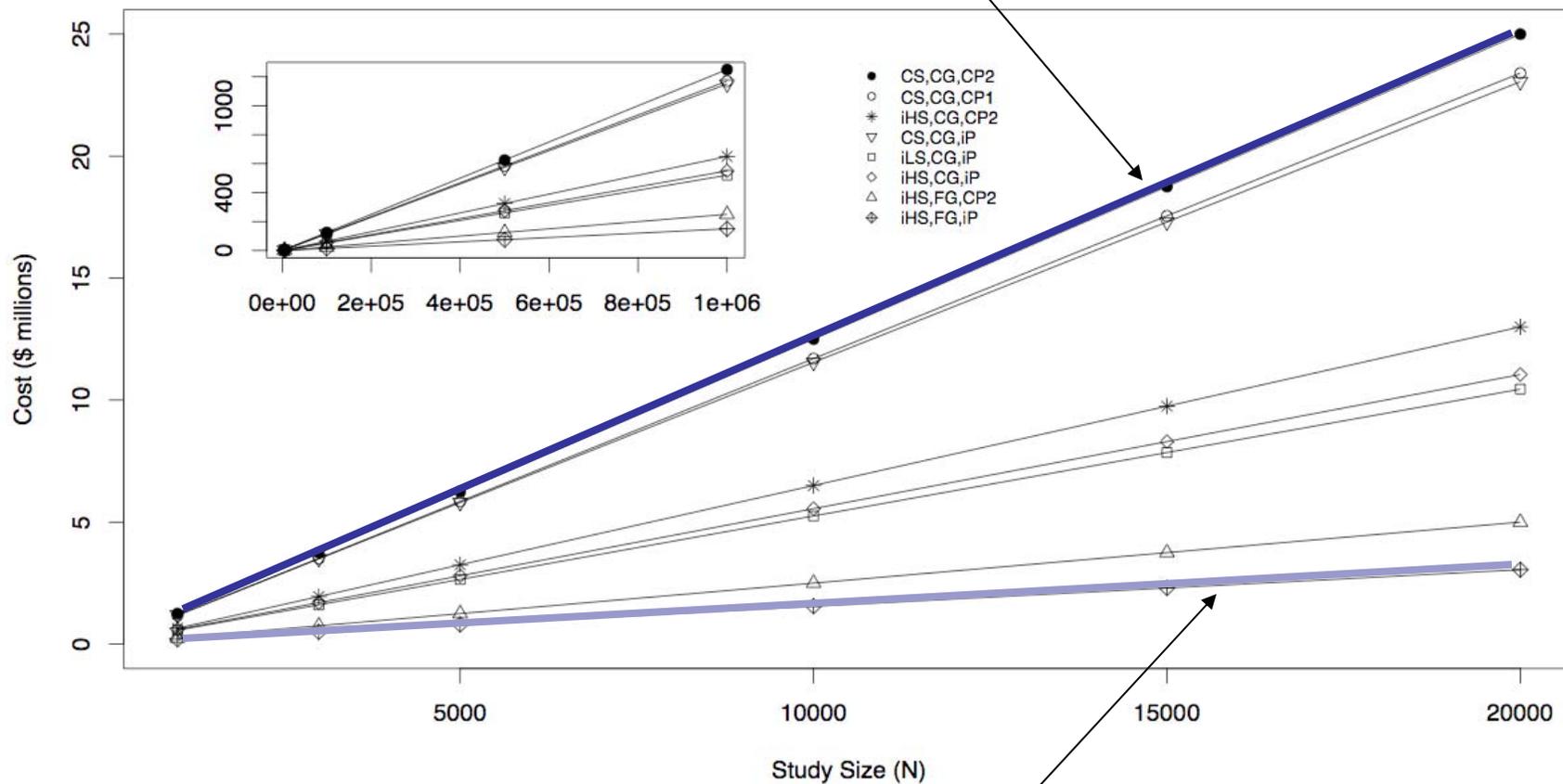
## Cost and time benefit of Instrumenting with Sample Collection for Modest-size Study with 10,000 subjects (cases + controls)

Old vs. New	Cost (\$)	Time
1 chart review per patient (CP1)	\$20	15 minutes/subject
<b>High-throughput phenotyping (iP) through RPDR and i2b2</b>	<b>\$50K Total</b>	<b>1 month total (conservative high estimate)</b>
Sample acquisition through primary care provider (CP)	\$650	3-5 subjects/week <sup>1</sup>
<b>High-throughput sample acquisition through RPDR and BETR/Crimson.</b>	<b>\$20</b>	<b>50-200 subjects /week<sup>2</sup></b>

= \$6.7 million/study vs. \$250 thousand/study

# Escalating cost and time benefit of Instrumenting with Sample Collection

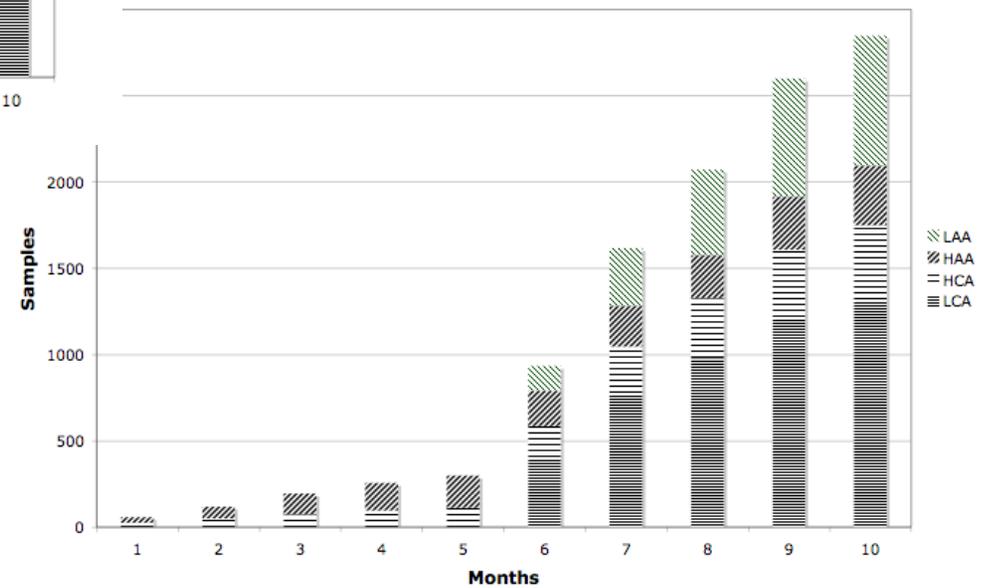
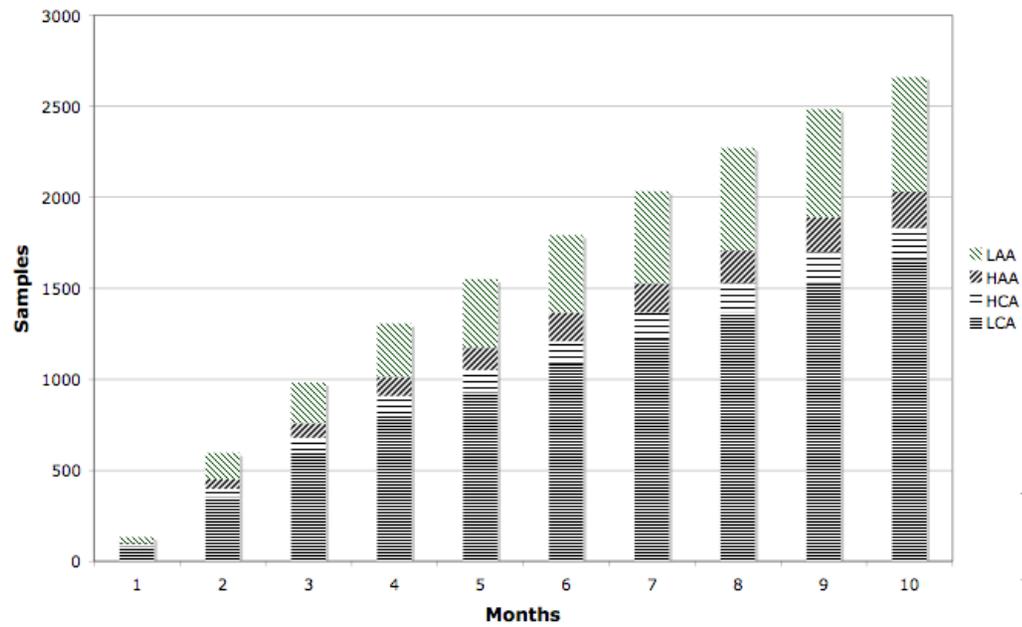
Previous model for collecting specimens



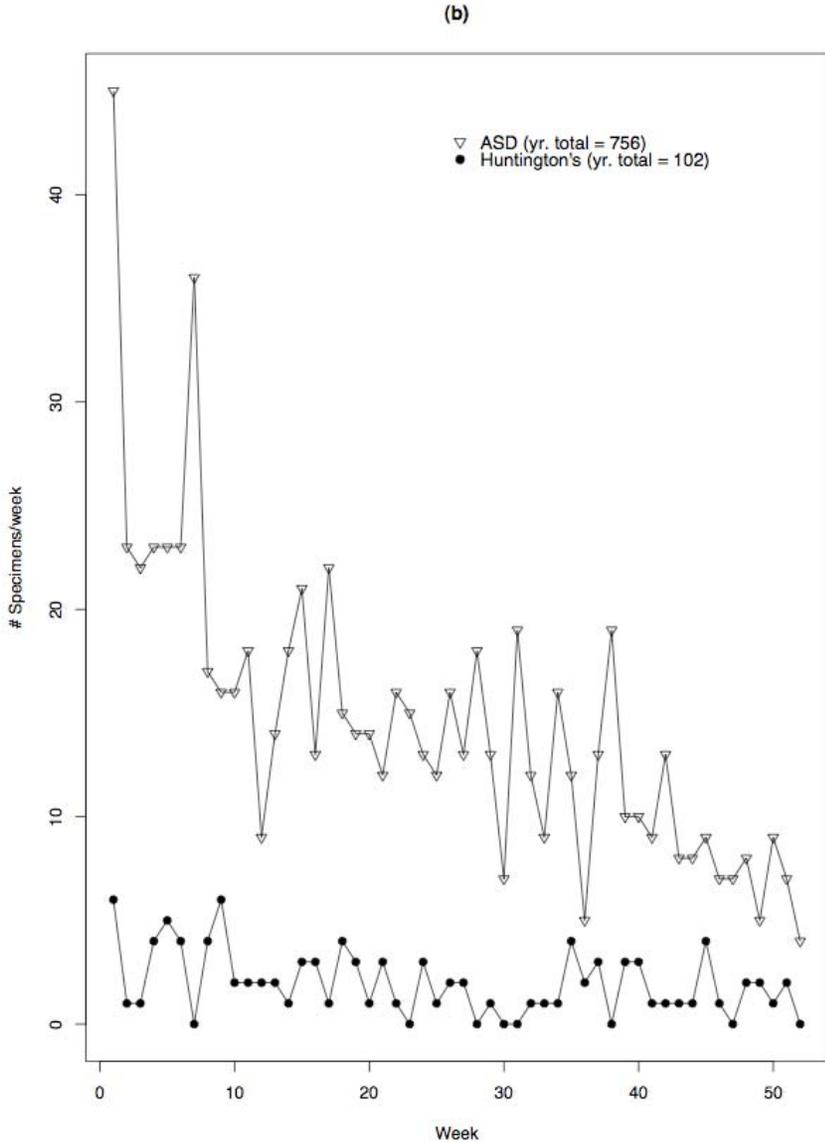
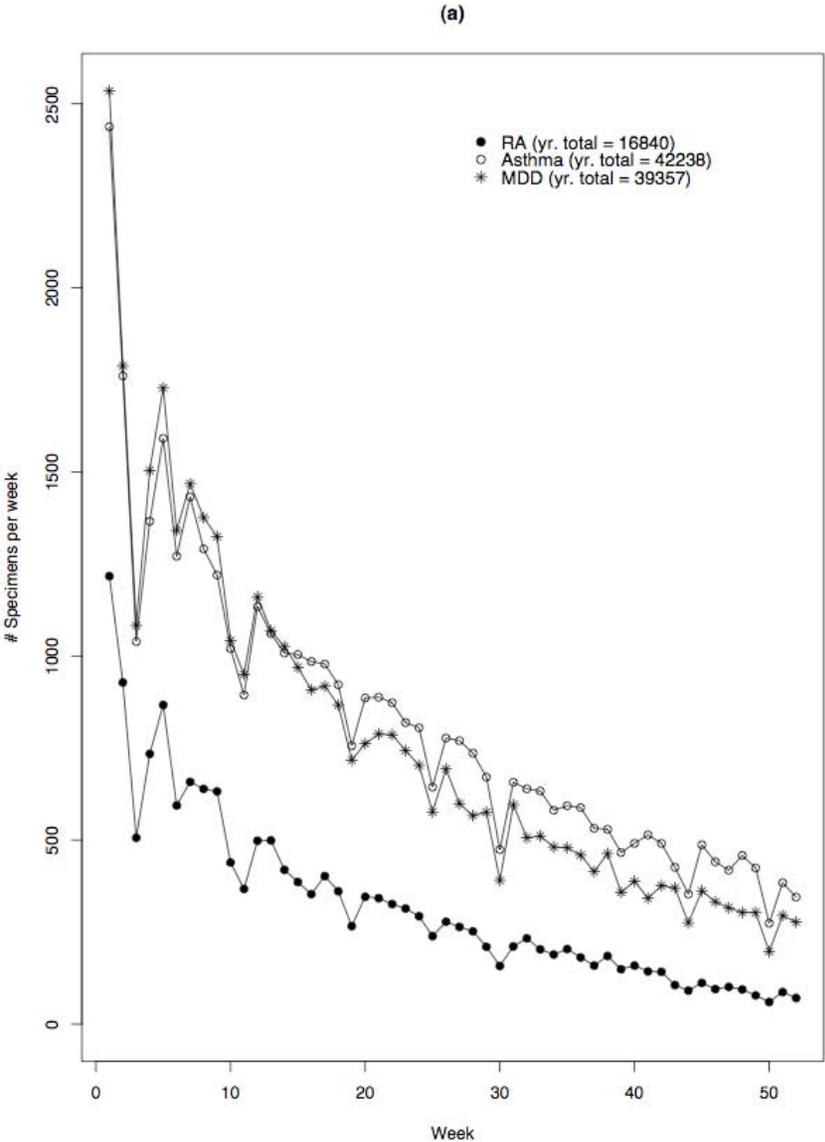
New model for collecting specimens

# Meeting Expectations

(a)



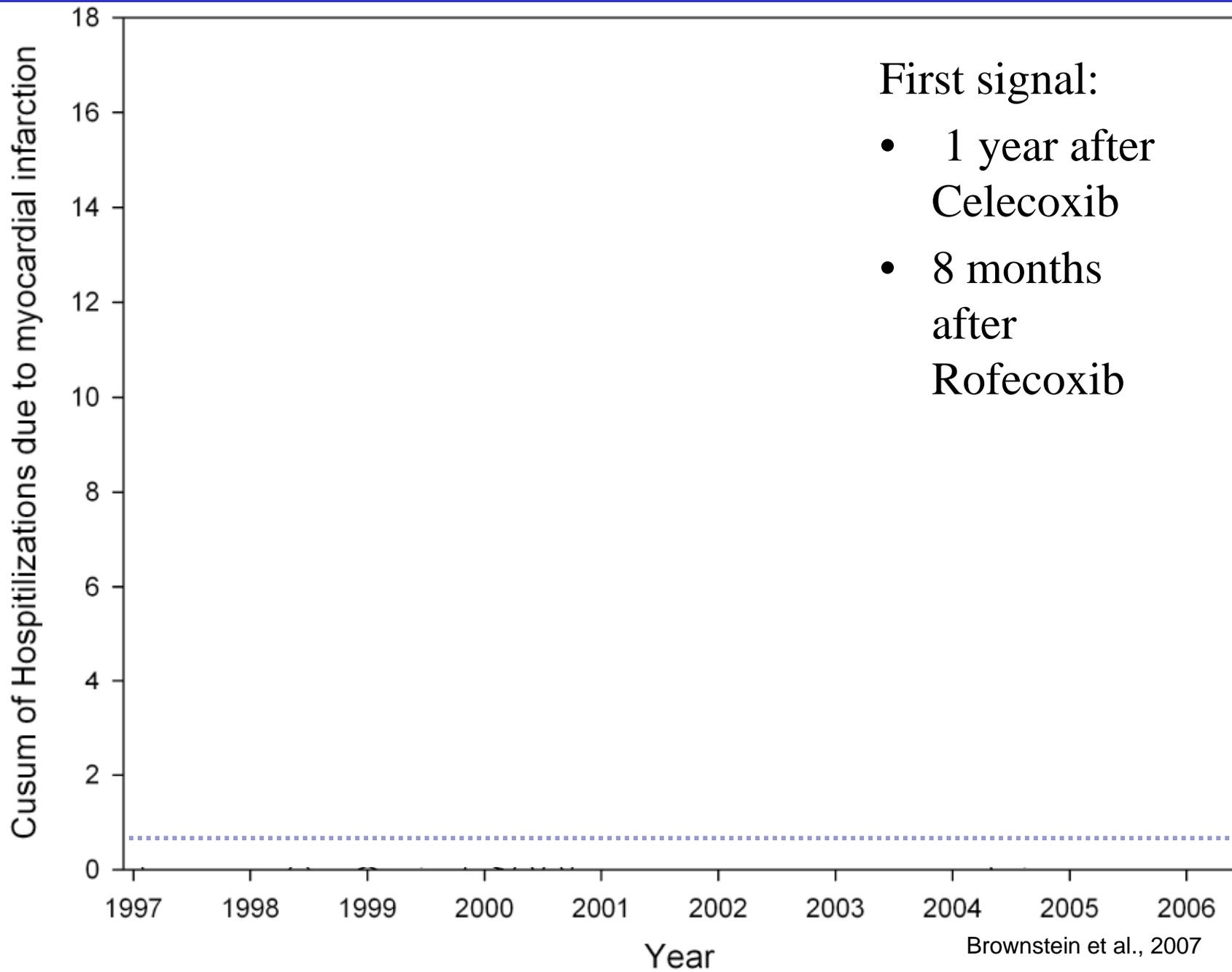
# Accrual Rates



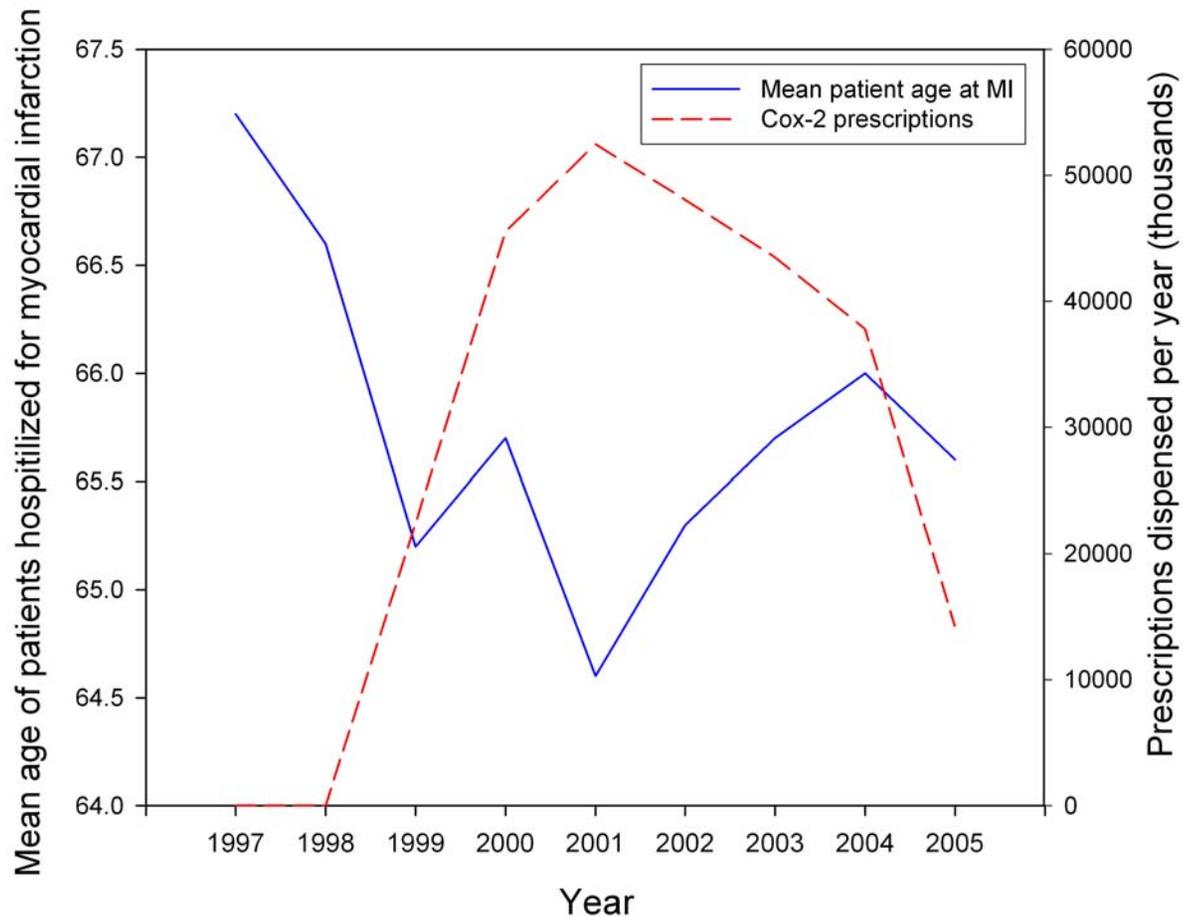


# High Throughput Method for supporting Research

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research
  - John Brownstein
  - Judy Colecchi



# Effect on patient age



- Negative association between mean age at MI and prescription volume
- Spearman correlation  $-0.67$ ,  $P < 0.05$

# Centrally supported Automated Discovery Projects

## Clinical trials performed in-silico

- Performing an observational, phase IV study is an expensive and complex process that can be potentially modeled in a retrospective database using groups of patients available in the large amounts of highly organized medical data.
- Fundamental problems complicate this approach:
  - Patients drift in and out of the system. Sophisticated statistical models using adequate control populations are necessary to compensate.
  - Confounding variables are not found in the database. Sophisticated natural language processing is needed to extract the confounders from textual reports to allow these confounders to be controlled.
  - Missing data disrupts typical statistical approaches

# Odds Ratios for Diseases expressed in comparing Rosiglitazone vs. Pioglitazone

The screenshot displays the i2b2 Workbench interface for a 'Demo Group'. The main window is titled 'Query Tool' and shows a query named 'Thiazolidinediones'. The query tool includes a 'Query Name' field, 'Reset Groups' button, and three columns for 'Group 1', 'Group 2', and 'Group 3'. Each column has fields for 'Dates', 'Occurs > 0x', and 'Exclude'. Below the columns are instructions: 'The terms of this group are joined then intersected with other groups', 'Drag terms from Navigate, Find and Workplace into this group', and 'Drag terms from N Find and Workplace into this group'. A 'Run Query' button and 'Patient(s) returned:' field are at the bottom.

The 'Workplace' pane on the right shows a tree view of the data source, including folders like 'demo', 'Cardiac Tests', 'Cardiovascular agents', 'Definitions (Queries)', 'Definitions (Query Groups)', 'Demogra-Diagnos@02:31:36 [1-11-2009]', 'Diabetes mellit@22:57:00 [1-15-2009]', 'Diagnoses', 'IRON (LOINC:2498-4)', 'Iron (G-ALT (5G@09:08:27 [12-15-2009]', 'Observations', 'PATIENT:1000000025', 'PATIENTSET\_8971\_Ischemic', and 'Patient Sets'.

The 'Pharmacovigilance Analysis' pane is active, showing a treemap visualization. The title is 'Analysis Title: Avandia vs. Pioglitazone Calculation for SNMO on 3-04-2009'. The treemap shows a hierarchical structure of diseases, with 'Circulatory system (390-459)' being the largest category. Other categories include 'Symptoms, signs, and ill-defined conditions (780-849)', 'Endocrine disorders (240-299)', 'Metabolic and immun...', 'Gastrointestinal system (530-579)', 'Respiratory system (460-519)', 'Musculoskeletal and connective tissue (710-729)', 'Digestive system (530-579)', 'Neurologic (840-899)', 'Skin (860-919)', 'Hematologic (550-609)', 'Genitourinary system (550-625)', 'Nutritional deficiencies (260-285)', 'Infectious and parasitic disease (040-099)', 'Mental Disorders (290-319)', 'Neoplasms (140-239)', 'Injury and poisoning (880-999)', 'zz E-codes', 'zz Y-codes', and 'zz V-codes'. The treemap uses a color scale from green to red to represent the magnitude of the odds ratios. The 'Strategy' is set to 'Squarified' and the 'Color Provider' is 'HCB linear'. The 'TREX params' are 'TotalPatients'.

The 'Previous Queries' pane at the bottom left shows a list of recent queries, including 'Female-Asthma@04:54:07 [02-16-2009]', 'Female-Asthma [02-16-2009] [demo]', 'Immunology@01:38:28 [02-15-2009] [d]', 'Demographics@18:15:23 [02-14-2009] [d]', '0-9 yea-Male@22:21:20 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', '0-9 yea-Male@22:21:20 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', '0-9 yea-Male@22:21:20 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', '0-9 yea-Male@22:21:20 [02-14-2009] [c]', 'Demographics@18:15:23 [02-14-2009] [c]', 'Demographics@18:15:23 [02-13-2009] [c]', '0-9 yea-Male@22:21:20 [02-13-2009] [c]', 'Demographics@18:15:23 [02-13-2009] [c]', and '0-9 yea-Male@22:21:20 [02-13-2009] [c]'. At the bottom of the treemap, there are buttons for 'Open File ..', 'Save File ..', and 'Open in Excel'.

## Take-away points from i2b2

- 1) Enable Enterprise use of Patients for Research
- 2) Power comes from Numbers of Patients Recruited
- 3) Extensible Architecture for Developers
- 4) Enable Scientist Workflow
- 5) Enable Teamwork between Informaticians and Researchers
- 6) Communicate through Visualizations
- 7) Enable Natural Language Processing
- 8) High throughput Tissue acquisition for Genomic Research
- 9) Enable Health Surveillance
- 10) Enable Data Sharing

# References

- Murphy, S. N., Morgan, M. M., Barnett, G. O., Chueh, H. C. Optimizing Healthcare Research Data Warehouse Design through Past COSTAR Query Analysis. Proc AMIA Symp. 1999;892-
- Kimball, R. The Data Warehousing Toolkit. New York: John Wiley, 1997.
- Johnson SB, Paut T., Khenima A.; Generic database design for patient management information. Proc AMIA Annu Fall Symp 1997:22-26.
- Nadkarni, P.M., Brandt, C. Data Extraction and Ad Hoc Query of an Entity-Attribute-Value Database. J Am Med Inform Assoc. 1998; 5:511-7.
- Murphy, S.N., Chueh, H. A Security Architecture for Query Tools Used to Access Large Biomedical Databases. AMIA, Fall Symp. 2002, pages 552-556.
- Murphy, S.N., Gainer, V.S., Chueh, H. A Visual Interface Designed for Novice Users to find Research Patient Cohorts in a Large Biomedical Database. AMIA, Fall Symp. 2003: 489-493.
- Murphy, SN; Mendis, M, K. Hackett; et al.; Architecture of the Open-source Clinical Research Chart from Informatics for Integrating Biology and the Bedside; AMIA, Fall Symp. 2007.
- Qing T Zeng, Sergey Goryachev, Scott Weiss, et. al., Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system, BMC Medical Informatics and Decision Making, 2006.
- Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ. 2009 Jan 27;338:b81