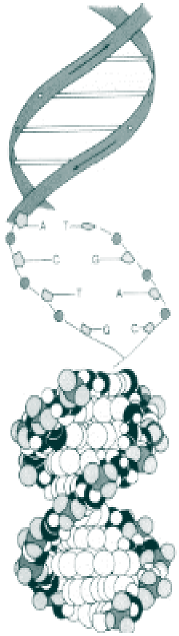


Overall Description of i2b2

Shawn Murphy MD, Ph.D.

Mass General Brigham and Harvard Medical School

November 14, 2020

“Informatics for Integrating Biology and the Bedside (i2b2)” what is it?

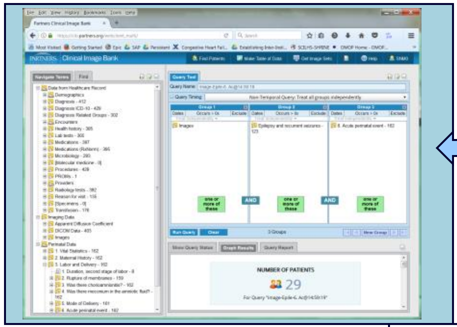
- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
  - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
  - Software is built in a modular pattern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

i2b2 used for Big Clinical Data

1) Queries for aggregate patient numbers

- Warehouse of in & outpatient clinical data
- 6.7 million Partners Healthcare patients
- 3.1 billion diagnoses, medications, genomics, procedures, 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



De-identified Data Warehouse

Z731984X
Z74902XX
…
…

Encrypted identifiers

2) Returns detailed patient data

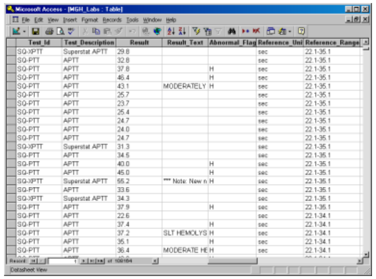
- Start with list of specific patients, usually from (1)
- Authorized use by IRB Protocol
- Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary,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

0000004
2185793
…
…

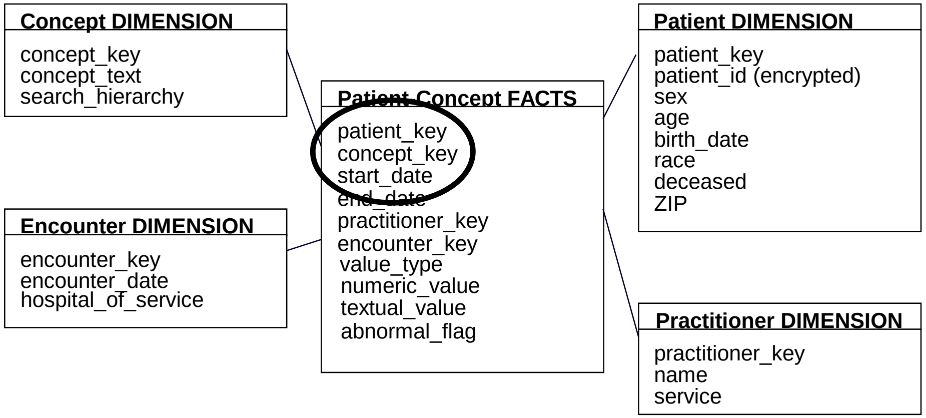
OR

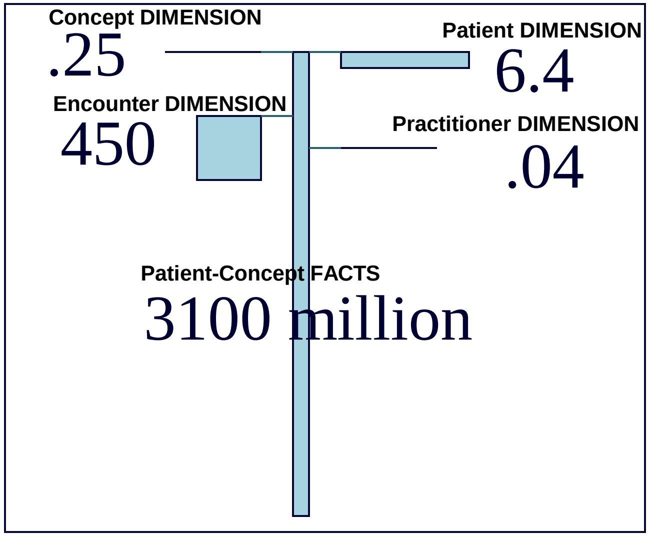
0000004
2185793
…
…

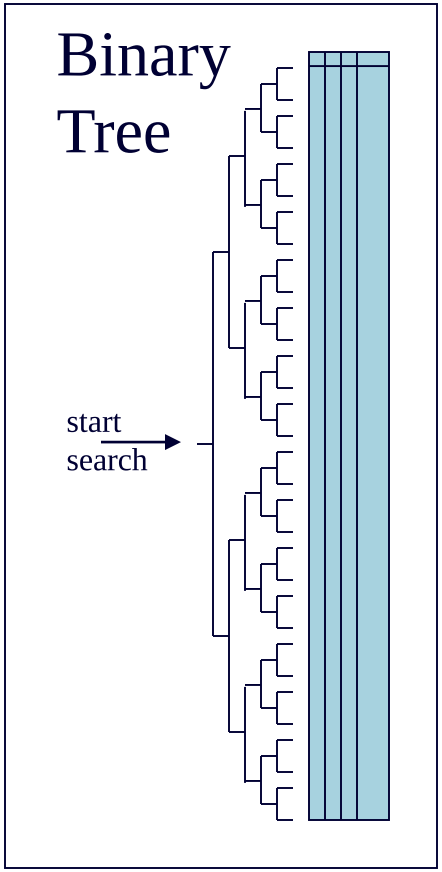
Real identifiers



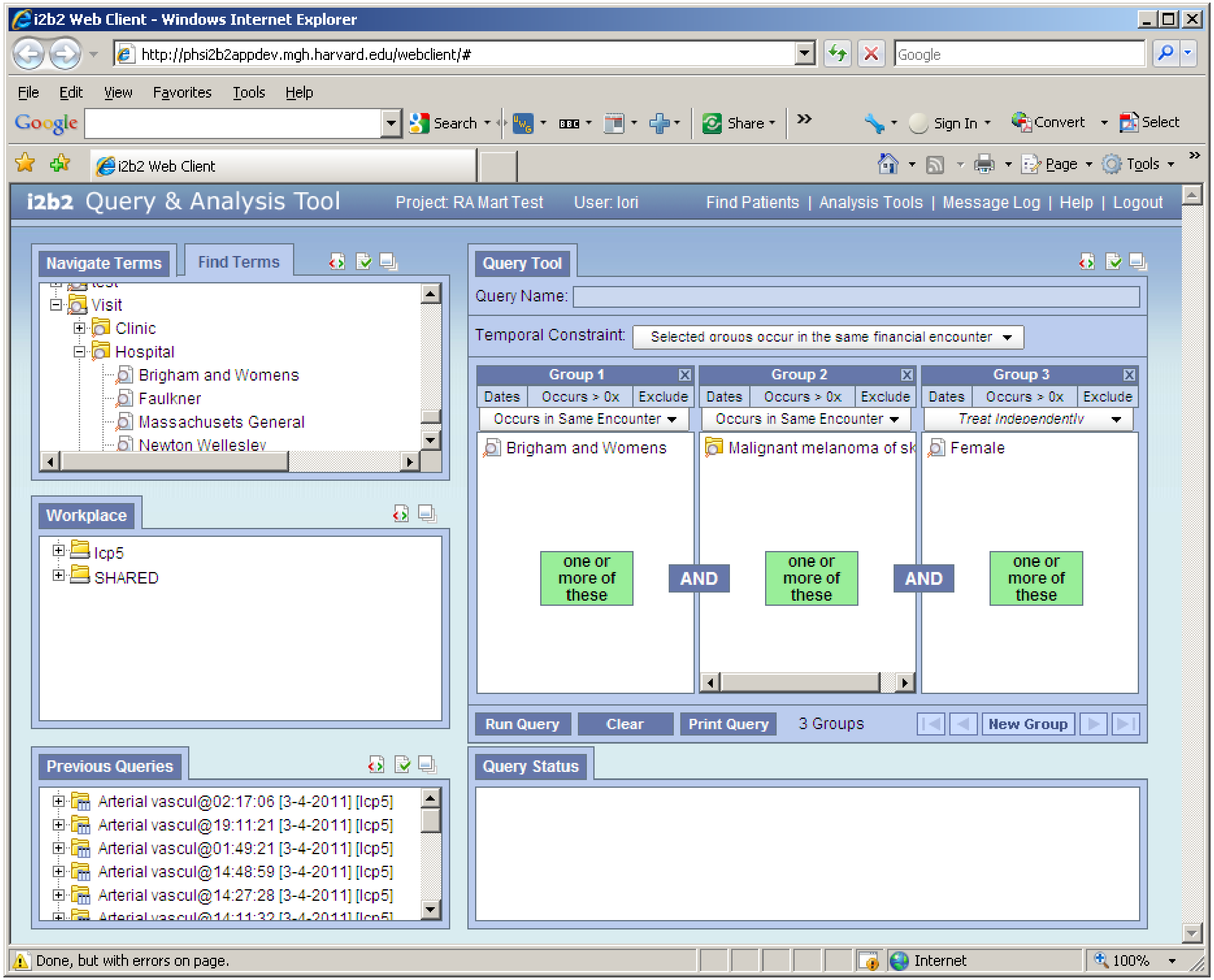
Enabled by Star Schema



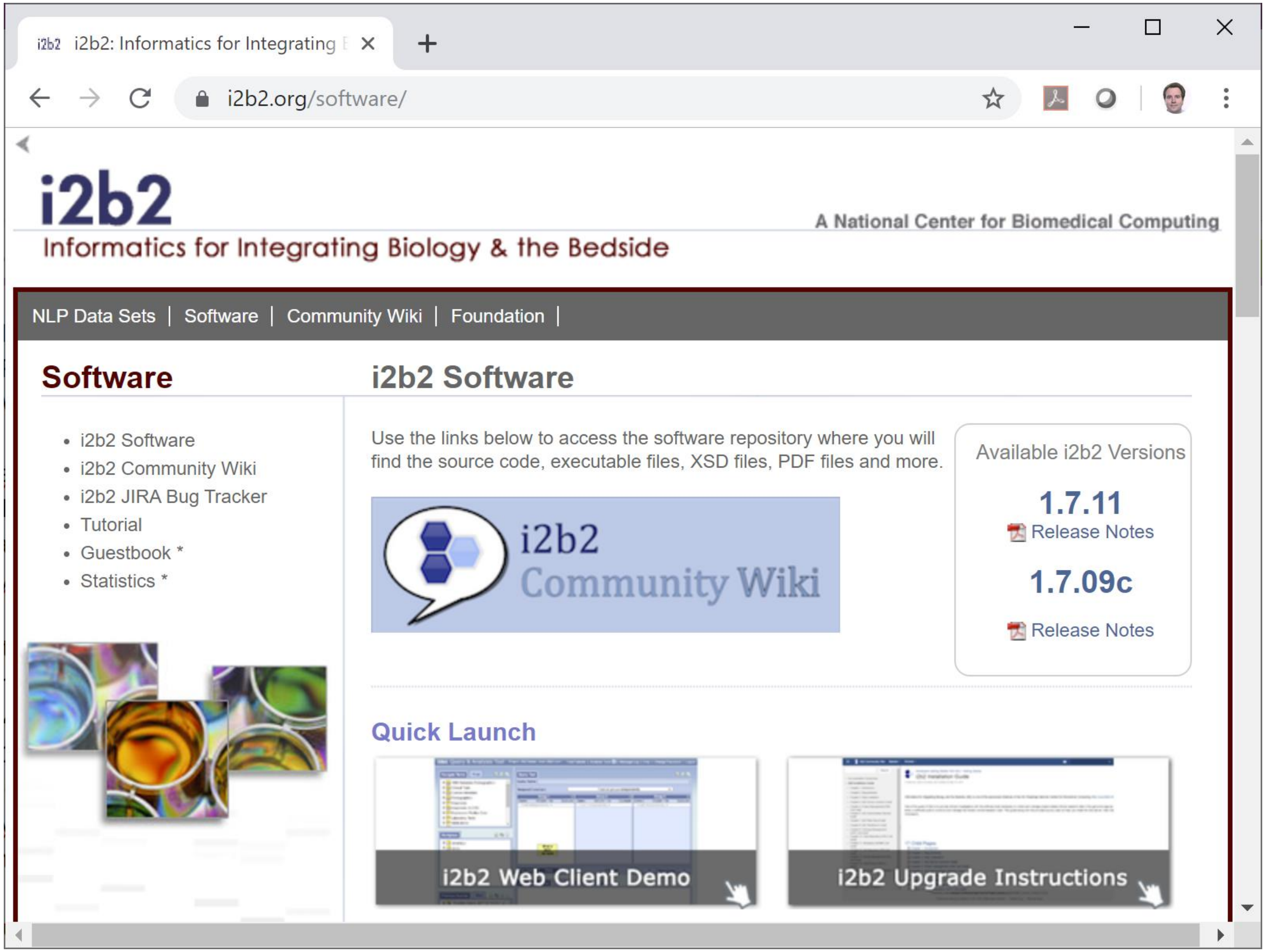




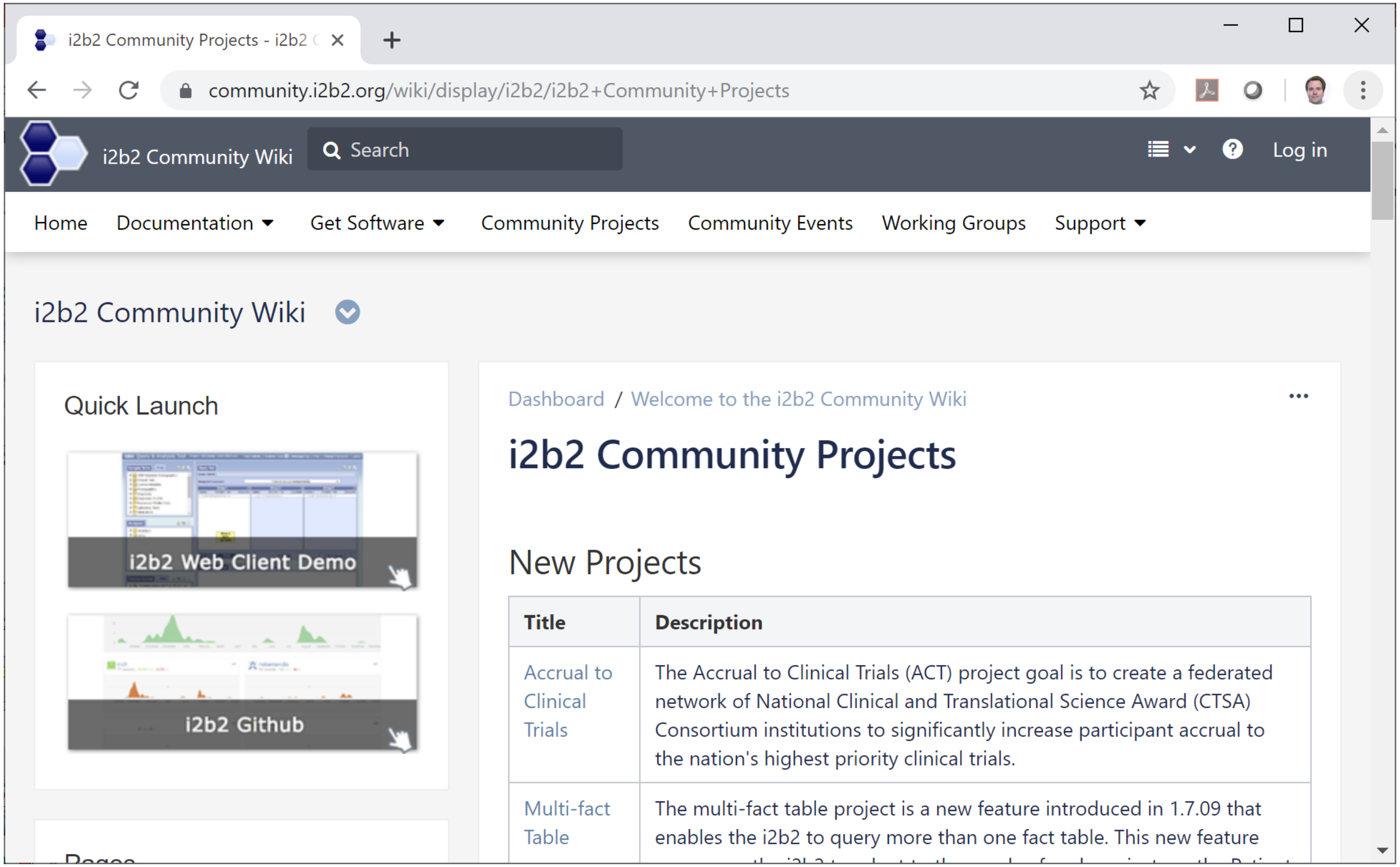
Interogation can occur through i2b2 web client



I2b2 Software components are distributed as open source

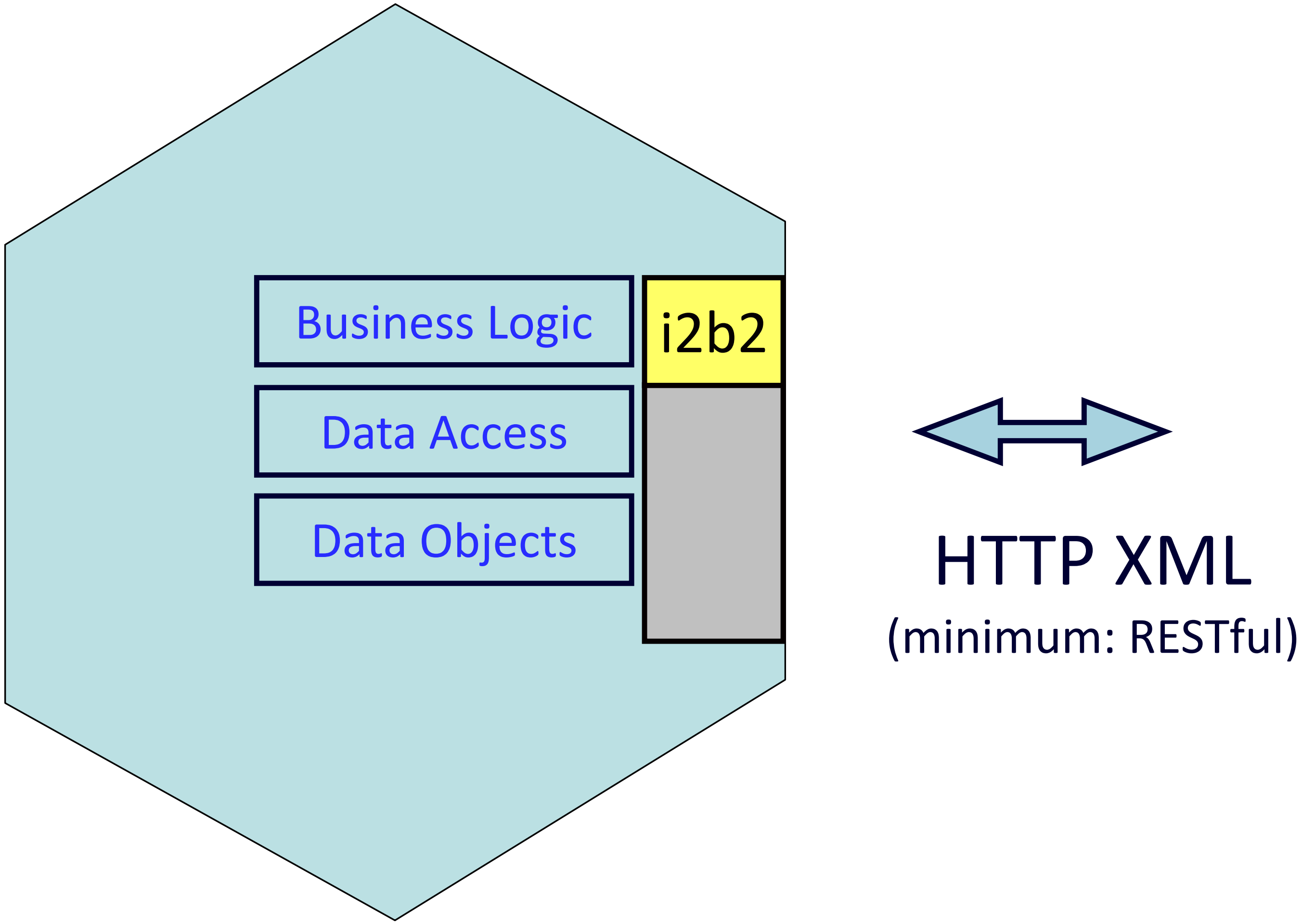


I2b2 Community Software Modules contributed as “Cells”

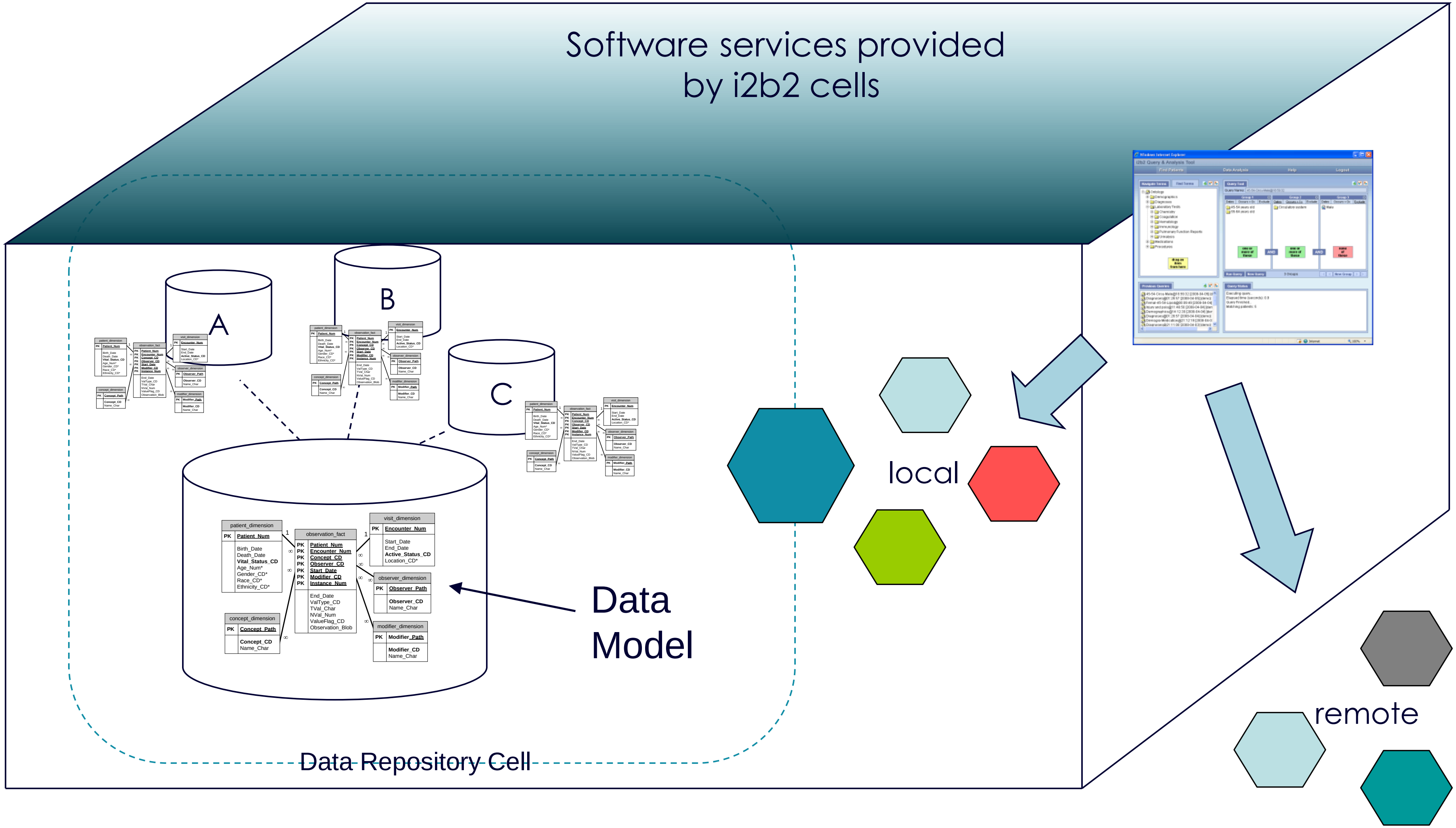


https://community.i2b2.org/wiki/display/i2b2/i2b2+Community+Projects

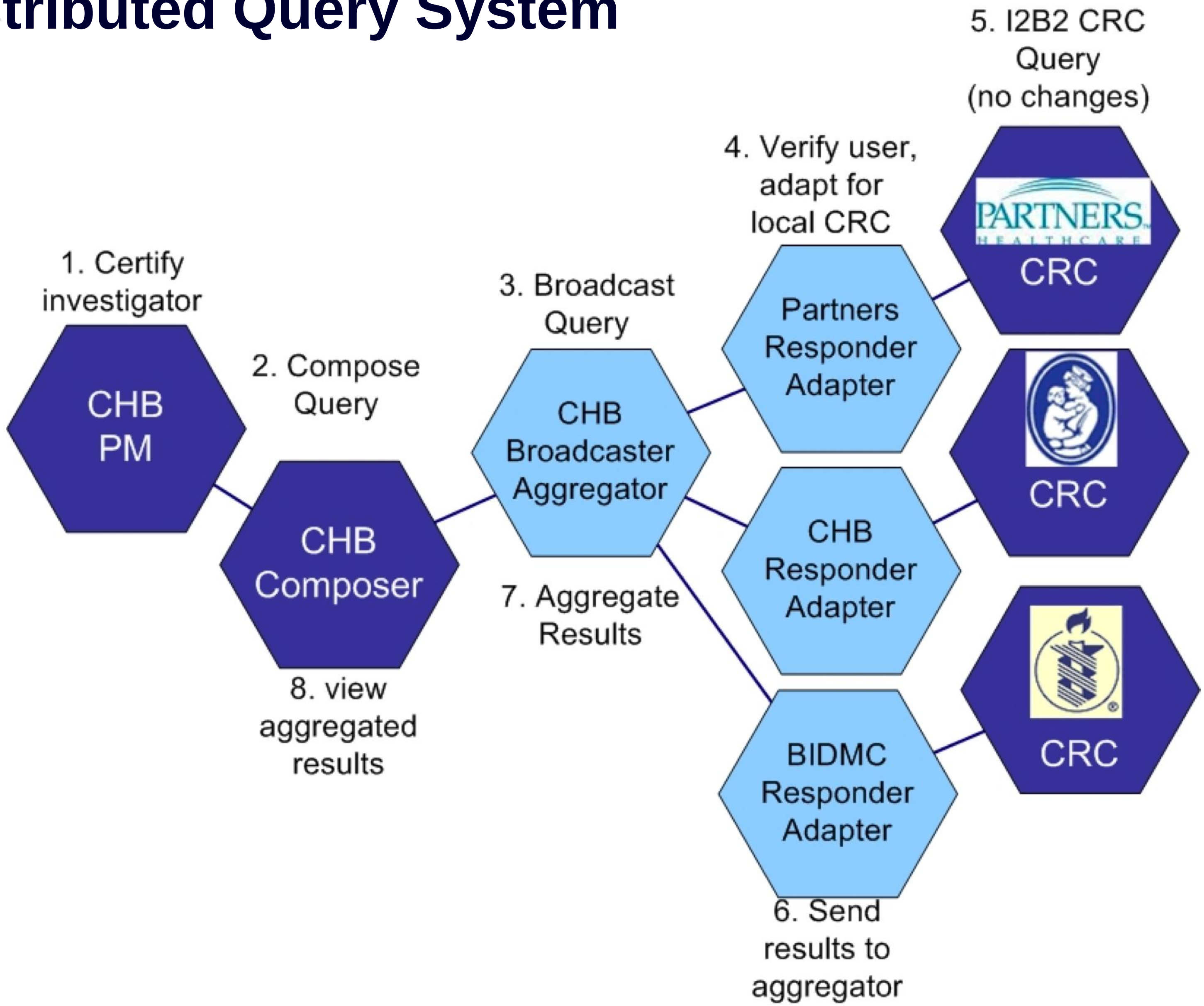
i2b2 Cell: The Canonical Software Module



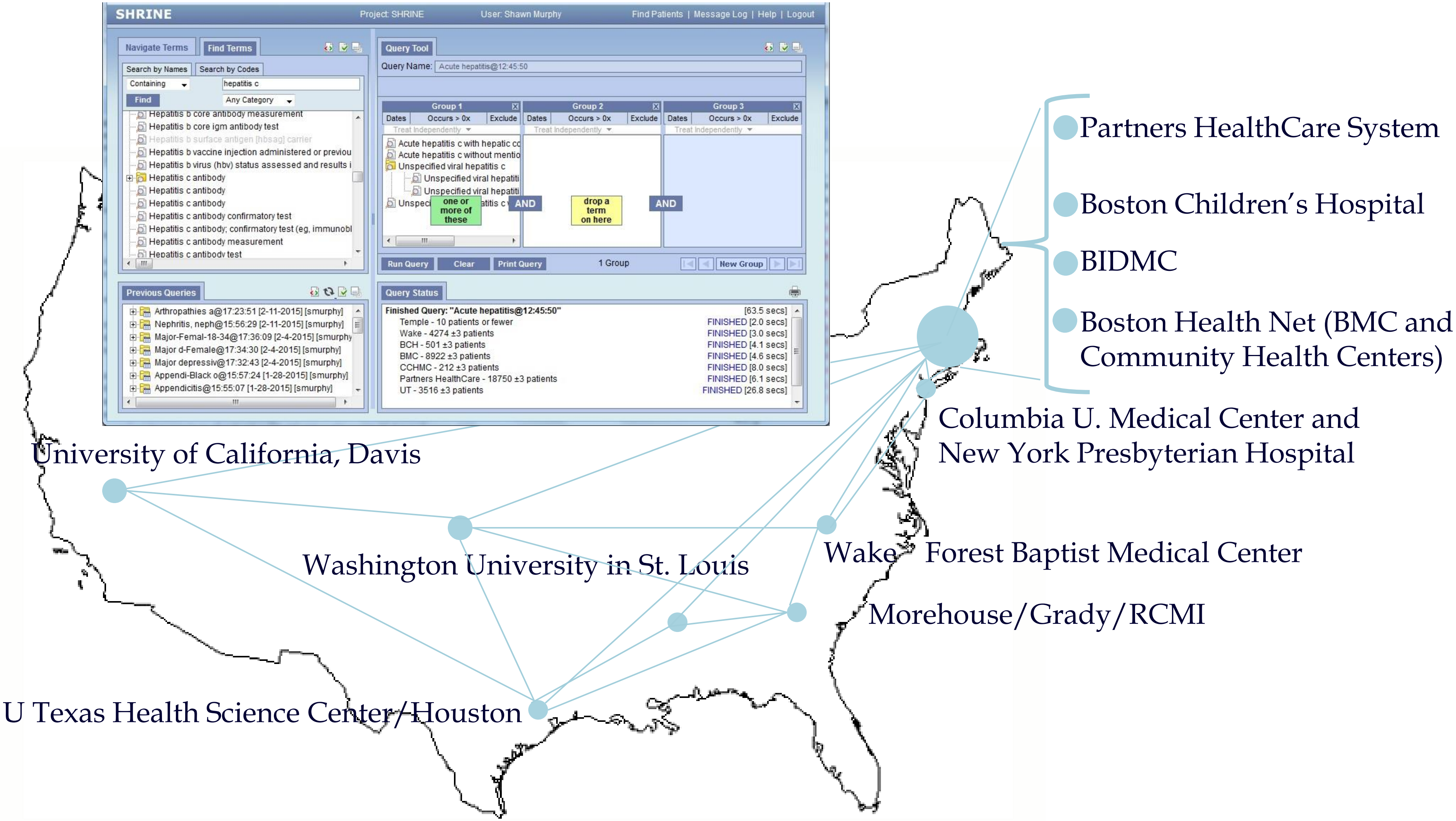
An i2b2 Environment is built from i2b2 Cells



Distributed Query System



Federated Queries



Implementations

CTSA's

- Boston University
- Case Western Reserve University (*including Cleveland Clinic*)
- Children's National Medical Center (GWU), Washington D.C.
- Duke University
- Emory University (*including Morehouse School of Medicine and Georgia Tech*)
- Harvard University (*including Beth Israel Deaconness Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Dana Farber Cancer Center, Joslin Diabetes Center, Massachusetts General Hospital*)
- Medical University of South Carolina
- Medical College of Wisconsin
- Oregon Health & Science University
- Penn State MIlton S. Hershey Medical Center
- Tufts University
- University of Alabama at Birmingham
- University of Arkansas for Medical Sciences
- University of California Davis
- University of California, Irvine
- University of California, Los Angeles*
- University of California, San Diego*
- University of California San Francisco
- University of Chicago
- University of Cincinnati (including *Cinncinati Children's Hospital Medical Center*)
- University of Colorado Denver (including *Children's Hospital Colorado*)
- University of Florida
- University of Kansas Medical Center
- University of Kentucky Research Foundation
- University of Massachusetts Medical School, Worcester
- University of Michigan
- University of Pennsylvania (including *Children's Hospital of Philadelphia*)
- University of Pittsburgh (including their *Cancer Institute*)
- University of Rochester School of Medicine and Dentistry
- University of Texas Health Sciences Center at Houston
- University of Texas Health Sciences Center at San Antonio
- University of Texas Medical Branch (Galveston)
- University of Texas Southwestern Medical Center at Dallas
- University of Utah
- University of Washington
- University of Wisconsin - Madison (*including Marshfield Clinic*)
- Virginia Commonwealth University
- Weill Cornell Medical College

Academic Health Centers (does not include AHCs that are part of a CTSA):

- Arizona State University
- City of Hope, Los Angeles
- Georgia Health Sciences University, Augusta
- Hartford Hospital, CN
- HealthShare Montana
- Massachusetts Veterans Epidemiology Research and Information Center (MAVERICK), Boston
- Nemours
- Phoenix Children's Hospital
- Regenstrief Institute
- Thomas Jefferson University
- University of Connecticut Health Center
- University of Missouri School of Medicine
- University of Tennessee Health Sciences Center
- Wake Forest University Baptist Medical Center

HMOs:

- Group Health Cooperative
- Kaiser Permanente

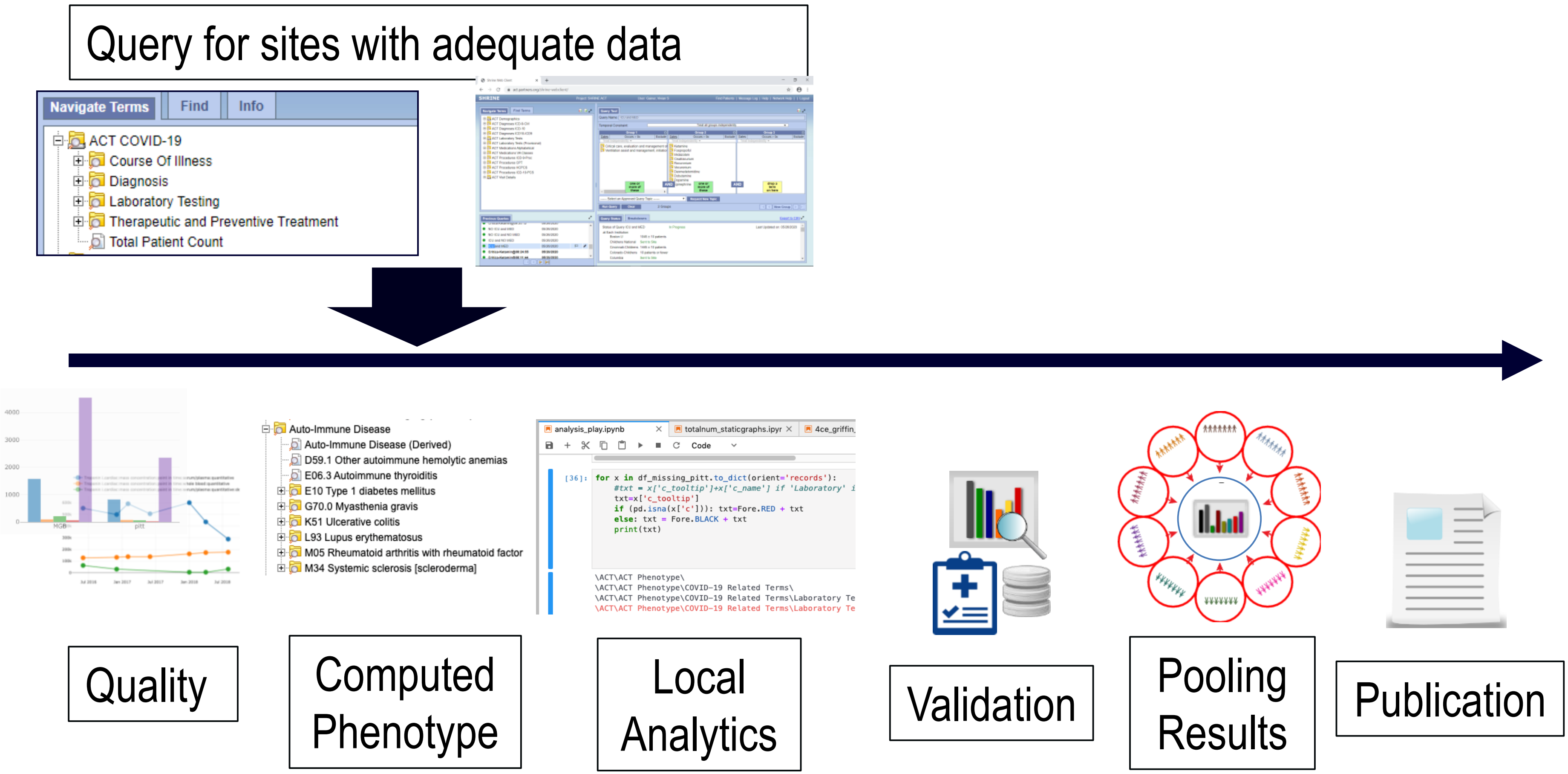
International:

- Georges Pompidou Hospital, Paris, France
- Hospital of the Free University of Brussels, Belgium
- Inserm U936, Rennes, France
- Institute for Data Technology and Informatics (IDI), NTNU, Norway
- Institute for Molecular Medicine Finland (FIMM)
- Karolinska Institute, Sweden
- Landspitali University Hospital, Reykjavik, Iceland
- Tokyo Medical and Dental University, Japan
- University of Bordeau Segalen, France
- University of Erlangen-Nuremberg, Germany
- University of Goettingen, Goettingen, Germany
- University of Leicester and Hospitals, England (Biomed. Res. Informatics Ctr. for Clin. Sci)
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

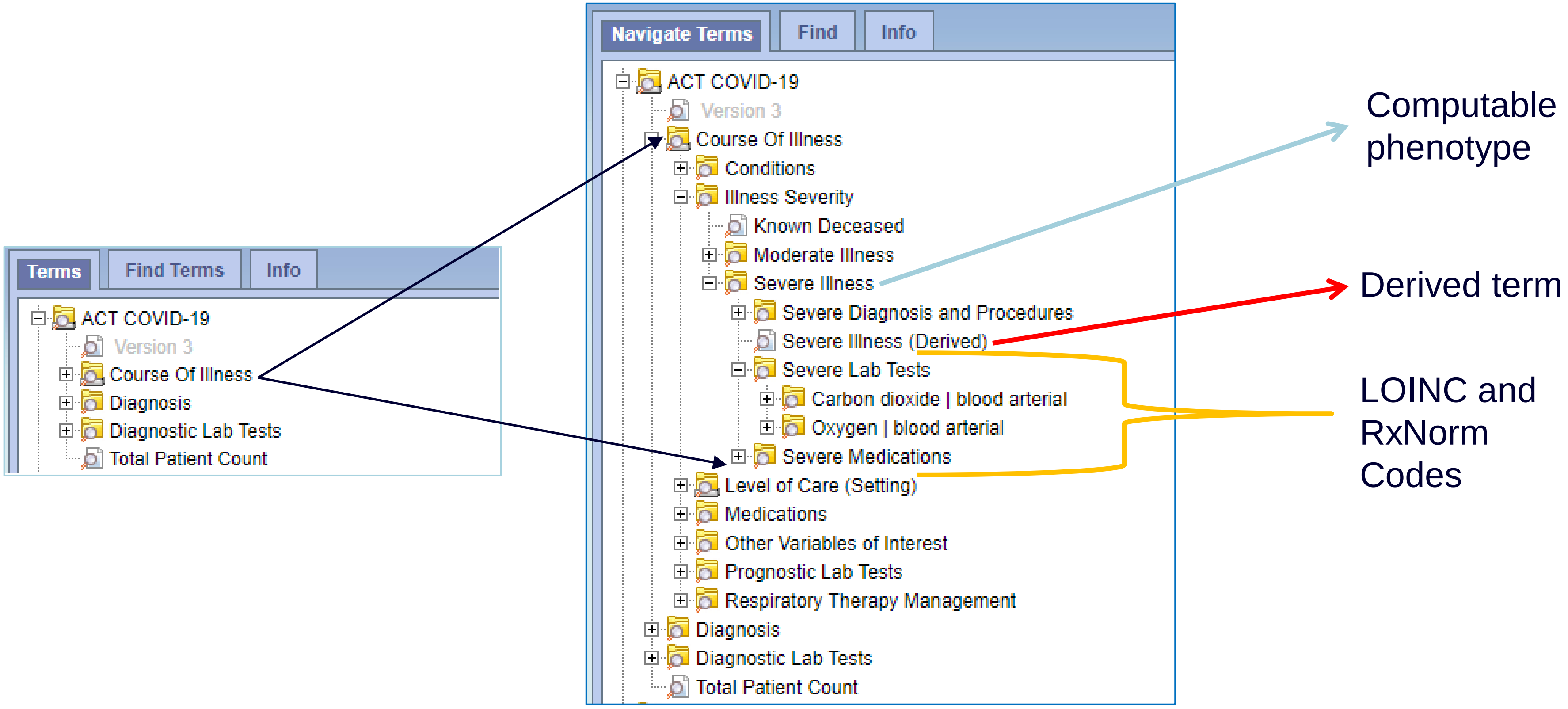
Companies:

- Johnson and Johnson (TransMART)
- GE Healthcare Clinical Data Services

Research Pipeline for ACT



Computable phenotype: Severe illness in COVID-19



Simple Analytic Table Generator for i2b2

To generate a request template, drag a previous query to the **Patient(s)** box below. Any concepts found in the query will be automatically added. Additional concepts can be added by dragging them to the **Concept(s)** box below. Each concept in the list will be a column in the table request. Use the **Aggregation Option** column to select how columns of data will be aggregated.

Patient(s): 45-54-Diabe-Encou@12:39:58 **[Patient Count: 20995]**

Include concepts from the Previous Query

Concept(s): Drag & Drop additional concepts here from *Navigate Terms* or a *Previous Query*

Append concepts to the list below

Concept	Constraints	Aggregation Option	Include In Request
Patient Number		Value	
Gender		Value	
Age		Value	
Race		Value	
45-54 years old		Existence (Yes/No)	
Diabetes mellitus (>766000)	[Set Date]	Date (First)	
Encounter-based (>7632000)	[Set Date]	Count	

Generates One-Subject-Per-Row Tables

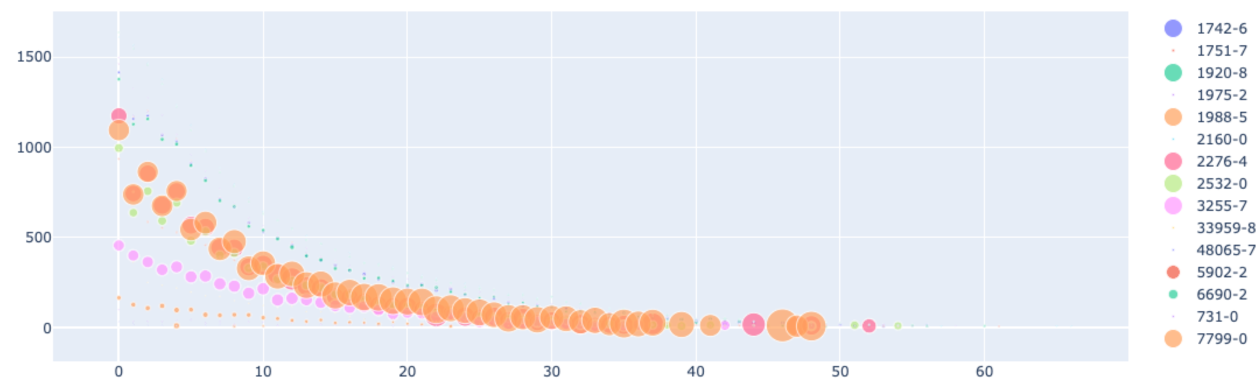
Patient ID	Date of First DX Diabetes	Most Recent A1C	Average A1C	Name of Antidiabetic Agent	…
104	3/4/2002	5.4	7.2	Miglitol	
1829	9/11/2013	12.1	9.4	Insulin	
2161	4/23/2000	4.1	6.2	Glipizide	

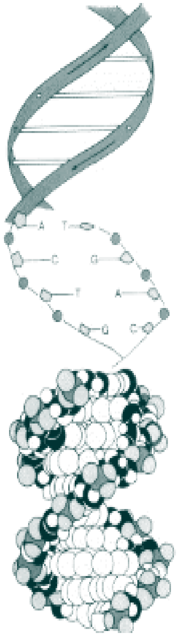
Then Shared Analytics Scripts run on Generated Table

analysis_play.ipynb | totalnum_staticgraphs.ipyr | 4ce_griffin_analytics.ipynb

Code | Python 3

```
[22]: # Plot labs - 4CE 1.1
      #traces = []
      #traces.append(go.Scatter(x=griffin_labs))
      griffin_labs = griffin_labs.replace({-99:np.nan}).fillna(0)
      px.scatter(griffin_labs.reset_index(),x='days_since_admission',y='num_patients_all',color='loinc',size='mean_value_all',title='Labs quantity',
                 labels={'x':'Days Since Positive'},hover_data=['stdev_value_all'])
```





Improving Quality of i2b2 Queries through Machine Learning

Psychological Medicine (2012), **42**, 41–50. © Cambridge University Press 2011
doi:10.1017/S0033291711000997

ORIGINAL ARTICLE

Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model

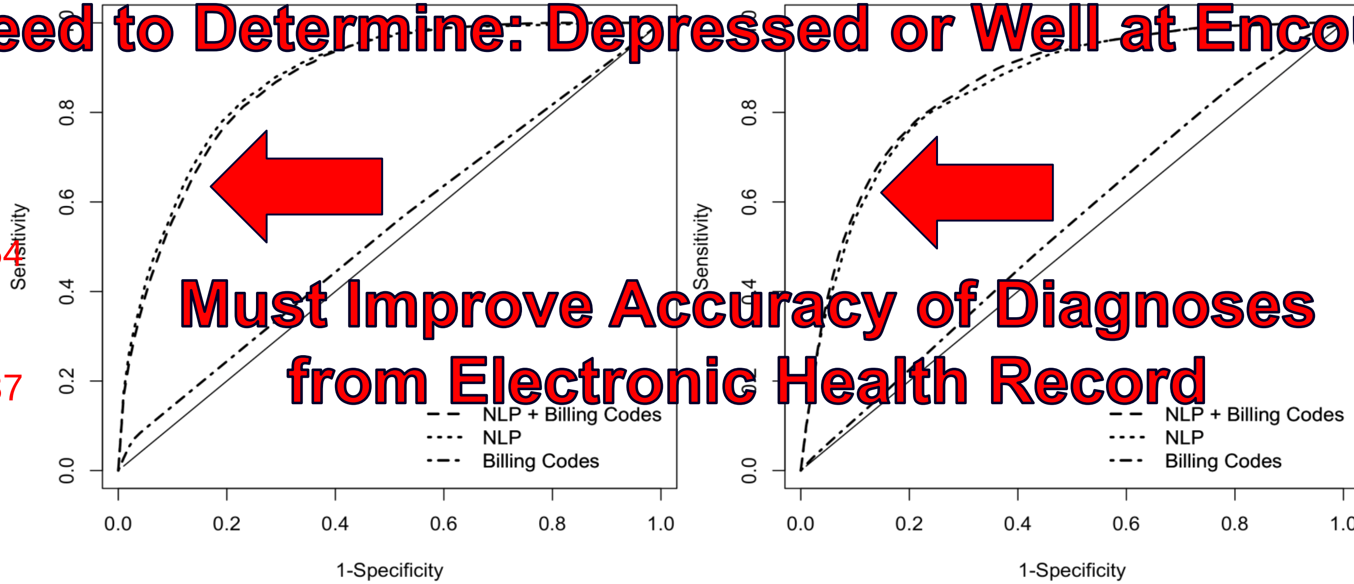
R. H. Perlis[1,2]*, D. V. Iosifescu[1,3], V. M. Castro[4], S. N. Murphy[5], V. S. Gainer[4], J. Minnier[6], T. Cai[6], S. Goryachev[4], Q. Zeng[7], P. J. Gallagher[2], M. Fava[1], J. B. Weilburg[1], S. E. Churchill[8], I. S. Kohane[9] and J. W. Smoller[2]

Use Phenotyping Algorithms to define cohorts of treatment-resistant and treatment-responsive depression

Need to Determine: Depressed or Well at Encounter

Initially:
AUC = 0.54

Finally:
AUC = 0.87



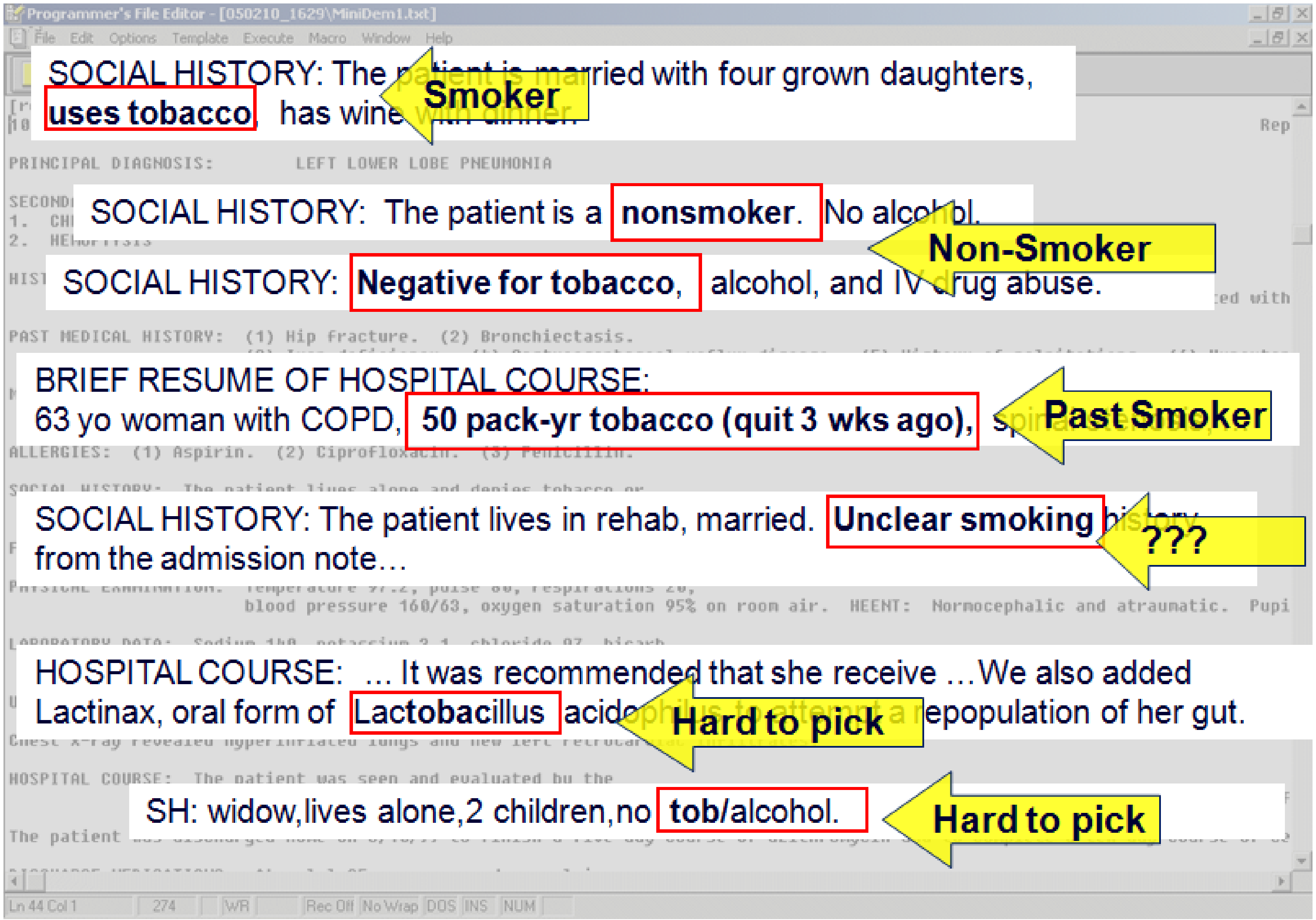
Must Improve Accuracy of Diagnoses from Electronic Health Record

Initially:
AUC = 0.55

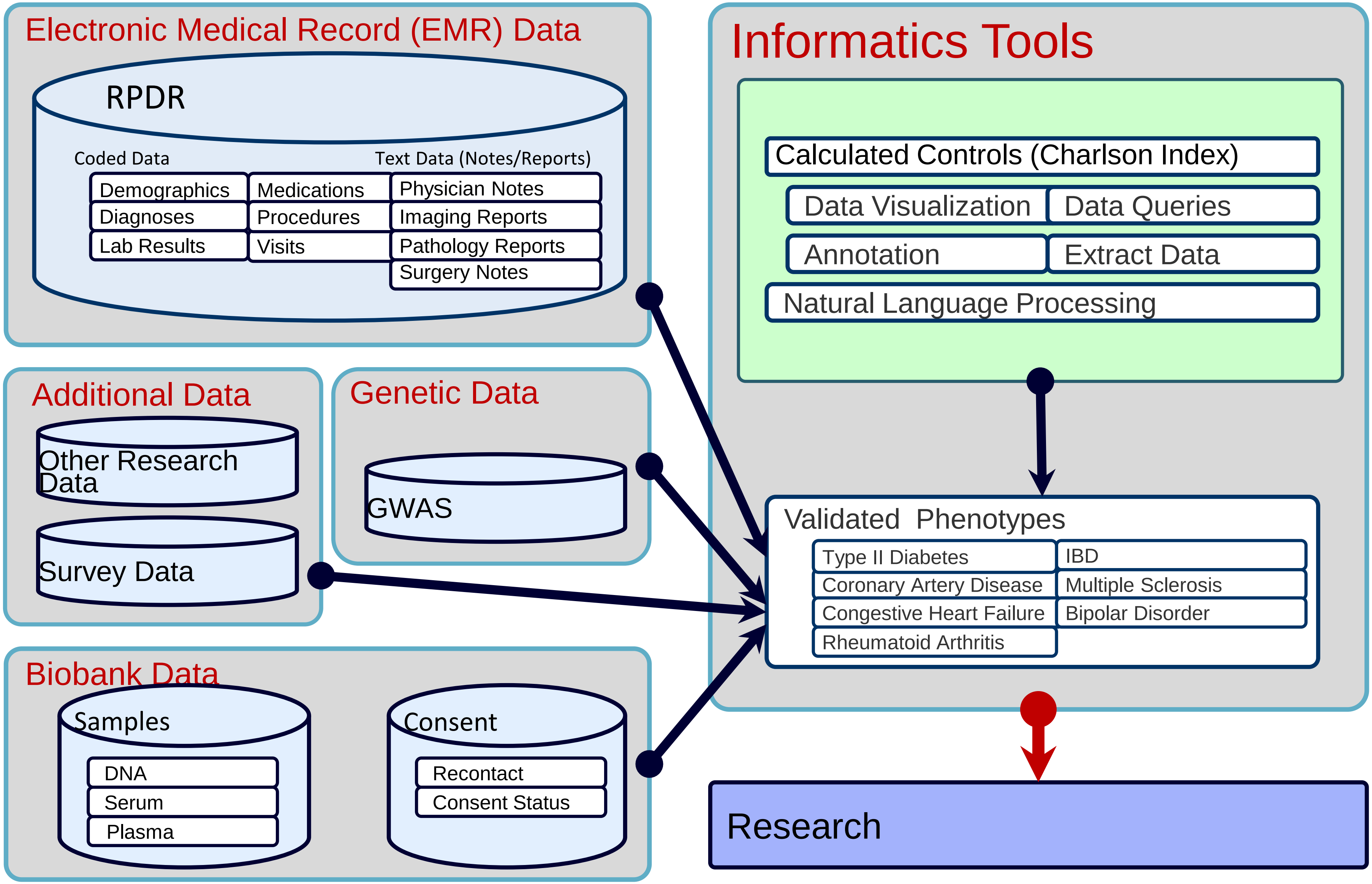
Finally:
AUC = 0.86

Clinical Status	Model	Specificity	Sensitivity	Precision	AUC
Depressed	Billing Codes	0.95	0.09 (0.03)	0.57 (0.14)	0.54 (0.02)
Depressed	NLP	0.95	0.42 (0.05)	0.78 (0.02)	0.88 (0.02)
Depressed	NLP + Billing Codes	0.95	0.39 (0.06)	0.78 (0.02)	0.87 (0.02)
Well	Billing Codes	0.95	0.06 (0.02)	0.26 (0.27)	0.55 (0.03)
Well	NLP	0.95	0.37 (0.06)	0.86 (0.02)	0.85 (0.02)
Well	NLP + Billing Codes	0.95	0.39 (0.07)	0.85 (0.02)	0.86 (0.02)

Use NLP to extract the relevant features from the set of patient notes.

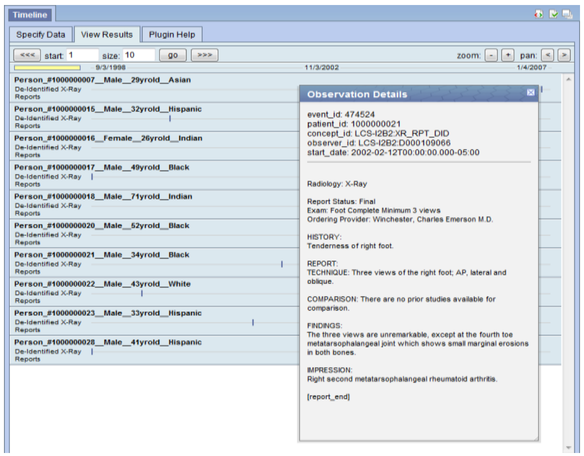


Data Integration in Biobank Portal

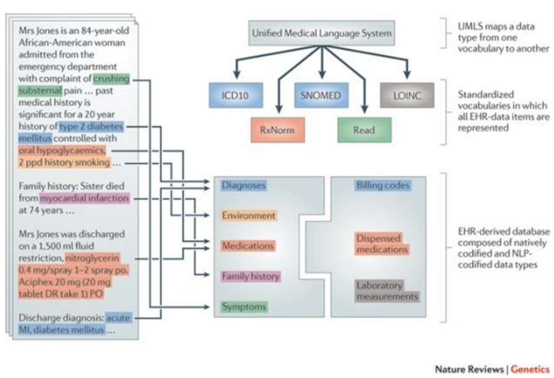


Curating a Disease Algorithm with a Gold Standard

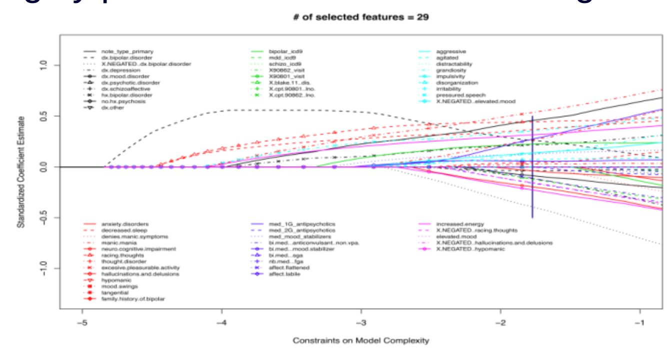
1. **Create a gold standard training set.**



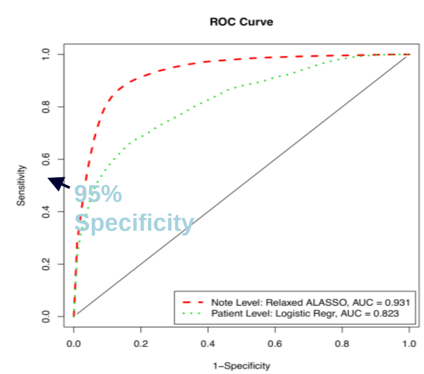
2. **Create a comprehensive list of features** from patient's electronic data that describe the disease of interest



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

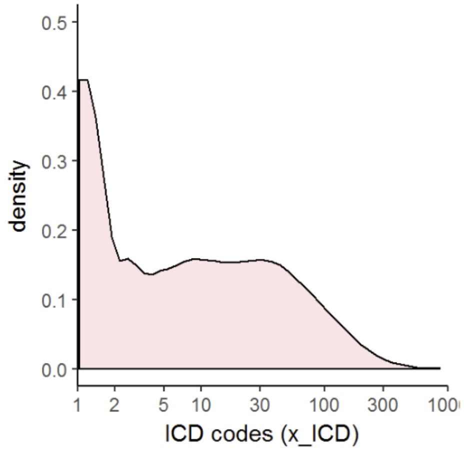


4. **Apply the algorithm to all subjects** in the superset and assign each subject a probability of having the phenotype

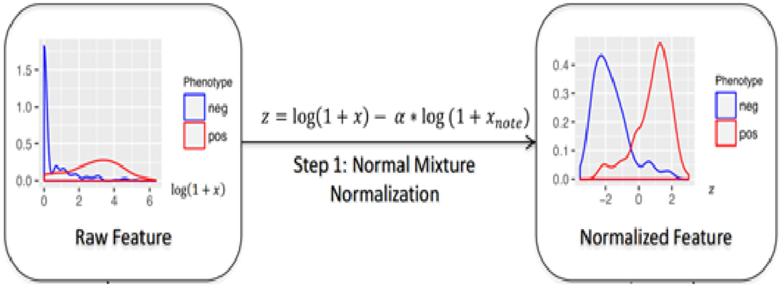


Curating a Disease Algorithm with a Silver Standard

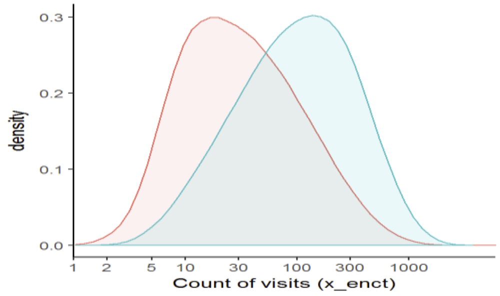
1. **Query for total number of mentions of disease**



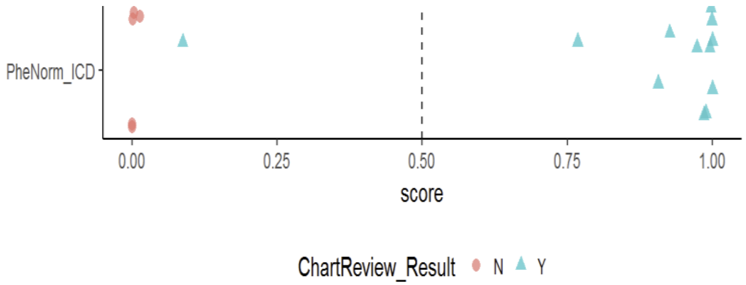
2. **Fit the mentions of disease to two curves normalizing for # of visits**



3. **Resolve the curves and separate into two groups, these are actually equal to patients with and without the disease**



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



Computable Phenotype Dashboard

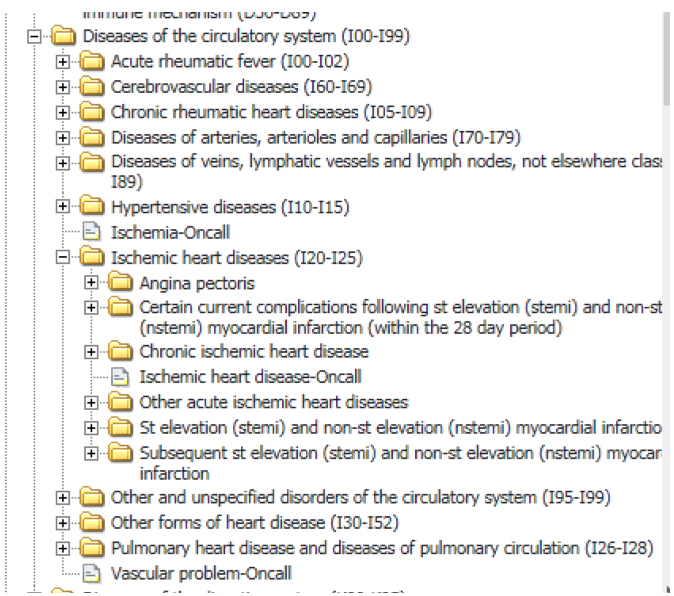
Phenotypes Methods About us

	category	PheWAS_code	abbr	PheWAS_name	model	ICD_PPV	ICD_AUC	AUC	PPV	TPR
1	ONC	PheWAS:189.21	BLCA	Bladder cancer	PheNorm_ICD	0.80	0.903	1.000	1.00	0.42
2	ONC	PheWAS:204	LEUK	Leukemia	PheNorm_ICD	0.73	1.000	1.000	1.00	0.91
3	PSYCH	PheWAS:297.1	SI	Suicidal ideation	PheNorm_ICDNLP	0.93	0.786	1.000	1.00	0.43
4	PSYCH	PheWAS:305.2	EATD	Eating disorder	PheNorm_ICDNLP	0.53	0.482	1.000	1.00	1.00
5	NEURO	PheWAS:327.4	INSOM	Insomnia	PheNorm_ICDNLP	0.93	0.821	1.000	1.00	0.50
6	CARDIO	PheWAS:452.2	DVT	Deep vein thrombosis	PheNorm_ICDNLP	0.87	0.692	1.000	1.00	1.00
7	NEURO	PheWAS:817	CONC	Concussion	PheNorm_NLP	0.73	0.682	1.000	1.00	0.27
8	METAB	PheWAS:250.1	T1DM	Type 1 diabetes	PheNorm_ICD	0.17	0.882	0.984	0.91	0.91
9	ONC	PheWAS:184.11	OVCA	Ovarian cancer	PheNorm_ICDNLP	0.60	0.926	0.981	1.00	0.67
10	ONC	PheWAS:182	UTCA	Uterine cancer	PheNorm_ICD	0.50	0.867	0.980	1.00	0.86
11	GI	PheWAS:555.1	CD	Crohn's disease	PheNorm_mean	0.54	0.961	0.980	0.90	0.97

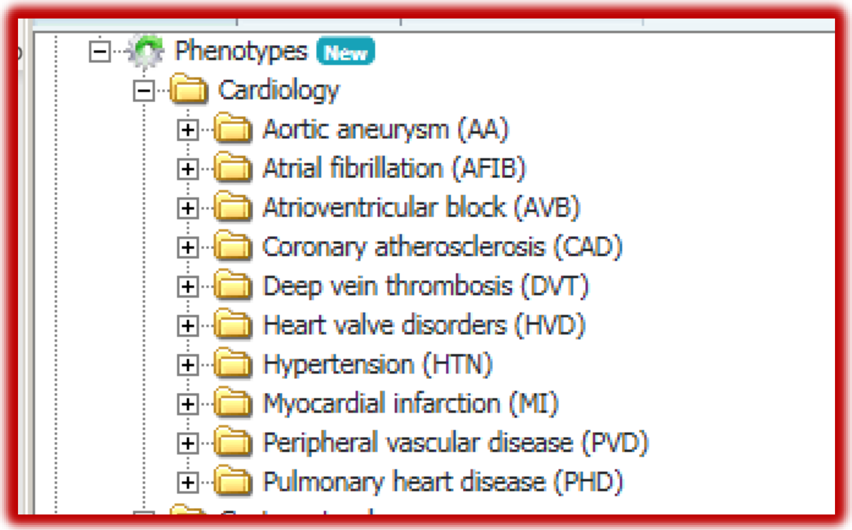
RESULT

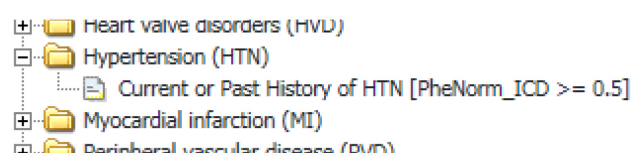
Accurate and Simple Disease Labels for Queries

Complicated

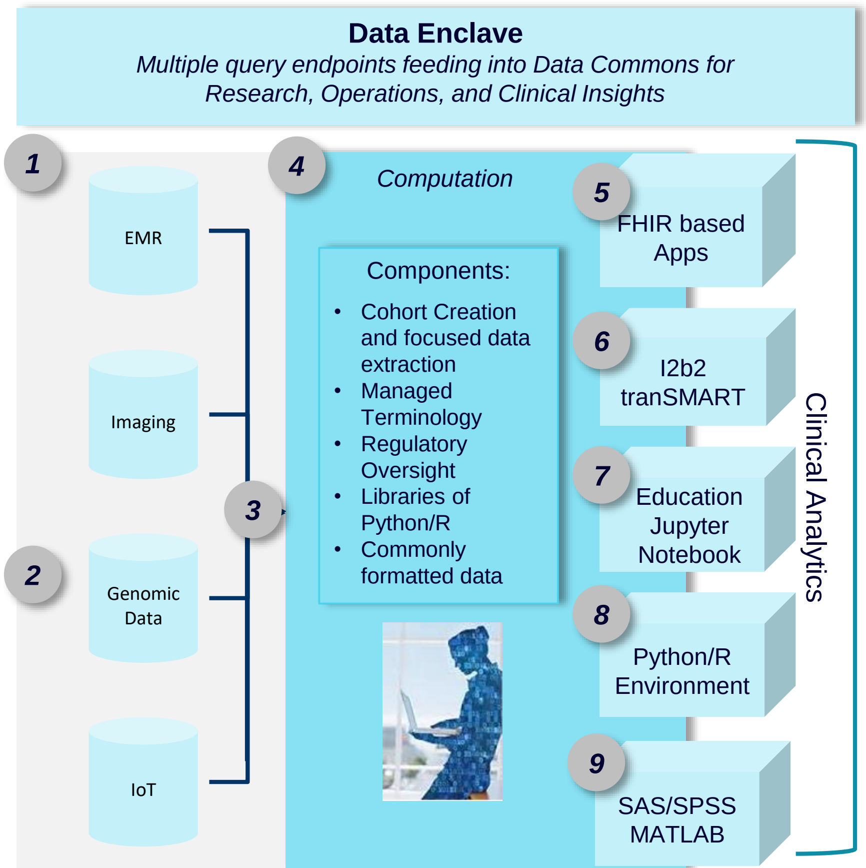


Simple



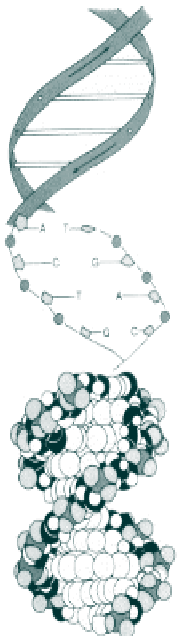


Digital Twin for Continuously Assessing Patient



Data Enclave	
1	Enable Data Extracts, perhaps some are Federated
2	Combine and Link data, put in common OMOP/i2b2 format
3	Conduit to Data Enclave
Technical Solution Development	
4	Healthcare Ready Bundle
Research and Clinical Application Projects	
5	FHIR based SMART Apps
6	I2b2 tranSMART with Fractalis plugin (next version of SmartR plugin)
7	Jupyter Notebook with AI Visualizations – code can advance to production
8	Python/R Environment full interactive development in Data Lake

PARTNERS HEALTHCARE



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)

I2b2 tranSMART Site (https://i2b2transmart.org/)

NIH/NCBC/BD2K; /NIMH; /NCATS; /NIBIB; /NHGRI

NIH R01 EB014947
NIH U54 LM008748
NIH U01 HG008685
PCORI 282364.5077585.0007

NIH U54 HG007963
NIH R01 AT006364
NIH R01 AT005280
NIH P01 AT006663

THANK YOU