

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?

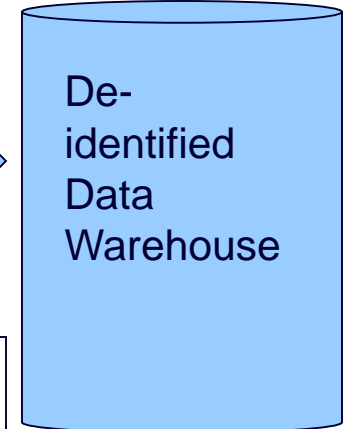
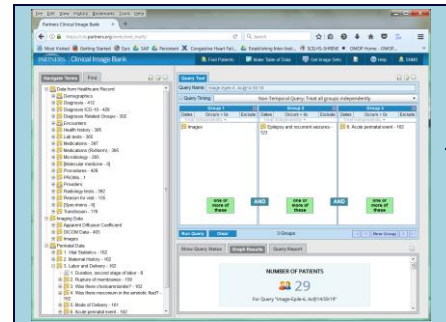
- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
 - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
 - Software is built in a modular pattern 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)

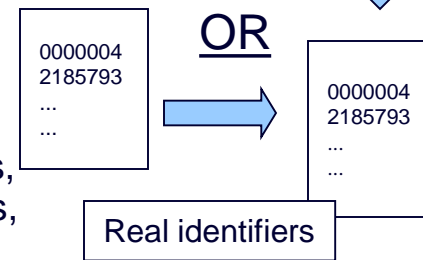
Query construction in web tool



Z731984X
Z74902XX
...
Encrypted identifiers

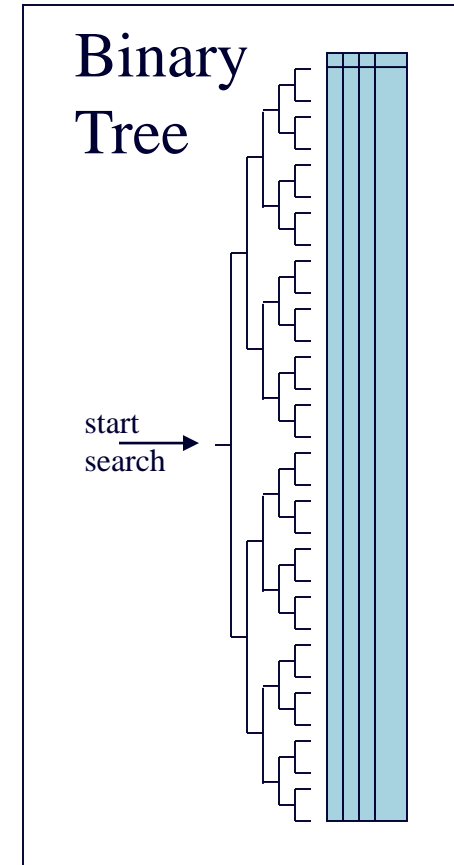
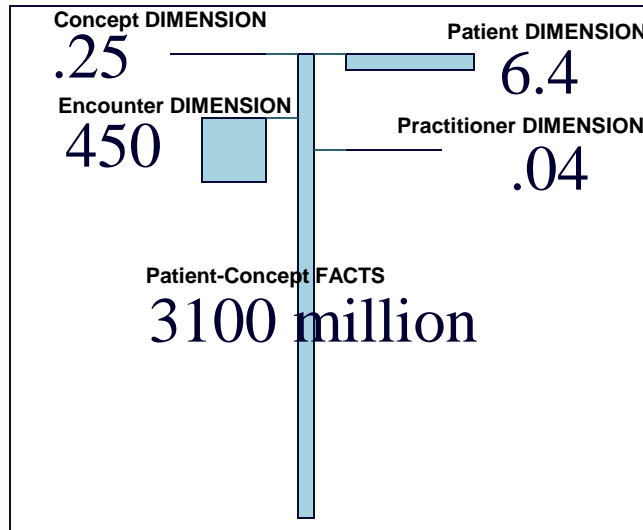
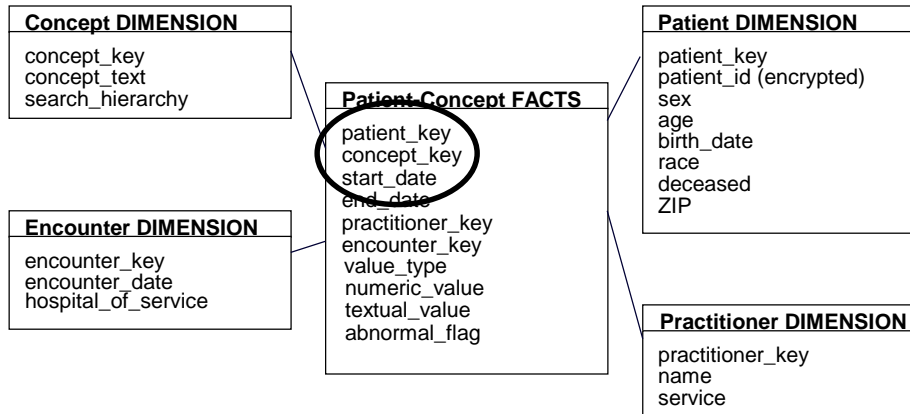
2) Returns detailed patient data

- 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.



| Task ID | Task Description | Result | Result Text | Abnormal Flag | Reference | Unit Reference | Range |
|---------|------------------|--------|-----------------|---------------|-----------|----------------|-----------|
| 50-PTT | Superior APTT | 23.8 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 23.8 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 46.4 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 43.1 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 25.7 | MODERATELY H | | | sec | 22.1-35.1 |
| 50-PTT | APTT | 23.7 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 35.4 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 24.7 | | N | | sec | 22.1-35.1 |
| 50-PTT | Superior APTT | 51.3 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 34.8 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 40.0 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 46.0 | | N | | sec | 22.1-35.1 |
| 50-PTT | Superior APTT | 55.2 | Note: New in H | | | sec | 22.1-35.1 |
| 50-PTT | APTT | 33.6 | | N | | sec | 22.1-35.1 |
| 50-PTT | Superior APTT | 34.3 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 37.8 | | N | | sec | 22.1-35.1 |
| 50-PTT | APTT | 22.6 | | N | | sec | 22.1-34.1 |
| 50-PTT | APTT | 37.4 | | N | | sec | 22.1-34.1 |
| 50-PTT | APTT | 37.2 | SLT HEMOLYSIS H | | | sec | 22.1-34.1 |
| 50-PTT | APTT | 38.1 | | N | | sec | 22.1-34.1 |
| 50-PTT | APTT | 38.4 | MODERATE HEH | | | sec | 22.1-34.1 |

Enabled by Star Schema



Interrogation can occur through i2b2 web client

The screenshot displays the i2b2 Query & Analysis Tool web client interface within a Windows Internet Explorer browser window. The browser's address bar shows the URL: `http://phsi2b2appdev.mgh.harvard.edu/webclient/#`. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The browser's toolbar contains various icons for navigation and utility. The i2b2 interface itself has a header with the title "i2b2 Query & Analysis Tool" and navigation links for "Project: RA Mart Test", "User: Iori", "Find Patients", "Analysis Tools", "Message Log", and "Help | Logout".

The main interface is divided into several panels:

- Navigate Terms:** A tree view showing a hierarchy of terms: Visit, Clinic, Hospital, Brigham and Womens, Faulkner, Massachusetts General, and Newton Welleslev.
- Workplace:** A panel showing a file structure with folders "Icp5" and "SHARED".
- Previous Queries:** A list of recent queries, all starting with "Arterial vascul@02:17:06 [3-4-2011] [Icp5]".
- Query Tool:** The central panel for building queries. It includes a "Query Name" field, a "Temporal Constraint" dropdown set to "Selected groups occur in the same financial encounter", and three query groups:
 - Group 1:** Contains the term "Brigham and Womens".
 - Group 2:** Contains the term "Malignant melanoma of sk".
 - Group 3:** Contains the term "Female".Each group has a dropdown menu for "Occurs in Same Encounter", "Occurs in Same Encounter", and "Treat Independently". The groups are connected by "AND" operators. Below the groups, there are buttons for "Run Query", "Clear", "Print Query", and "New Group".
- Query Status:** A panel at the bottom right for displaying the results of the query.

The browser's status bar at the bottom shows a warning icon and the text "Done, but with errors on page." The system tray includes the Internet Explorer icon and a zoom level of 100%.

I2b2 Software components are distributed as open source

The screenshot shows a web browser window with the URL `i2b2.org/software/`. The page header includes the i2b2 logo and the text "A National Center for Biomedical Computing" and "Informatics for Integrating Biology & the Bedside". A navigation bar contains links for "NLP Data Sets", "Software", "Community Wiki", and "Foundation".

Software

- i2b2 Software
- i2b2 Community Wiki
- i2b2 JIRA Bug Tracker
- Tutorial
- Guestbook *
- Statistics *

i2b2 Software

Use the links below to access the software repository where you will find the source code, executable files, XSD files, PDF files and more.

Available i2b2 Versions

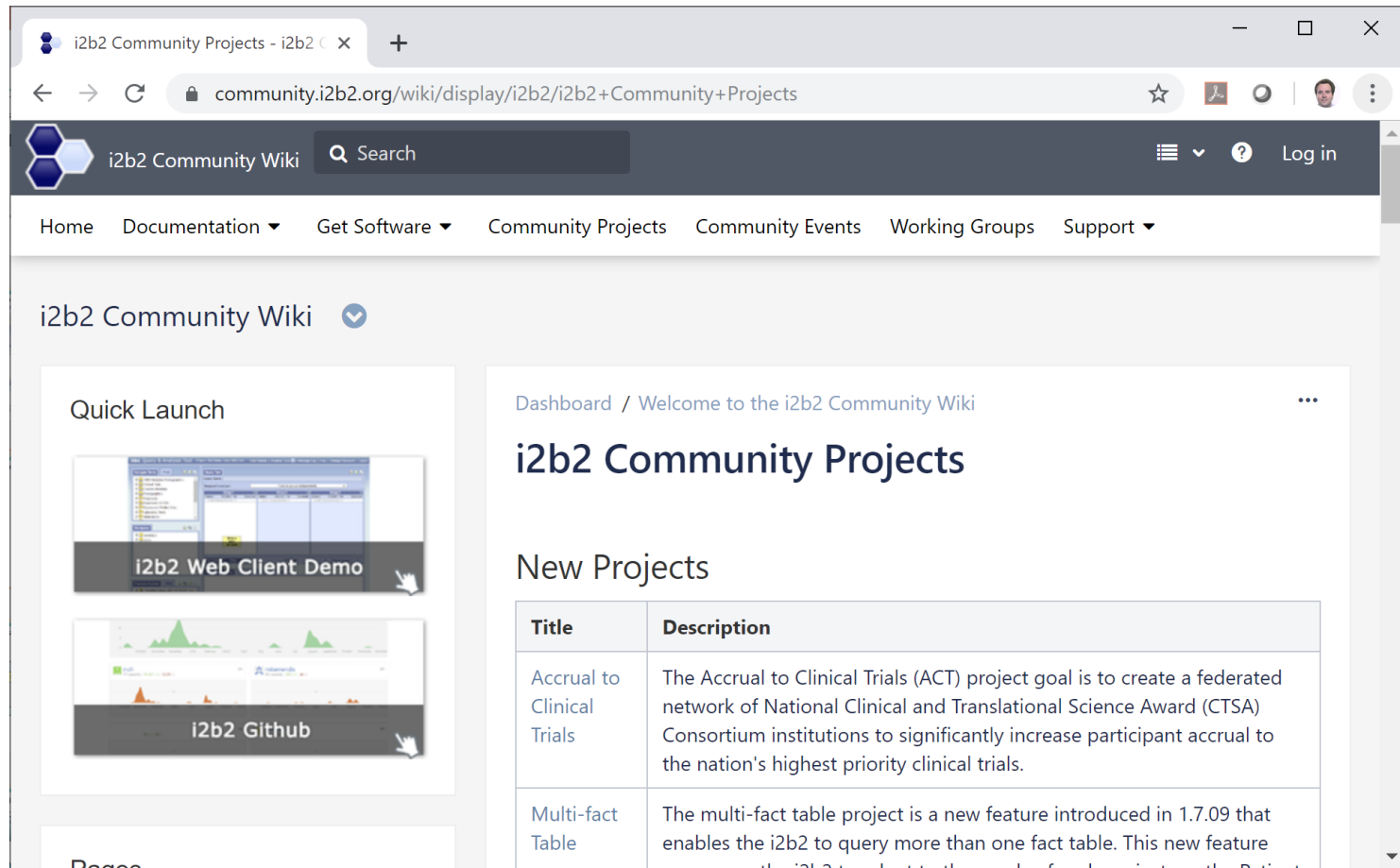
- 1.7.11**
Release Notes
- 1.7.09c**
Release Notes

Quick Launch

- i2b2 Web Client Demo**
- i2b2 Upgrade Instructions**

The page also features a "Community Wiki" logo and a collage of laboratory petri dishes.

I2b2 Community Software Modules contributed as “Cells”

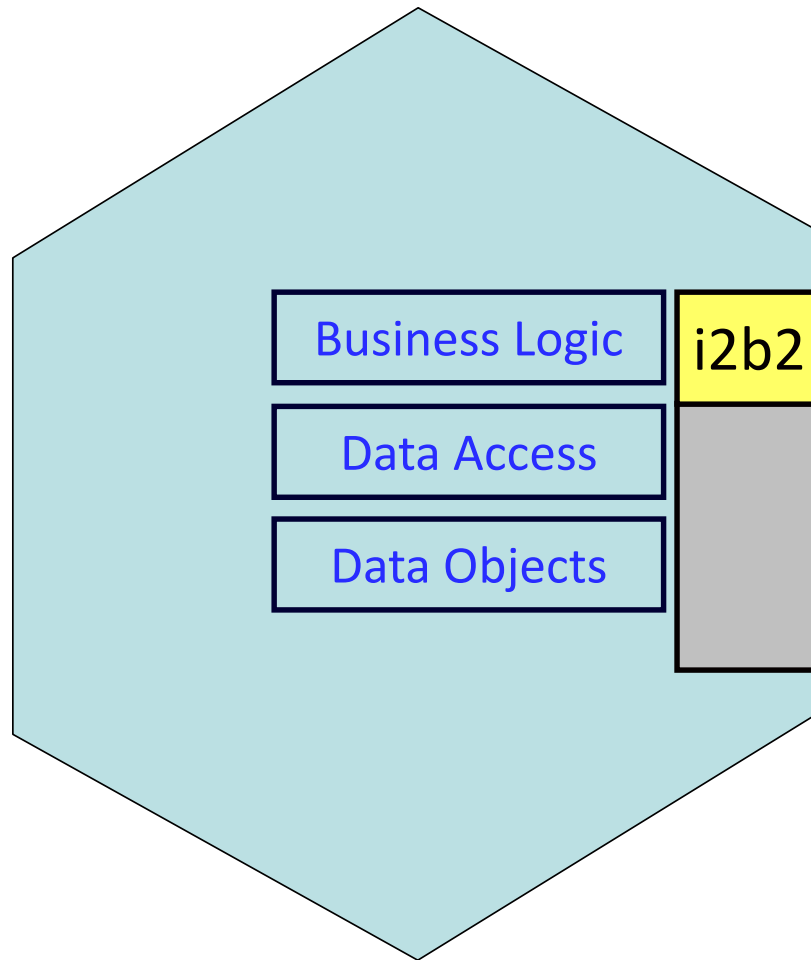


The screenshot shows a web browser window displaying the I2b2 Community Wiki page. The page features a navigation menu with items like Home, Documentation, Get Software, Community Projects, Community Events, Working Groups, and Support. The main content area is titled "i2b2 Community Wiki" and includes a "Quick Launch" section with two cards: "i2b2 Web Client Demo" and "i2b2 Github". To the right, there is a "New Projects" section with a table listing projects.

| Title | Description |
|----------------------------|--|
| Accrual to Clinical Trials | The Accrual to Clinical Trials (ACT) project goal is to create a federated network of National Clinical and Translational Science Award (CTSA) Consortium institutions to significantly increase participant accrual to the nation's highest priority clinical trials. |
| Multi-fact Table | The multi-fact table project is a new feature introduced in 1.7.09 that enables the i2b2 to query more than one fact table. This new feature |

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

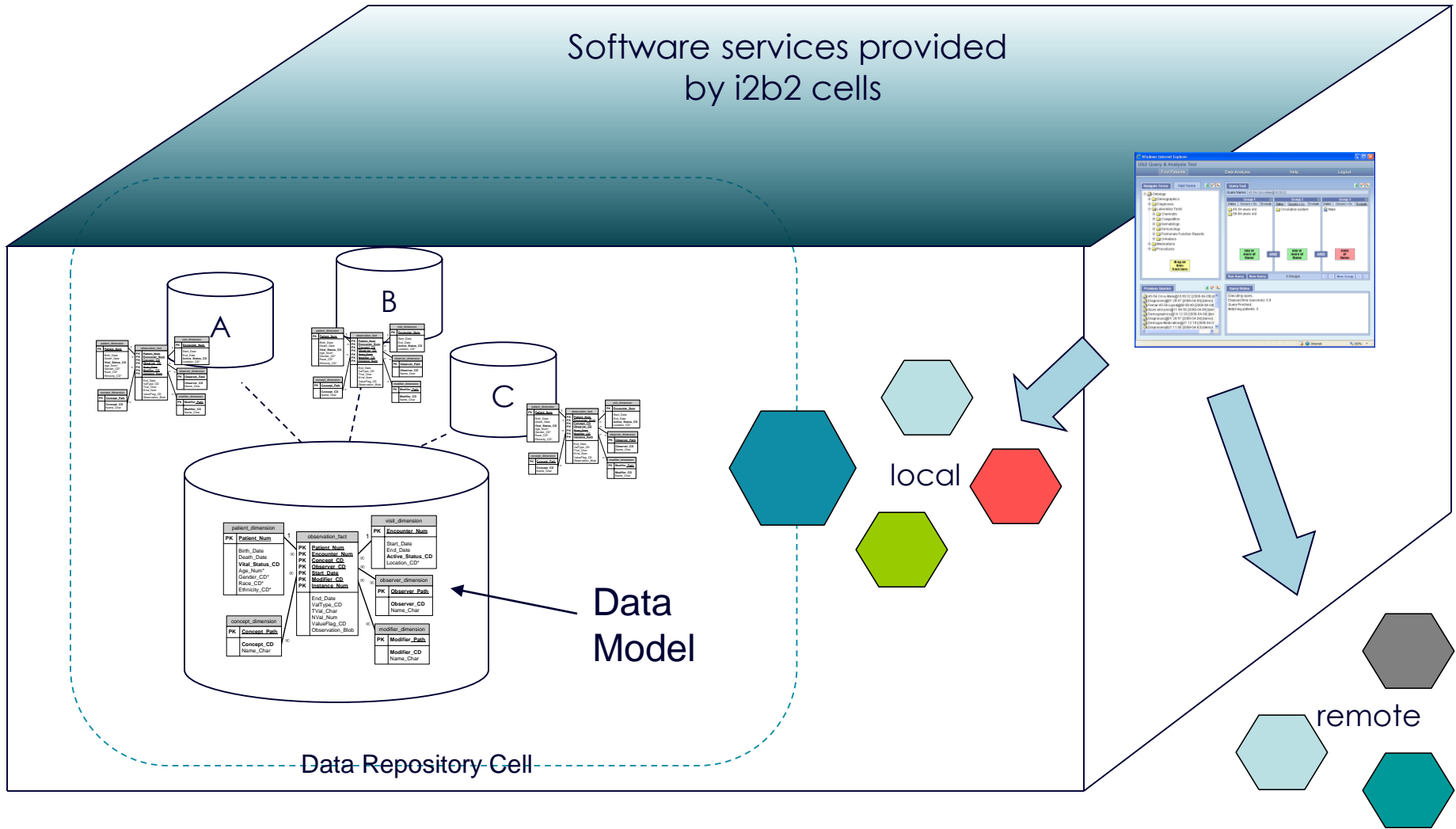
i2b2 Cell: The Canonical Software Module



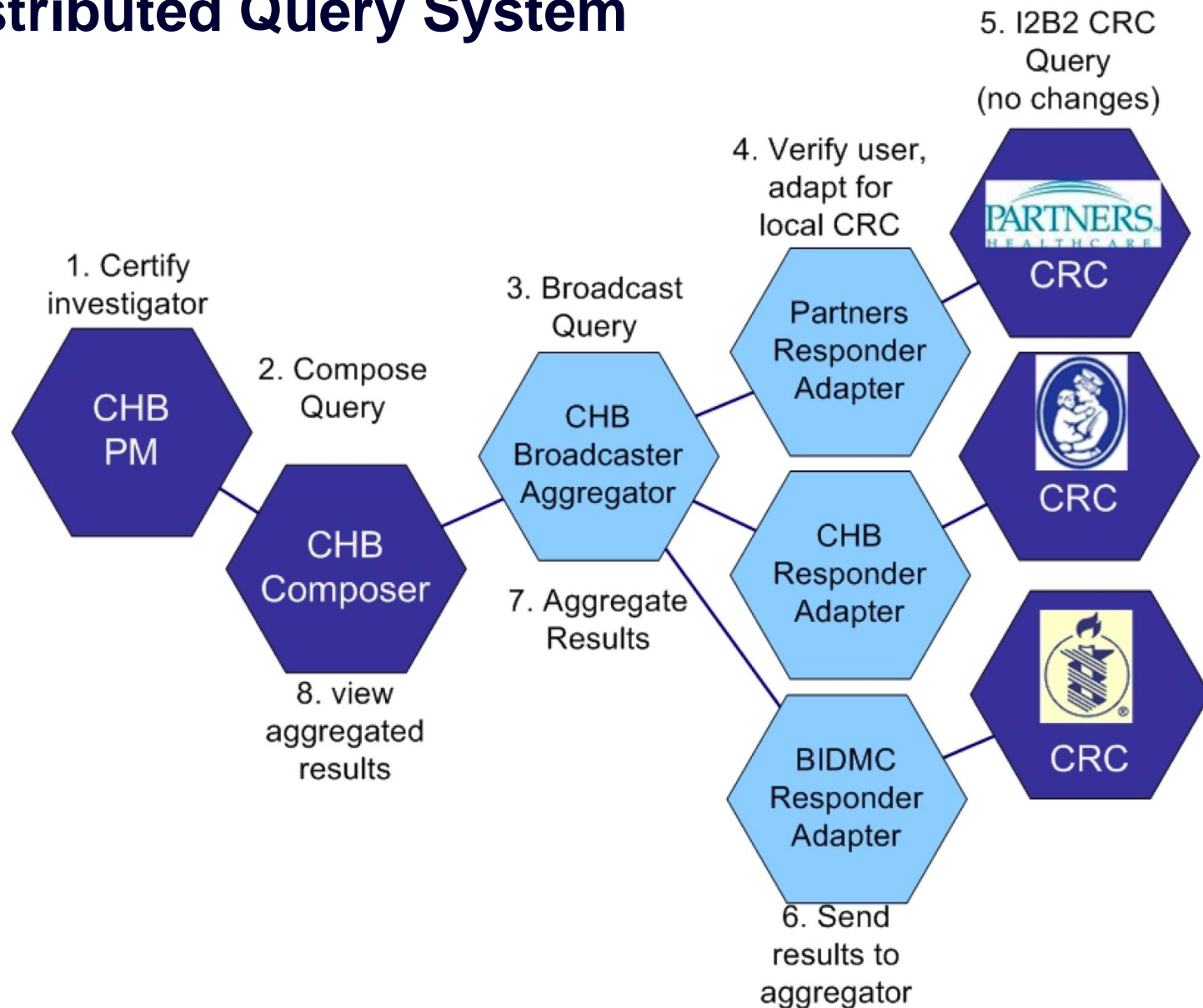
HTTP XML
(minimum: RESTful)

An i2b2 Environment is built from i2b2 Cells

Software services provided by i2b2 cells



Distributed Query System



Federated Queries

SHRINE Project: SHRINE User: Shawn Murphy Find Patients | Message Log | Help | Logout

Navigate Terms **Find Terms**

Search by Names Search by Codes

Containing hepatitis c

Find Any Category

- Hepatitis b core antibody measurement
- Hepatitis b core igm antibody test
- Hepatitis b surface antigen [hbsag] carrier
- Hepatitis b vaccine injection administered or previous
- Hepatitis b virus (hbv) status assessed and results i
- Hepatitis c antibody
- Hepatitis c antibody
- Hepatitis c antibody confirmatory test
- Hepatitis c antibody confirmatory test (eg, immunobi
- Hepatitis c antibody measurement
- Hepatitis c antibody test

Query Tool

Query Name: Acute hepatitis@12:45:50

| Group 1 | | | Group 2 | | | Group 3 | | |
|-----------------------------------|-------------|---------|----------------------------------|-------------|---------|---------------------|-------------|---------|
| Dates | Occurs > 0x | Exclude | Dates | Occurs > 0x | Exclude | Dates | Occurs > 0x | Exclude |
| Treat Independently | | | Treat Independently | | | Treat Independently | | |
| Acute hepatitis c with hepatic cd | | | Acute hepatitis c without mentio | | | | | |
| Unspecified viral hepatitis c | | | Unspecified viral hepatitis | | | | | |
| Unspecified viral hepatitis c | | | Unspecified viral hepatitis c | | | | | |
| Unspecified viral hepatitis c | | | Unspecified viral hepatitis c | | | | | |
| Unspecified viral hepatitis c | | | Unspecified viral hepatitis c | | | | | |

Run Query Clear Print Query 1 Group New Group

Query Status

Finished Query: "Acute hepatitis@12:45:50" [63.5 secs]

| | |
|---|----------------------|
| Temple - 10 patients or fewer | FINISHED [2.0 secs] |
| Wake - 4274 ±3 patients | FINISHED [3.0 secs] |
| BCH - 501 ±3 patients | FINISHED [4.1 secs] |
| BMC - 8922 ±3 patients | FINISHED [4.6 secs] |
| CC-HMC - 212 ±3 patients | FINISHED [8.0 secs] |
| Partners HealthCare - 18750 ±3 patients | FINISHED [6.1 secs] |
| UT - 3516 ±3 patients | FINISHED [26.8 secs] |

- Partners HealthCare System
- Boston Children's Hospital
- BIDMC
- Boston Health Net (BMC and Community Health Centers)

University of California, Davis

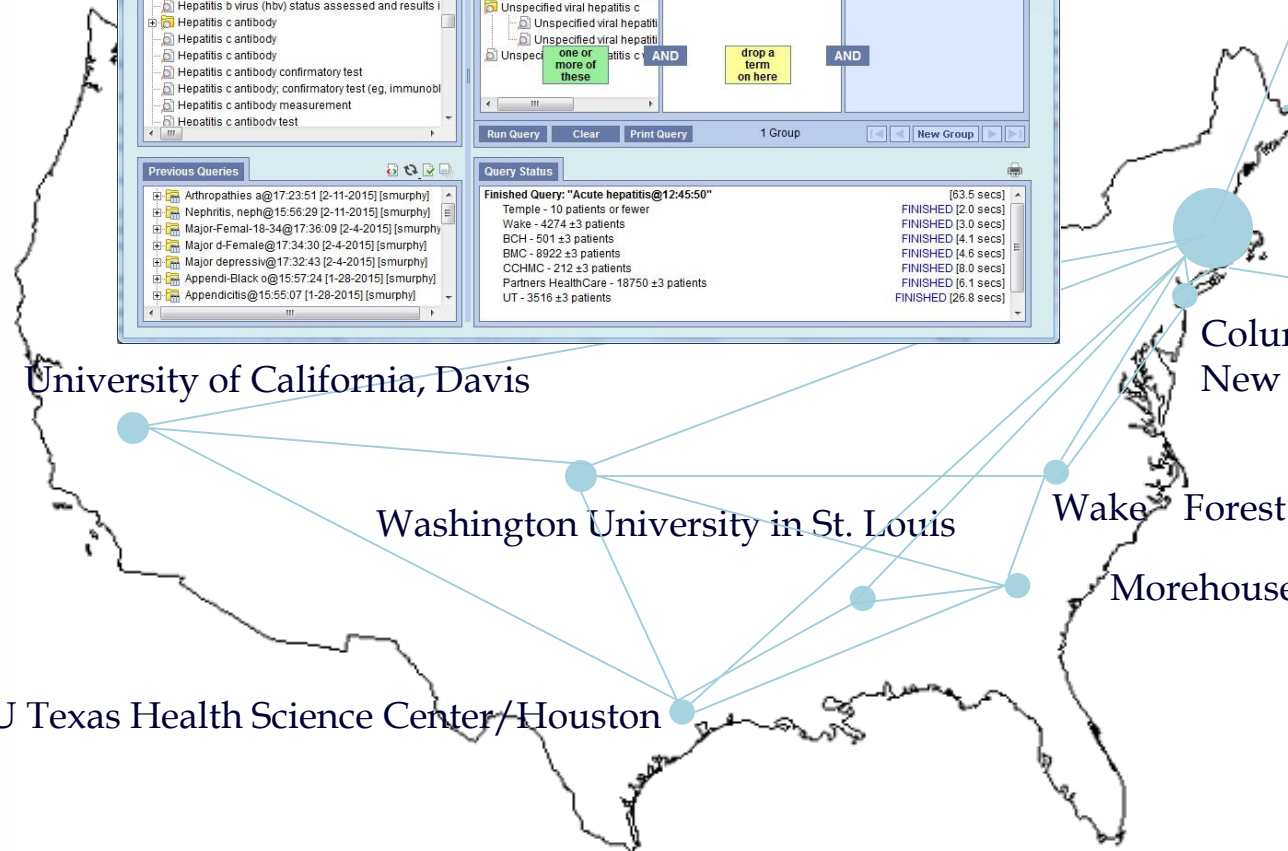
Washington University in St. Louis

Wake Forest Baptist Medical Center

Morehouse/Grady/RCMI

Columbia U. Medical Center and New York Presbyterian Hospital

U Texas Health Science Center/Houston



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 Deaconess 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 Cincinnati 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:

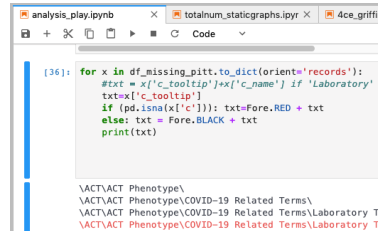
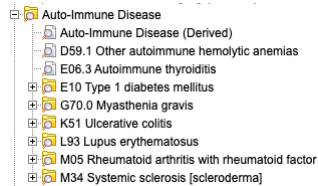
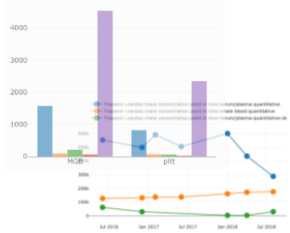
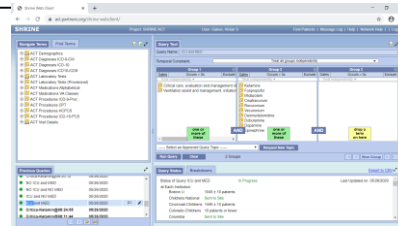
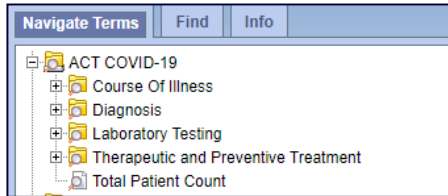
- Georges Pompidou Hospital, Paris, France
- Hospital of the Free University of Brussels, Belgium
- Inserm U936, Rennes, France
- Institute for Data Technology and Informatics (IDI), NTNU, Norway
- Institute for Molecular Medicine Finland (FIMM)
- Karolinska Institute, Sweden
- Landspítali University Hospital, Reykjavik, Iceland
- Tokyo Medical and Dental University, Japan
- University of Bordeaux 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:

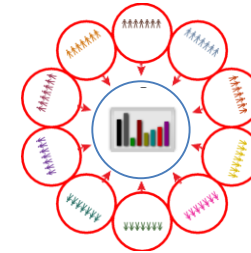
- Johnson and Johnson (TransMART)
- GE Healthcare Clinical Data Services

Research Pipeline for ACT

Query for sites with adequate data



```
[36]: for x in df_missing_pitt.to_dict(orient='records'):  
      #txt = x['c_tooltip']*x['c_name'] if 'Laboratory' in  
      txt=x['c_tooltip']  
      if (pd.isna(x['c'])): txt=Fore.RED + txt  
      else: txt = Fore.BLACK + txt  
      print(txt)  
  
\\ACT\ACT Phenotype\  
\\ACT\ACT Phenotype\COVID-19 Related Terms\  
\\ACT\ACT Phenotype\COVID-19 Related Terms\Laboratory Te  
\\ACT\ACT Phenotype\COVID-19 Related Terms\Laboratory Te
```



Quality

Computed
Phenotype

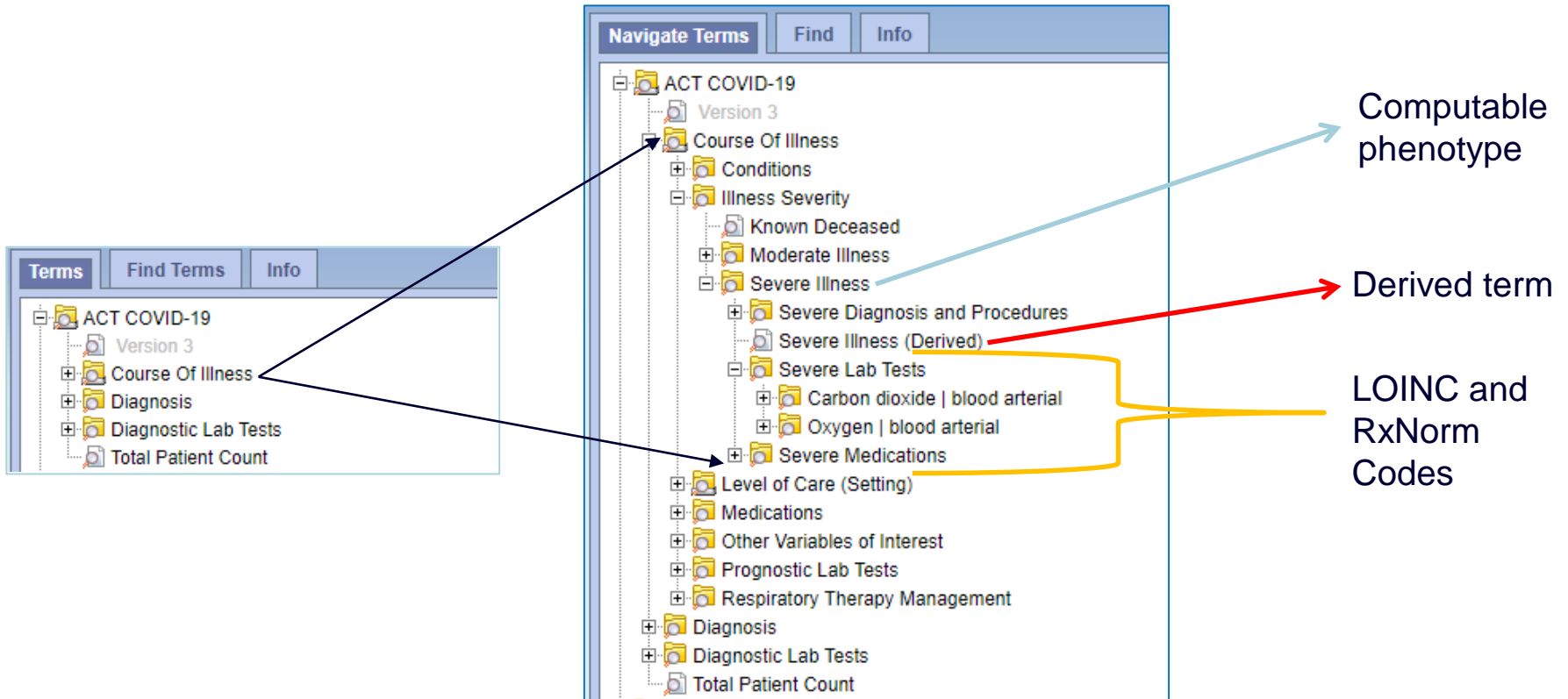
Local
Analytics

Validation

Pooling
Results

Publication

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*

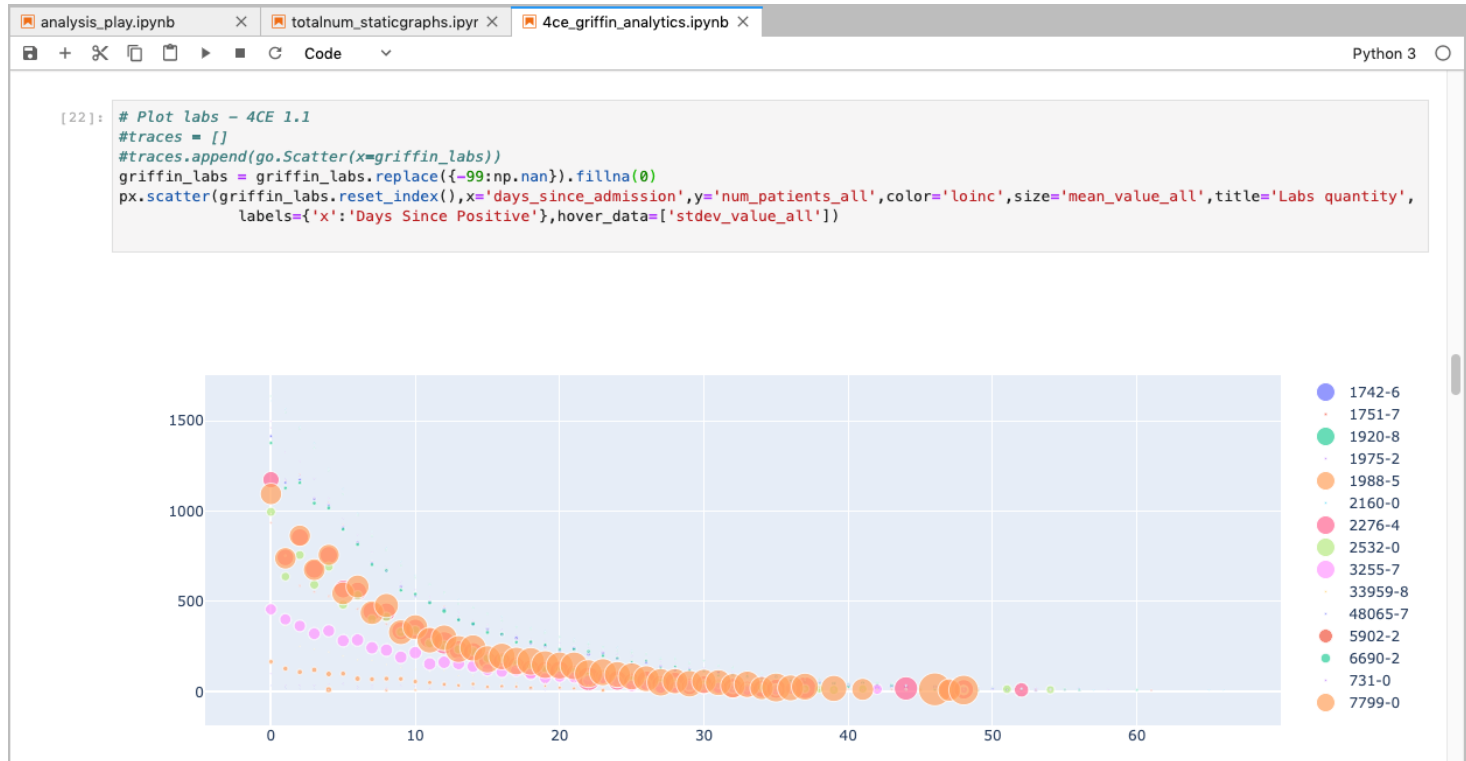
Append concepts to the list below

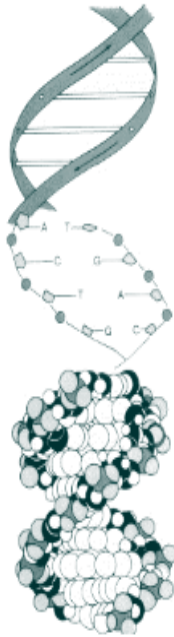
| Concept | Constraints | Aggregation Option | Include In Request |
|-----------------------------|-------------|--------------------|-------------------------------------|
| Patient Number | | Value | <input type="checkbox"/> |
| Gender | | Value | <input checked="" type="checkbox"/> |
| Age | | Value | <input checked="" type="checkbox"/> |
| Race | | Value | <input checked="" type="checkbox"/> |
| 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





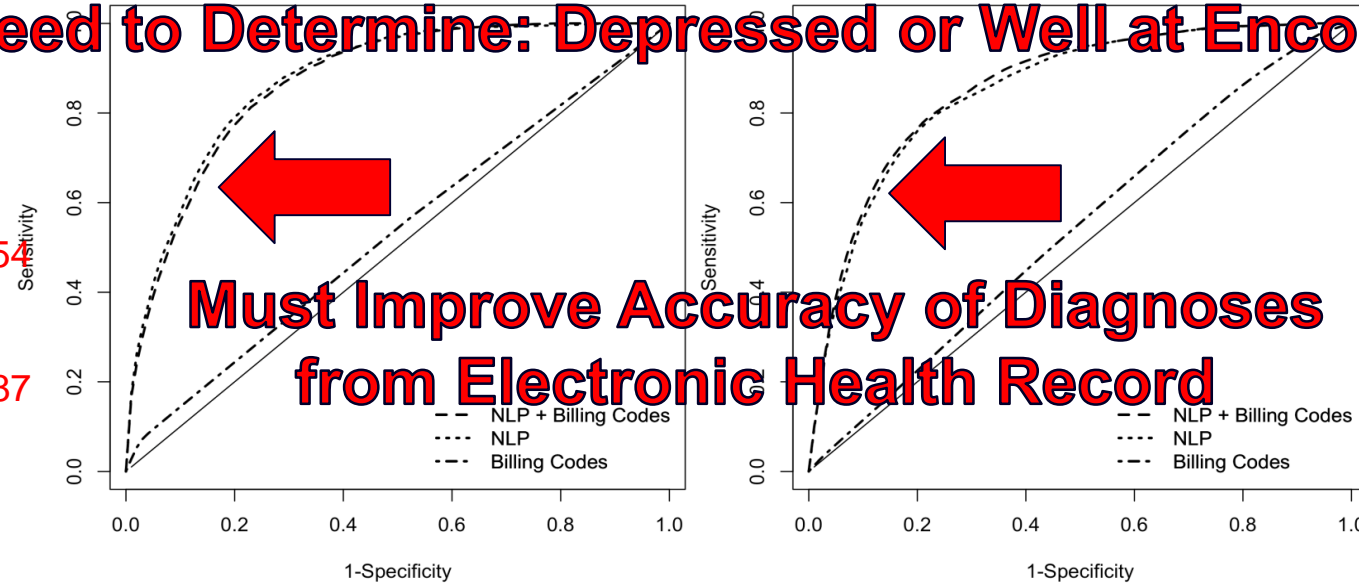
Improving Quality of i2b2 Queries through Machine Learning

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. Castro⁴, S. N. Murphy⁵, V. S. Gainer⁴, J. Minnier⁶, T. Cai⁶,
S. Goryachev⁴, Q. Zeng⁷, P. J. Gallagher², M. Fava¹, J. B. Weillburg¹, S. E. Churchill⁸,
I. S. Kohane⁹ and J. W. Smoller²

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

Need to Determine: Depressed or Well at Encounter



| 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.

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. CHRONIC BRONCHITIS
2. HEART FAILURE

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

HISTORY OF PRESENT ILLNESS: **Negative for tobacco**, alcohol, and IV drug abuse.

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

BRIEF RESUME OF HOSPITAL COURSE:
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 alcohol use. **Unclear smoking history ???**

PHYSICAL EXAMINATION: Temperature 97.2, pulse 66, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils equal and reactive to light.

LABORATORY DATA: Sodium 148, potassium 2.1, chloride 87, bicarbonate 24, glucose 108, creatinine 1.2, BUN 18, Hgb 12.5, Hct 38, WBC 12,000, platelets 250,000.

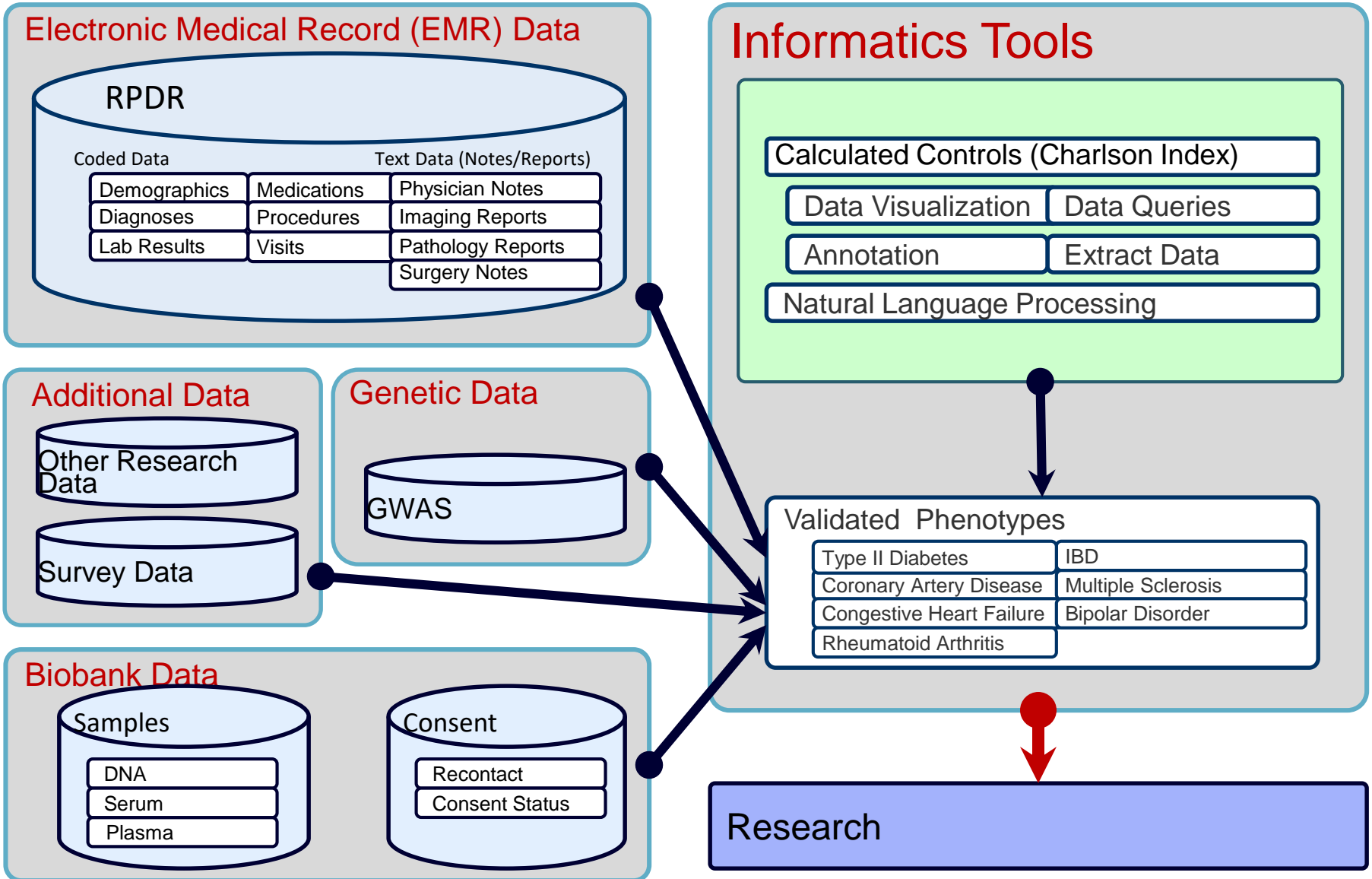
HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of **Lactobacillus acidophilus** to attempt a repopulation of her gut. **Hard to pick**

HOSPITAL COURSE: The patient was seen and evaluated by the physician on 10/10/77. She was discharged home on 10/10/77 to return to a 14-day course of treatment.

The patient is a widow, lives alone, 2 children, no **tob/alcohol**. **Hard to pick**

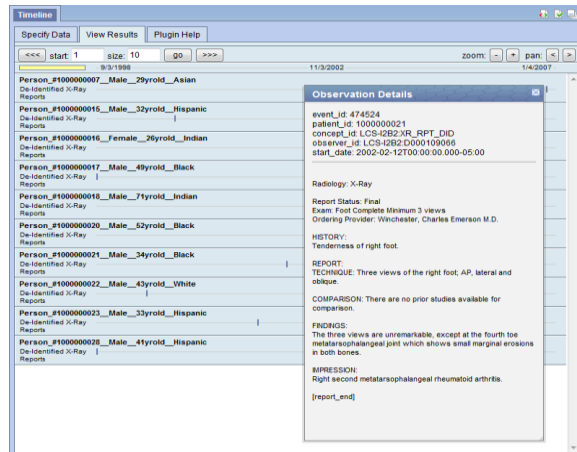
Ln 44 Col 1 | 274 | |WR| |Rec Off|No Wrap|DOS|INS|NUM

Data Integration in Biobank Portal

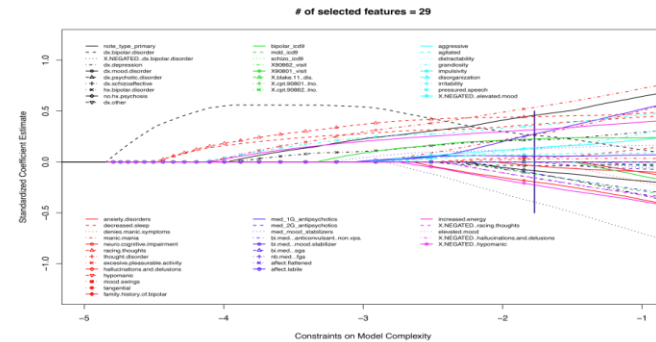


Curating a Disease Algorithm with a Gold Standard

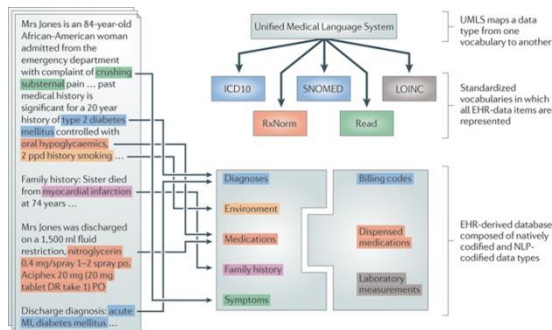
1. Create a gold standard training set.



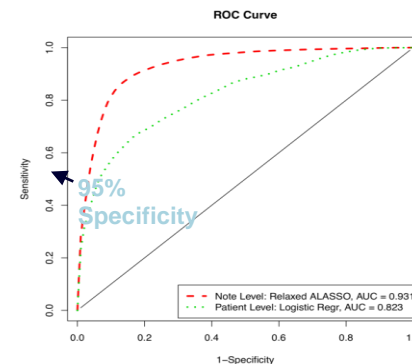
3. Develop the classification algorithm. Using the data analysis file and the training set from step 1, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm



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

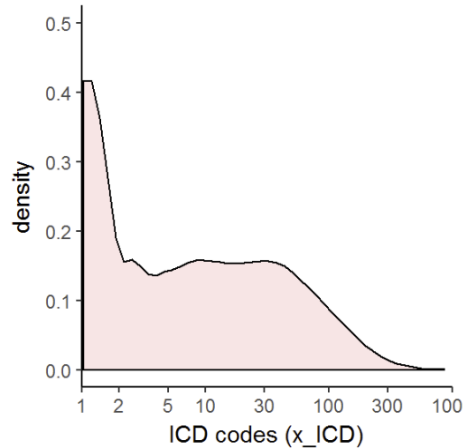


4. Apply the algorithm to all subjects in the superset and assign each subject a probability of having the phenotype

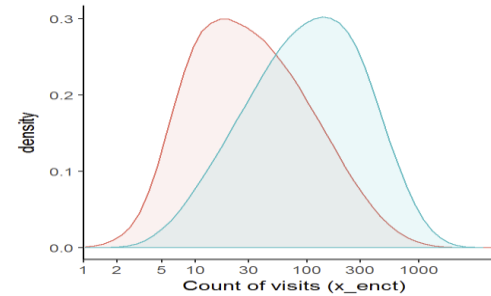


Curating a Disease Algorithm with a Silver Standard

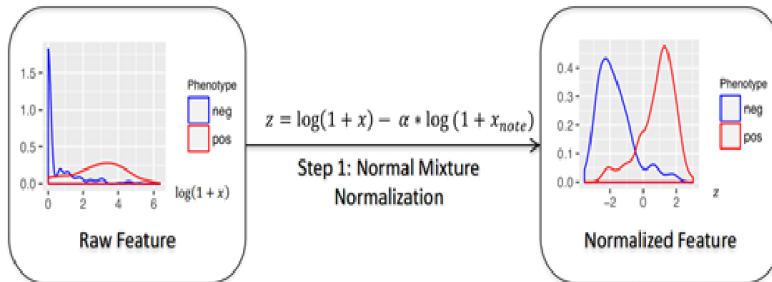
1. Query for total number of mentions of disease



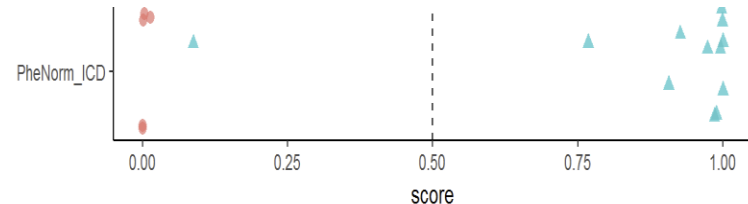
3. Resolve the curves and separate into two groups, these are actually equal to patients with and without the disease



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



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





Computable Phenotype Dashboard

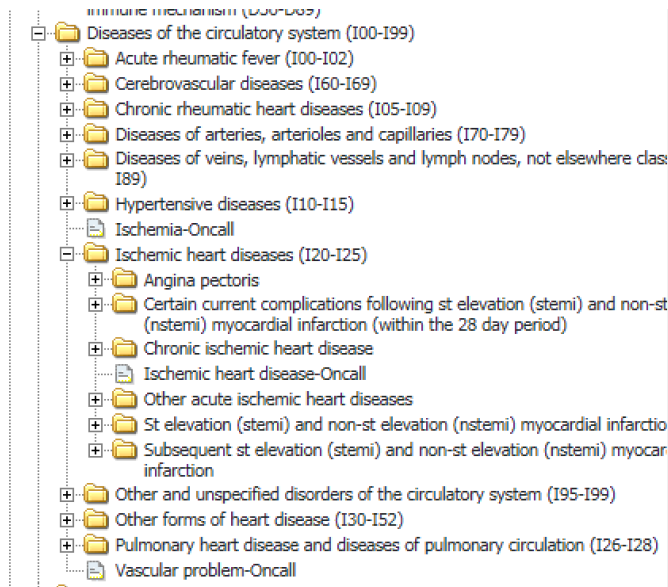
Phenotypes [Methods](#) [About us](#)

| | 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 |

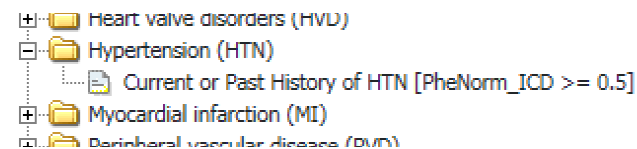
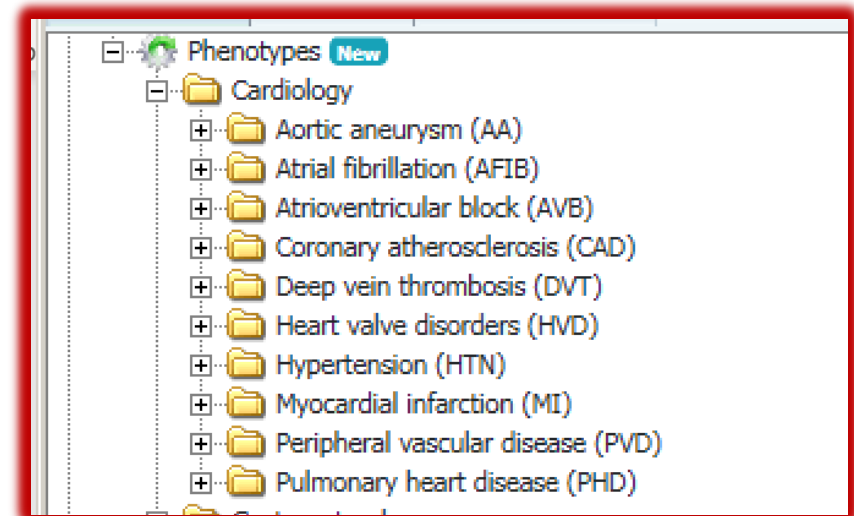
RESULT

Accurate and Simple Disease Labels for Queries

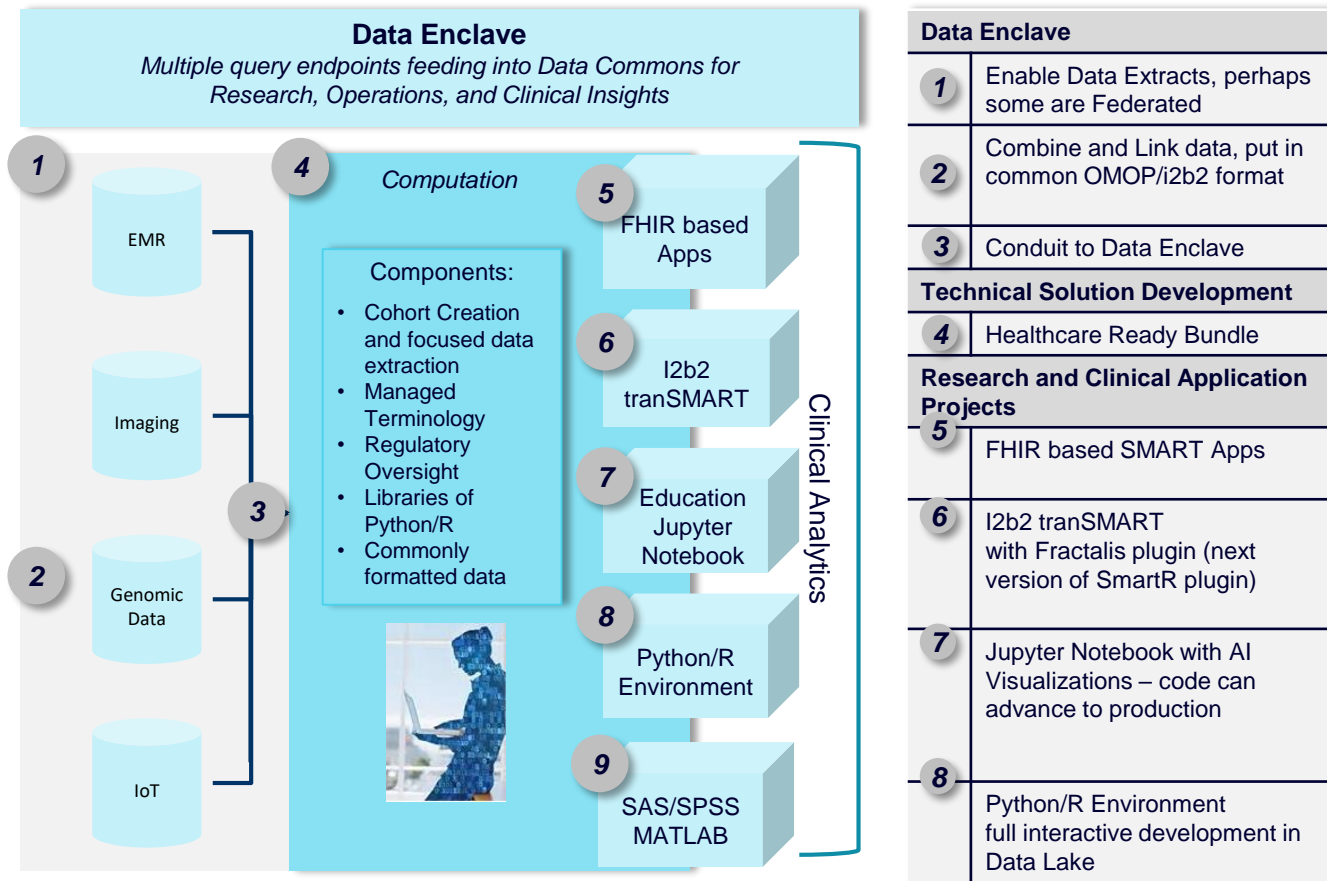
Complicated

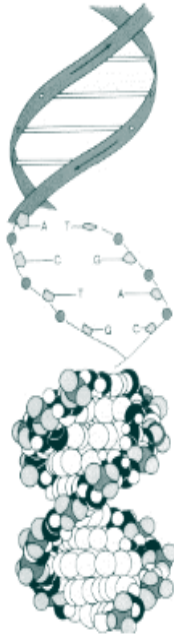


Simple



Digital Twin for Continuously Assessing Patient





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