

i2b2 Clinical Research Chart

Shawn Murphy MD, Ph.D.

Griffin Weber MD, Ph.D.

Michael Mendis

Vivian Gainer MS

Lori Phillips MS

Rajesh Kuttan

Wensong Pan MS

Henry Chueh MD

Susanne Churchill Ph.D.

John Glaser Ph.D.

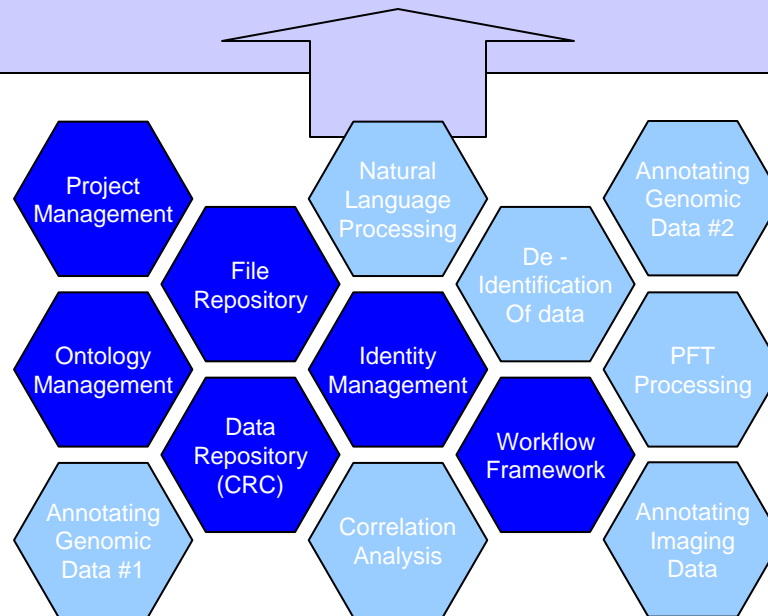
Isaac Kohane MD, Ph.D.



The National Center for Biomedical Computing entitled Informatics for Integrating Biology and the Bedside (i2b2) Clinical Research Chart, what is it?

- Explicitly organized and transformed person-oriented clinical data optimized for clinical genomics research
- Formed as a collection of loosely coupled interoperable web services provided by i2b2 Cells
- An architecture that allows different studies to come together seamlessly
- An integration of clinical data, trials data, genotypic data, and knowledge annotation

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

Enterprise web client

The screenshot displays the i2b2 Query & Analysis Tool interface within a Windows Internet Explorer browser window. The interface is organized into several main sections:

- Navigation and Menu:** At the top, there are tabs for "Find Patients", "Data Analysis", "Help", and "Logout". Below these are "Navigate Terms" and "Find Terms" buttons.
- Ontology Tree:** A hierarchical tree on the left lists various medical categories such as Demographics, Diagnoses, Laboratory Tests, Chemistry, Coagulation, Hematology, Immunology, Pulmonary Function Reports, Urinalysis, Medications, and Procedures. A yellow callout box says "drag an item from here".
- Query Tool:** The central area is titled "Query Tool" and shows a "Query Name" field with the text "45-54-Circu-Male@18:59:32". It features three columns for building a query:
 - Group 1:** Contains "45-54 years old" and "55-64 years old". A green box below it says "one or more of these".
 - Group 2:** Contains "Circulatory system". A green box below it says "one or more of these".
 - Group 3:** Contains "Male". A red box below it says "none of these".Blue "AND" boxes connect the groups. At the bottom of this section are buttons for "Run Query", "New Query", and "New Group", along with a "3 Groups" indicator.
- Previous Queries:** A list on the bottom left shows recent queries with details like "45-54-Circu-Male@18:59:32 [2008-04-05] [demo]".
- Query Status:** A box on the bottom right displays the execution progress: "Executing query...", "Elapsed time (seconds): 0.9", "Query Finished...", and "Matching patients: 5".

The browser's taskbar at the bottom shows the "Internet" icon and a 100% zoom level.

i2b2 Query & Analysis Tool

Find Patients

Analysis Tools

Logout

Plugin Viewer

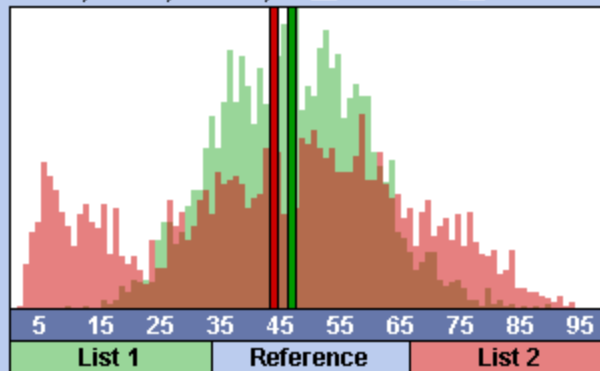
Two List Demographics Comparison

Number of patients in Morbid Obesity = 2829

Number of patients in Allergic Asthma = 2407

Age Comparison

1y 5y 10y means reference



Reference list includes all patients in the database.

Done

Internet

i2b2 Query & Analysis Tool

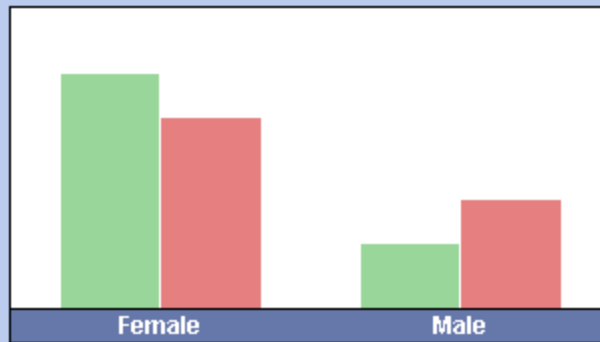
Find Patients

Analysis Tools

Logout

Plugin Viewer

Sex Race Marital Deceased

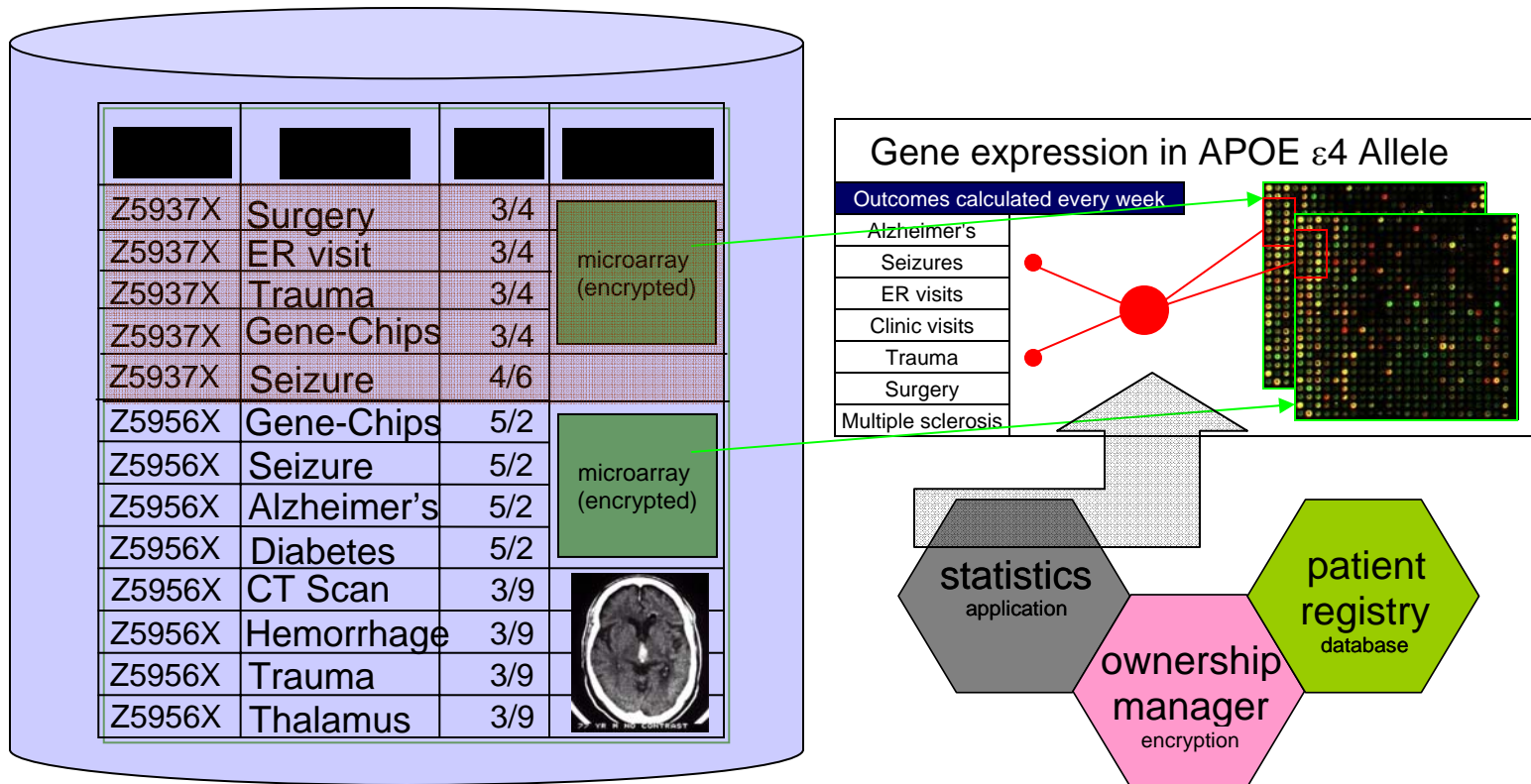


(ChiSq values are gray if there are not enough patients.)

	Demographic	% List1	% List2	Odds Ratio	CI 95% OR	Relative Risk	CI 95% RR	Pearson ChiSq	Yate's ChiSq
1	Sex = F	78.15	63.48	2.06	(1.82,2.32)	1.23	(1.19,1.28)	137.1	136.4
2	Sex = M	21.85	36.52	0.49	(0.43,0.55)	0.6	(0.55,0.65)	137.1	136.4
3	Race = @	3.5	14.13	0.22	(0.17,0.28)	0.25	(0.2,0.31)	191.2	189.8
4	Race = Amer. Indian	0.18	0.12	1.42	(0.34,5.94)	1.42	(0.34,5.93)	0.2	0

Relevance Networks – unsupervised learning techniques

- Associates phenotypes and genotypes to generate novel hypothesis.



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	Vector
97_s_at	29100	0.8636	0.8831	1390
D0_s_at	28971	0.8581	0.8891	1390
96_s_at	84060	0.795	0.765	1390
D0_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
D0_s_at	28971	0.5928	0.5447	1390
D0_s_at	28971	0.5615	0.5783	1390
D2_s_at	9214	0.5565	0.9631	1390

Graph Options

Show metric

Negative metric colour ■

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

Navigate Terms

- Ontology
 - Demographics
 - Diagnoses
 - Expression Profiles Data
 - Affymetrix HG-U133
 - 221591_s_at (54478)
 - 221592_at (11138)
 - 221593_s_at (6160)
 - 221594_at (84060)
 - 221595_at (84060)
 - 221596_s_at (84060)
 - 221597_s_at (29100)
 - 221598_s_at (9442)
 - 221599_s_at (28971)
 - 221600_s_at (28971)
 - 221601_s_at (9214)
 - 221602_s_at (9214)
 - 221603_at (9409)
 - 221604_s_at (9409)
 - 221605_s_at (51268)
 - 221606_s_at (79366)

Previous Queries

- 221596_s_at (84060) [03-20-2008] [v]
- 222320_at (7957@03:28:43 [03-19-2008] [v]
- Demographics@03:28:23 [03-19-2008] [vv05
- 222320_at (7957@10:34:45 [03-18-2008] [v]
- 222320_at (7957@10:34:37 [03-18-2008] [v]
- Demographics@08:19:15 [03-18-2008] [vv05
- Demographics@08:17:12 [03-18-2008] [vv05
- Demographics@08:16:15 [03-18-2008] [vv05

Correlation Analysis Cell

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

Parameters

Valid intervals that are not set manually will use the precalculated method:

Treat all diagnoses as time-independent

Calculation method:

MIC Bins:

Calculate Stop

Patient Set: Max Number of Patients Includ:

Cluster Utility Cell

Current Cluster connection: **RCCLU.partners.org**

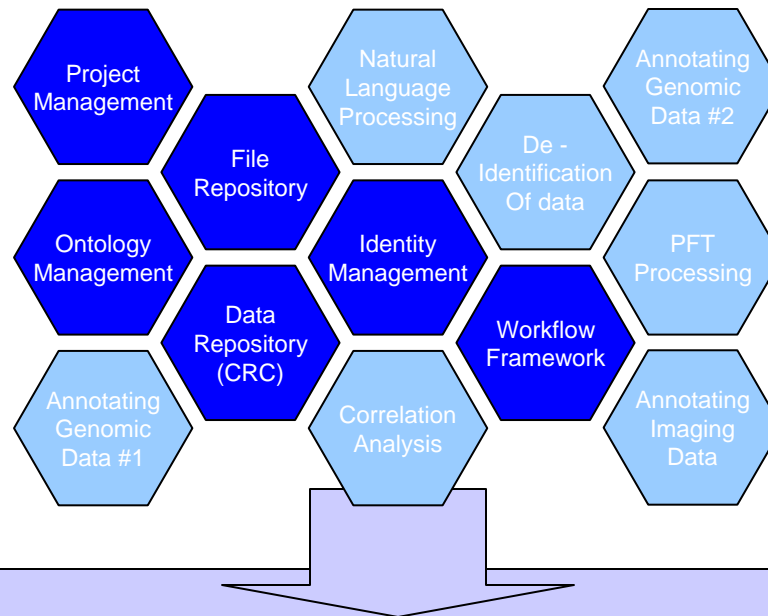
Command:

PAIR_ID	TYPE_1	TYPE_2	MIC	PEARSON	NPERM	VECTORCNT
1	221591_s_at	221592_at	0.3352	0.3291	0	1390
2	221591_s_at	221593_s_at	0.4721	0.5422	0	1390
3	221591_s_at	221594_at	0.2533	0.299	0	1390
4	221591_s_at	221595_at	0.2768	0.329	0	1390
5	221591_s_at	221596_s_at	0.4408	0.32	0	1390
6	221591_s_at	221597_s_at	0.6344	0.6279	0	1390
7	221591_s_at	221598_s_at	0.3379	0.6047	0	1390
8	221592_at	221593_s_at	0.2826	0.2633	0	1390
9	221592_at	221594_at	0.3572	0.4299	0	1390
10	221592_at	221595_at	0.2837	0.3301	0	1390
11	221592_at	221596_s_at	0.3699	0.3145	0	1390
12	221592_at	221597_s_at	0.3314	0.3373	0	1390

Last Command Results

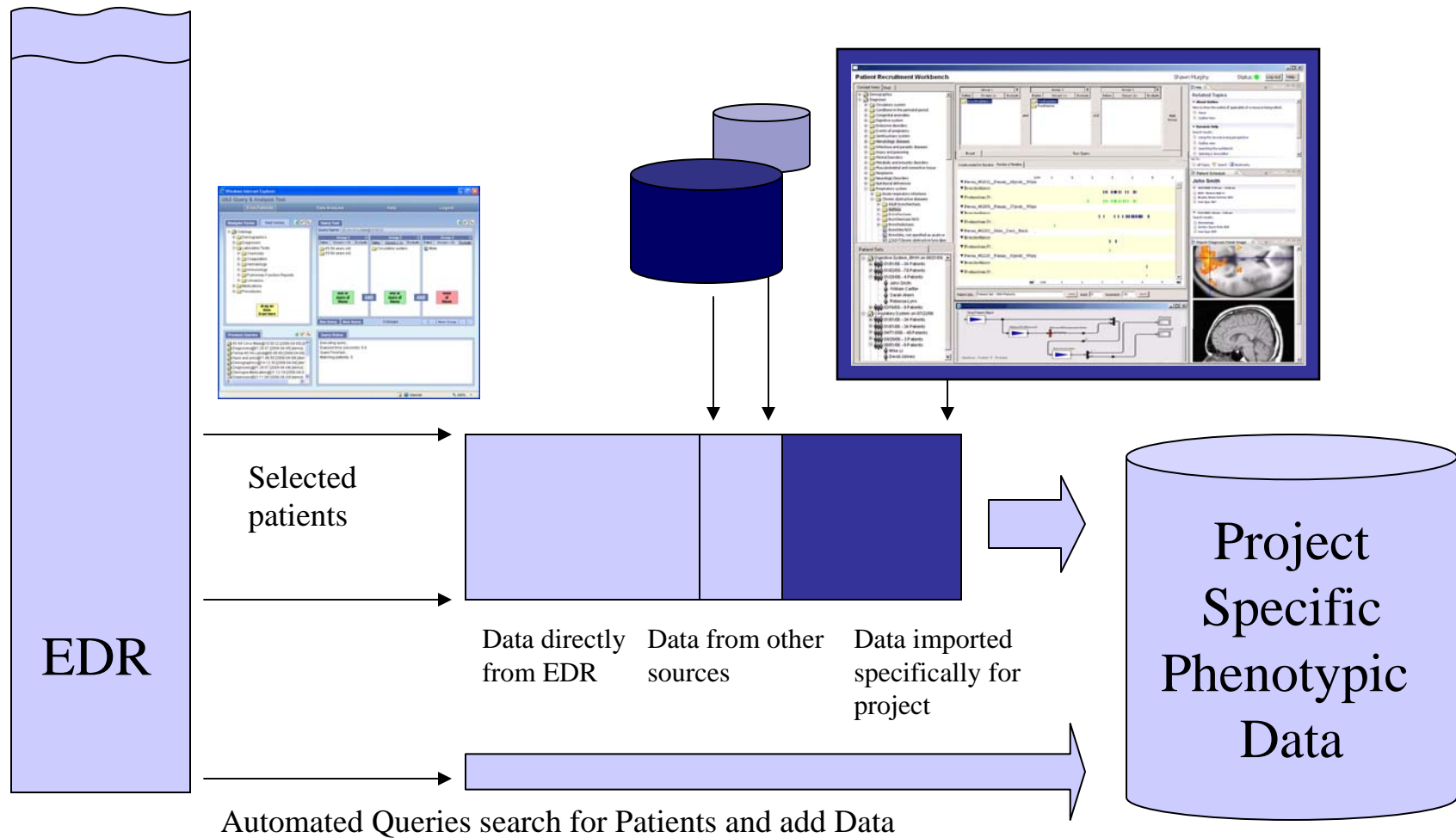
PAIR_ID	TYPE_1	TYPE_2	MIC	PEARSON	NPERM	VECTORCNT
1	221591_s_at	221592_at	0.3352	0.3291	0	1390 2008-04-07 15:12:14.012
2	221591_s_at	221593_s_at	0.4721	0.5422	0	1390 2008-04-07 15:12:14.012
3	221591_s_at	221594_at	0.2533	0.299	0	1390 2008-04-07 15:12:14.012
4	221591_s_at	221595_at	0.2768	0.329	0	1390 2008-04-07 15:12:14.012
5	221591_s_at	221596_s_at	0.4408	0.32	0	1390 2008-04-07 15:12:14.012
6	221591_s_at	221597_s_at	0.6344	0.6279	0	1390 2008-04-07 15:12:14.012
7	221591_s_at	221598_s_at	0.3379	0.6047	0	1390 2008-04-07 15:12:14.012
8	221592_at	221593_s_at	0.2826	0.2633	0	1390 2008-04-07 15:12:14.012
9	221592_at	221594_at	0.3572	0.4299	0	1390 2008-04-07 15:12:14.012
10	221592_at	221595_at	0.2837	0.3301	0	1390 2008-04-07 15:12:14.012
11	221592_at	221596_s_at	0.3699	0.3145	0	1390 2008-04-07 15:12:14.012
12	221592_at	221597_s_at	0.3314	0.3373	0	1390 2008-04-07 15:12:14.012
13	221592_at	221598_s_at	0.1343	0.2281	0	1390 2008-04-07 15:12:14.012

- Repurpose medical record information for research studies
 - I2b2 Workbench
 - Natural language processing
- Enable genomic studies
 - Tissue/blood selection
 - Genetic data integration



Use of medical record data in clinical studies
focused upon genomics and pharmacology

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



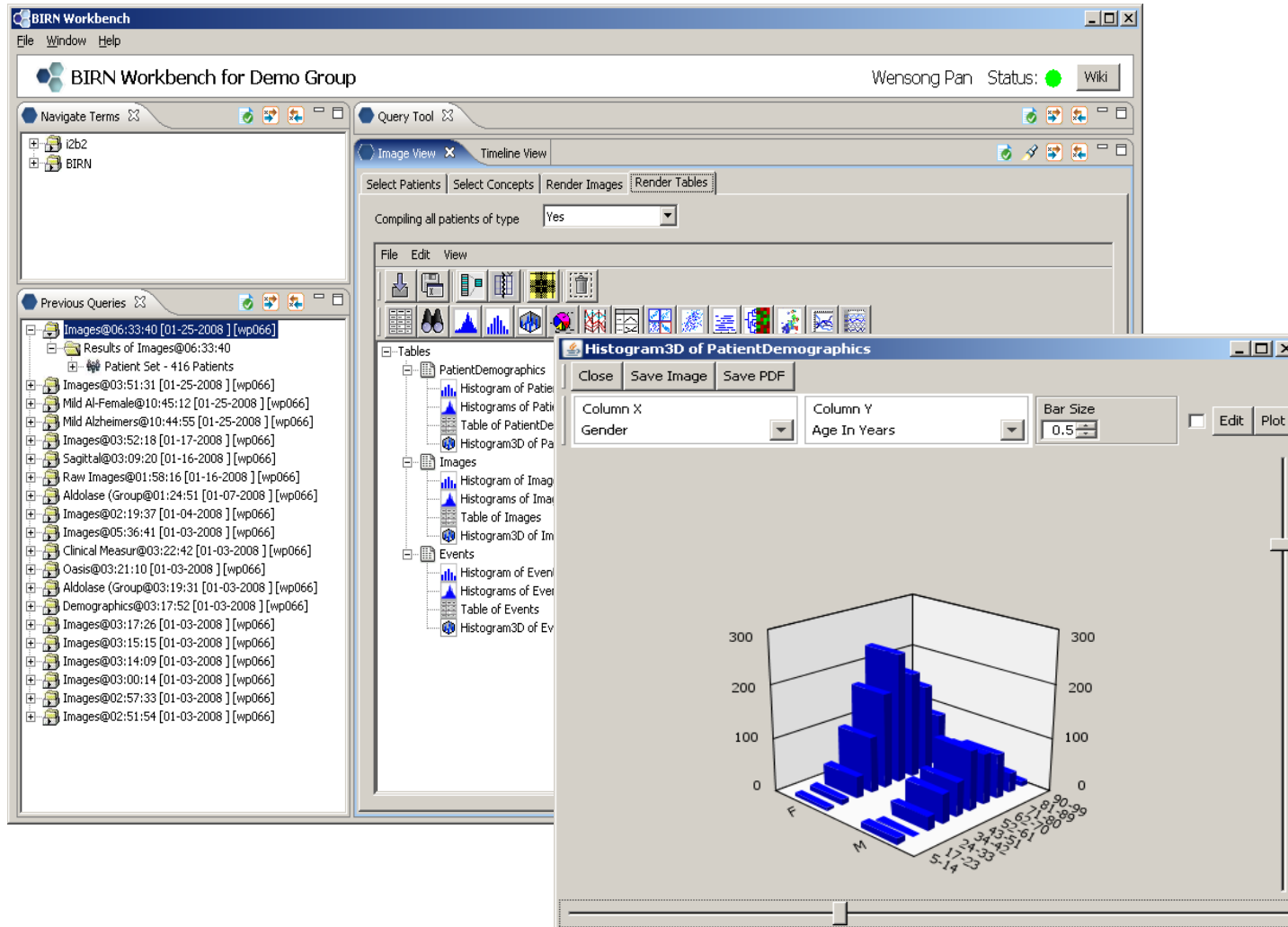
Patient population medical record review

The screenshot displays the BIRN Workbench interface for a demo group. The main window shows a 'Query Tool' with a table of patient data. The table has columns for Row #, Decision, PSet #, Patient ID, Patient Name, Gender, Race, Date of Birth, Age, and MRNs. The data is as follows:

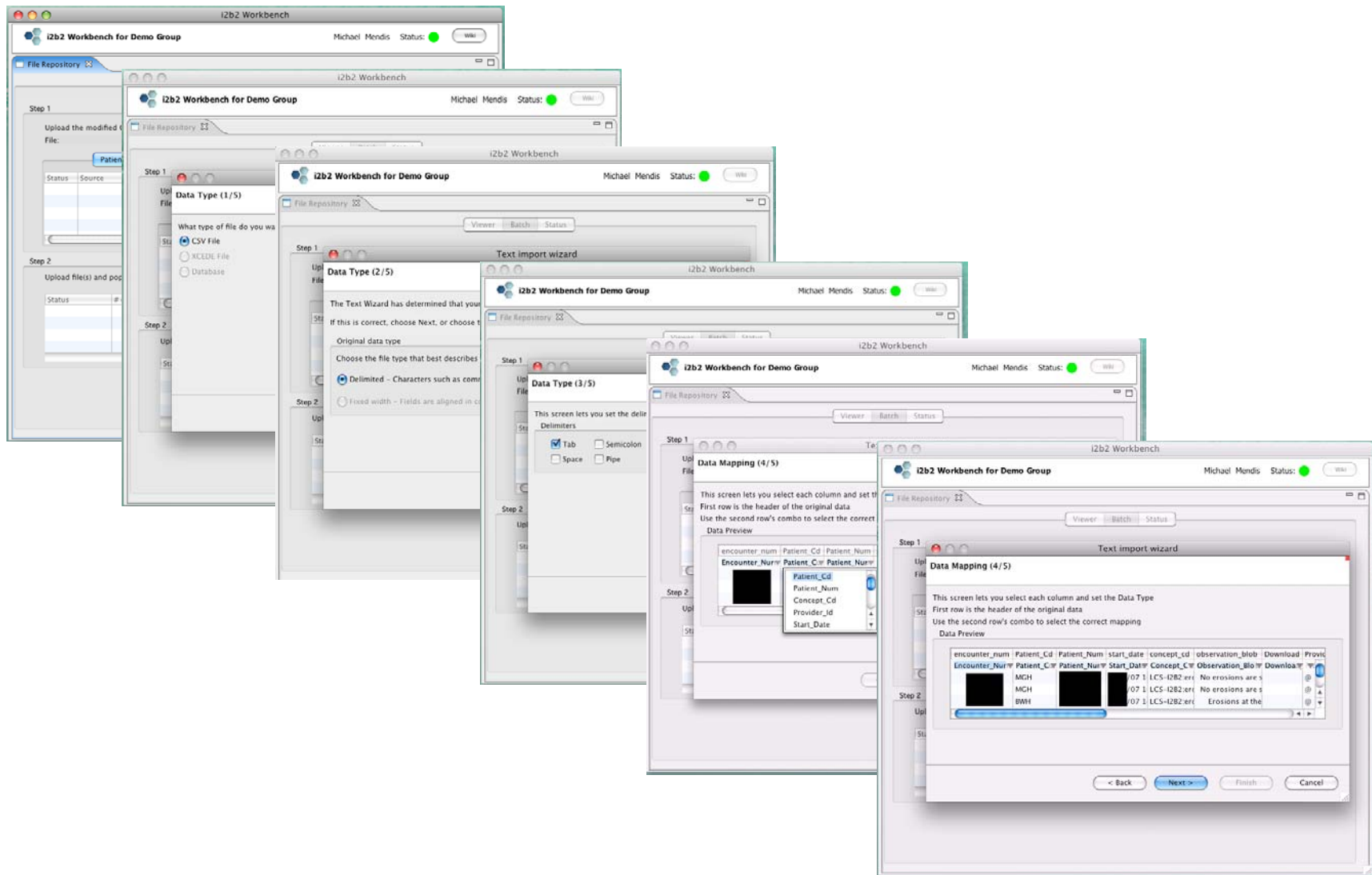
Row #	Decision	PSet #	Patient ID	Patient Name	Gender	Race	Date of Birth	Age	MRNs
1	Yes	1	100000001	xxxxxx, xxxxxx	F	black		74	
2	Yes	1	100000002	xxxxxx, xxxxxx	F	white		55	
3	Yes	1	100000003	xxxxxx, xxxxxx	M	asian		73	
4	UnD	1	100000004	xxxxxx, xxxxxx	M	black		28	
5	No								
6	UnD								
7	No								
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' list on the left, with the selected query being 'Images@06:33:40 [01-25-2008] [wp066]'. A detailed view of a patient's medical record is shown on the right, including a list of events and a sagittal MRI scan. The patient ID is 100000001. The event list shows four sagittal MRI scans on 2003-05-03. The MRI scan is labeled 'OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-1_anon.img'.

Exploration of targeted population



Import of batch data sets through i2b2 Workbench

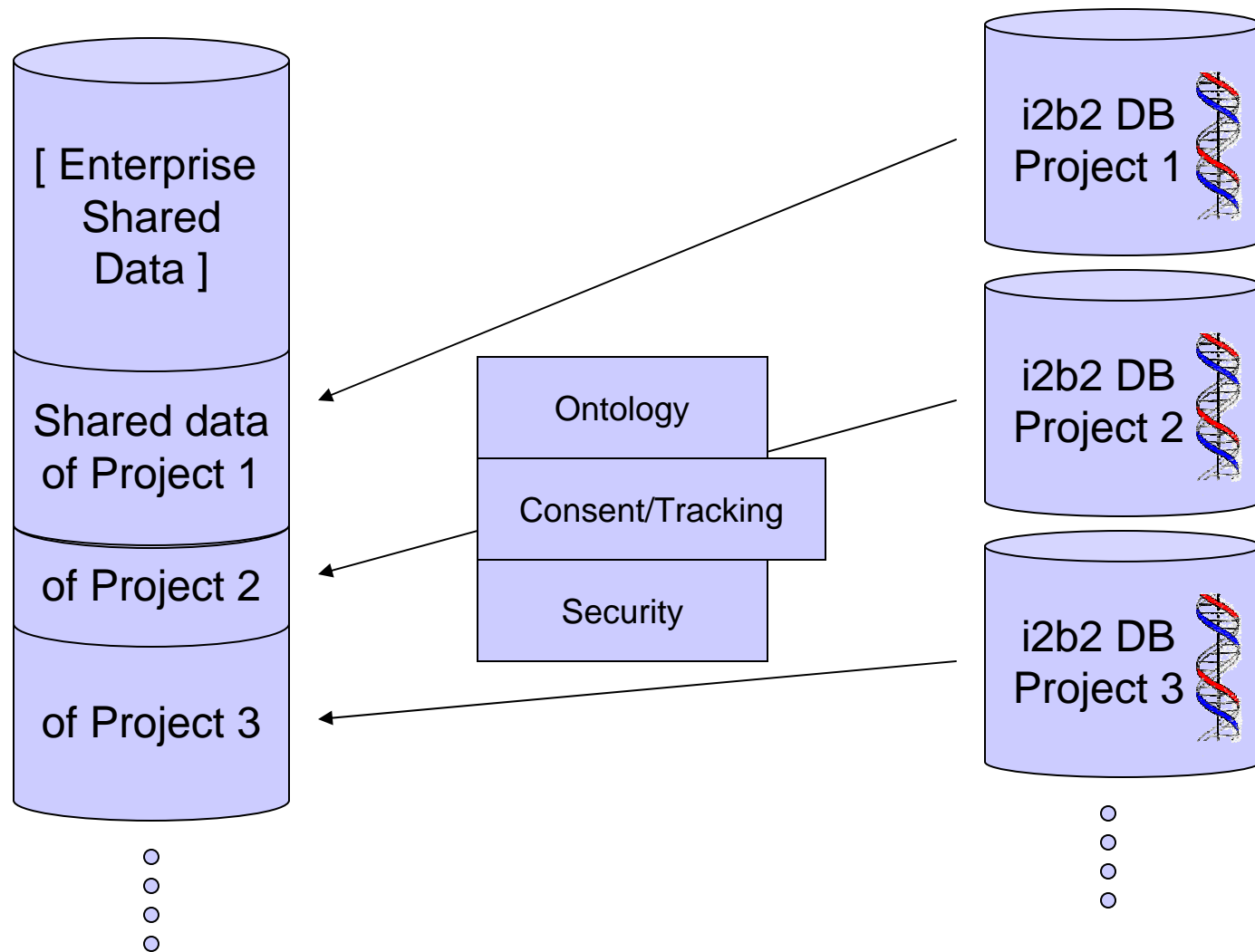


Export of batch data sets for specific applications

The screenshot displays the BIRN Workbench interface. The main window, titled 'Table of Images', contains a table with the following columns: Patient ID, Concept Name, Start Date, End Date, and Text Value. The table lists 40 rows of data, each representing a patient's image set. The 'Text Value' column contains file paths such as 'OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-1_anon.img'. Below the main table, a smaller window titled 'ImageList - Worksheet' provides a detailed view of the data, including columns for Patient ID, Concept Name, Start Date, End Date, Modifier Name, Value Type Code, Text Value, and Numeric Value. A blue arrow points from the 'Table of Images' window to the 'ImageList - Worksheet' window, indicating the export process.

Patient ID	Concept Name	Start Date	End Date	Text Value
1000000001	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-1_anon.img
1000000001	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-2_anon.img
1000000001	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-3_anon.img
1000000001	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0001_MR1/RAW/OAS1_0001_MR1_mpr-4_anon.img
1000000002	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0002_MR1/RAW/OAS1_0002_MR1_mpr-1_anon.img
1000000002	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0002_MR1/RAW/OAS1_0002_MR1_mpr-2_anon.img
1000000002	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0002_MR1/RAW/OAS1_0002_MR1_mpr-3_anon.img
1000000002	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0002_MR1/RAW/OAS1_0002_MR1_mpr-4_anon.img
1000000003	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0003_MR1/RAW/OAS1_0003_MR1_mpr-1_anon.img
1000000003	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0003_MR1/RAW/OAS1_0003_MR1_mpr-2_anon.img
1000000003	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0003_MR1/RAW/OAS1_0003_MR1_mpr-3_anon.img
1000000003	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0003_MR1/RAW/OAS1_0003_MR1_mpr-4_anon.img
1000000004	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0004_MR1/RAW/OAS1_0004_MR1_mpr-1_anon.img
1000000004	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0004_MR1/RAW/OAS1_0004_MR1_mpr-2_anon.img
1000000004	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0004_MR1/RAW/OAS1_0004_MR1_mpr-3_anon.img
1000000004	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0004_MR1/RAW/OAS1_0004_MR1_mpr-4_anon.img
1000000005	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0005_MR1/RAW/OAS1_0005_MR1_mpr-1_anon.img
1000000005	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0005_MR1/RAW/OAS1_0005_MR1_mpr-2_anon.img
1000000005	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0005_MR1/RAW/OAS1_0005_MR1_mpr-3_anon.img
1000000005	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0005_MR1/RAW/OAS1_0005_MR1_mpr-4_anon.img
1000000006	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0006_MR1/RAW/OAS1_0006_MR1_mpr-1_anon.img
1000000006	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0006_MR1/RAW/OAS1_0006_MR1_mpr-2_anon.img
1000000006	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0006_MR1/RAW/OAS1_0006_MR1_mpr-3_anon.img
1000000006	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0006_MR1/RAW/OAS1_0006_MR1_mpr-4_anon.img
1000000007	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0007_MR1/RAW/OAS1_0007_MR1_mpr-1_anon.img
1000000007	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0007_MR1/RAW/OAS1_0007_MR1_mpr-2_anon.img
1000000007	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0007_MR1/RAW/OAS1_0007_MR1_mpr-3_anon.img
1000000007	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0007_MR1/RAW/OAS1_0007_MR1_mpr-4_anon.img
1000000008	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0008_MR1/RAW/OAS1_0008_MR1_mpr-1_anon.img
1000000008	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0008_MR1/RAW/OAS1_0008_MR1_mpr-2_anon.img
1000000008	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0008_MR1/RAW/OAS1_0008_MR1_mpr-3_anon.img
1000000008	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0008_MR1/RAW/OAS1_0008_MR1_mpr-4_anon.img
1000000009	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0009_MR1/RAW/OAS1_0009_MR1_mpr-1_anon.img
1000000009	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0009_MR1/RAW/OAS1_0009_MR1_mpr-2_anon.img
1000000009	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0009_MR1/RAW/OAS1_0009_MR1_mpr-3_anon.img
1000000009	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0009_MR1/RAW/OAS1_0009_MR1_mpr-4_anon.img
1000000010	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0010_MR1/RAW/OAS1_0010_MR1_mpr-1_anon.img
1000000010	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0010_MR1/RAW/OAS1_0010_MR1_mpr-2_anon.img
1000000010	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0010_MR1/RAW/OAS1_0010_MR1_mpr-3_anon.img
1000000010	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0010_MR1/RAW/OAS1_0010_MR1_mpr-4_anon.img
1000000011	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0011_MR1/RAW/OAS1_0011_MR1_mpr-1_anon.img
1000000011	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0011_MR1/RAW/OAS1_0011_MR1_mpr-2_anon.img
1000000011	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0011_MR1/RAW/OAS1_0011_MR1_mpr-3_anon.img
1000000011	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0011_MR1/RAW/OAS1_0011_MR1_mpr-4_anon.img
1000000012	Sagittal	2003-05-03_00:00:00	2003-05-03_00:00:00	OAS1_0012_MR1/RAW/OAS1_0012_MR1_mpr-1_anon.img

Project data can be added back to Enterprise Repository





Enterprise Querytool allows:

- Selection of patient populations through complex queries
- Specifically targeted population analysis
- Privacy Management



Project Workbench allows:

- Patient population medical record review and exploration of data in a targeted population
- Use of Natural Language Processing
- Import of batch data sets
- Workflow support to simulate sequences of operations
- Export of batch outputs for specific applications
- Privacy management

Identity Management

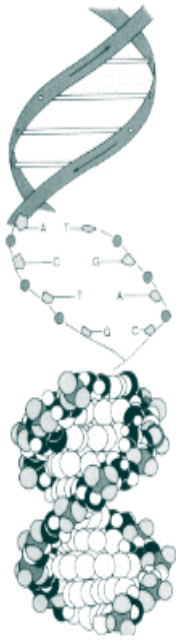
- Integration into existing medical record systems is an important requirement
- Levels of de-identification need to be supported
 - 1 – aggregate data only (obfuscate overly identifying items)
 - 2 – random records from a limited data set
 - 3 – full limited data set
 - 3.1 – de-identified text reports
 - 4 – fully identified data set
- Auditing to be tied into Project Management Cell
- Encryption key policy is an important difference between locals

Help outside of i2b2 and AUG

- Ontology management
 - NCBO
- Imaging Research
 - BIRN / XNAT
- Workflow management
 - [several options]
- Genomics integration
 - Vista
- Tissue management
 - Crimson

Organizing help within AUG

- New Web Site
- Issue Tracking
- Open Source Management, with a long term goal to help the following be supported:
 - - ETL
 - - Epidemiology workflows
 - - GWAS support



i2b2 Clinical Research Chart

Shawn Murphy MD, Ph.D.

Griffin Weber MD, Ph.D.

Michael Mendis

Vivian Gainer MS

Lori Phillips MS

Rajesh Kuttan

Wensong Pan MS

Henry Chueh MD

Susanne Churchill Ph.D.

John Glaser Ph.D.

Isaac Kohane MD, Ph.D.