

Improving EHR Data Quality with Automated Phenotyping

Shawn Murphy MD, Ph.D.

Partners Healthcare and Harvard Medical School

October 10th, 2019

Integration strategy from “Informatics for Integrating Biology and the Bedside (i2b2)” sponsored by the National Institutes of Health, 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>

Interrogation can occur through i2b2 web client

The screenshot displays the i2b2 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 toolbar contains various icons for navigation and search, including a Google search bar. The i2b2 Web Client interface is titled "i2b2 Query & Analysis Tool" and includes a navigation pane on the left with sections for "Navigate Terms" (showing a tree structure of Visit, Clinic, Hospital, Brigham and Womens, Faulkner, Massachusetts General, and Newton Wellesley) and "Workplace" (showing folders for lcp5 and SHARED). The main content area is divided into several sections: "Query Tool" (containing a "Query Name" field, a "Temporal Constraint" dropdown set to "Selected groups occur in the same financial encounter", and three query groups: Group 1 (Brigham and Womens), Group 2 (Malignant melanoma of sk), and Group 3 (Female)), "Previous Queries" (listing several queries related to "Arterial vascul@02:17:06 [3-4-2011] [lcp5]"), and "Query Status" (empty). The interface also includes a "Run Query" button, a "Clear" button, a "Print Query" button, and a "New Group" button. The status bar at the bottom indicates "Done, but with errors on page."

Running a Query

The screenshot displays the Partners Clinical Image Bank web application interface. The browser address bar shows the URL https://cib.partners.org/webclient_multi/. The application header includes the Partners Clinical Image Bank logo and navigation links: Find Patients, Make Table of Data, Get Image Sets, Help, and SNM0.

The main interface is divided into two main sections: a left sidebar for navigating terms and a right pane for the query tool.

Left Sidebar (Navigate Terms):

- Data from Healthcare Record
 - Demographics
 - Diagnosis - 412
 - Diagnosis ICD-10 - 429
 - Diagnosis Related Groups - 302
 - Encounters
 - Health history - 385
 - Lab tests - 365
 - Medications - 397
 - Medications (RxNorm) - 395
 - Microbiology - 293
 - [Molecular medicine - 0]
 - Procedures - 429
 - PROMs - 1
 - Providers
 - Radiology tests - 392
 - Reason for visit - 135
 - [Specimens - 0]
 - Transfusion - 176
- Imaging Data
 - Apparent Diffusion Coefficient
 - DICOM Data - 403
 - Images
- Perinatal Data
 - 1. Vital Statistics - 162
 - 2. Maternal History - 162
 - 3. Labor and Delivery - 162
 - 1. Duration, second stage of labor - 8
 - 2. Rupture of membranes - 159
 - 3. Was there chorioamnionitis? - 162
 - 4. Was there meconium in the amniotic fluid? - 162
 - 5. Mode of Delivery - 161
 - 6. Acute perinatal event - 162

Right Pane (Query Tool):

Query Name: Image-Epile-6. Ac@14:59:19

Query Timing: Non-Temporal Query: Treat all groups independently

The query is structured into three groups, each with a table of criteria:

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
Images			Epilepsy and recurrent seizures - 123			6. Acute perinatal event - 162		


Below the groups, there are green boxes indicating "one or more of these" for each group, connected by "AND" operators.

Buttons: Run Query, Clear, 3 Groups, New Group

Bottom Section (Query Results):

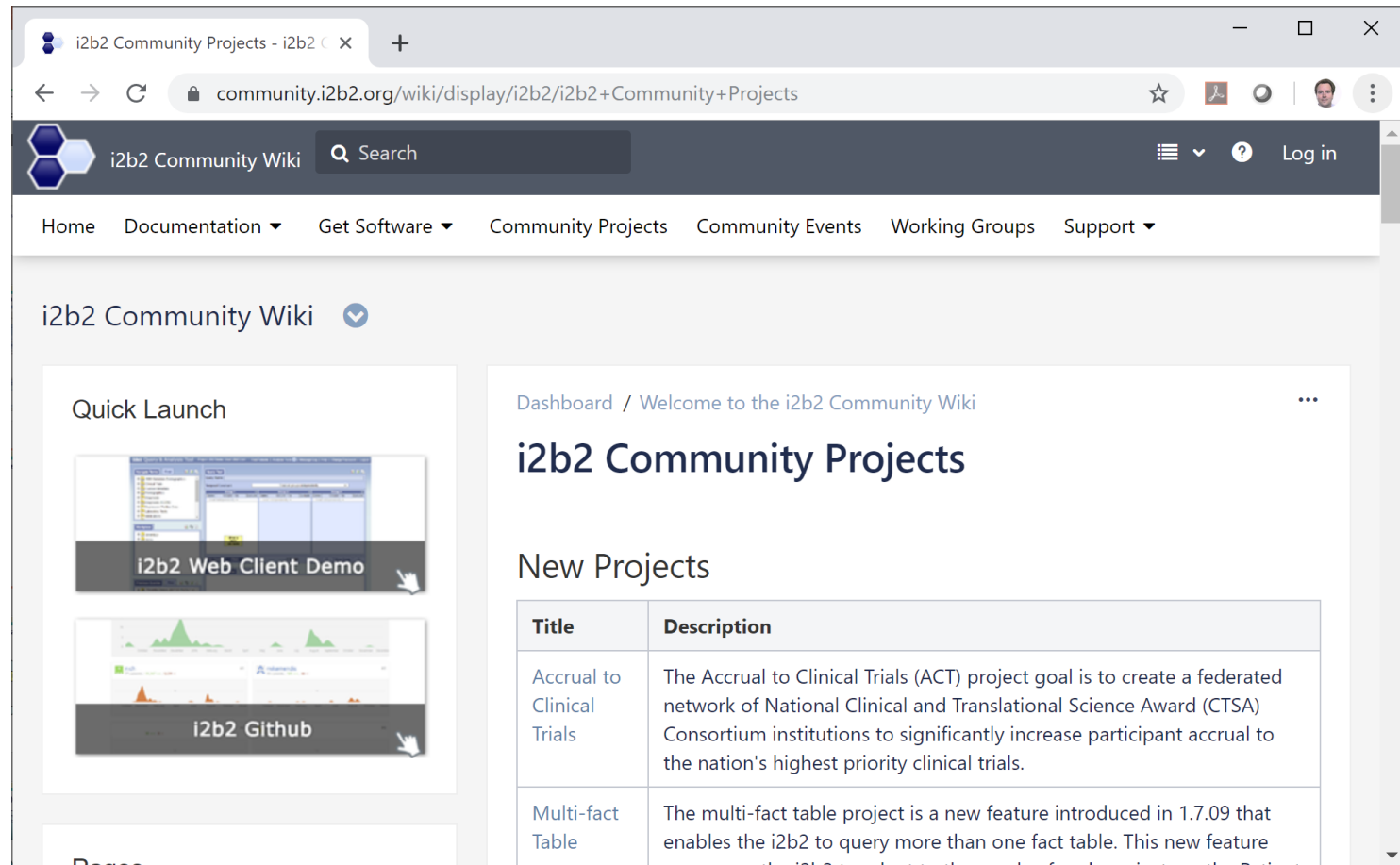
Show Query Status | Graph Results | Query Report

NUMBER OF PATIENTS

 **29**

For Query "Image-Epile-6. Ac@14:59:19"

I2b2 Community Software Modules contributed as “Cells”

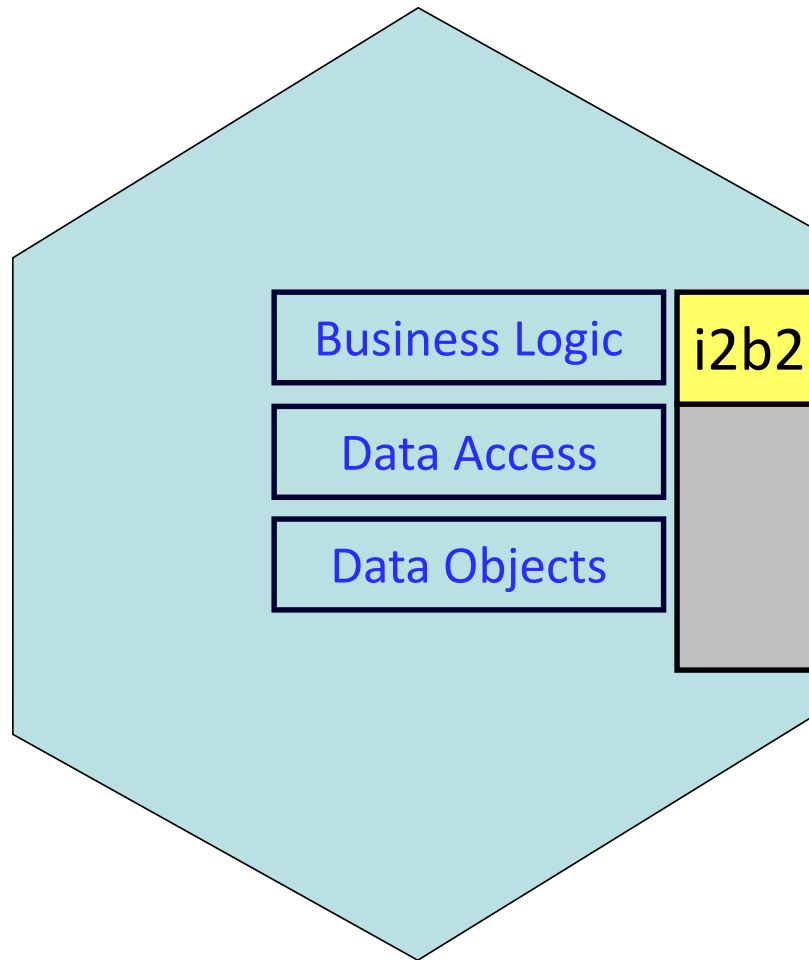


The screenshot shows a web browser window displaying the I2b2 Community Wiki. The address bar shows the URL community.i2b2.org/wiki/display/i2b2/i2b2+Community+Projects. The page header includes the I2b2 logo, a search bar, and a navigation menu with links: Home, Documentation, Get Software, Community Projects, Community Events, Working Groups, and Support. The main content area is titled 'i2b2 Community Wiki' and features a 'Quick Launch' section with two tiles: 'i2b2 Web Client Demo' and 'i2b2 Github'. To the right, there is a section titled 'i2b2 Community Projects' with a subtitle 'New Projects' and 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

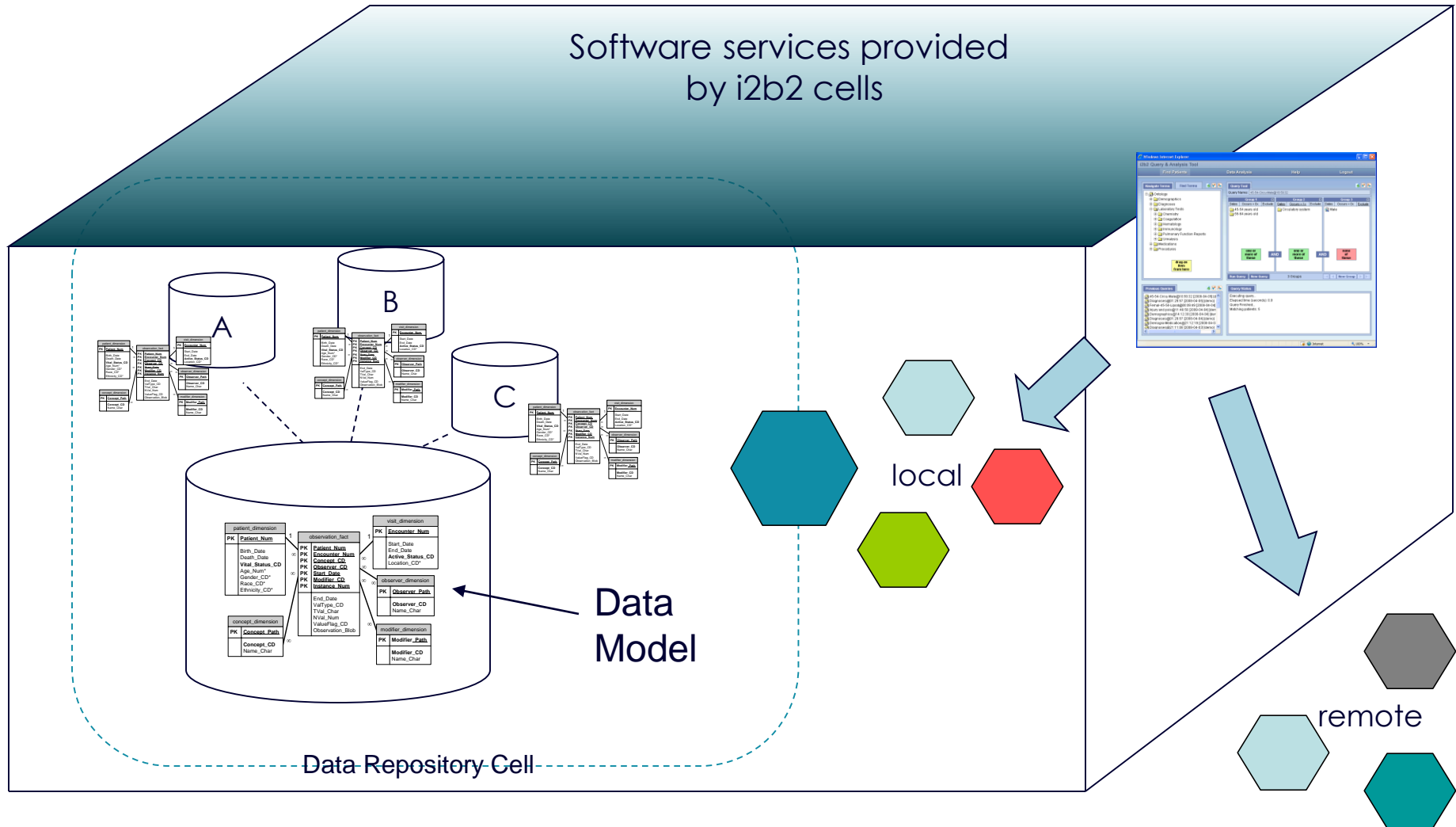
<https://community.i2b2.org/wiki/display/i2b2/i2b2+Community+Projects>

i2b2 Cell: The Canonical Software Module

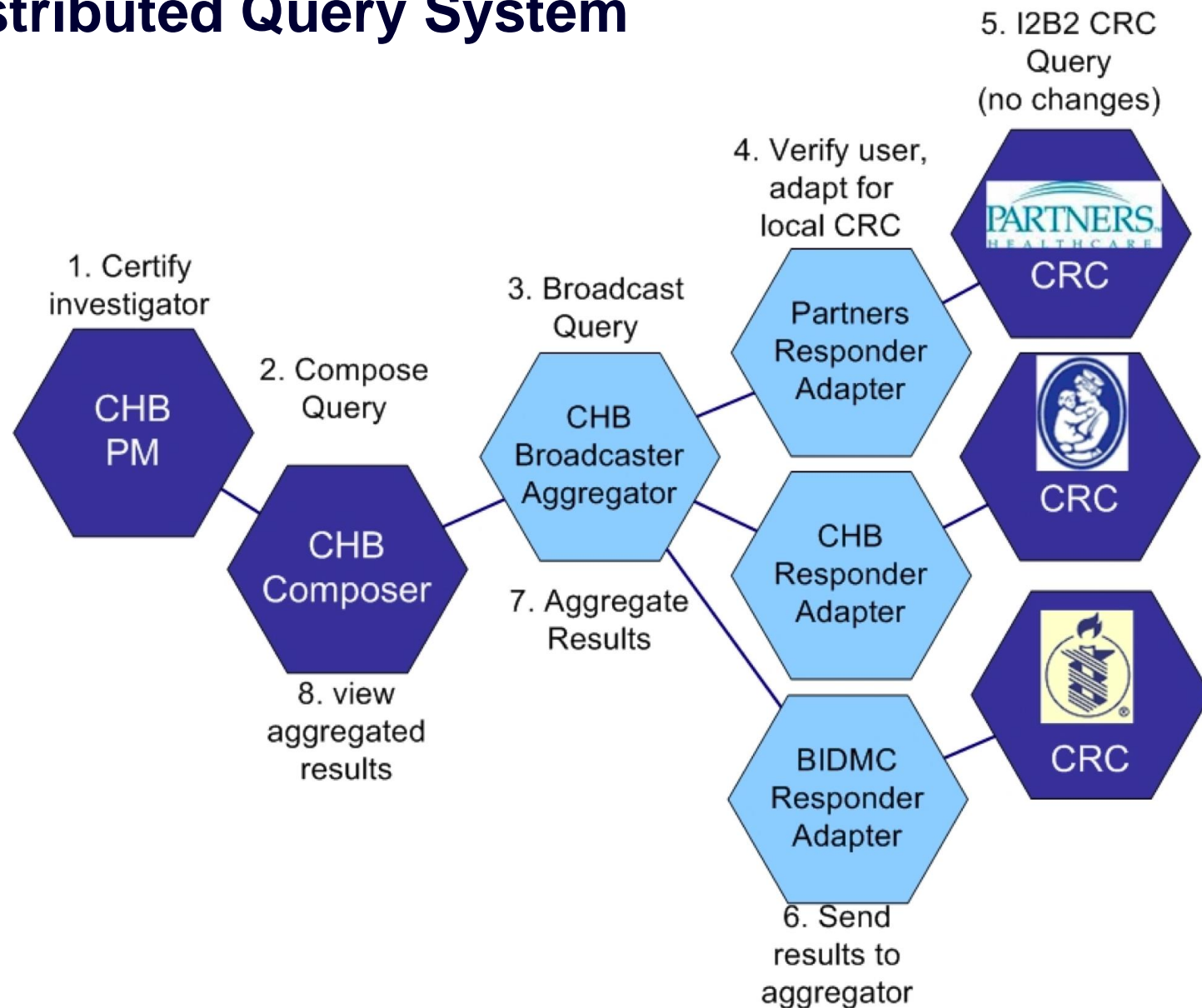


↔
HTTP XML
(minimum: RESTful)

An i2b2 Environment is built from i2b2 Cells



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 Any Category

Find

- 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

Previous Queries

- Arthropathies a@17:23:51 [2-11-2015] [smurphy]
- Nephritis, neph@15:56:29 [2-11-2015] [smurphy]
- Major-Femal-18-34@17:36:09 [2-4-2015] [smurphy]
- Major d-Female@17:34:30 [2-4-2015] [smurphy]
- Major depressiv@17:32:43 [2-4-2015] [smurphy]
- Append-Black o@15:57:24 [1-28-2015] [smurphy]
- Appendicitis@15:55:07 [1-28-2015] [smurphy]

Query Tool

Query Name: Acute hepatitis@12:45:50

Group 1	Group 2	Group 3
Dates	Dates	Dates
Occurs > 0x	Occurs > 0x	Occurs > 0x
Exclude	Exclude	Exclude
Treat independently	Treat independently	Treat independently
Acute hepatitis c with hepatic co		
Acute hepatitis c without mentio		
Unspecified viral hepatitis c		
Unspecified viral hepatitis		
Unspec one or more of these	AND	
	drop a term on here	AND

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 [14.6 secs]
COH-MC - 212 ±3 patients	FINISHED [9.0 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)

Columbia U. Medical Center and New York Presbyterian Hospital

University of California, Davis

Washington University in St. Louis

Wake Forest Baptist Medical Center

Morehouse/Grady/RCMI

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

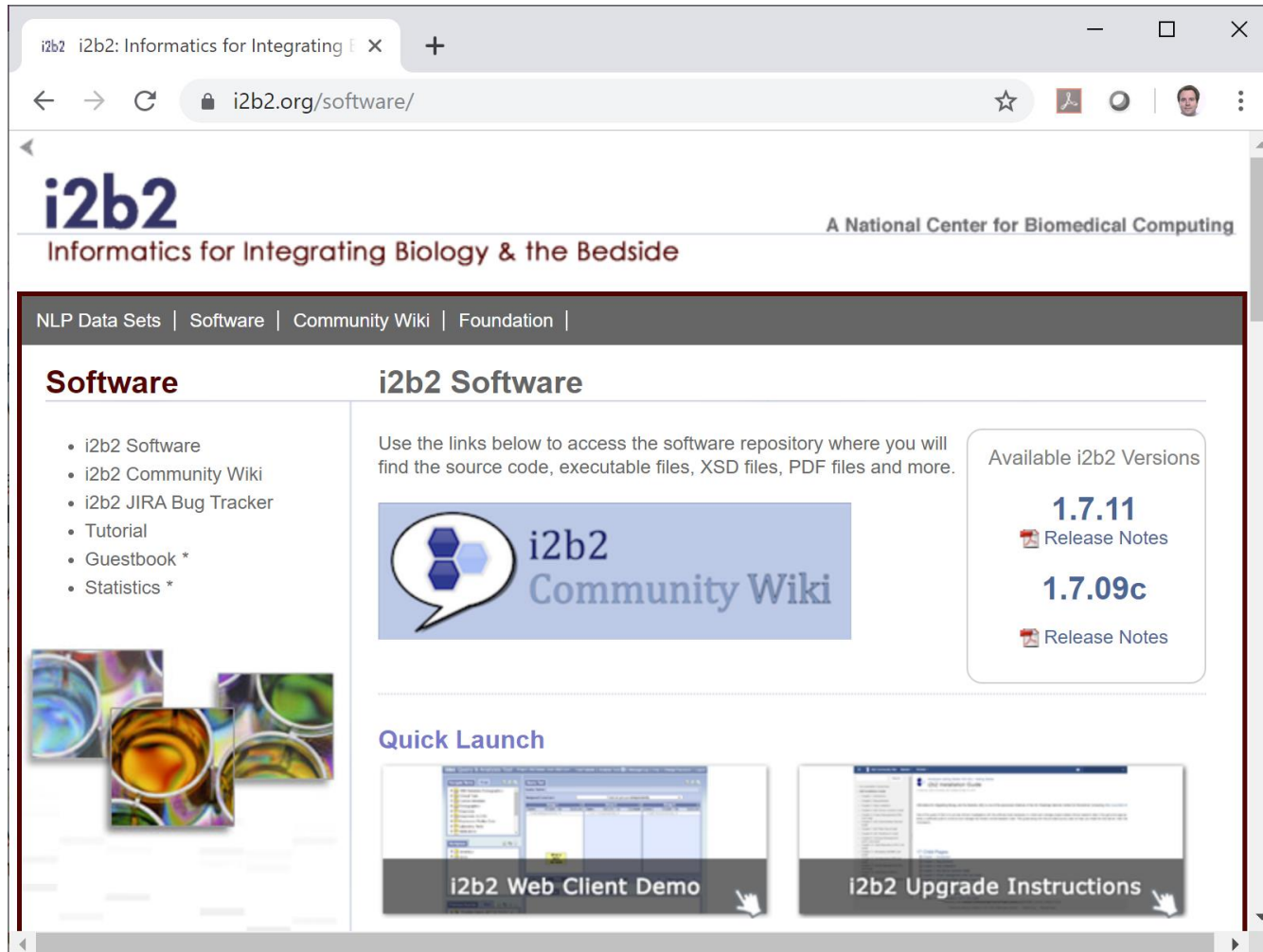
International:

- 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

I2b2 Software components are distributed as open source



The screenshot shows a web browser window with the address bar displaying "i2b2: Informatics for Integrating" and "i2b2.org/software/". The page features the i2b2 logo and the tagline "A National Center for Biomedical Computing". A navigation bar includes links for "NLP Data Sets", "Software", "Community Wiki", and "Foundation". The main content area is titled "i2b2 Software" and includes a list of software components, a description of the software repository, a list of available versions (1.7.11 and 1.7.09c), and a "Quick Launch" section with links to the "i2b2 Web Client Demo" and "i2b2 Upgrade Instructions".

i2b2
Informatics for Integrating Biology & the Bedside

A National Center for Biomedical Computing


NLP Data Sets | Software | Community Wiki | Foundation |

Software



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

i2b2 Software

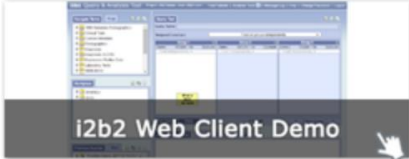

Use the links below to access the software repository where you will find the source code, executable files, XSD files, PDF files and more.

 i2b2 Community Wiki

Available i2b2 Versions

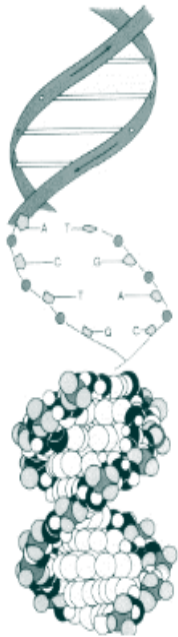
- 1.7.11**
 Release Notes
- 1.7.09c**
 Release Notes

Quick Launch

-  i2b2 Web Client Demo
-  i2b2 Upgrade Instructions

Recent Community Project Releases

- Incorporation of FHIR into i2b2
 - FHIR Cell to allow single patient's data to be returned in FHIR
 - FHIR Ontology and ETL to allow direct import of FHIR into i2b2
 - FHIR Cell to allow i2b2 to extend to FHIR query endpoints
- I2b2-based system for Accruing Patients for Clinical Trials
 - Extensions to SHRINE for management of clinical trials
 - Web Client Plug-ins to extend SHRINE queries to local i2b2-based patient recruitment
- i2b2-based system to query and return data from Observation-Fact tables contained in multiple different i2b2 Hives



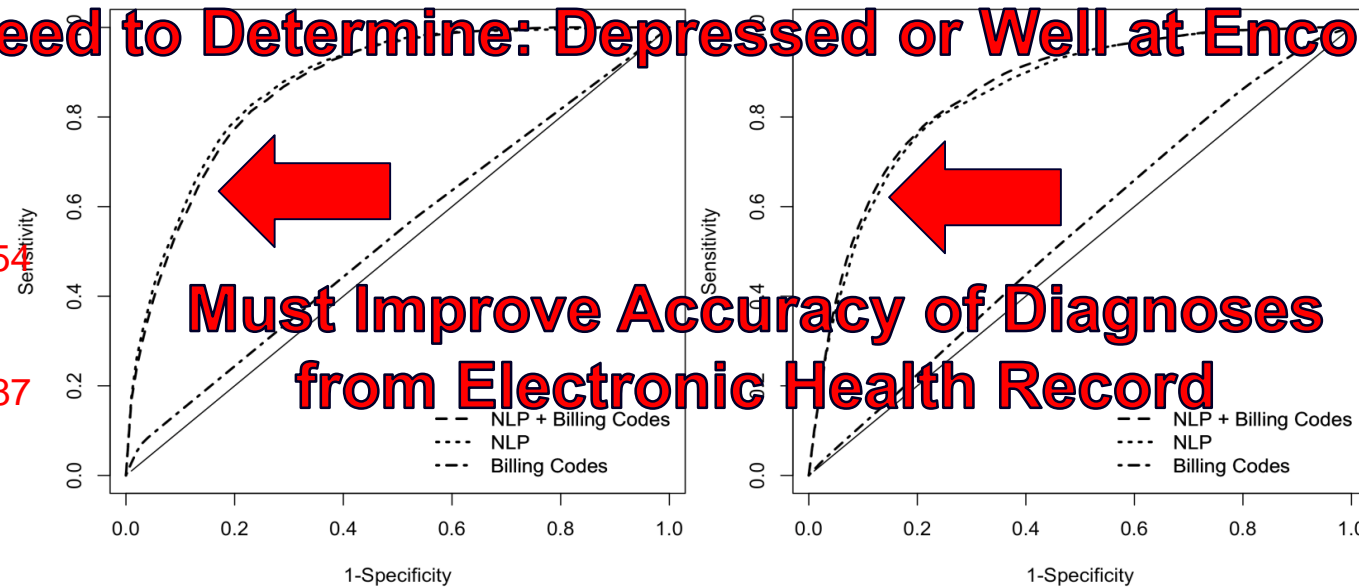
Improve Quality of i2b2 Queries through Machine Learning

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



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.

The image shows a screenshot of a text editor window titled "Programmer's File Editor - [050210_1629\MiniDem1.txt]". The text contains several patient notes with various annotations. Red boxes highlight specific phrases, and yellow arrows point to them with labels indicating the NLP extraction results.

Annotation 1: A red box highlights "uses tobacco" in the text "SOCIAL HISTORY: The patient is married with four grown daughters, uses tobacco has wine with dinner." A yellow arrow points to it with the label "Smoker".

Annotation 2: A red box highlights "nonsmoker." in the text "SOCIAL HISTORY: The patient is a nonsmoker. No alcohol." A yellow arrow points to it with the label "Non-Smoker".

Annotation 3: A red box highlights "Negative for tobacco," in the text "SOCIAL HISTORY: Negative for tobacco, alcohol, and IV drug abuse." A yellow arrow points to it with the label "Non-Smoker".

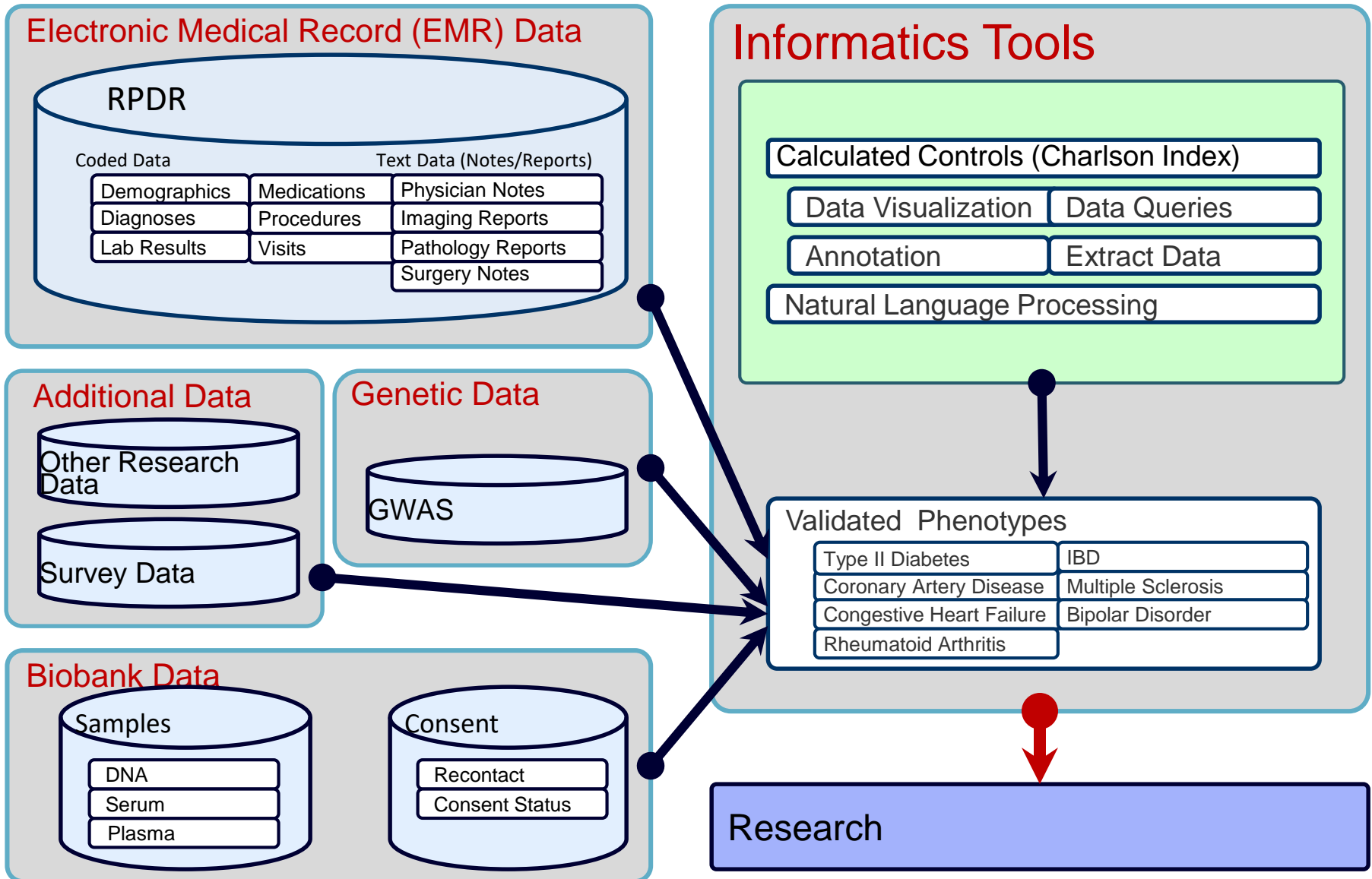
Annotation 4: A red box highlights "50 pack-yr tobacco (quit 3 wks ago)," in the text "BRIEF RESUME OF HOSPITAL COURSE: 63 yo woman with COPD, 50 pack-yr tobacco (quit 3 wks ago), spir..." A yellow arrow points to it with the label "Past Smoker".

Annotation 5: A red box highlights "Unclear smoking history" in the text "SOCIAL HISTORY: The patient lives in rehab, married. Unclear smoking history from the admission note..." A yellow arrow points to it with the label "???".

Annotation 6: A red box highlights "Lactobacillus" in the text "HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of Lactobacillus acidophilus to attempt a repopulation of her gut." A yellow arrow points to it with the label "Hard to pick".

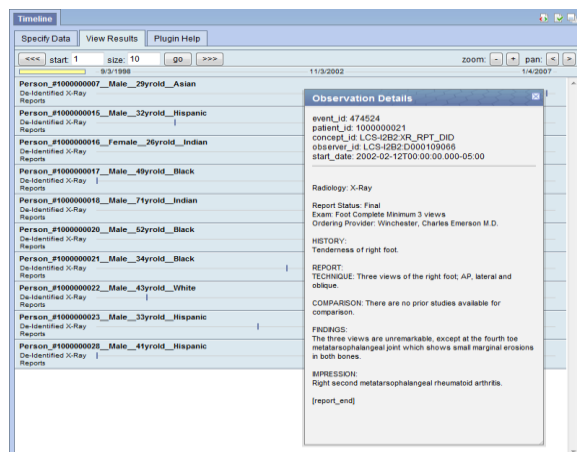
Annotation 7: A red box highlights "tob/alcohol." in the text "SH: widow,lives alone,2 children,no tob/alcohol." A yellow arrow points to it with the label "Hard to pick".

Data Integration in Biobank Portal

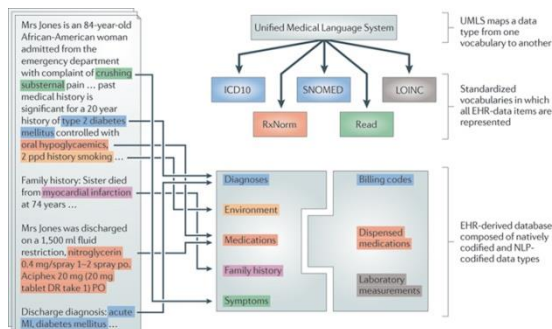


Curating a Disease Algorithm with a Gold Standard

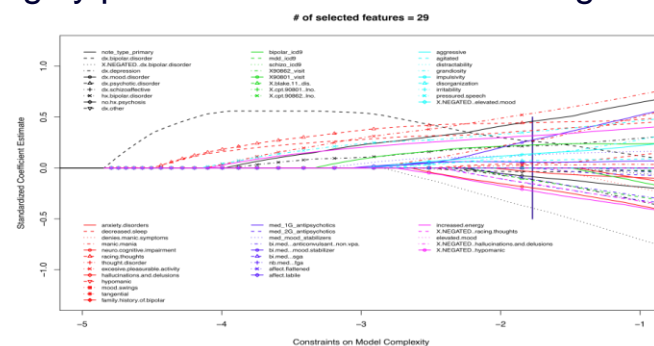
- ## 1. Create a gold standard training set.



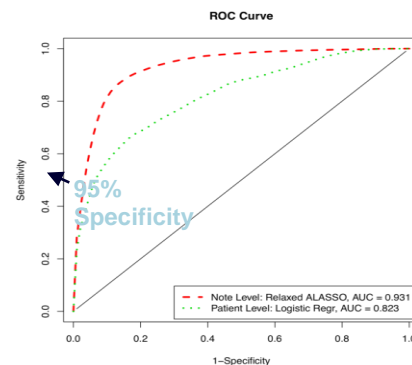
- ## 2. Create a comprehensive list of features from patient's electronic data that describe the disease of interest



- 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

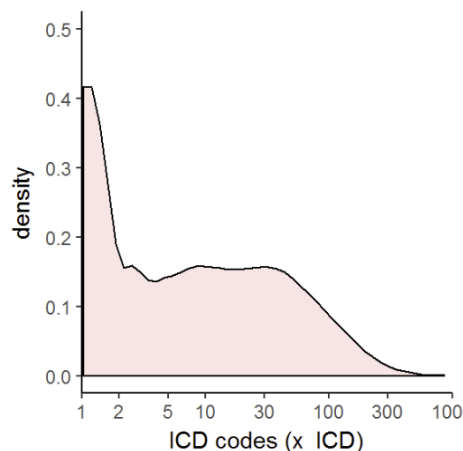


- 4. Apply the algorithm to all subjects** in the superset and assign each subject a probability of having the phenotype

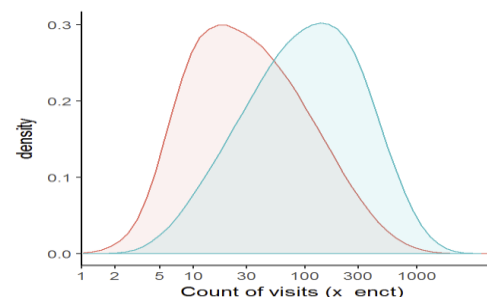


Curating a Disease Algorithm with a Silver Standard

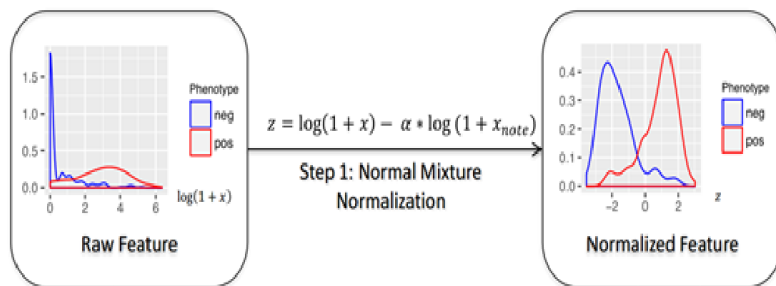
1. Query for total number of mentions of disease



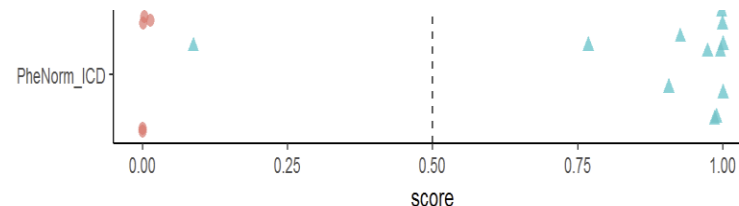
3. Resolve the curves and separate into two groups, these are actually equal to patients with and without the disease



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

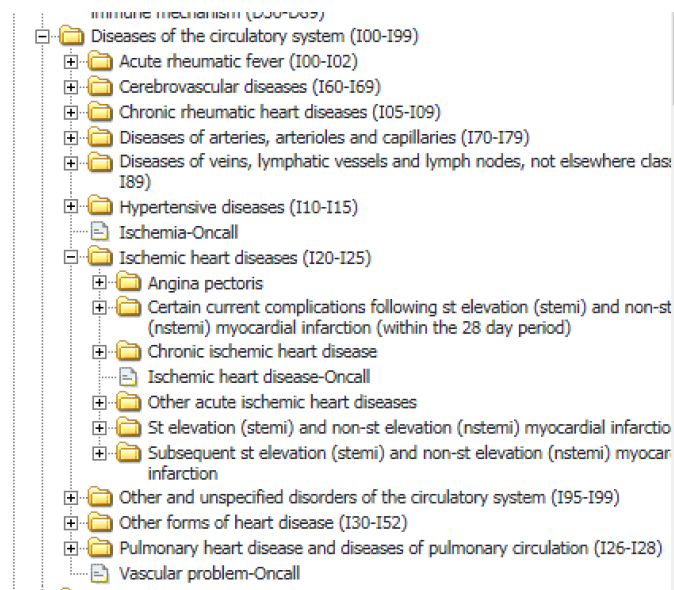
[Phenotypes](#)[Methods](#)[About us](#)

	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

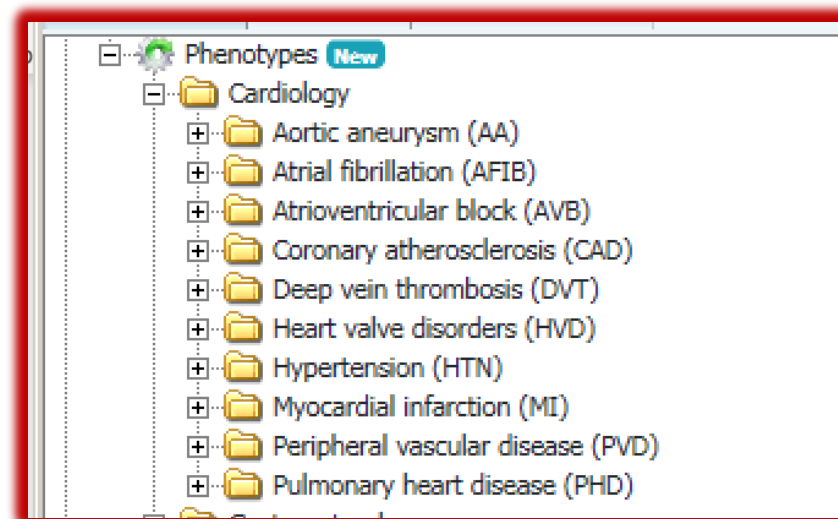
RESULT

Accurate and Simple Disease Labels for Queries

Complicated



Simple



Heart valve disorders (HVD)
Hypertension (HTN)
Current or Past History of HTN [PheNorm_ICD >= 0.5]
Myocardial infarction (MI)
Peripheral vascular disease (PVD)

In Digitally Driven Healthcare, Disease Labels determine Algorithms for Managing Patient

 Mary

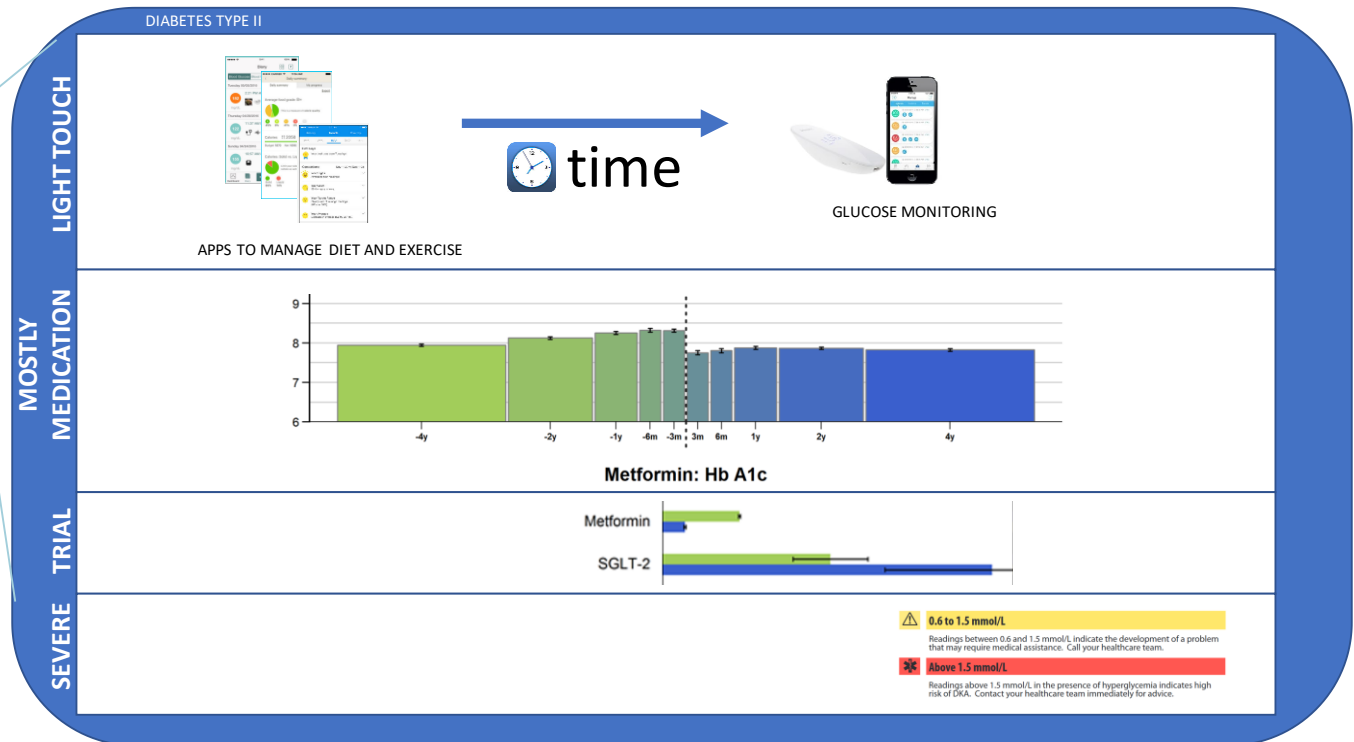
Followed for:

Prevention

Diabetes Type II

Hypertension

Heart Failure



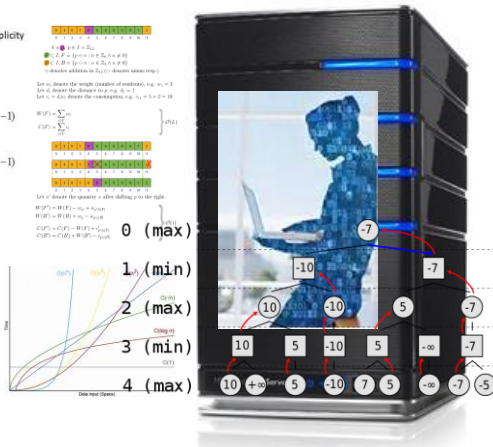
Innovations Enabling Digital Care

Digital and IoT devices continuously output Patient Data

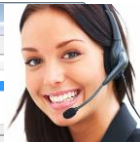
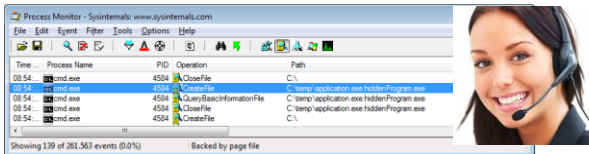


```
1. for ( i = 1; i ≤ a1.length; i++ )
  1.1. j = 1
  1.2. while ( a1[i] != a2[j] )
    if ( j ≥ a1.length )
      1. return false
    1.1. j++
  1.3. tmp = a2[j]
  1.4. a2[j] = a2[i]
  1.5. a2[i] = tmp
2. return true
```

```
cost
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20
C21
C22
C23
C24
C25
C26
C27
C28
C29
C30
C31
C32
C33
C34
C35
C36
C37
C38
C39
C40
C41
C42
C43
C44
C45
C46
C47
C48
C49
C50
C51
C52
C53
C54
C55
C56
C57
C58
C59
C60
C61
C62
C63
C64
C65
C66
C67
C68
C69
C70
C71
C72
C73
C74
C75
C76
C77
C78
C79
C80
C81
C82
C83
C84
C85
C86
C87
C88
C89
C90
C91
C92
C93
C94
C95
C96
C97
C98
C99
C100
C101
C102
C103
C104
C105
C106
C107
C108
C109
C110
C111
C112
C113
C114
C115
C116
C117
C118
C119
C120
C121
C122
C123
C124
C125
C126
C127
C128
C129
C130
C131
C132
C133
C134
C135
C136
C137
C138
C139
C140
C141
C142
C143
C144
C145
C146
C147
C148
C149
C150
C151
C152
C153
C154
C155
C156
C157
C158
C159
C160
C161
C162
C163
C164
C165
C166
C167
C168
C169
C170
C171
C172
C173
C174
C175
C176
C177
C178
C179
C180
C181
C182
C183
C184
C185
C186
C187
C188
C189
C190
C191
C192
C193
C194
C195
C196
C197
C198
C199
C200
C201
C202
C203
C204
C205
C206
C207
C208
C209
C210
C211
C212
C213
C214
C215
C216
C217
C218
C219
C220
C221
C222
C223
C224
C225
C226
C227
C228
C229
C230
C231
C232
C233
C234
C235
C236
C237
C238
C239
C240
C241
C242
C243
C244
C245
C246
C247
C248
C249
C250
C251
C252
C253
C254
C255
C256
C257
C258
C259
C260
C261
C262
C263
C264
C265
C266
C267
C268
C269
C270
C271
C272
C273
C274
C275
C276
C277
C278
C279
C280
C281
C282
C283
C284
C285
C286
C287
C288
C289
C290
C291
C292
C293
C294
C295
C296
C297
C298
C299
C300
C301
C302
C303
C304
C305
C306
C307
C308
C309
C310
C311
C312
C313
C314
C315
C316
C317
C318
C319
C320
C321
C322
C323
C324
C325
C326
C327
C328
C329
C330
C331
C332
C333
C334
C335
C336
C337
C338
C339
C340
C341
C342
C343
C344
C345
C346
C347
C348
C349
C350
C351
C352
C353
C354
C355
C356
C357
C358
C359
C360
C361
C362
C363
C364
C365
C366
C367
C368
C369
C370
C371
C372
C373
C374
C375
C376
C377
C378
C379
C380
C381
C382
C383
C384
C385
C386
C387
C388
C389
C390
C391
C392
C393
C394
C395
C396
C397
C398
C399
C400
C401
C402
C403
C404
C405
C406
C407
C408
C409
C410
C411
C412
C413
C414
C415
C416
C417
C418
C419
C420
C421
C422
C423
C424
C425
C426
C427
C428
C429
C430
C431
C432
C433
C434
C435
C436
C437
C438
C439
C440
C441
C442
C443
C444
C445
C446
C447
C448
C449
C450
C451
C452
C453
C454
C455
C456
C457
C458
C459
C460
C461
C462
C463
C464
C465
C466
C467
C468
C469
C470
C471
C472
C473
C474
C475
C476
C477
C478
C479
C480
C481
C482
C483
C484
C485
C486
C487
C488
C489
C490
C491
C492
C493
C494
C495
C496
C497
C498
C499
C500
C501
C502
C503
C504
C505
C506
C507
C508
C509
C510
C511
C512
C513
C514
C515
C516
C517
C518
C519
C520
C521
C522
C523
C524
C525
C526
C527
C528
C529
C530
C531
C532
C533
C534
C535
C536
C537
C538
C539
C540
C541
C542
C543
C544
C545
C546
C547
C548
C549
C550
C551
C552
C553
C554
C555
C556
C557
C558
C559
C560
C561
C562
C563
C564
C565
C566
C567
C568
C569
C570
C571
C572
C573
C574
C575
C576
C577
C578
C579
C580
C581
C582
C583
C584
C585
C586
C587
C588
C589
C590
C591
C592
C593
C594
C595
C596
C597
C598
C599
C600
C601
C602
C603
C604
C605
C606
C607
C608
C609
C610
C611
C612
C613
C614
C615
C616
C617
C618
C619
C620
C621
C622
C623
C624
C625
C626
C627
C628
C629
C630
C631
C632
C633
C634
C635
C636
C637
C638
C639
C640
C641
C642
C643
C644
C645
C646
C647
C648
C649
C650
C651
C652
C653
C654
C655
C656
C657
C658
C659
C660
C661
C662
C663
C664
C665
C666
C667
C668
C669
C670
C671
C672
C673
C674
C675
C676
C677
C678
C679
C680
C681
C682
C683
C684
C685
C686
C687
C688
C689
C690
C691
C692
C693
C694
C695
C696
C697
C698
C699
C700
C701
C702
C703
C704
C705
C706
C707
C708
C709
C710
C711
C712
C713
C714
C715
C716
C717
C718
C719
C720
C721
C722
C723
C724
C725
C726
C727
C728
C729
C730
C731
C732
C733
C734
C735
C736
C737
C738
C739
C740
C741
C742
C743
C744
C745
C746
C747
C748
C749
C750
C751
C752
C753
C754
C755
C756
C757
C758
C759
C760
C761
C762
C763
C764
C765
C766
C767
C768
C769
C770
C771
C772
C773
C774
C775
C776
C777
C778
C779
C780
C781
C782
C783
C784
C785
C786
C787
C788
C789
C790
C791
C792
C793
C794
C795
C796
C797
C798
C799
C800
C801
C802
C803
C804
C805
C806
C807
C808
C809
C810
C811
C812
C813
C814
C815
C816
C817
C818
C819
C820
C821
C822
C823
C824
C825
C826
C827
C828
C829
C830
C831
C832
C833
C834
C835
C836
C837
C838
C839
C840
C841
C842
C843
C844
C845
C846
C847
C848
C849
C850
C851
C852
C853
C854
C855
C856
C857
C858
C859
C860
C861
C862
C863
C864
C865
C866
C867
C868
C869
C870
C871
C872
C873
C874
C875
C876
C877
C878
C879
C880
C881
C882
C883
C884
C885
C886
C887
C888
C889
C890
C891
C892
C893
C894
C895
C896
C897
C898
C899
C900
C901
C902
C903
C904
C905
C906
C907
C908
C909
C910
C911
C912
C913
C914
C915
C916
C917
C918
C919
C920
C921
C922
C923
C924
C925
C926
C927
C928
C929
C930
C931
C932
C933
C934
C935
C936
C937
C938
C939
C940
C941
C942
C943
C944
C945
C946
C947
C948
C949
C950
C951
C952
C953
C954
C955
C956
C957
C958
C959
C960
C961
C962
C963
C964
C965
C966
C967
C968
C969
C970
C971
C972
C973
C974
C975
C976
C977
C978
C979
C980
C981
C982
C983
C984
C985
C986
C987
C988
C989
C990
C991
C992
C993
C994
C995
C996
C997
C998
C999
C1000
```

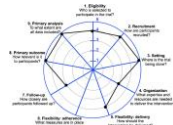


PATIENT



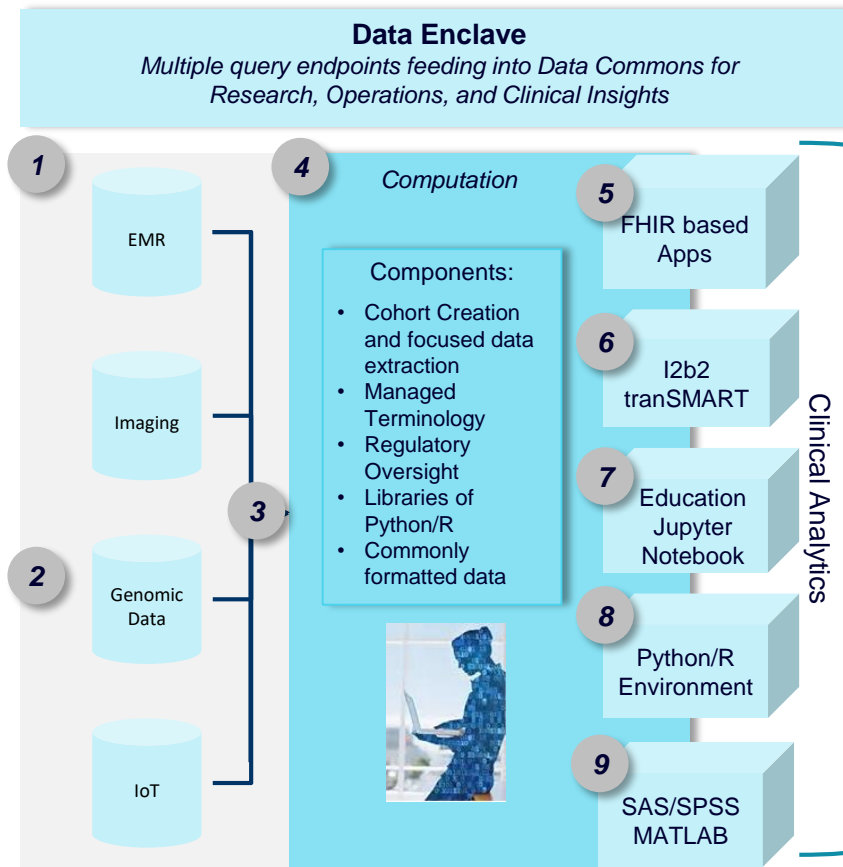
Navigator Model dramatically increases Frequency and Convenience for Patient Communication

Digital Twin of patient enables continuous assessment of patient with Real Time Algorithms

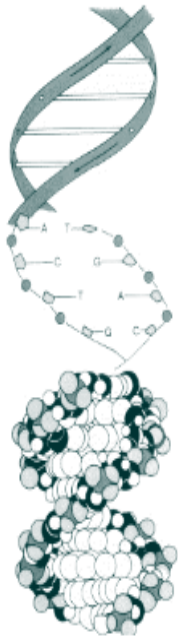


System drives Pragmatic Clinical Trials Leading to Continuous Process Improvement

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



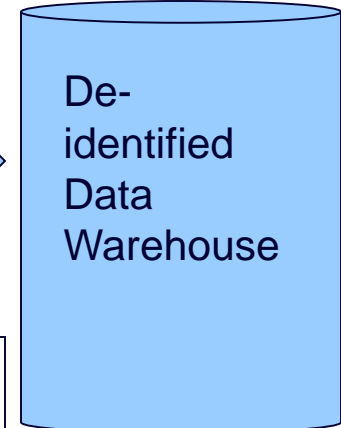
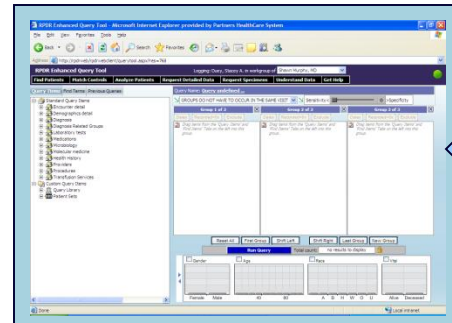
Combining i2b2 with tranSMART

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



Encrypted identifiers

Z731984X
Z74902XX
...

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

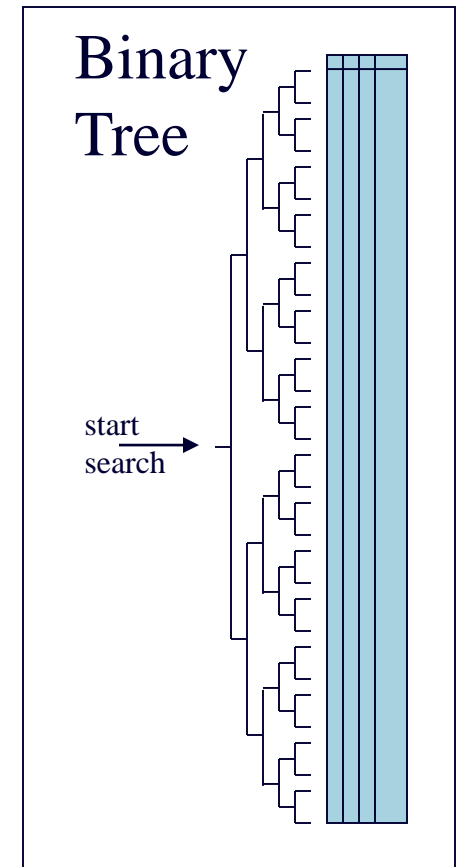
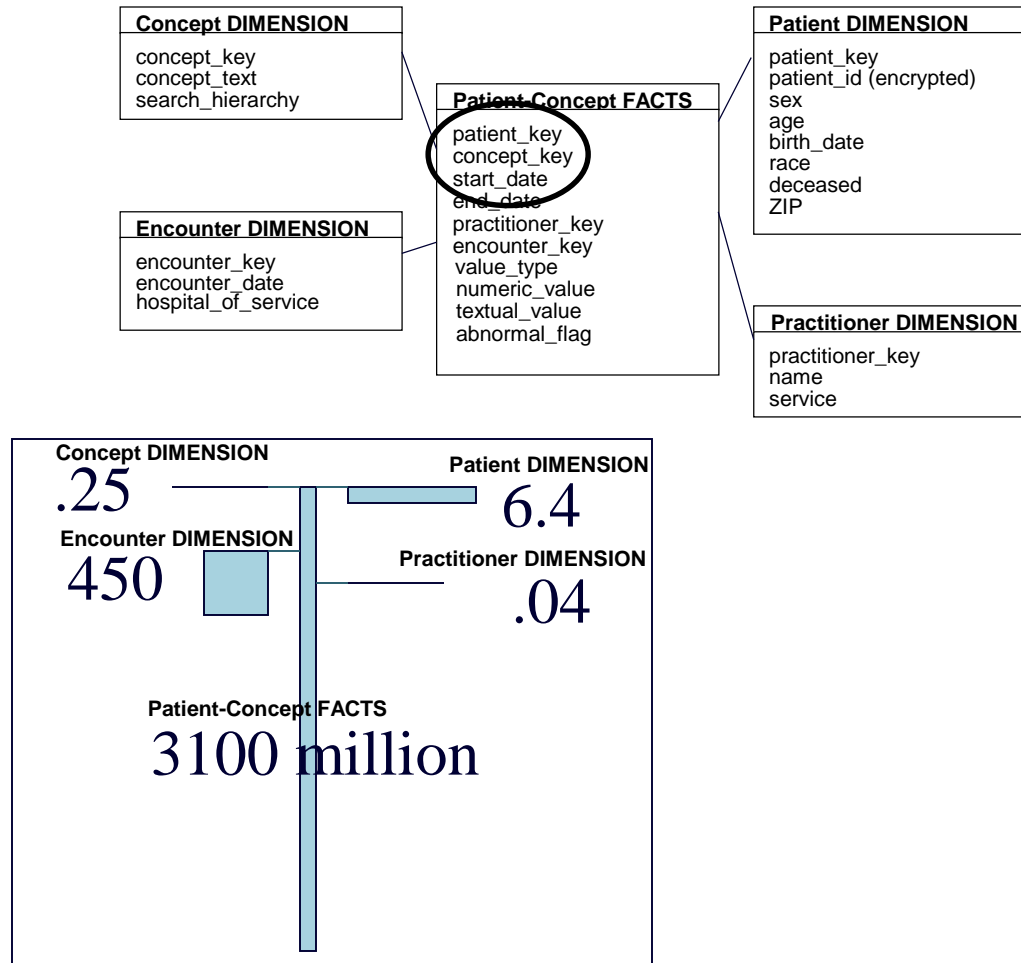
0000004
2185793
...

Real identifiers

A screenshot of a Microsoft Access database showing a table of patient data. The table has columns for Patient ID, Patient Name, Date of Birth, Sex, Race, Ethnicity, and Address. The data is sorted by Patient ID.

Patient ID	Patient Name	Date of Birth	Sex	Race	Ethnicity	Address
0000004	2185793

Enabled by Star Schema

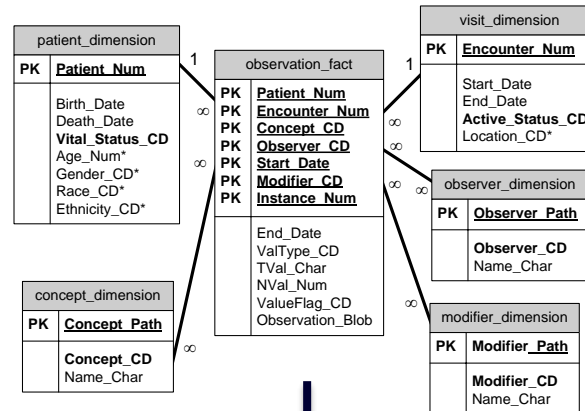


Start Schema enables both i2b2 and tranSMART

i2b2

tranSMART

CLINICAL
HOSPITAL
DATA



CLINICAL
TRIAL
DATA

REDCap

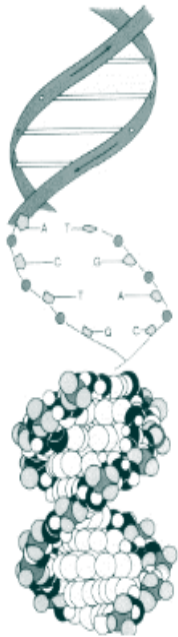
Fractalis

Risks with i2b2 - tranSMART common Database Instance

- Privacy Models different
- Database collisions from simultaneous actions
 - Principle risk is updating
- Index optimization different for different use cases
- Inherent assumptions about data are different
 - "One fact" per study visit regardless of time

Tribute to...

- Jeff Klann
- Michael Mendis
- Kavi Waghlikar
- Lori Phillips
- Isaac Kohane
- Kenneth Mandl
- Joshua Mandel
- Griffin Weber
- Paul Avillach
- Christopher Herrick
- Vivian Gainer
- Victor Castro
- Nich Wattanasin
- Wayne Chan
- David Wang
- Andrew Cagan
- Bhaswati Ghosh
- Retta Metta
- Adam Landman
- Willian Gordon



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>)

SMART Platforms Homepage (<http://smarthealthit.org>)

Partners Healthcare, 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