

Ontology-Based Data Interchange

Jeff Klann Ph.D.

Lori Phillips MS

Alyssa Goodson MS

Christopher Herrick MBA

Shawn Murphy MD, Ph.D.

Use Ontology to Drive:

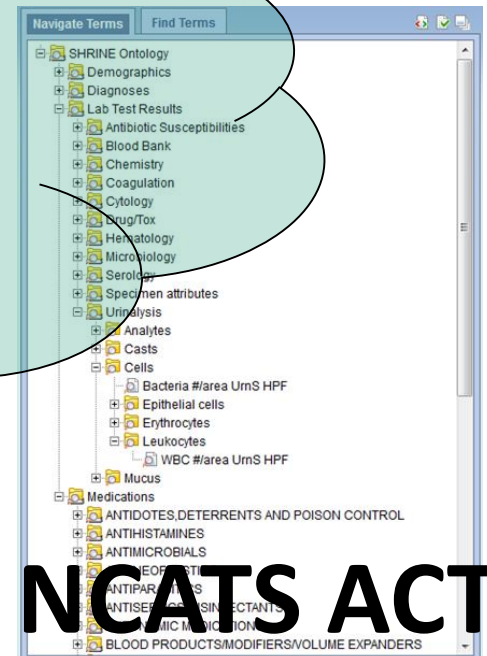
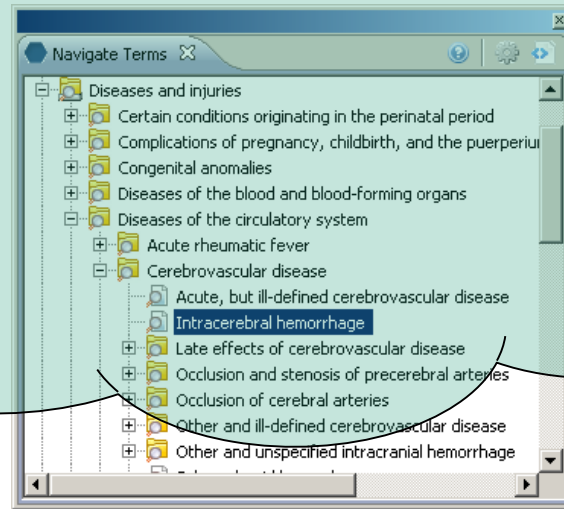
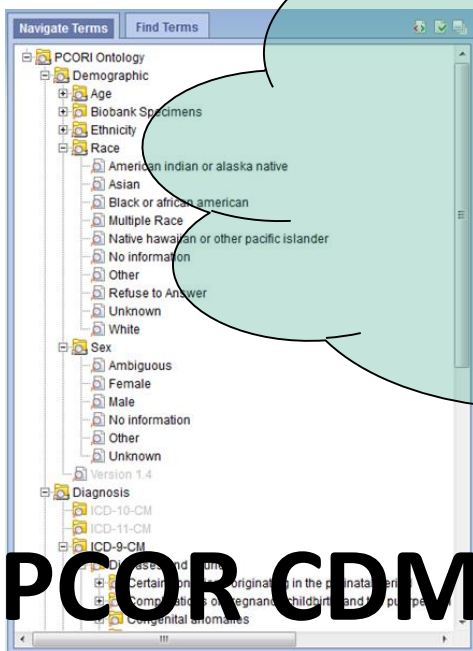
- Creation of Alternative Data Models from i2b2
 - Desirable when supporting computation directed to various data models
- Allowing there to be Multiple Fact tables in i2b2
 - Desirable when want i2b2 on top of other data models
 - Desirable when accumulating large static fact tables
- Distribution of Queries to Multiple i2b2 Hives
 - Desirable when Query Endpoint is not SQL

Creation of Alternative Data Models from i2b2

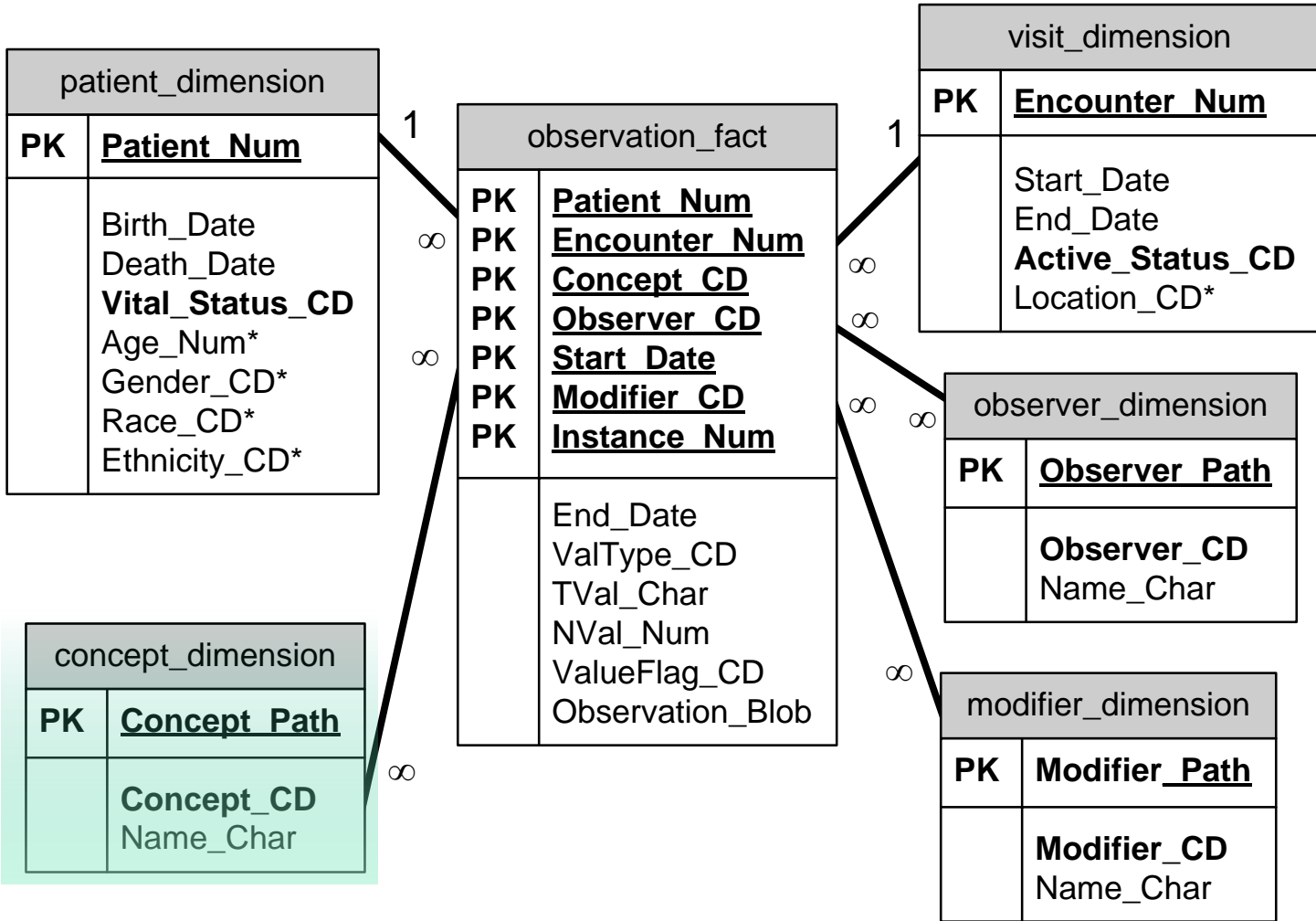
Jeff Klann Ph.D.

I2b2 is a Giant Data Sponge

Generally Represents Patient Data



i2b2 Star Schema



Concept_dimension “look” ...

CONCEPT_PATH	CONCEPT_CD	NAME_CHAR
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.4	Toxic myopathy
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.5	Myopathy in endocrine diseases ...
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.6	Symptomatic inflammatory myop...
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.8	Other myopathies
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.81	Critical illness myopathy
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.89	Other myopathies
\i2b2\Diagnoses\Neurologic Disorders (320-389)\Peripheral nerve disorde...	ICD9:359.9	Myopathy, unspecified
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(260) Kwashiorkor\	ICD9:260	Hypoproteinosis
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(261) Nutritional maras...	ICD9:261	Marasmus
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(262) Other severe pr...	ICD9:262	Other severe protein-calorie mal...
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263	Other and unspecified protein-c...
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263.0	Malnutrition of moderate degree
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263.1	Malnutrition of mild degree
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263.2	Dwarfism, nutritional
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263.8	Other protein-calorie malnutrition
\i2b2\Diagnoses\Nutritional deficiencies (260-269)\(263) Other and unspe...	ICD9:263.9	Protein-calorie undernutrition

New Information Model Ontology

Consensus Ontology
can live alongside
other ontologies

(For example:
PCORNet CDM
ontology and the i2b2
demo ontology in this
case)

- + Clinical Trials
- + Custom Metadata
- + Demographics
- + Diagnoses
- + Diagnoses (ICD10)
- + Expression Profiles Data
- + Laboratory Tests
- + Medications
- + PCORnet Core
- + PCORnet Demographics
- + PCORnet Diagnoses
- + PCORnet Encounters
- + PCORnet Enrollment
- + PCORnet Procedures
- + PCORnet Vital Signs
- + Procedures
- + Providers
- + Reports

Adapting i2b2 to PCORNet Data Model

1. Ontology-Driven Physical Transformation into PCORNet Common Data Model with a Generalizable approach to adapt to other Common Data Models

Data interchange using i2b2

RECEIVED 31 July 2015
REVISED 26 October 2015
ACCEPTED 31 October 2015

J Am Med Inform Assoc 2016;0:1–8. doi:10.1093/jamia/ocv188, Research and Applications

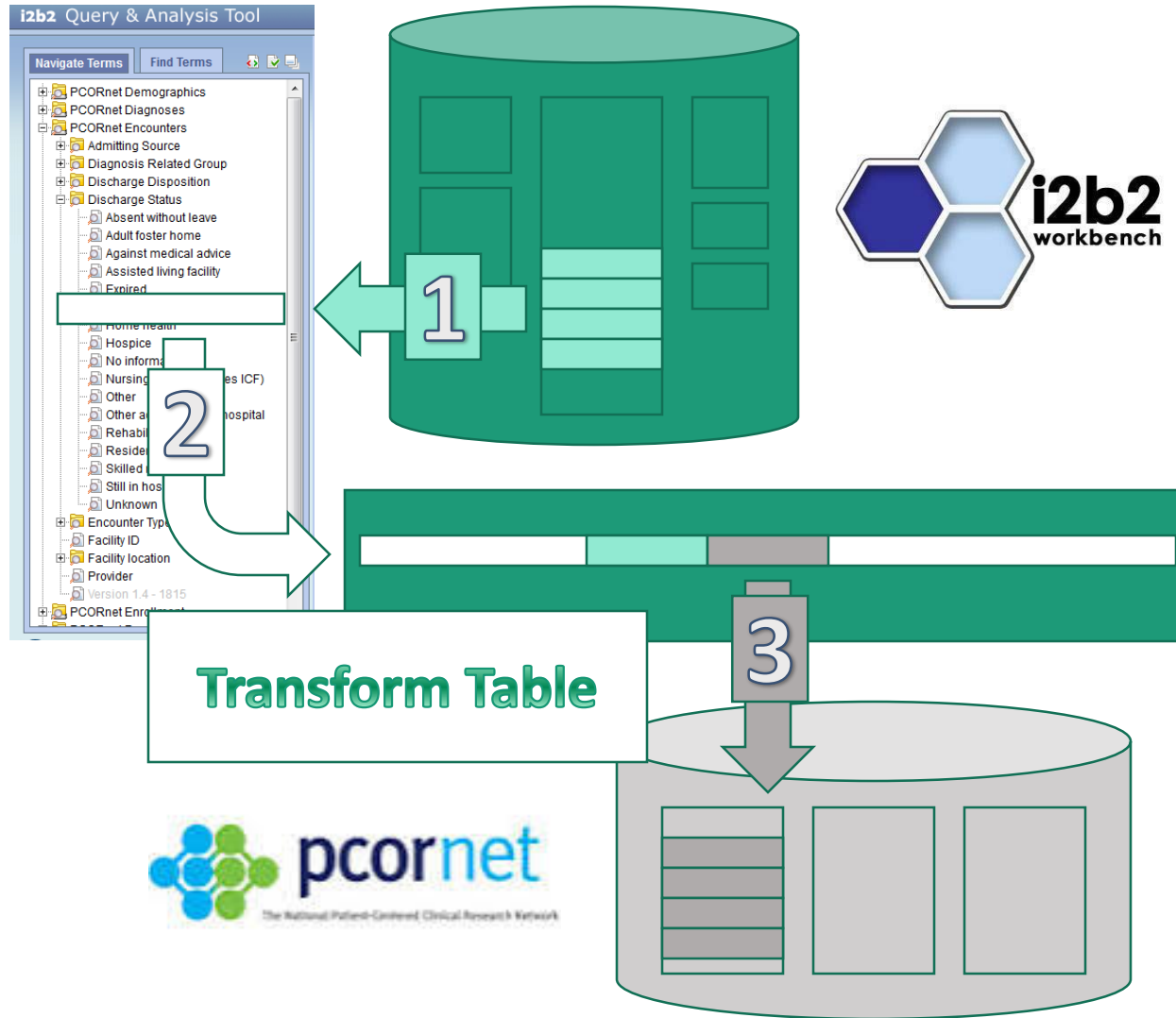
Jeffrey G Klann^{1,2,3}, Aaron Abend⁴, Vijay A Raghavan², Kenneth D Mandl^{2,5}, and
Shawn N Murphy^{1,2,3}



a peer-to-peer network

on in

Physical data transformations to non-i2b2 formats ontology driven



PCORnet Common Data Model v3.0

New to v3.0

DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS
DISPENSING
DISPENSINGID
PATID
PRESCRIBINGID (optional)
DISPENSE_DATE
NDC
DISPENSE_SUP
DISPENSE_AMT
DEATH
PATID
DEATH_DATE
DEATH_DATE_IMPUTE
DEATH_SOURCE
DEATH_MATCH_CONFIDENCE
DEATH_CONDITION
PATID
DEATH_CAUSE
DEATH_CAUSE_CODE
DEATH_CAUSE_TYPE
DEATH_CAUSE_SOURCE
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

VITAL
VITALID
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE
CONDITION
CONDITIONID
PATID
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
CONDITION
CONDITION_TYPE
CONDITION_SOURCE
PRO_CM
PRO_CM_ID
PATID
ENCOUNTERID (optional)
PRO_ITEM
PRO_LOINC
PRO_DATE
PRO_TIME
PRO_RESPONSE
PRO_METHOD
PRO_MODE
PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

ENCOUNTER
ENCOUNTERID
PATID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE
DIAGNOSIS
DIAGNOSISID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX
PROCEDURES
PROCEDURESID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX_DATE
PX
PX_TYPE
PX_SOURCE

Data captured from healthcare delivery, direct encounter basis

LAB_RESULT_CM
LAB_RESULT_CM_ID
PATID
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
RESULT_DATE
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND
PRESCRIBING
PRESCRIBINGID
PATID
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

PCORNET_TRIAL
PATID
TRIALID
PARTICIPANTID
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

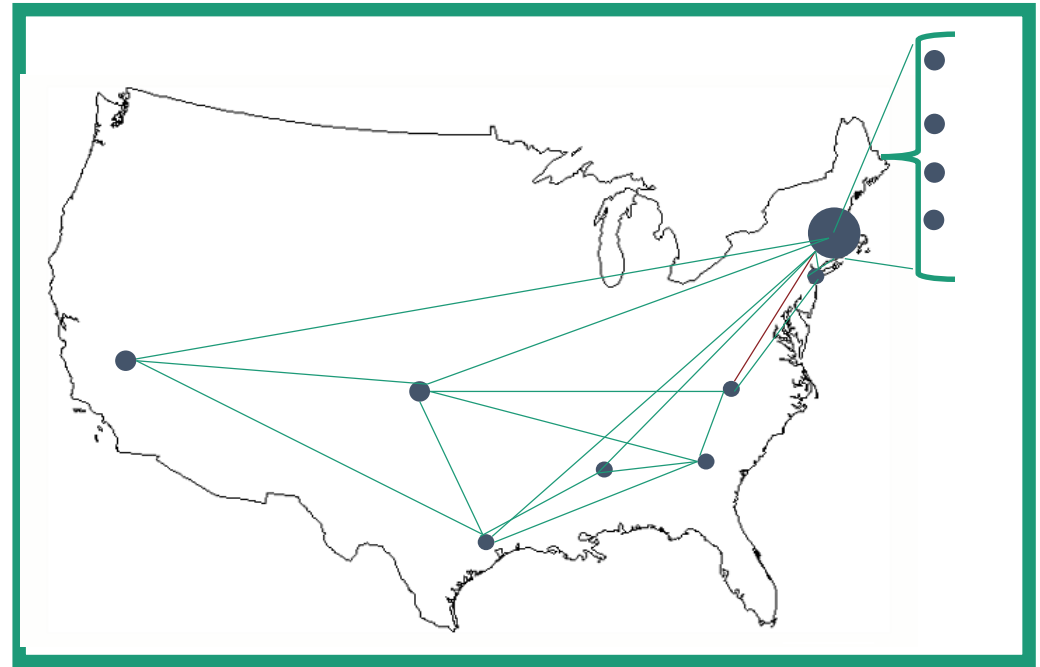
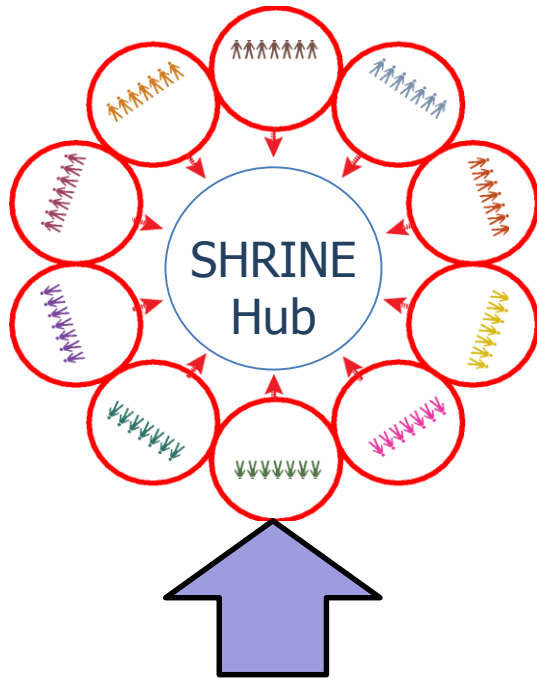
Associations with PCORnet clinical trials

HARVEST
NETWORKID
NETWORK_NAME
DATAMARTID
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

Process-related data

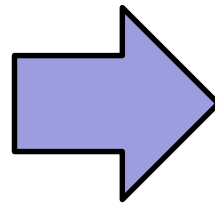
Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.

ARCH Network Participation



- + PCORnet Demographics
- + PCORnet Diagnoses
- + PCORnet Encounters
- + PCORnet Enrollment
- + PCORnet Procedures
- + PCORnet Vital Signs

SCILHS CDM
Information Model



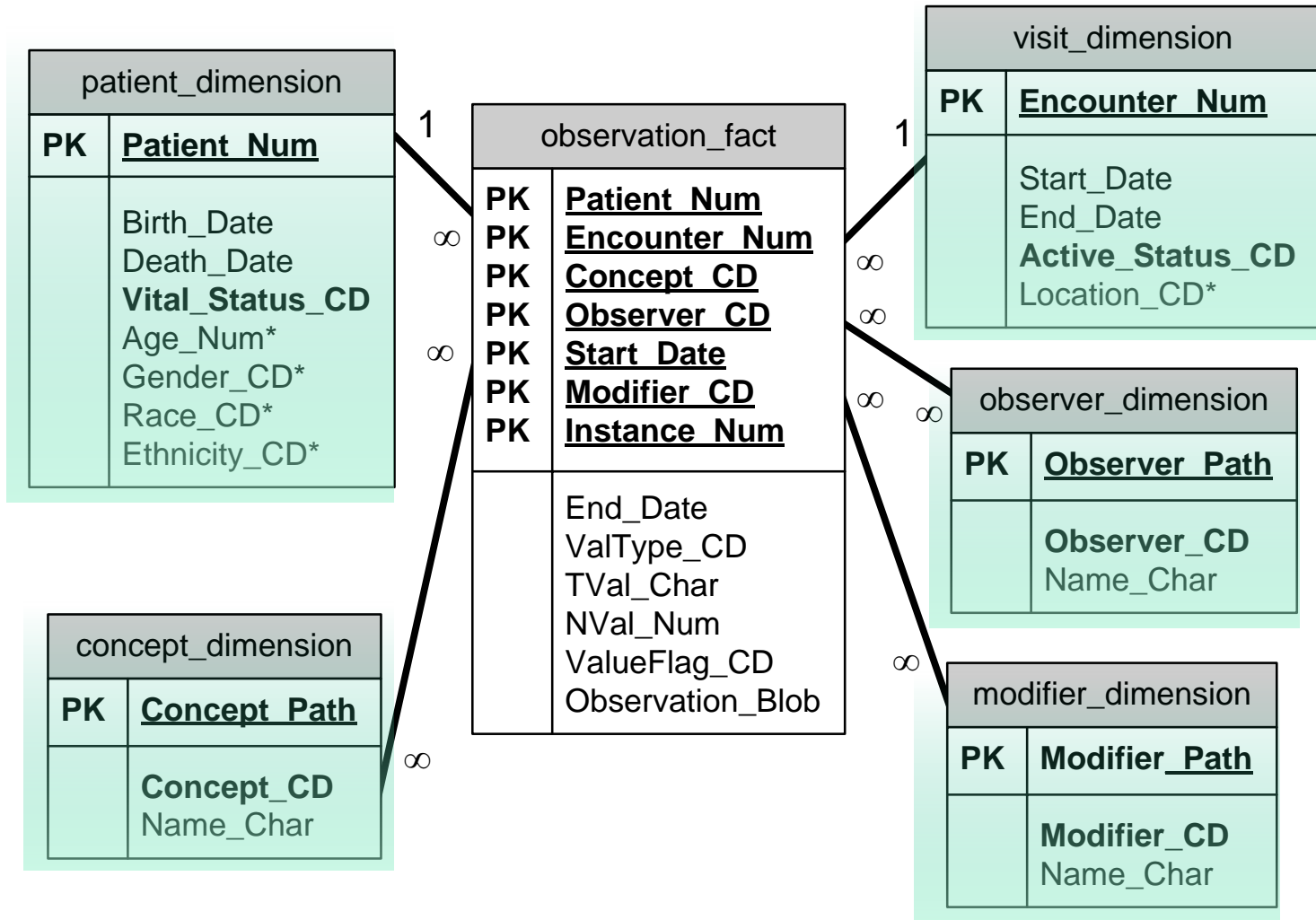
pcornet

The National Patient-Centered Clinical Research Network

Extending i2b2 for Multiple Fact Tables

Lori Phillips MS

i2b2 Star Schema



patient_dimension	
PK	<u>Patient Num</u>
	Birth_Date Death_Date Vital_Status_CD Age_Num* Gender_CD* Race_CD* Ethnicity_CD*

observation_fact	
PK	<u>Patient Num</u>
PK	<u>Encounter Num</u>
PK	<u>Concept CD</u>
PK	<u>Observer CD</u>
PK	<u>Start Date</u>
PK	<u>Modifier CD</u>
PK	<u>Instance Num</u>
	End_Date ValType_CD TVal_Char NVal_Num ValueFlag_CD Observation_Blob

observation_fact	
PK	<u>Patient Num</u>
PK	<u>Encounter Num</u>
PK	<u>Concept CD</u>
PK	<u>Observer CD</u>
PK	<u>Start Date</u>
PK	<u>Modifier CD</u>
PK	<u>Instance Num</u>
	End_Date ValType_CD TVal_Char NVal_Num ValueFlag_CD Observation_Blob

visit_dimension	
PK	<u>Encounter Num</u>
	Start_Date End_Date Active_Status_CD Location_CD*

observer_dimension	
PK	<u>Observer Path</u>
	Observer_CD Name_Char

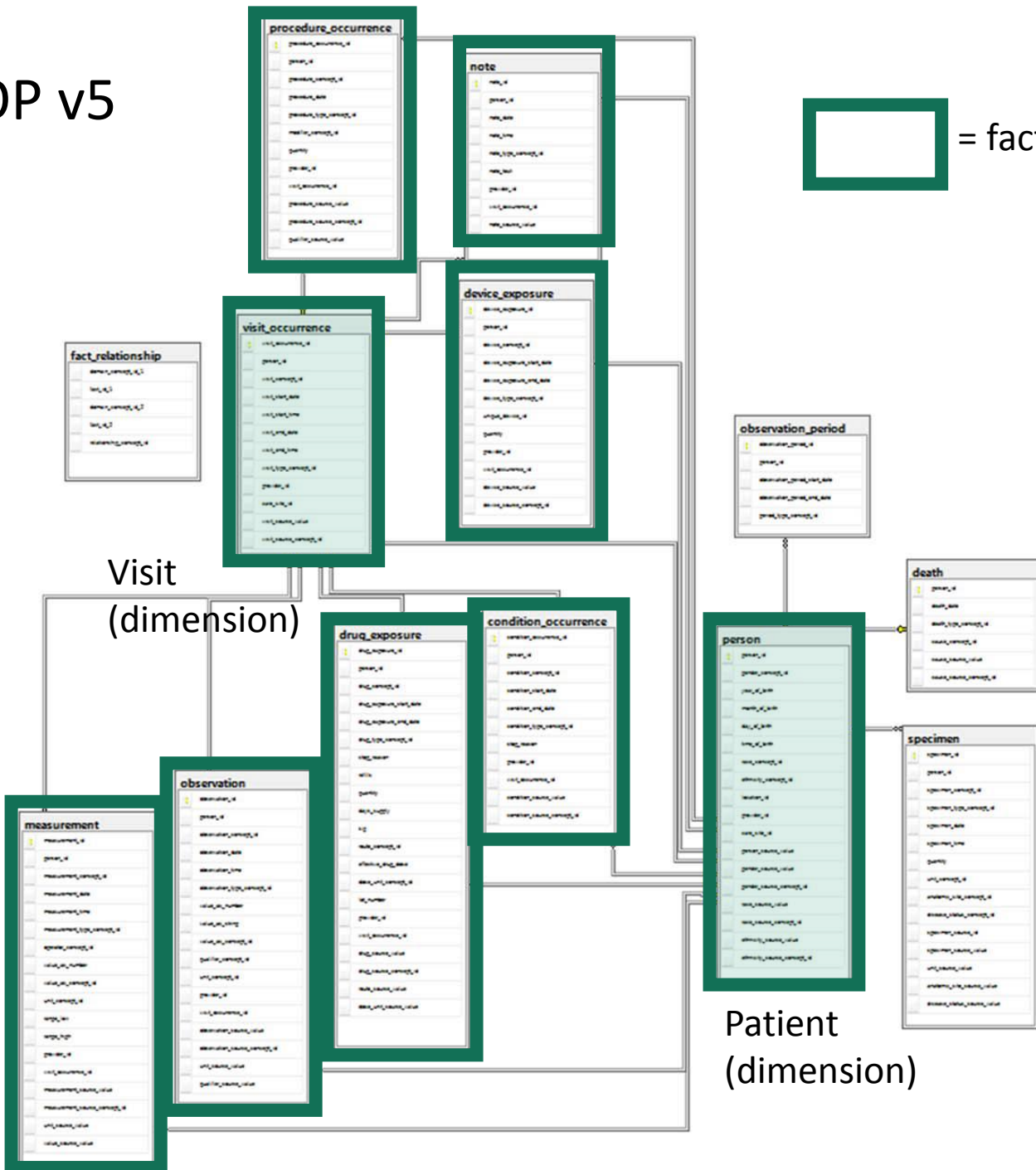
concept_dimension	
PK	<u>Concept Path</u>
	Concept_CD Name_Char

modifier_dimension	
PK	<u>Modifier Path</u>
	Modifier_CD Name_Char



OMOP v5

 = fact tables



Visit
(dimension)

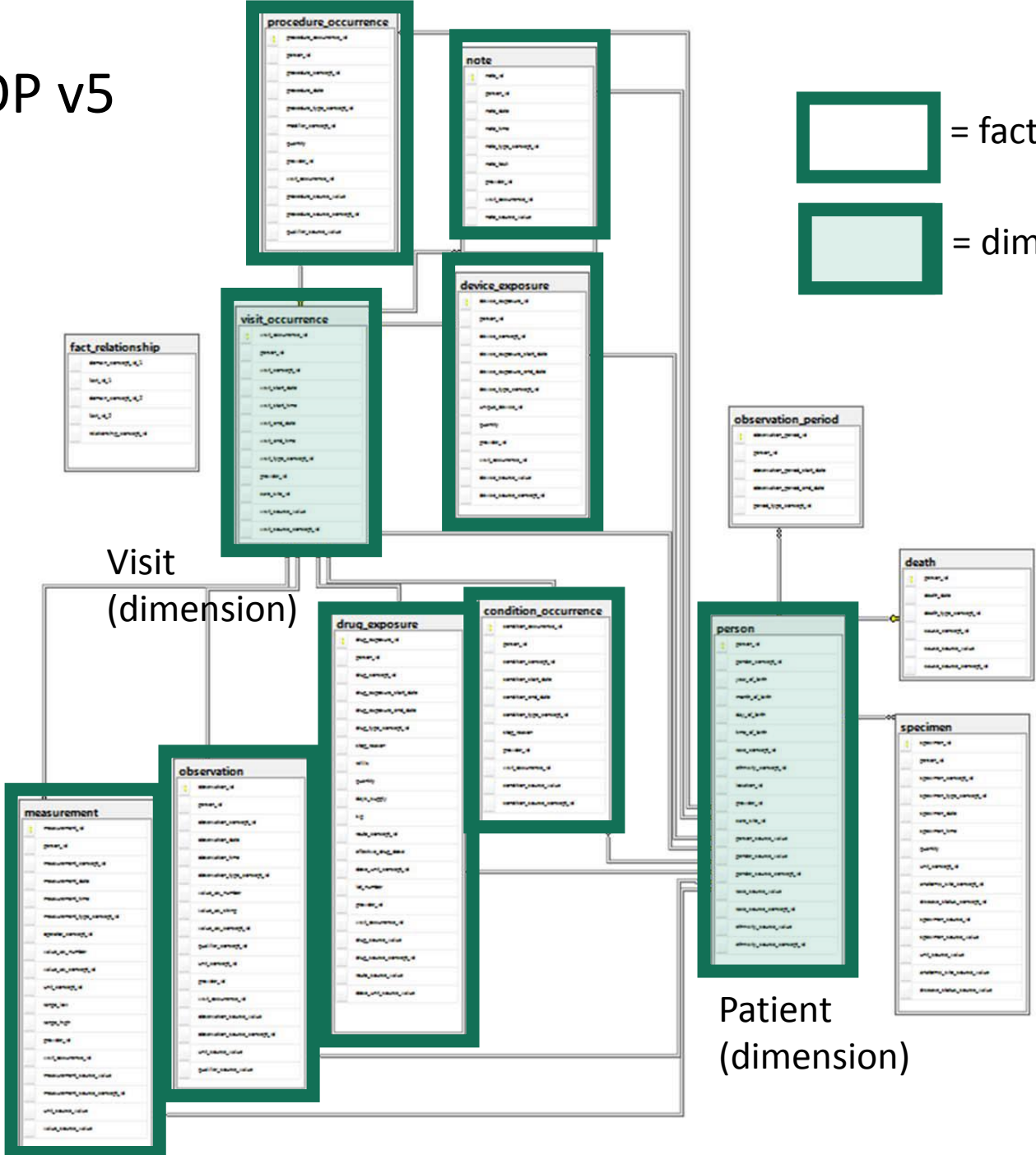
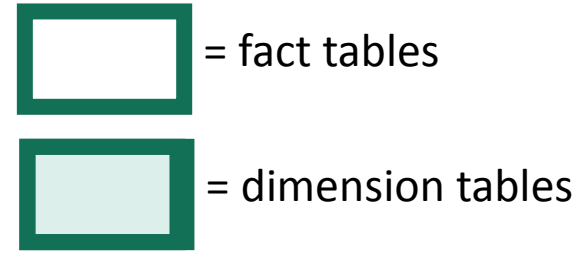
Patient
(dimension)

observation_fact	
PK	<u>Patient Num</u>
PK	<u>Encounter Num</u>
PK	<u>Concept CD</u>
PK	<u>Observer CD</u>
PK	<u>Start Date</u>
PK	<u>Modifier CD</u>
PK	<u>Instance Num</u>
	End_Date
	ValType_CD
	TVal_Char
	NVal_Num
	ValueFlag_CD
	Observation_Blob

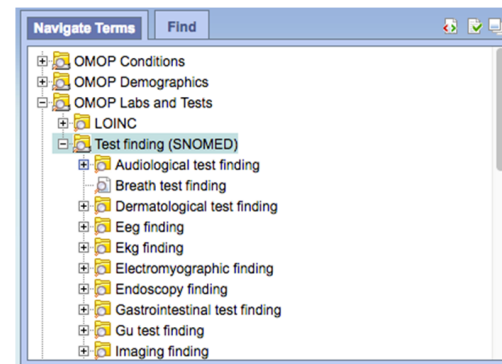
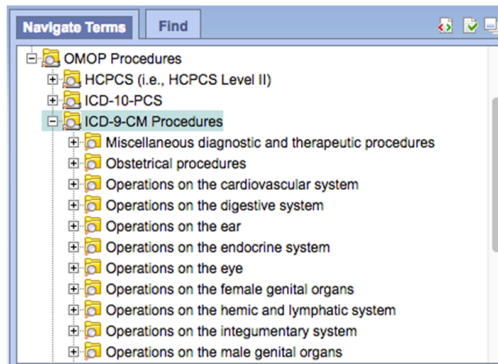
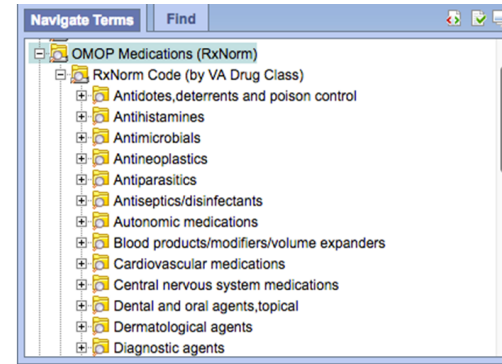
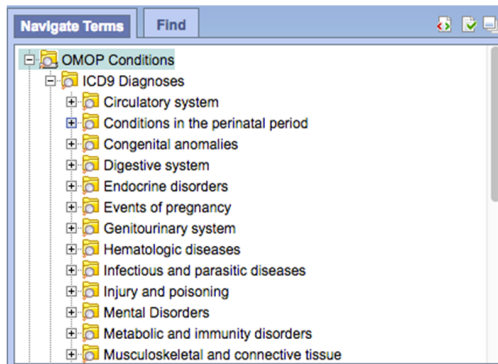
Field	Required	Type	Description
procedure_occurrence_id	Yes	integer	A system-generated unique identifier for each procedure occurrence
person_id	Yes	integer	A foreign key identifier to the person who is subjected to the procedure. The demographic details of that person are stored in the person table.
procedure_concept_id	Yes	integer	A foreign key that refers to a standard procedure concept identifier in the Standardized Vocabularies.
procedure_date	Yes	date	The date on which the procedure was performed.
procedure_type_concept_id	Yes	integer	A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of source data from which the procedure record is derived.
modifier_concept_id	No	integer	A foreign key to a standard concept identifier for a modifier to the procedure (e.g. bilateral)
quantity	No	integer	The quantity of procedures ordered or administered.
provider_id	No	integer	A foreign key to the provider in the provider table who was responsible for carrying out the procedure
visit_occurrence_id	No	integer	A foreign key to the visit in the visit table during which the procedure was carried out
procedure_source_value	No	varchar(50)	The source code for the procedure as it appears in the source data. This code is mapped to a standard procedure concept in the Standardized Vocabularies and the original code is, stored here for reference. Procedure source codes are typically ICD-9-Proc, CPT-4, HCPCS or OPCS-4 codes.
procedure_source_concept_id	No	integer	A foreign key to a procedure concept that refers to the code used in the source.
qualifier_source_value	No	varchar(50)	The source code for the qualifier as it appears in the source data.



OMOP v5



Ontology Tables Need to be Created - Build ontology of OMOP standard concepts



Ontologies covering the condition, procedures, drug, measurement and observation domains. All terms are mapped to standard concepts using OMOP's mapping tables

Create views for OMOP Fact Tables

	OBSERVATION_FACT	CONDITION
PK	PATIENT_NUM	PERSON_ID
PK	ENCOUNTER_NUM	VISIT_OCCURENCE_ID
PK	CONCEPT_CD	CONDITION_CONCEPT_ID
PK	PROVIDER_NUM	PROVIDER_ID
PK	START_DATE	CONDITION_START_DATE
PK	MODIFIER_CD	
PK	INSTANCE_NUM	
	End_date	CONDITION_END_DATE
	Valtype_cd	
	Tval_char	
	Nval_num	
	Valueflag_cd	
	Observation_blob	

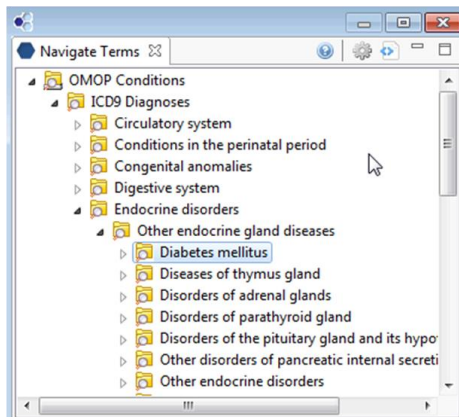
```

CREATE VIEW [dbo].[CONDITION_VIEW]
(patient_num, encounter_num, concept_cd, provider_id,
start_date, modifier_cd, instance_num, end_date,
valtype_cd, tval_char, nval_num, valueflag_cd,
observation_blob)
AS
SELECT person_id ,
       visit_occurrence_id ,
       cast(condition_concept_id as varchar),
       cast(provider_id as varchar),
       condition_start_date,
       '@',
       1,
       condition_end_date,
       cast(null as varchar),
       cast(null as varchar),
       cast(null as decimal),
       cast(null as varchar),
       cast(null as varchar(max)
FROM condition
  
```

Use Ontology Tables to direct Queries to proper Fact Table view

Prepend c_facttablecolumn with OMOP domain view and modify CRC to parse into 'domain_view' and 'c_facttablecolumn'

c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Diabetes mellitus	condition_view.concept_cd	concept_dimension	concept_path	LIKE	\i2b2\Diagnoses\Endocrine disorders (240-259)\Other endocrine gland



```
select patient_num from
condition_view where concept_cd
IN
(select
concept_cd from concept_dimension where
concept_path like '\i2b2\Diagnoses\Endocrine
disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus\%')
```

Queries can be performed in i2b2

The screenshot displays the i2b2 Query Tool interface. The query name is "Circu-Cardi-Ekg f@14:13:17". The temporal constraint is set to "Treat Independently". The query is composed of three groups:

- Group 1: Circulatory system
- Group 2: Cardiovascular medications
- Group 3: Ekg finding

The groups are connected by "AND" operators. Each group has a "one or more of these" button. The interface shows "3 Groups" and a "Run Query" button. Below the query, the results are displayed as a graph showing the number of patients, which is 87. The text below the graph reads: "Number of patients 87 For Query 'Circu-Cardi-Ekg f@14:13:17'".

Query across 3 domains: condition, medication, measurement.

CDM populated with OMOP's Synthetic Public Use File data (synPUF)² for 1000 patients.

Successful OMOP Queries:

- Query types included
 - Multi-panel, multi-domain queries
 - Date constrained queries
 - Occurs > x queries
 - Value constrained queries
 - Temporal queries
- Queries not fully worked out
 - Modifier queries
 - Ancillary tables
 - Cover all OMOP Ontologies

Same Approach to PCORNet CDM

PCORnet Common Data Model v3.0

New to v3.0

DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS

DISPENSING
DISPENSINGID
PATID
PRESCRIBINGID (optional)
DISPENSE_DATE
NDC
DISPENSE_SUP
DISPENSE_AMT

DEATH
PATID
DEATH_DATE
DEATH_DATE_IMPUTE
DEATH_SOURCE
DEATH_MATCH_CONFIDENCE

DEATH_CONDITION
PATID
DEATH_CAUSE
DEATH_CAUSE_CODE
DEATH_CAUSE_TYPE
DEATH_CAUSE_SOURCE
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

VITAL
VITALID
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE

CONDITION
CONDITIONID
PATID
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
CONDITION
CONDITION_TYPE
CONDITION_SOURCE

PRO_CM
PRO_CM_ID
PATID
ENCOUNTERID (optional)
PRO_ITEM
PRO_LOINC
PRO_DATE
PRO_TIME
PRO_RESPONSE
PRO_METHOD
PRO_MODE
PRO_CAT

contexts: healthcare delivery, registry activity, or directly from patients

ENCOUNTER
ENCOUNTERID
PATID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

DIAGNOSIS
DIAGNOSISID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

PROCEDURES
PROCEDURESID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX_DATE
PX
PX_TYPE
PX_SOURCE

Data captured from healthcare delivery, direct encounter basis

LAB_RESULT_CM
LAB_RESULT_CM_ID
PATID
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
RESULT_DATE
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND

PRESCRIBING
PRESCRIBINGID
PATID
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

PCORNET_TRIAL
PATID
TRIALID
PARTICIPANTID
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

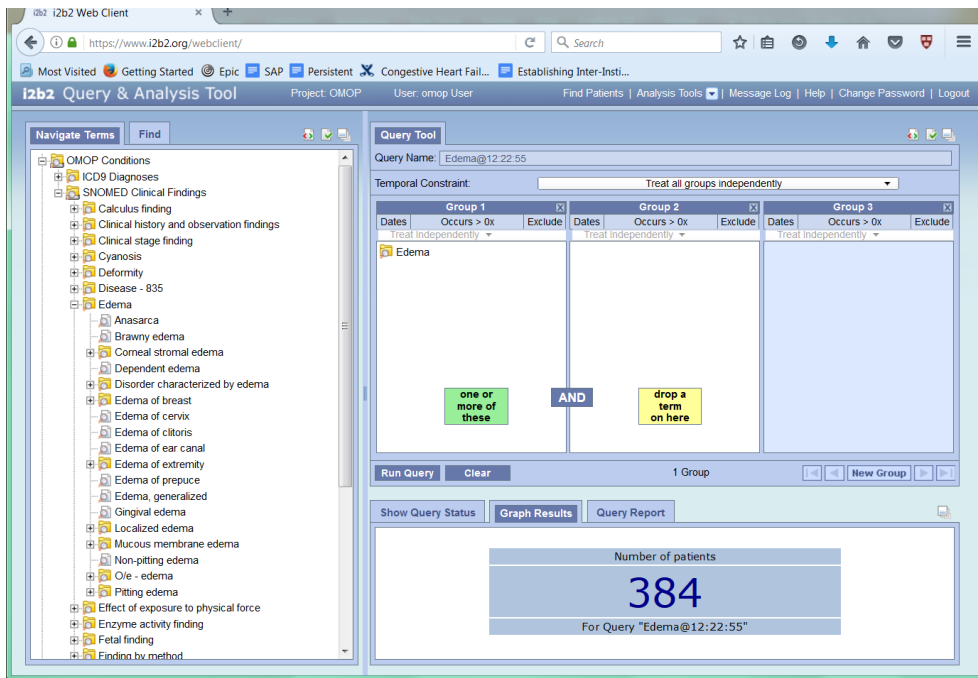
Associations with PCORnet clinical trials

HARVEST
NETWORKID
NETWORK_NAME
DATAMARTID
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

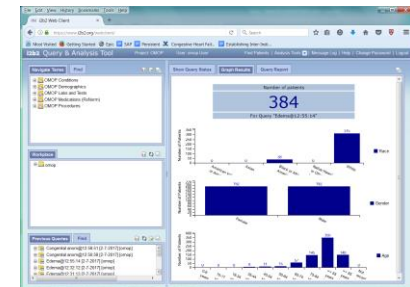
Process-related data

Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.

Demo of linking OMOP and i2b2 web services



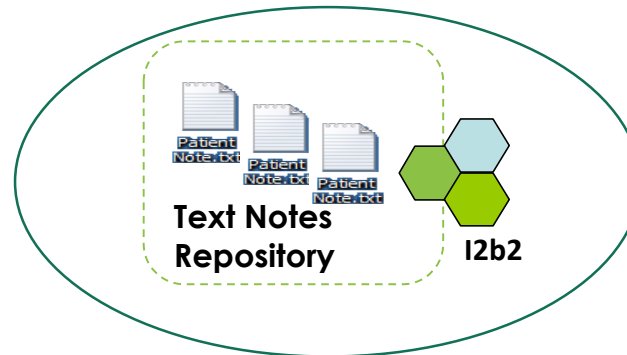
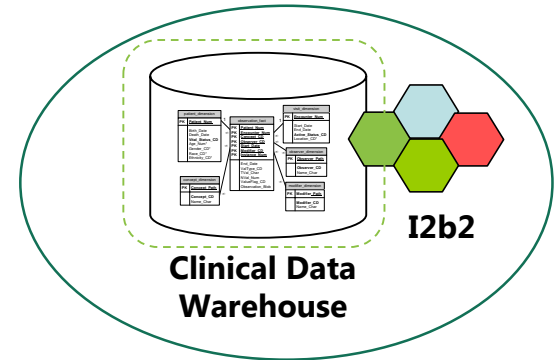
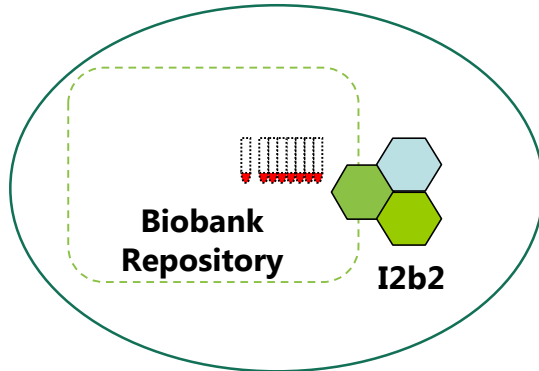
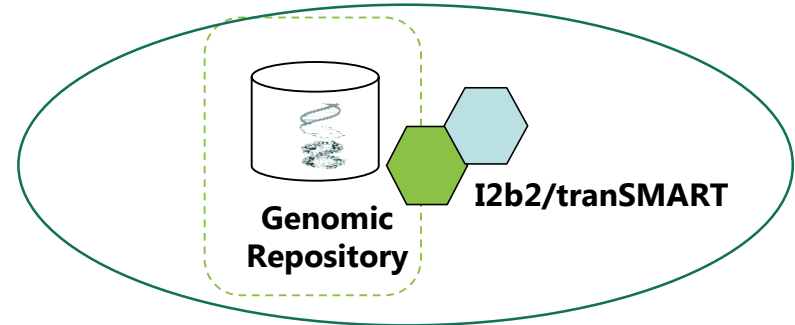
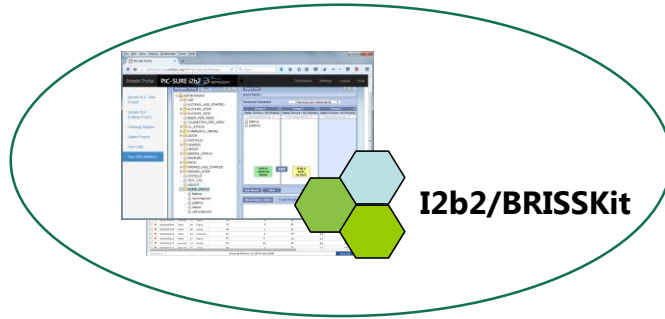
- Medicare Claims Synthetic Public Use Files (SynPUFs) in OMOP v5 CDM is background data set
- <https://www.i2b2.org/webclient/>
 - Username: omop
 - Password: demouser



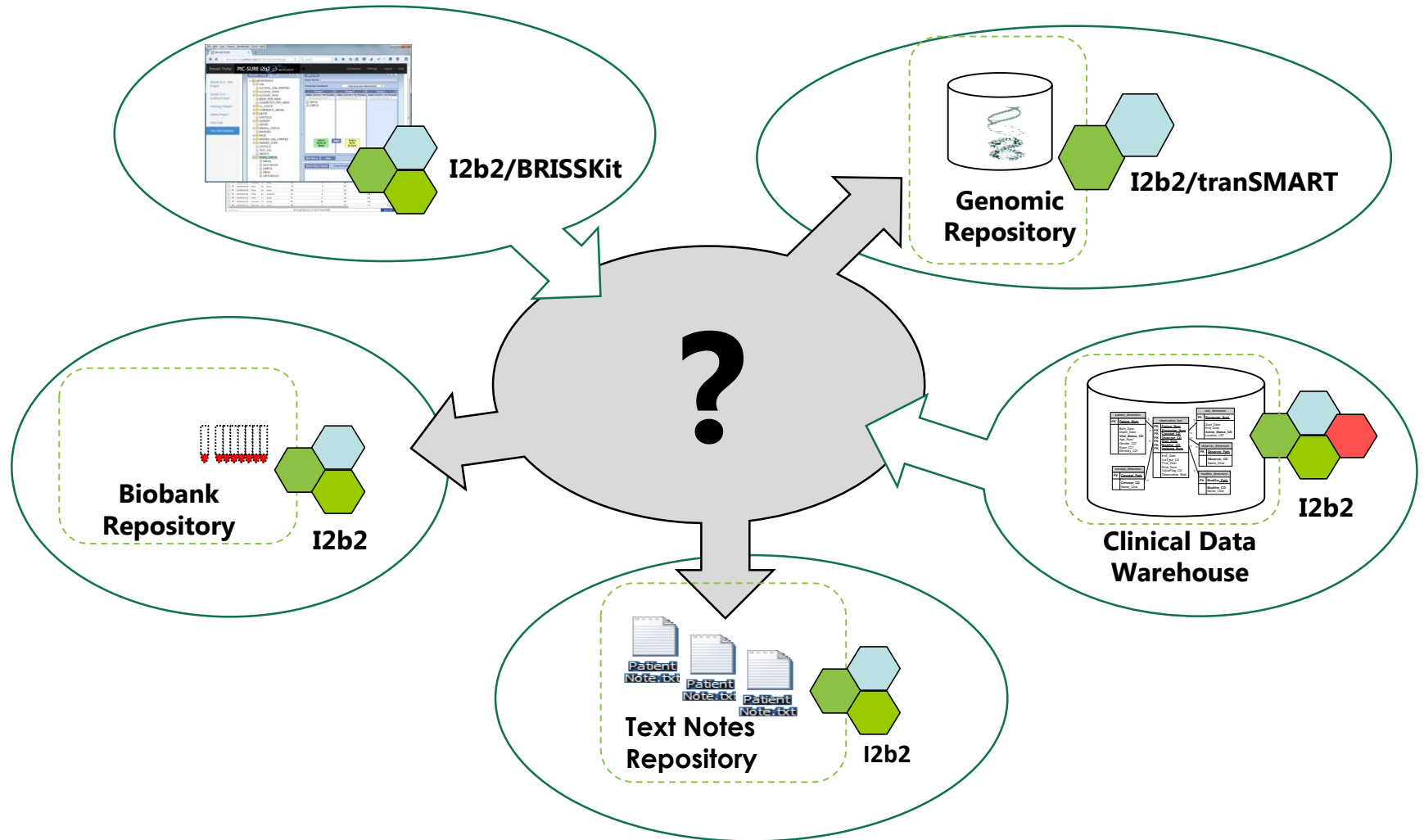
Distribution of Queries to Multiple i2b2 Hives

Christopher Herrick MBA

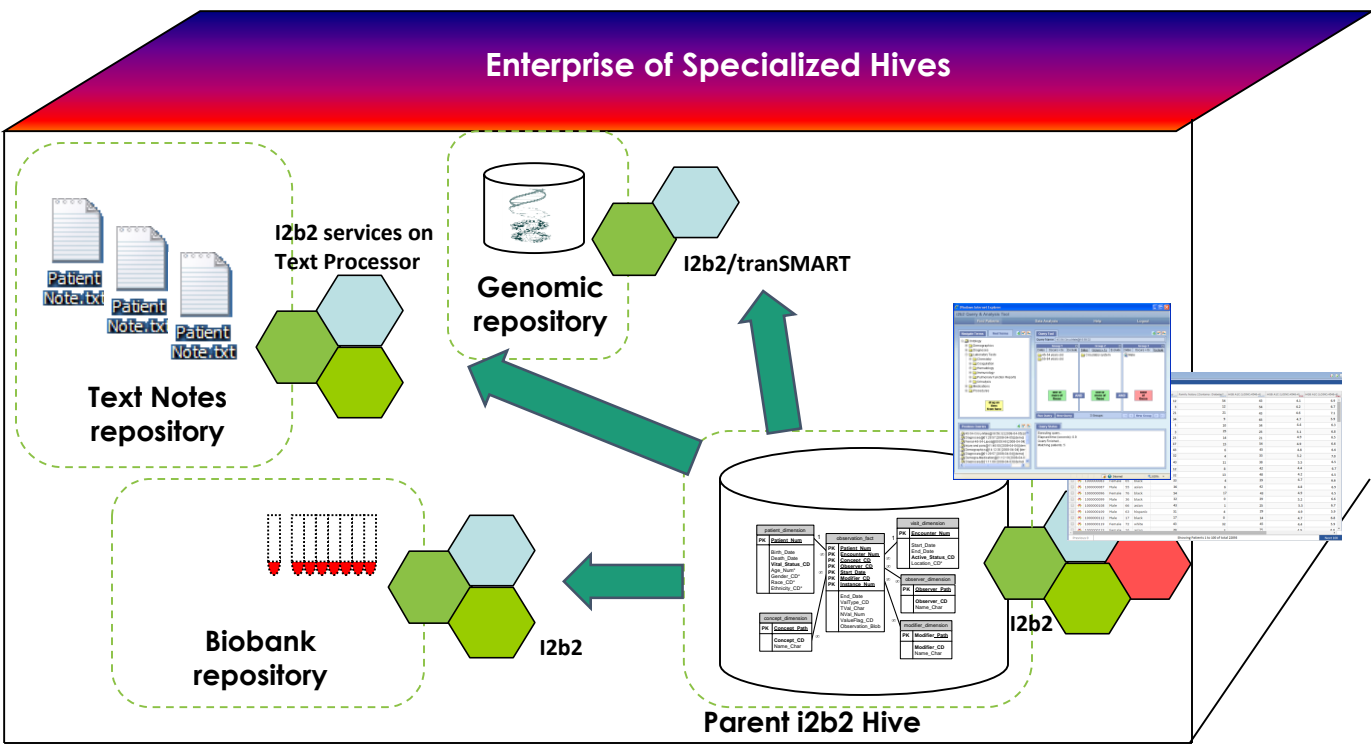
Hives Distributed in a Network



Hives Distributed in a Network

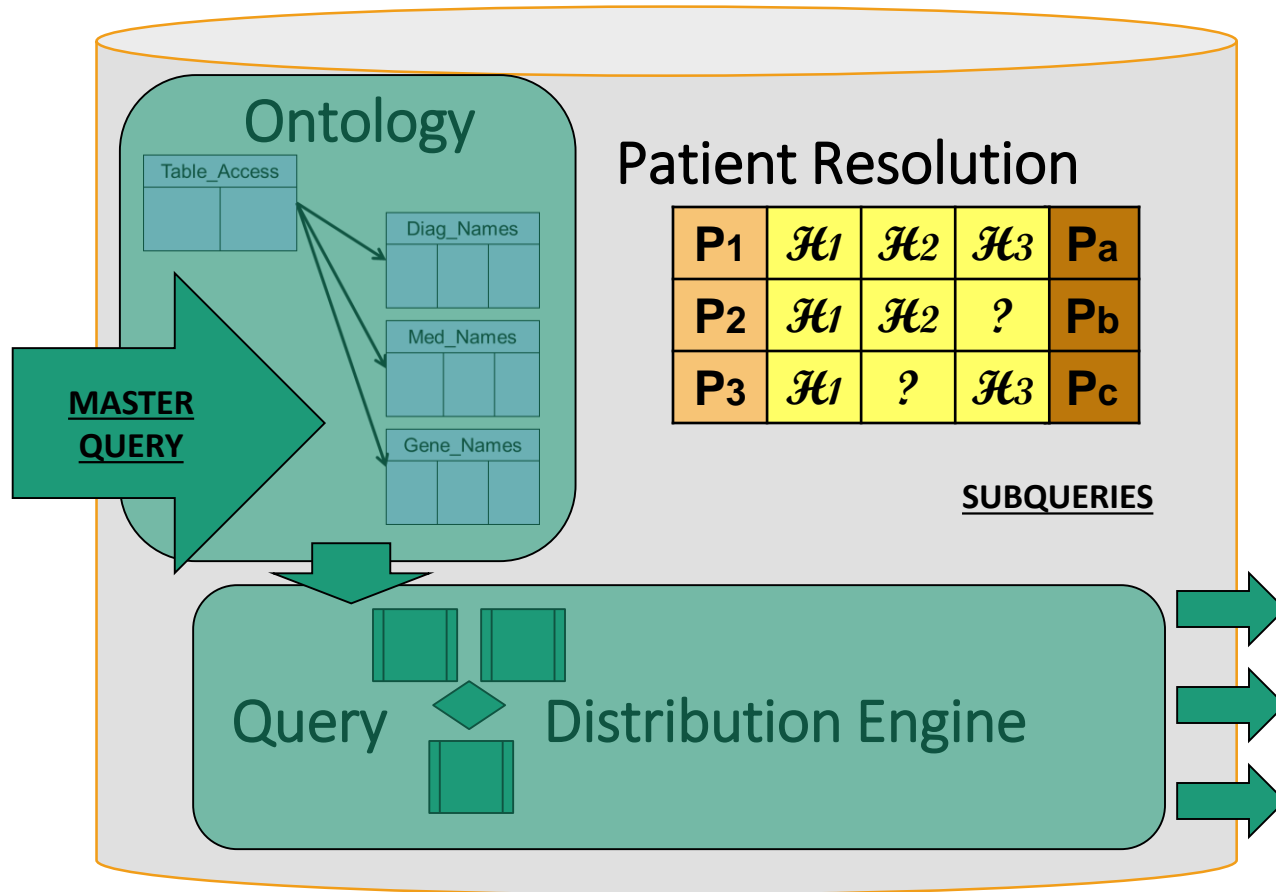


A Patient Information Commons from Specialized i2b2 Hives

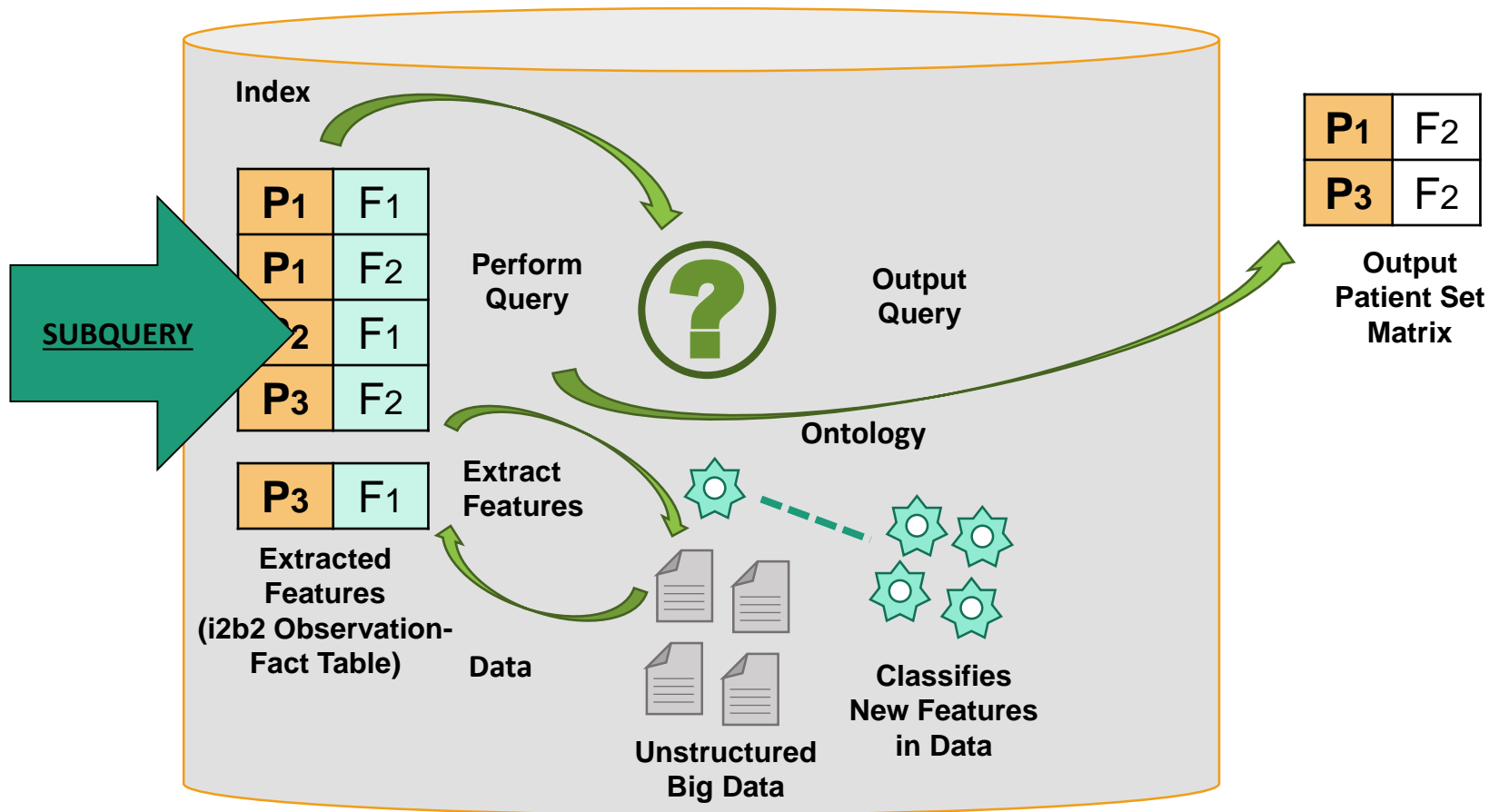


Concept ID	Parent Concept ID	Source	Active	Status	CD
1000000000000001	1000000000000000	ICD9CM	Y	U	1000000000000001
1000000000000002	1000000000000000	ICD9CM	Y	U	1000000000000002
1000000000000003	1000000000000000	ICD9CM	Y	U	1000000000000003
1000000000000004	1000000000000000	ICD9CM	Y	U	1000000000000004
1000000000000005	1000000000000000	ICD9CM	Y	U	1000000000000005

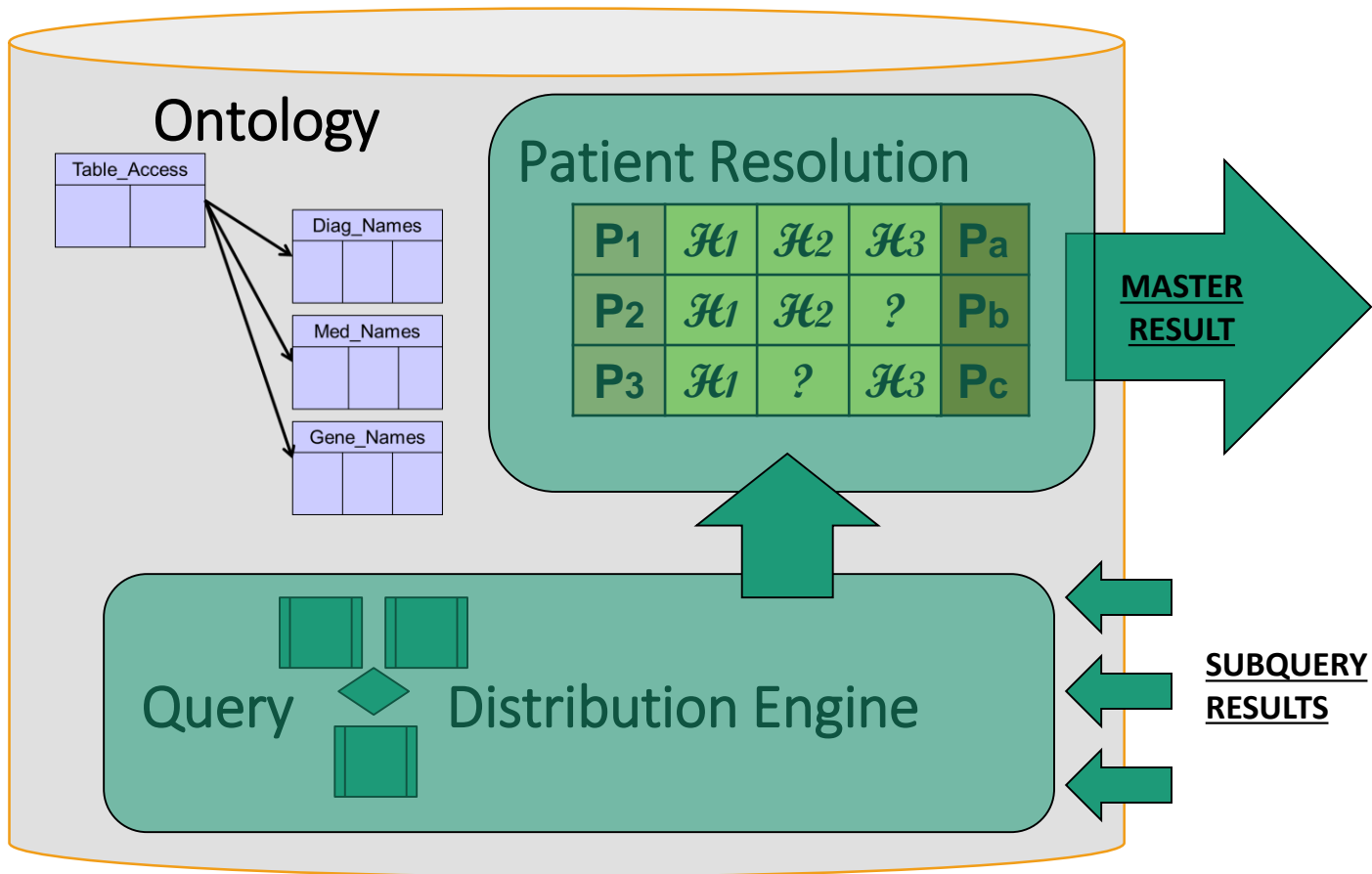
The Parent Hive Distributes Queries



The Child Hives Return Queries



The Parent Hive Returns Results



Common Sequence for Research Registries

- 1) A researcher creates a registry of patients
- 2) Data is collected on the patients
 - Abstracted from clinical chart as summary data and imaging
 - Questionnaires are given and/or Interviews with patients are performed
- 3) Data is analyzed and published
- Opportunity is lost – many researchers wish to combine with fresh clinical data and data from other registries

Two Approaches for Connecting Data

Enterprise Centric

Data is shared with all researchers across the enterprise. This is similar to how the Research Patient Data Registry (RPDR) currently displays and shares data with investigators across all of Partners.

Registries that become part of the enterprise centric view allow their data to be easily tied and queried with other enterprise wide data sources in the Big Data Commons.

This mode requires researchers to have the proper consents in place for their data to be queried from an enterprise level. Access to the identified data would still be controlled by the individual registry groups



Registry Centric

Data is imported into a registry for easier querying and analysis of patient cohorts; however, that data is not made readily available to the greater enterprise

Registries can supplement their project specific data by connecting with enterprise available datasets that are part of the Big Data Commons network. Access to the enterprise sets allow investigators to fill in important data gaps they may have with their own data

This mode is important if researchers have not collected the proper patient consents or, for other reasons, are not able to make their data available to the broader enterprise. Investigators would still be able to grant access to individual researchers who wish to collaborate.



An Enterprise Centered Data Network

Genomic Data

Genomic data collected through the Biobank lives in a separate repository, but is made available for connecting with clinical data. All patients within the Biobank are accessible



Research Repository

Broad repository of clinical data made available for research is the center point for all querying. Contains the entire Partners patient population.



Imaging Repository

DICOM Metadata is extracted from images downloaded from mi2b2. This may be supplemented by a limited amount of tags on all images given to us by Radiology group. Contains references to all patients from who we have imaging data



Notes & Reports

Notes and reports on all patients are collected and put into a separate data repository that can be full text indexed. Specific security precautions are used to limit the PHI that can be queried directly



Project Registry

Individual research groups may contribute their data or findings back to the Partners enterprise for querying and use by all researchers across the organization. Data is used for the greater good.



A Registry Centered Data Network

Genomic Data

Genomic data collected through the Biobank lives in a separate repository, but is made available for connecting with clinical data. **Only patients contained in the project registry can be queried within this network**



Project Registry

The central access point for this type of data network is the project specific registry. All queries will be limited to the patients that are part of this project.



Imaging Repository

DICOM Metadata is extracted from images downloaded from mi2b2. This may be supplemented by a limited amount of tags on all images given to us by Radiology group. Contains references to all patients from who we have imaging data. **Only patients contained in the project registry can be queried within this network**



Notes & Reports

Notes and reports on all patients are collected and put into a separate data repository that can be full text indexed. Specific security precautions are used to limit the PHI that can be queried directly. **Only patients contained in the project registry can be queried within this network**



Clinical Data

Repository that contains most clinical data from legacy systems as well as Epic for all patients across the enterprise. **Only patients contained in the project registry can be queried within this network**



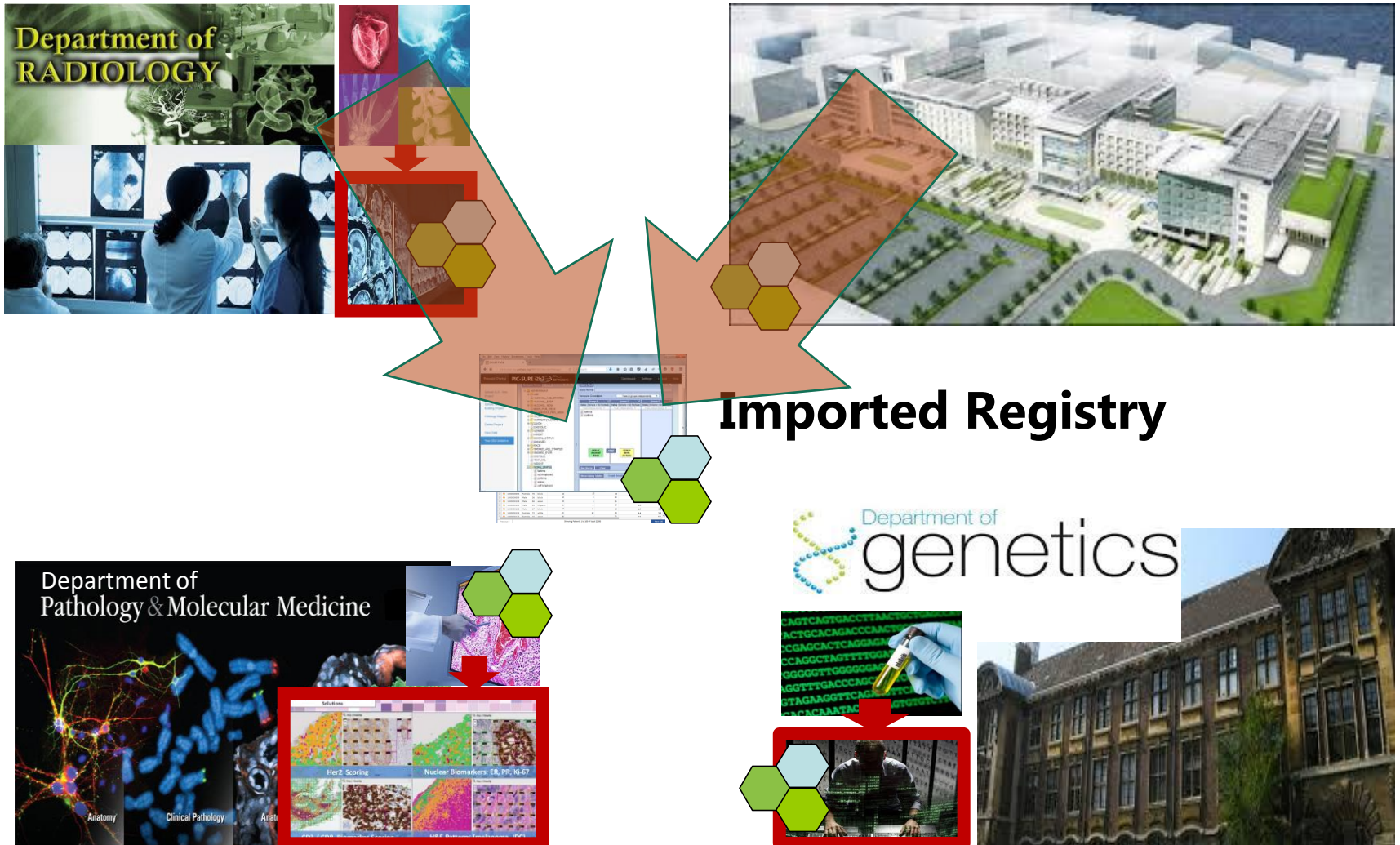
The screenshot shows the BRISKit web portal with a spreadsheet overlay. The spreadsheet is titled 'project2_1_optional_fields-1.xls' and is in 'Read-Only' mode. The active cell is A4, containing the value 'pr-00001'. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H
1	ID	OBS_START_DATE	AGE	GENDER	RACE	DEATH	MARITAL_STATUS	WORK_STATUS
2	null	null	Age	Sex	Race	Death	Marital Status	Work Status
3	null	null	age	Sex	Race	death	marital status	work status
4	pr-00001	02-06-2014 12:00:00	35 M	Hispanic	1	single	fulltime	
5	pr-00002	02-06-2014 12:30:00	46 F	Asian	1	married	not employed	
6	pr-00003	03-06-2014 13:23:00	54 M	White	1	widowed	parttime	
7	pr-00004	04-06-2014 11:45:00	33 F	White	1	separated	self employed	
8	pr-00005	null	65 M	Asian	0	divorced	retired	

tab Vital Status is mapped to the column Death in the 'Data' tab.
 This spreadsheet also contains 5 optional columns which are OBS_START_DATE, AGE, GENDER, RACE and DEATH.
 OBS_START_DATE is a special column which marks the start

- BRISKit is open source Biomedical Research Software as a Service Kit
- Developed by University of Leicester
- Allows spreadsheets of data to be auto imported into an i2b2 hive

Instantly Connected Databases in the Big Data Commons



File Edit View History Bookmarks Tools Help

Brisskit Portal

i2b2brisket.dipr.partners.org:8090/i2b2UploaderWebapp/i

Search

Brisskit Portal PIC-SURE i2b2 on BRISSKIT Dashboard Settings Logout Help

Upload XLS - New Project

Upload XLS - Existing Project

Ontology Mapper

Delete Project

View Data

Your i2b2 instance

Navigate Terms Find

- bd2kdistributed
 - AGE
 - ALCOHOL_AGE_STARTED
 - ALCOHOL_EVER
 - ALCOHOL_NOW
 - BEER_PER_WEEK
 - CIGARETTES_PER_WEEK
 - CL_STATUS
 - CURRENTLY_SMOKE
 - DEATH
 - WORK_STATUS
 - fulltime
 - not employed
 - parttime
- Biobank Genomics
 - People with genomic data (PC:9358 FC:9358 CC:1)
 - Illumina Multi-ethnic Genotyping Array (PC:4930 FC:4930 CC:1)
 - Illumina Multi-ethnic Genotyping Array b2_May2016 (PC:4428 FC:4428 CC:1)
 - Biobank Genomics MEGA Search
 - Illumina Multi Ethnic Genotyping Array (PC:1698262937 FC:1698262937 CC:6)
 - dbSNP rs Identifier (PC:4841 FC:1694143211)
 - indel (PC:4841 FC:4119726 CC:1)
 - SNP (PC:4841 FC:1694143211)
 - Gene (PC:4841 FC:1698262937 CC:1)
 - indel (PC:4841 FC:4119726 CC:1)
 - SNP (PC:4841 FC:1694143211)

Query Tool

Query Name:

Temporal Constraint: Treat all groups independently

Date	Code	Dates	Occurs > 0x	Exclude
				Treat independently

From Registry

From Biobank Portal

Show Query Status Graph Results Query Report

Services

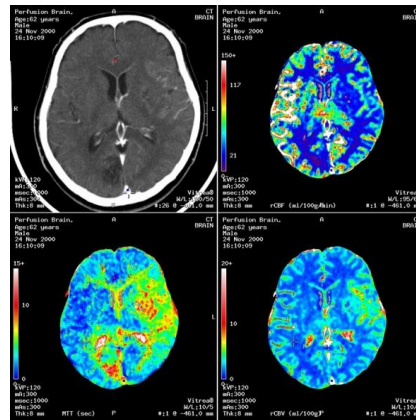
Perform Queries

The screenshot shows the BRISQ Portal interface. On the left, there is a sidebar with navigation options like 'Upload XLS - New Project', 'Upload XLS - Existing Project', 'Ontology Mapper', 'Delete Project', and 'View Data'. The main area is titled 'PIC-SURE i2b2 BRISQ' and contains a 'Query Tool'. The 'Query Name' field is empty. Below it, there are three 'Group' boxes for 'Temporal Constraint', each with 'Date' and 'Occurs' options. A 'Lifetime' section is also visible. At the bottom, there are buttons for 'Run Query', 'Clear', 'Show Query Status', 'Graph Results', and 'Query Report'.

Obtain Summary Tables

ID	Gender	Age	Race	Diabetes mellitus	Family history (Contaminant: Diabetes)	HbA1c (LSDIC-HbA1c)	HbA1c (LSDIC-HbA1c)	HbA1c (LSDIC-HbA1c)
1000000003	Male	41	asian	12	54	43	4.1	6.5
1000000008	Male	23	hispanic	3	12	54	4.2	6.7
1000000011	Female	60	white	21	21	43	4.6	7.1
1000000013	Female	83	black	34	9	65	4.7	5.9
1000000015	Male	32	hispanic	1	10	34	4.4	6.3
1000000029	Male	25	hispanic	3	25	25	5.1	6.8
1000000043	Male	61	hispanic	23	14	21	4.9	6.5
1000000045	Female	89	asian	37	10	34	4.9	6.8
1000000054	Female	65	asian	43	6	45	4.8	6.8
1000000066	Female	22	black	32	4	33	5.2	7.0
1000000070	Male	38	black	43	11	38	5.5	6.5
1000000077	Male	13	black	12	8	42	4.4	6.7
1000000080	Male	33	indian	32	13	48	4.2	6.5
1000000083	Female	65	black	33	4	39	4.7	6.6
1000000087	Male	55	asian	36	6	42	4.8	6.9
1000000096	Female	76	black	54	17	43	4.9	6.5
1000000099	Male	36	black	32	0	39	5.2	6.8
1000000108	Male	65	asian	43	1	25	5.3	6.7
1000000109	Male	63	hispanic	31	4	29	4.9	5.9
1000000112	Male	17	black	17	0	14	4.7	6.8
1000000119	Female	72	white	43	12	45	4.4	5.9
1000000133	Exam site	16	asian	20	3	75	4.6	5.1

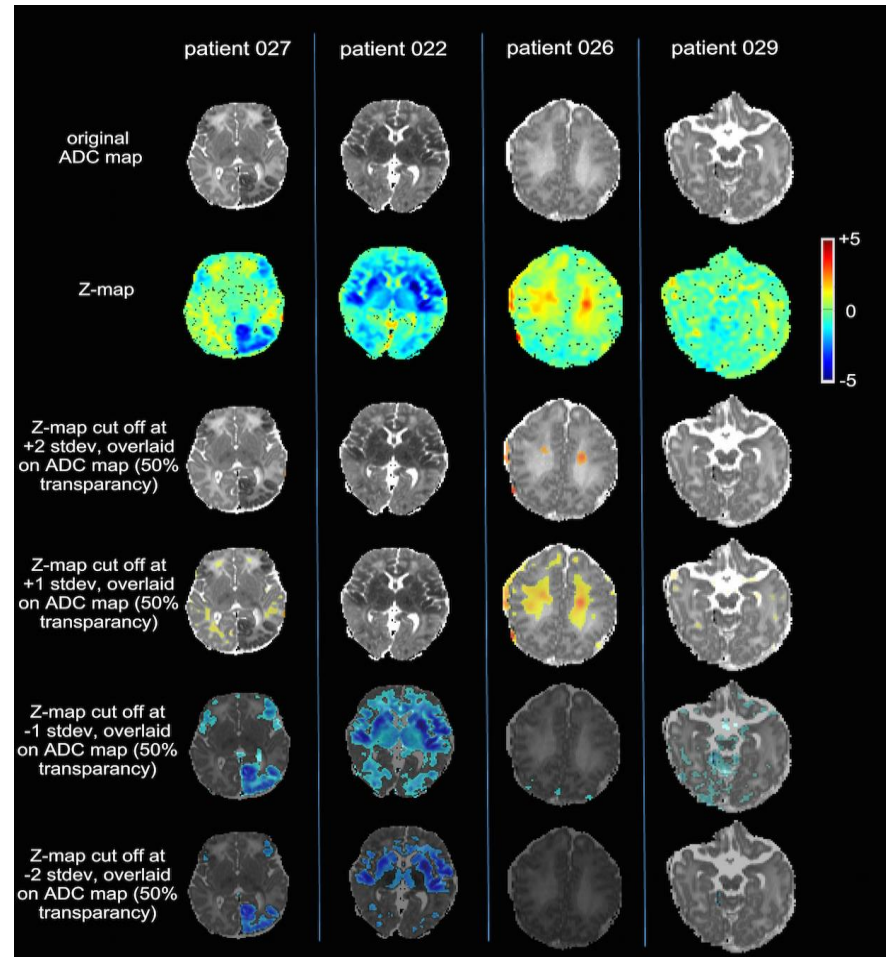
Link to Detailed Data



Atlases provide a visual guide for Radiology Decision Support, such as determining Perinatal Hypoxic Ischemic Encephalopathy

ADC map from 4 infants:
Each statistically compared to age matched atlas yields visual guide to pathology

**Quantitative analysis tools
+ large data sets = Great
insights for practicing
doctors**



Presentation of Baby Brains Registry in i2b2

The screenshot displays the i2b2 Query Tool interface within a web browser window titled "Partners Clinical Image Bank". The browser address bar shows the URL https://cib.partners.org/webclient_multi/. The interface includes a navigation pane on the left with a tree view of data categories such as "Data from Healthcare Record", "Imaging Data", and "Perinatal Data". The main area is the "Query Tool" window, which shows a query named "Image-Epile-6. Ac@14:59:19" with a "Non-Temporal Query" setting. The query is structured into three groups connected by "AND" operators. Each group contains a specific criterion: "Images", "Epilepsy and recurrent seizures - 123", and "6. Acute perinatal event - 162". Below the query groups, there are buttons for "Run Query", "Clear", and "New Group". At the bottom of the interface, a "Query Report" tab is active, displaying the results: "NUMBER OF PATIENTS" is 29, represented by a large number "29" with a person icon. The text below the number reads "For Query 'Image-Epile-6. Ac@14:59:19'".

Partners Clinical Image Bank

File Edit View History Bookmarks Tools Help

Partners Clinical Image Bank

Find Patients Make Table of Data Get Image Sets Help SNMO

Navigate Terms Find

Query Tool

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

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

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

one or more of these AND one or more of these AND one or more of these

Run Query Clear 3 Groups New Group

Show Query Status Graph Results Query Report

NUMBER OF PATIENTS

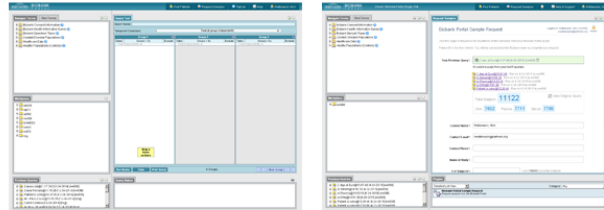
29

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

Sign the DUA



The Clinical Image Bank is a web-based query tool that allows Partners investigators to query and download data from curated phenotypically characterized pediatric patient cohorts. You can also view and make requests for medical images directly from the image bank.



Find patients using the advanced query tool

Request images from the Clinical Image Bank

Registration Form

Enter Your Information

Partners Username: *

E-mail address: *

Full Name (e.g. Doe, John): *

Sign Data Use Agreement

[\[Print this DUA \]](#) [\[View in new browser window \]](#)

Partners HealthCare Clinical Image Bank

Data Use Agreement - Clinical Image Bank Portal (September 11th, 2017)

Please read the terms of this Data Use Agreement ("Agreement") carefully before accessing the Partners HealthCare Clinical Image Bank Portal. The Clinical Image Bank Portal provides direct access to:

- (1) Limited Data Set health information, for purposes of running queries to determine the feasibility of conducting your potential research project(s) and the availability of clinical images for the project(s); and
- (2) Clinical Images and/or Limited Data Set health information, for purposes of conducting such project(s).

As used in this Agreement, "Limited Data Set" means health information that may include dates (such as admission dates and birth dates) and certain non-racial information (cities, states, and zip codes) but that excludes the

By checking this box, you agree to the Clinical Image Bank [Data Use Agreement](#).

[Complete Registration](#)

Definition to Extract Table for Analysis

The screenshot displays the Partners Clinical Image Bank web interface. The browser address bar shows https://cib.partners.org/webclient_multi/. The interface includes a navigation menu with options like 'Find Patients', 'Make Table of Data', 'Get Images', 'Help', and 'SNM0'. The main content area is titled 'Patient Set Viewer' and contains a 'Define View' tab. Below this, there is a section for 'Patients' with a selected query: 'Image-Epile-6. Ac@14:59:19 [Patient Count: 29]'. A 'Concepts' section allows for adding terms from a 'Navigate Terms' list. A table below lists various concepts with their constraints, aggregation options, and include checkboxes.

Concept	Constraints	Aggregation Option	Include
Patient Number		Value	<input type="checkbox"/>
Gender		Value	<input checked="" type="checkbox"/>
Age		Value	<input checked="" type="checkbox"/>
Race		Value	<input checked="" type="checkbox"/>
Images	[Set Date]	Images (Yes/No) [?]	<input type="checkbox"/>
Epilepsy and recurrent seizures	[Set Date]	Existence (Yes/No) [?]	<input type="checkbox"/>
6. Acute perinatal event	[Set Date]	Existence (Yes/No) [?]	<input type="checkbox"/>

Buttons at the bottom of the table include 'Review Patients' and 'Reset'.

Table Extracted

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 Healthcare logo and the title "Clinical Image Bank". Navigation tabs include "Find Patients", "Make Table of Data", "Get Images", "Help", and "SNM0".

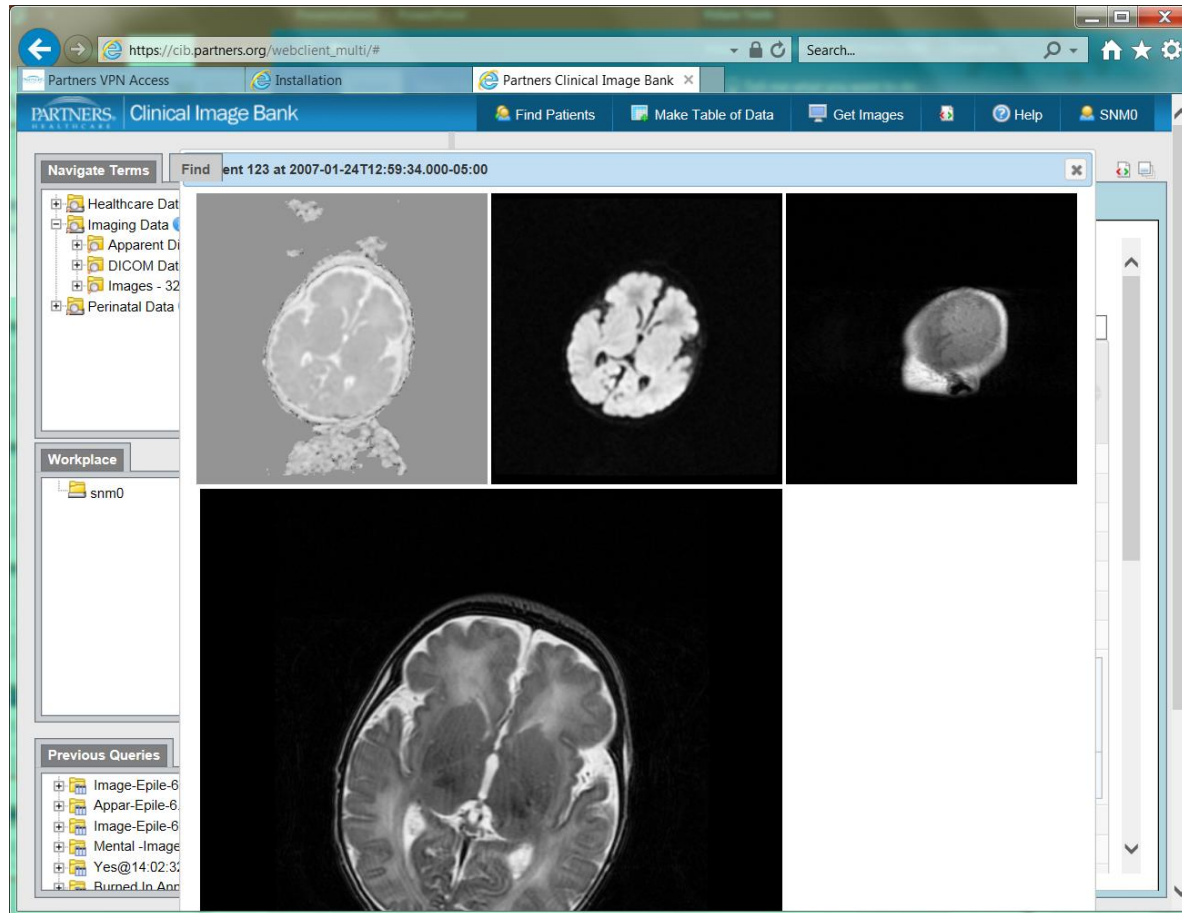
The main content area is titled "Patient Set Viewer" and contains a table of patient data. The table has the following columns: Gender, Age, Race, Images [Images (Yes/No)], Epilepsy and recurrent seizures [Existence (Yes/No)], and 6. Acute perinatal event [Existence (Yes/No)]. The table displays 29 rows of data, with the first 10 rows visible. The "View" link in the "Images" column of the 10th row is highlighted with a green box.

Below the table, there is a search bar and a "Download" button. The text above the table indicates "You are viewing rows 1-29 out of 29 rows, one row per patient." and "Tools: Download | Download Status:".

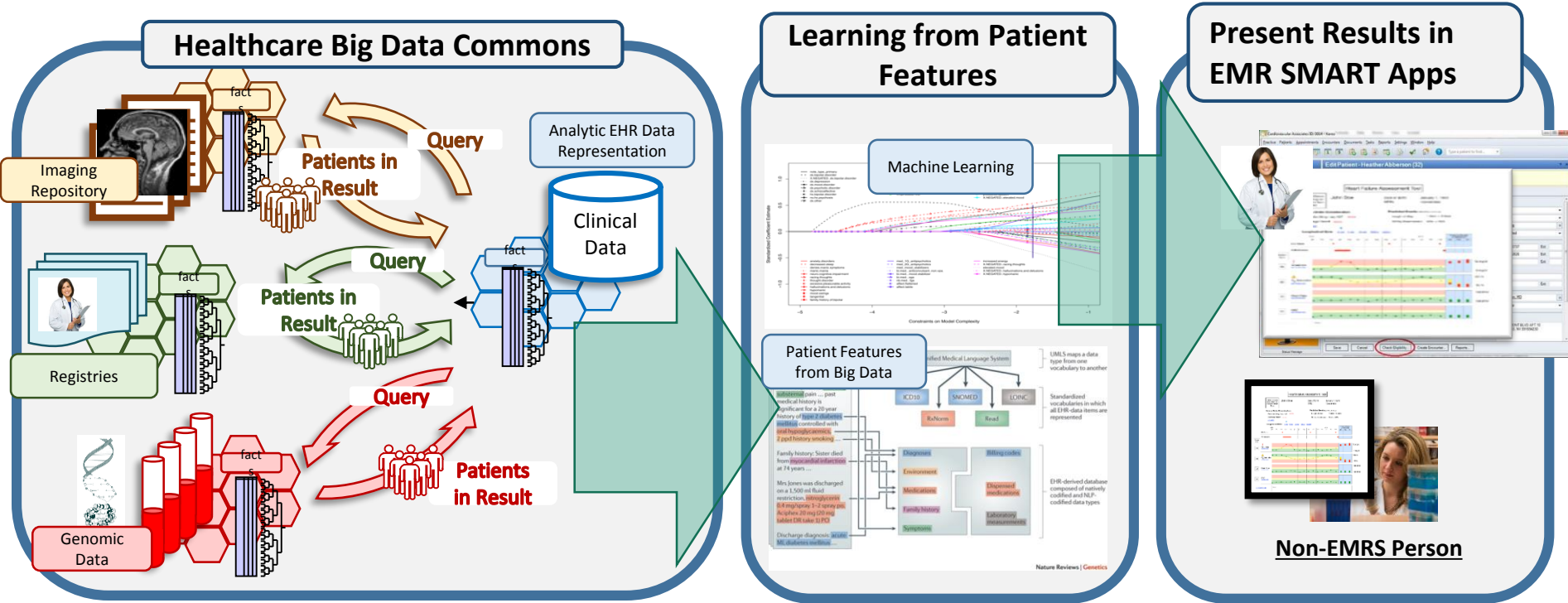
The left sidebar contains a "Navigate Terms" section with "Healthcare Data", "Imaging Data", and "Perinatal Data". Below it is a "Workplace" section with "snm0". At the bottom is a "Previous Queries" section with a list of queries including "Image-Epile-6. Ac@14:59:19 [9-25-2017] [snm0]", "Appar-Epile-6. Ac@14:58:12 [9-25-2017] [snm0]", "Image-Epile-6. Ac@14:56:22 [9-25-2017] [snm0]", "Mental -Images@14:55:00 [9-25-2017] [snm0]", "Yes@14:02:32 [9-22-2017] [snm0]", and ".Burned_In Annot@14:02:07 [9-22-2017] [snm0]".

Gender	Age	Race	Images [Images (Yes/No)]	Epilepsy and recurrent seizures [Existence (Yes/No)]	6. Acute perinatal event [Existence (Yes/No)]
F	13	White	View	Yes	Yes
F	9	Black	View	Yes	Yes
F	8	Unknown	View	Yes	Yes
F	7	White	View	Yes	Yes
F	4	Unknown	View	Yes	Yes
F	3	White	View	Yes	Yes
F	2	White	View	Yes	Yes
F	1	White	View	Yes	Yes
F	1	White	View	Yes	Yes
F	8	Unknown	View	Yes	Yes
F	0	Black	View	Yes	Yes
M	13	Unknown	View	Yes	Yes
M	14	White	View	Yes	Yes
M	11	White	View	Yes	Yes
M	10	White	View	Yes	Yes

Images Viewable



Flow of Healthcare Innovations



Tribute to...

■ I2b2/BD2K Core Team

- Issac Kohane
- Paul Avillach
- Griffin Weber
- Christopher Herrick
- Alyssa Goodson
- Lori Phillips
- Michael Mendis
- Victor Castro
- Janice Donahoe
- Nich Wattanasin
- Wayne Chan
- David Wang
- Mike Ollendieck
- Jeff Klann
- Andrew Cagan
- Bhaswati Ghosh
- Retta Metta

■ Biobank Team

- Natalie Boutin
- Scott Weiss
- Vivian Gainer

■ Innovation Team

- Randy Gollub
- Sandy Aronson
- Heidi Rehm
- Calum MacRea

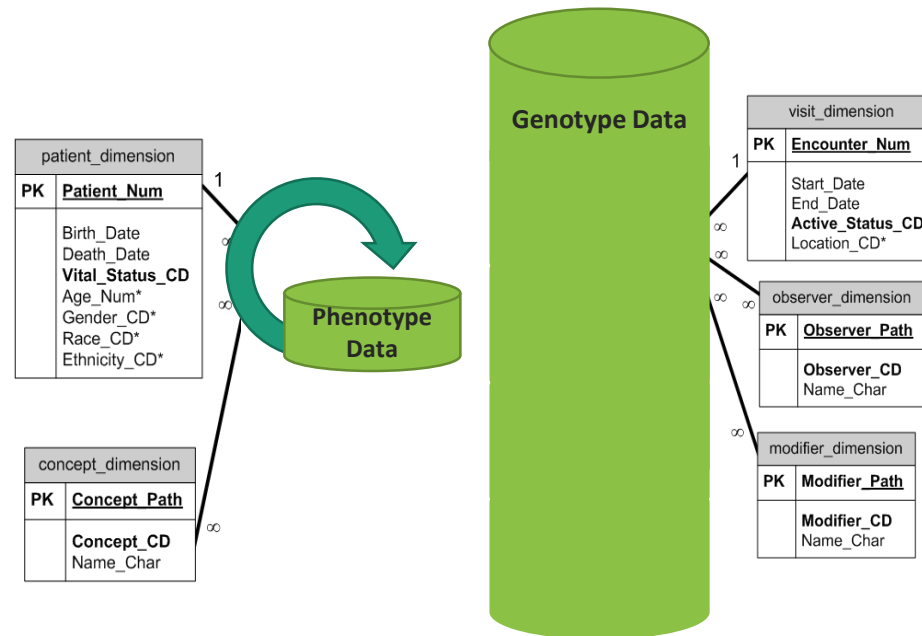
Thank You

Appendix

Multi-Fact Table in Genomic Use Case

Alyssa Goodson MS

Use of Static Genomic Fact Table



Variant Call Format (VCF)

- A specification maintained by the Global Alliance Data Working Group File Formats Task Team
- Used for describing genomic positions (loci)

metadata lines in the form of key=value pairs that define the data type and format of specific columns

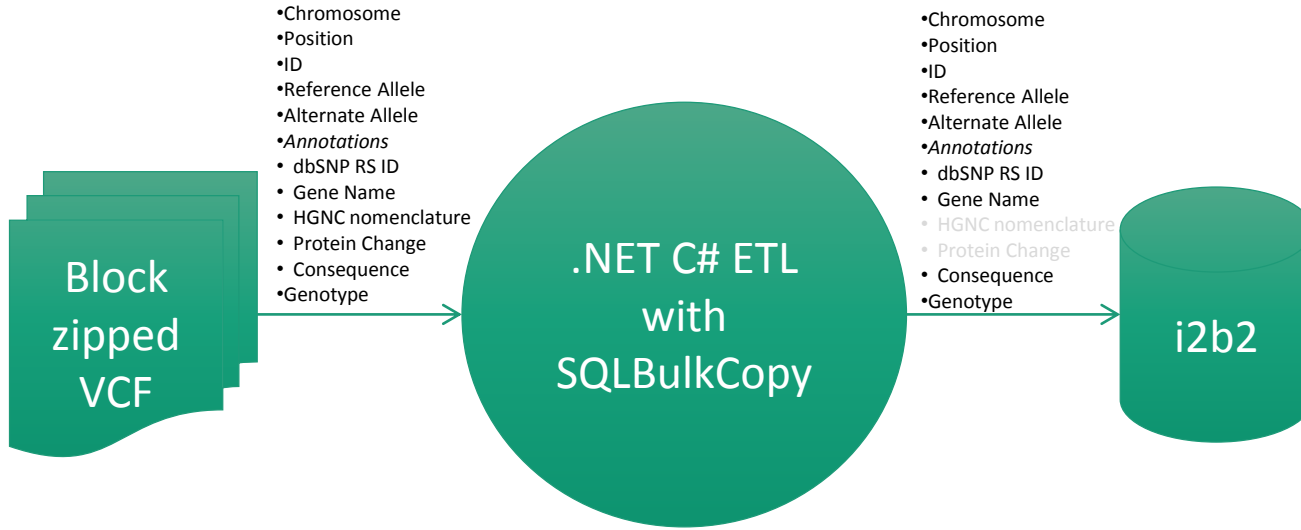
eight mandatory tab-delimited columns with the headers: #CHROM, POS, ID, REF, ALT, QUAL, FILTER, and INFO

```
##fileformat=VCFv4.2
##filedate=20160127
##source=PED
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=RSID,Number=.,Type=String,Description="dbSNP ID in build 142">
##INFO=<ID=VariantEffect,Number=.,Type=String,Description="Variant annotation (if present) in the following format: GeneSymbol|RefSeqTranscriptID:c.Nomenclature|RefSeqProteinID:p.Nomenclature|VariantEffect.">|
##contig=<ID=1,length=249250621>
##contig=<ID=2,length=243199373>
##contig=<ID=3,length=198022430>
##contig=<ID=4,length=191154276>
##contig=<ID=5,length=180915260>
...
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT 010061321010_R01C01-10007854
1 17538 JHU_1.17537 C A . . . . . RSID=rs200046632;VariantEffect=WASH7P|NR_024540.1:n.587+68G>T|p.?|intron GT 0/0
1 704251 JHU_1.704250 A G . . . . . GT 0/0
1 737263 JHU_1.737262 G A . . . . . RSID=rs369986014 GT 0/0
1 740243 JHU_1.740242 T C . . . . . RSID=rs11488375 GT 0/0
1 746189 JHU_1.746188 A G . . . . . RSID=rs139221807 GT 0/0
1 752566 rs3094315 G A . . . . . RSID=rs3094315;VariantEffect=FAM87B|NR_103536.1:n.-185G>A|p.=|upstream GT 1/1
1 776394 JHU_1.776393 G A . . . . . RSID=rs143876710;VariantEffect=LINC01128|NR_047519.1:n.288-6640G>A|p.?|intron GT 0/0
1 779286 rs149978434 C A . . . . . RSID=rs149978434;VariantEffect=LINC01128|NR_047519.1:n.288-3748C>A|p.?|intron GT 0/0
1 779744 JHU_1.779743 T G . . . . . RSID=rs145028227;VariantEffect=LINC01128|NR_047519.1:n.288-3290T>G|p.?|intron GT 0/0
1 783318 rs6686696 A G . . . . . RSID=rs6686696;VariantEffect=LINC01128|NR_047519.1:n.440+132A>G|p.?|intron GT 0/0
1 786796 JHU_1.786795 C T . . . . . RSID=rs148833734;VariantEffect=LINC01128|NR_047519.1:n.441-511C>T|p.?|intron GT 0/0
1 786949 JHU_1.786948 G T . . . . . RSID=rs140908266;VariantEffect=LINC01128|NR_047519.1:n.441-358G>T|p.?|intron GT 0/0
1 787217 JHU_1.787216 G A . . . . . RSID=rs369238012;VariantEffect=LINC01128|NR_047519.1:n.441-90G>A|p.?|intron GT 1/1
1 787428 JHU_1.787427 G A . . . . . RSID=rs575685800;VariantEffect=LINC01128|NR_047519.1:n.562G>A|exon GT 0/0
1 788713 JHU_1.788712 T C . . . . . RSID=rs2980306;VariantEffect=LINC01128|NR_047519.1:n.721-58T>C|p.?|intron GT 1/1
1 790760 JHU_1.790759 A G . . . . . RSID=rs80164748;VariantEffect=LINC01128|NR_047519.1:n.2710A>G|exon GT 1/1
1 791853 JHU_1.791852 G A . . . . . RSID=rs6684487;VariantEffect=LINC01128|NR_047519.1:n.3803G>A|exon GT 0/0
1 793521 JHU_1.793520 T G . . . . . RSID=rs376882577;VariantEffect=LINC01128|NR_047519.1:n.5471T>G|exon GT 0/0
1 793750 JHU_1.793749 G T . . . . . RSID=rs202099173;VariantEffect=LINC01128|NR_047519.1:n.5700G>T|exon GT 0/0
1 793770 JHU_1.793769 G A . . . . . RSID=rs147371531;VariantEffect=LINC01128|NR_047519.1:n.5720G>A|exon GT 0/0
1 794332 JHU_1.794331 G A . . . . . RSID=rs12127425;VariantEffect=LINC01128|NR_047519.1:n.6282G>A|exon GT 0/0
1 796775 JHU_1.796774 T C . . . . . RSID=rs111922608;VariantEffect=LINC01128|NR_047519.1:n.*1949T>C|p.=|downstream GT 0/0
1 797549 JHU_1.797548 G T . . . . . RSID=rs200288882 GT 0/0
```

- FORMAT column (optional) is used and contains the "GT" keyword to specify that genotype data exist
- The #CHROM, POS, REF and ALT fields are taken directly from the MEGA Consortium chip manifest file provided by Illumina
- ID field contains unique id for each position
- QUAL and FILTER fields are not utilized for these data
- Annotations for each position are stored in the INFO column

Final column = genotype of the individual at this genomic position

ETL



VCF

#CHROM	POS	ID	REF	ALT	...	INFO	...	SUBJECT_1
1	752566	rs3094315	G	A	...	RSID=rs3094315;VariantEffect=FAM87B NR_103536.1:n.-185G>A p.= upstream	...	1/1

FACT

PATIENT_NUM	CONCEPT_CD	INSTANCE_NUM	VALTYPE_CD	TVAL_CHAR	NVAL_NUM	OBSERVATION_BLOB
1	SO:0001483	338720	B	CHROM_1	752566	rs3094315,G_to_A,FAM87B,homozygous_ref,upstream,ID_rs3094315

Arrows in the original image connect the VCF row to the FACT row, showing the mapping of VCF fields to FACT columns.

I2b2 observation_fact table

CONCEPT_CD

- Two concepts with codes from Sequence Ontology: SNP (SO:0001483) or indel (SO:1000032)

INSTANCE_NUM

- The set of all SNPs for each patient will all have the same encounter number and date
- The concept codes will be the same for all SNPs (SO:0001483) and for all indels (SO:1000032).
- The set of all SNP facts will be enumerated in the instance_num field to make the primary key unique, as will the set of all indels.

VALTYPE_CD

- always equal "B" to indicate that data are stored in the observation_blob field and to trigger the full text search already existing in the i2b2 environment

TVAL_CHAR

- Chromosome

NVAL_NUM

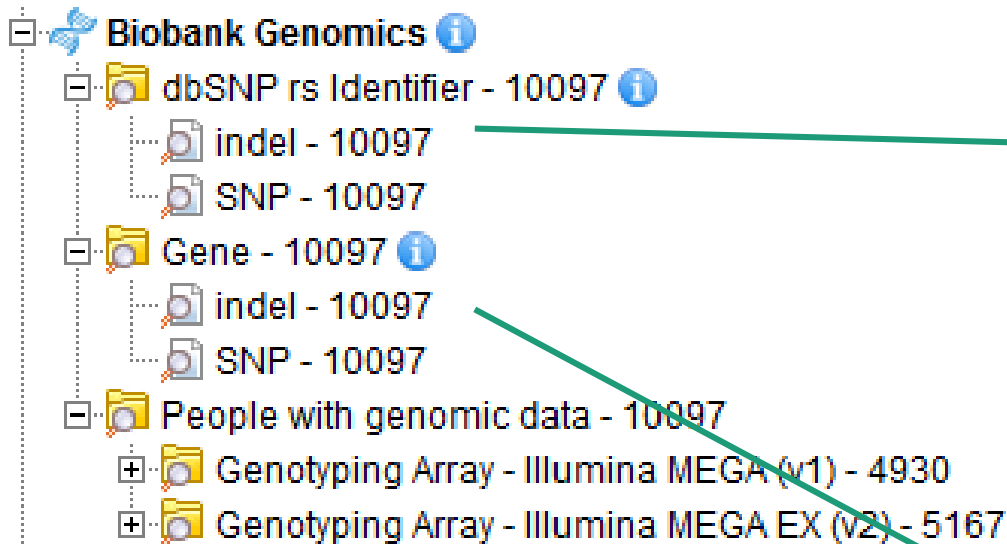
- Position

OBSERVATION_BLOB

<RSID | "missing_rsld">,<REF_TO_ALT>,<GENE_SYMBOL | "missing_gene">,<ZYGOSITY | "missing_zygosity">,<CONSEQUENCE | "missing_consequence"><CHIP_ID>

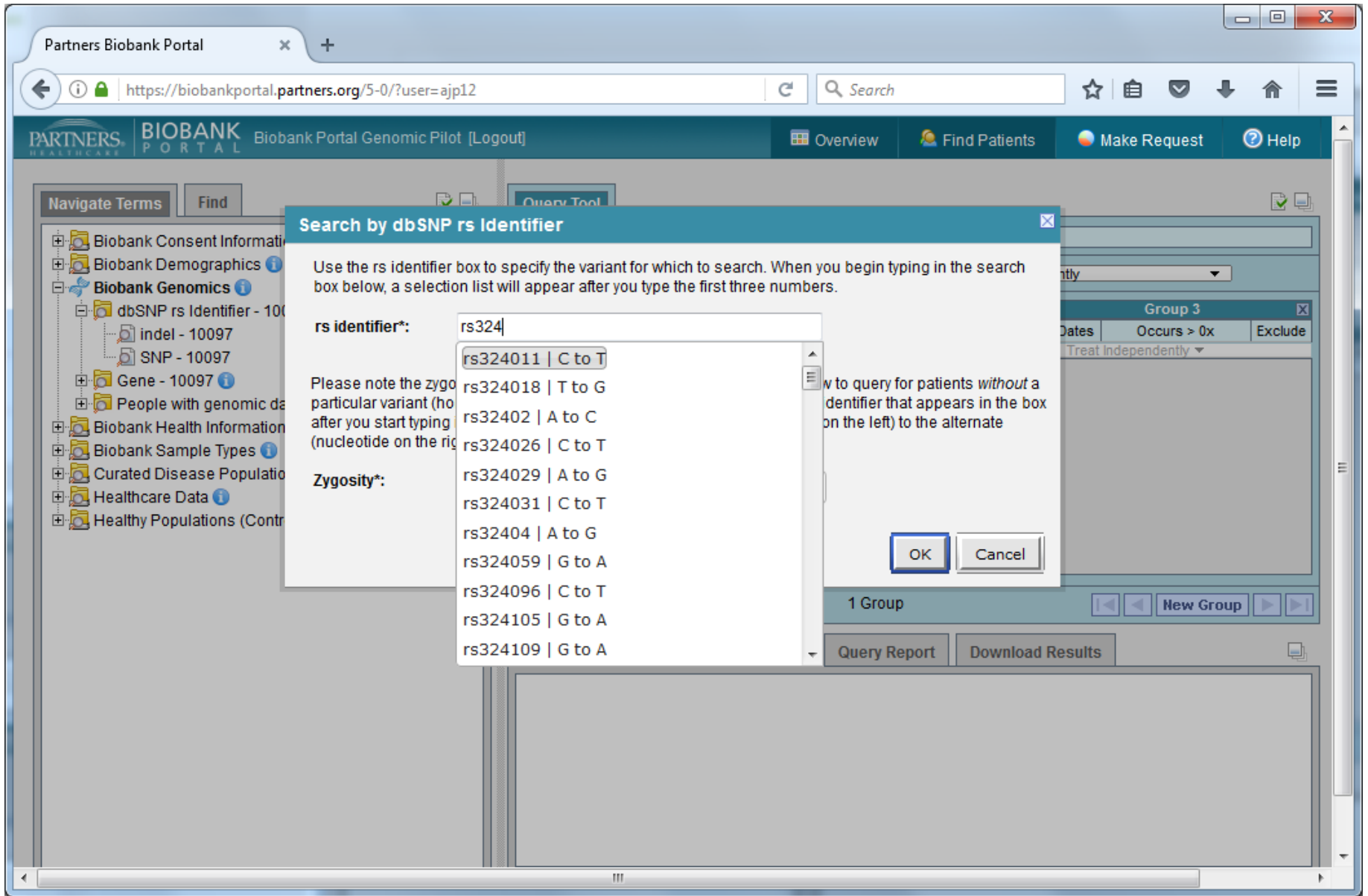
LARGESTRING search of OBSERVATION_BLOB			
CONCEPT_CD	INSTANCE_NUM	VALTYPE_CD	OBSERVATION_BLOB
SO:0001483	1	B	rs3094315,G_to_A,FAM87B,homozygous,upstream
SO:0001483	2	B	rs3131972,A_to_G,FAM87B,homozygous,upstream
SO:0001483	3	B	rs61770172,C_to_G,FAM87B,homozygous,exon
SO:0001483	4	B	rs3115860,C_to_A,FAM87B,homozygous,exon
SO:0001483	5	B	rs12567639,G_to_A,FAM87B,homozygous,downstream
SO:0001483	6	B	rs377214516,C_to_T,LINC01128,homozygous,upstream
SO:0001483	7	B	rs540936498,C_to_T,LINC00115,homozygous,exon

Ontology Formulation



```
<?xml version="1.0"?>
- <ValueMetadata>
  <Version>3.03</Version>
  <CreationDateTime>02/23/2016</CreationDateTime>
  <TestID/>
  <TestName>dbSNP rs Identifier</TestName>
  <DataType>GENOTYPE_RSID</DataType>
  <Oktousevalues/>
  <MaxStringLength>30</MaxStringLength>
  <EnumValues/>
  - <UnitValues>
    <NormalUnits/>
  </UnitValues>
</ValueMetadata>
```

```
<?xml version="1.0"?>
- <ValueMetadata>
  <Version>3.03</Version>
  <CreationDateTime>02/23/2016</CreationDateTime>
  <TestID/>
  <TestName>Gene</TestName>
  <DataType>GENOTYPE_GENE</DataType>
  <Oktousevalues/>
  <MaxStringLength>30</MaxStringLength>
  <EnumValues/>
  - <UnitValues>
    <NormalUnits/>
  </UnitValues>
</ValueMetadata>
```



Navigate Terms

Find

Query Tool

Search by dbSNP rs Identifier

Use the rs identifier box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first three numbers.

rs identifier*:

rs324

Please note the zygosity of the particular variant (shown after you start typing) (nucleotide on the right)

Zygosity*:

- rs324011 | C to T
- rs324018 | T to G
- rs32402 | A to C
- rs324026 | C to T
- rs324029 | A to G
- rs324031 | C to T
- rs32404 | A to G
- rs324059 | G to A
- rs324096 | C to T
- rs324105 | G to A
- rs324109 | G to A

How to query for patients *without* a particular identifier that appears in the box on the left) to the alternate

OK

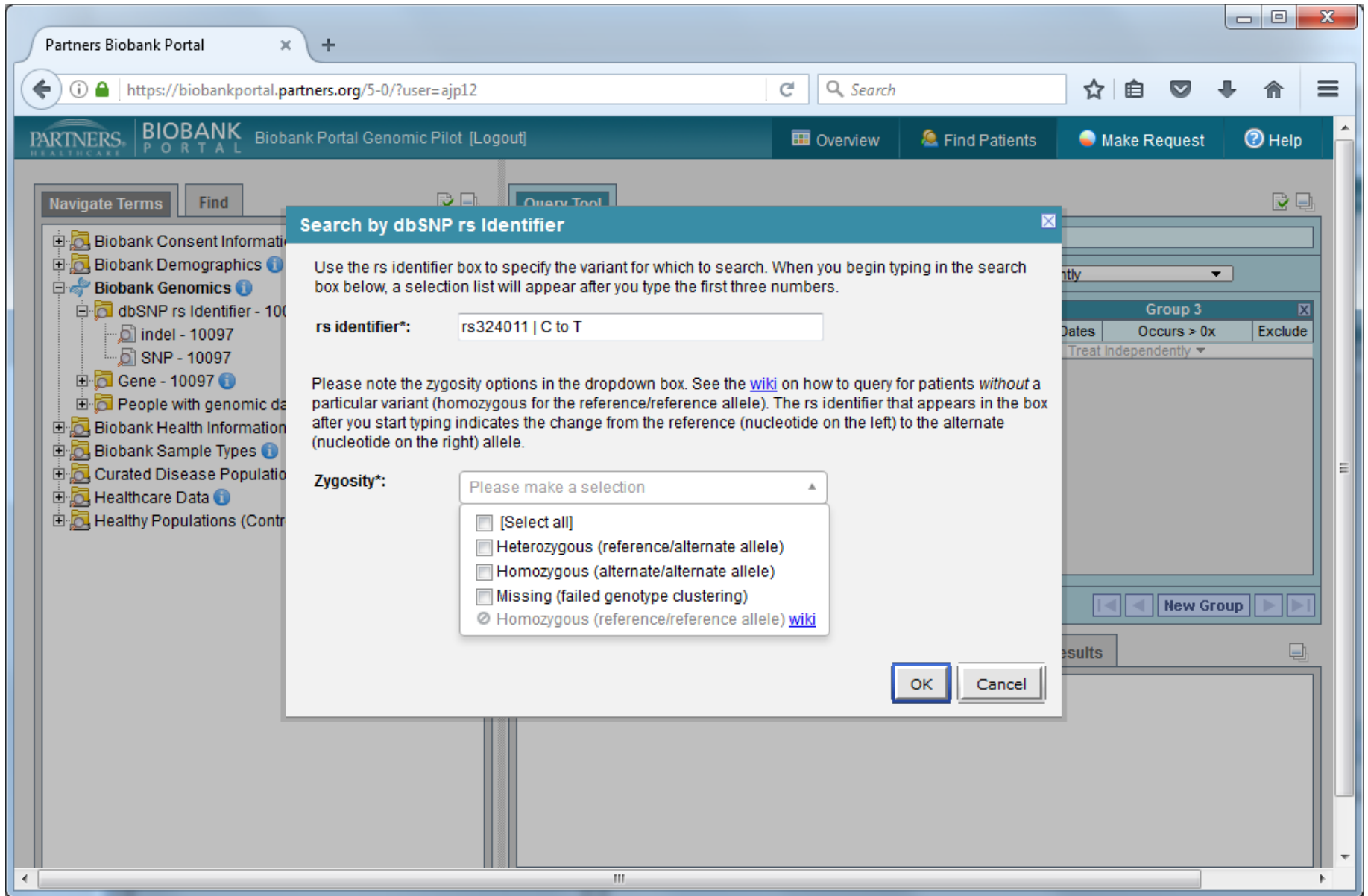
Cancel

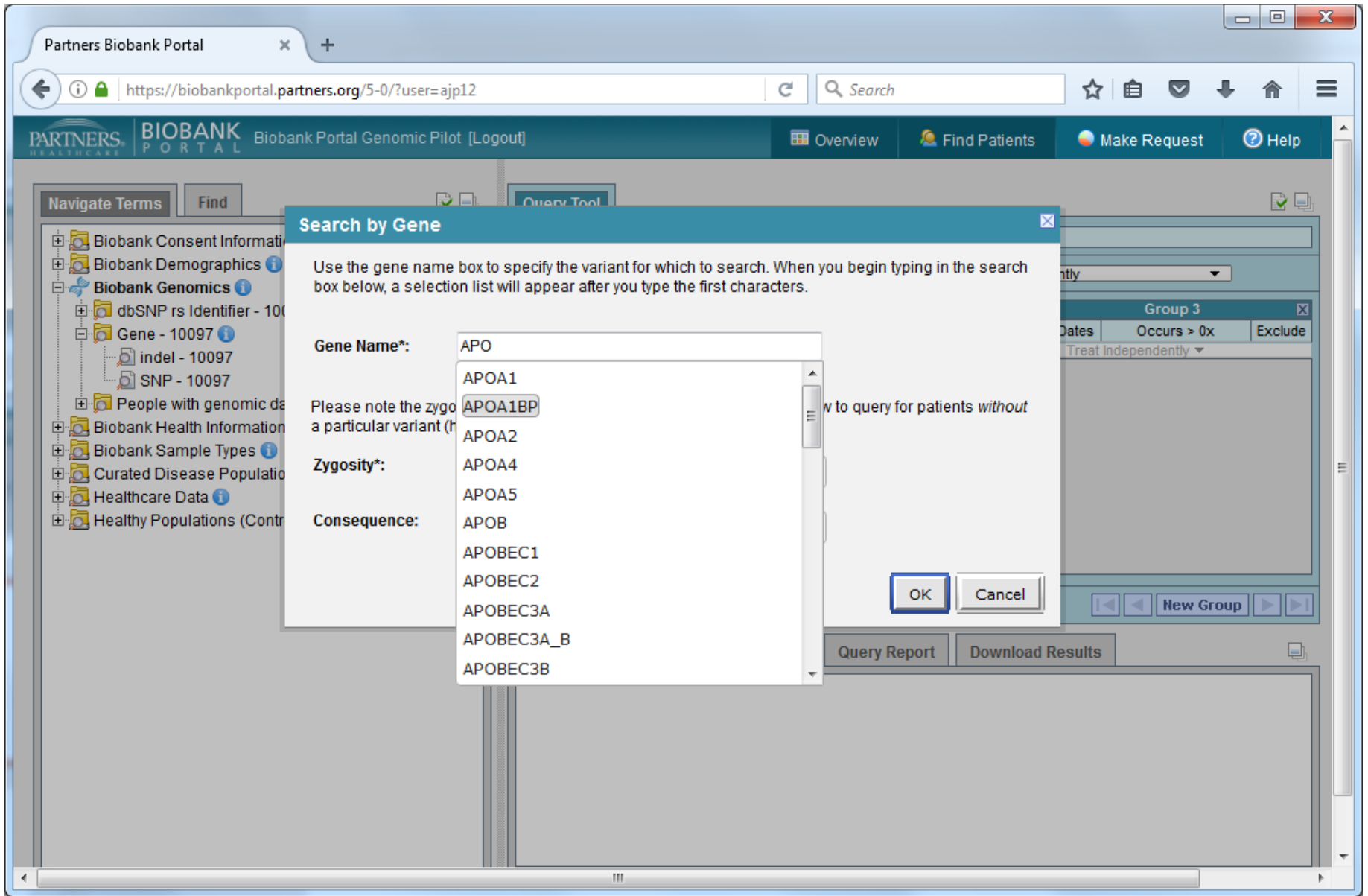
1 Group

New Group

Query Report

Download Results





Search by Gene

Use the gene name box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first characters.

Gene Name*:

APO

- APOA1
- APOA1BP
- APOA2
- APOA4
- APOA5
- APOB
- APOBEC1
- APOBEC2
- APOBEC3A
- APOBEC3A_B
- APOBEC3B

Please note the zygosity for a particular variant (if applicable)

Zygosity*:

Consequence:

OK Cancel

Navigate Terms Find

- Biobank Consent Information
- Biobank Demographics
- Biobank Genomics
 - dbSNP rs Identifier - 10097
 - Gene - 10097
 - indel - 10097
 - SNP - 10097
 - People with genomic data
- Biobank Health Information
- Biobank Sample Types
- Curated Disease Populations
- Healthcare Data
- Healthy Populations (Control)

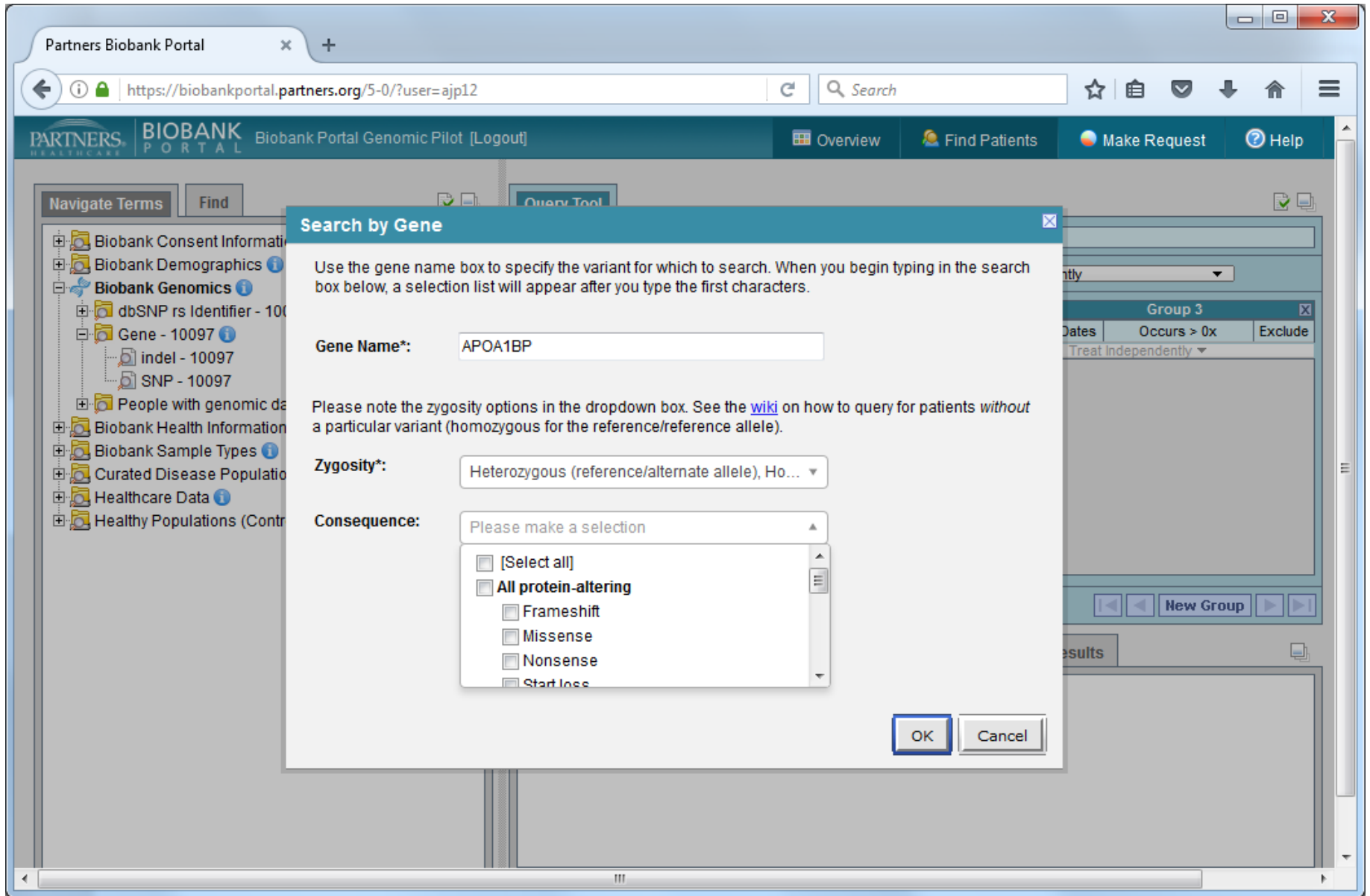
Overview Find Patients Make Request Help

Group 3

Dates	Occurs > 0x	Exclude
Treat Independently		

New Group

Query Report Download Results



Navigate Terms

Find

- Biobank Consent Information
- Biobank Demographics
- Biobank Genomics
 - dbSNP rs Identifier - 10097
 - Gene - 10097
 - indel - 10097
 - SNP - 10097
 - People with genomic data - 10097
- Biobank Health Information Survey
- Biobank Sample Types
- Curated Disease Populations
 - Asthma (AST)
 - Bipolar Disorder (BD)
 - Breast Cancer (BRCA)
 - Chronic Obstructive Pulmonary Disease (COPD)
 - Congestive Heart Failure (CHF)
 - Coronary Artery Disease (CAD)
 - CAD - current or past history (PPV 0.90) - 4963
 - CAD - current or past history (PPV 0.95) - 4711
 - CAD - current or past history (PPV 0.97) - 4462
 - CAD - no history (NPV 0.99) - 41381
 - Crohn's Disease (CD)
 - Depression (DEP)
 - Epilepsy (EPIL)
 - Cyst (COUT)

Query Tool

Query Name: Gene-CAD - c@13:06:52

Temporal Constraint:

Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
Gene [contains "EGFR AND (Heterozygous OR Homozygous) AND (Frameshift OR missense OR nonsense OR start_loss OR stop_loss)"]			CAD - current or past history (PPV 0.97) - 4462					
one or more of these			AND			one or more of these		
			AND			drop a term on here		

Run Query

Clear

2 Groups

New Group

Show Query Status

Graph Results

Query Report

Download Results

NUMBER OF PATIENTS


1009

For Query "Gene-CAD - c@13:06:52"

Query Formulation in SQL

dbSNP rs identifier

```
select count(distinct patient_num)
from observation_fact
where contains(observation_blob, 'FAM148 AND (stop_loss OR
missense)')
```

Gene Name

```
select count(distinct patient_num)
from observation_fact
where contains(observation_blob, 'rs183605470 AND
heterozygous')
```

Times to complete queries

