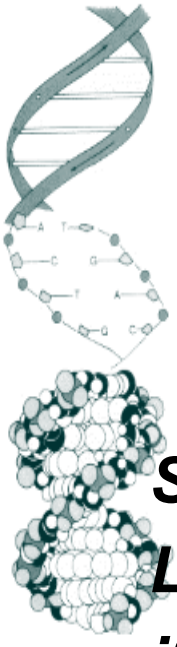


Ontology Approaches for PCORI

Shawn Murphy MD, Ph.D.
Lori Phillips MS

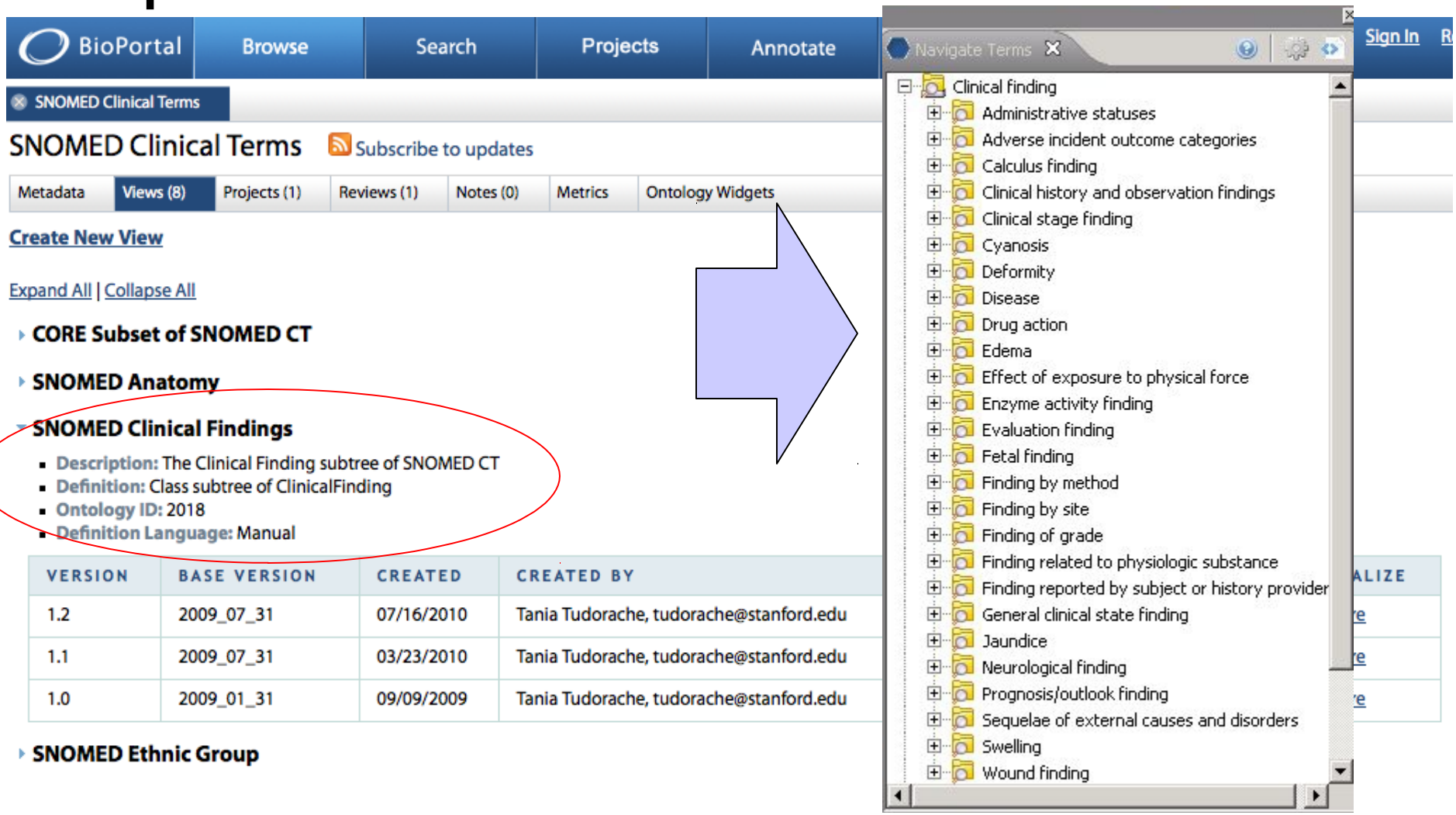
i2b2/SHRINE AUG

July 11, 2014

Approach – Expressing Data

- I2b2 does not require any specific ontology
  - Queries can be created using any system of local codes
  - Allows i2b2 to attach data of nearly any kind to patients, visits, and providers

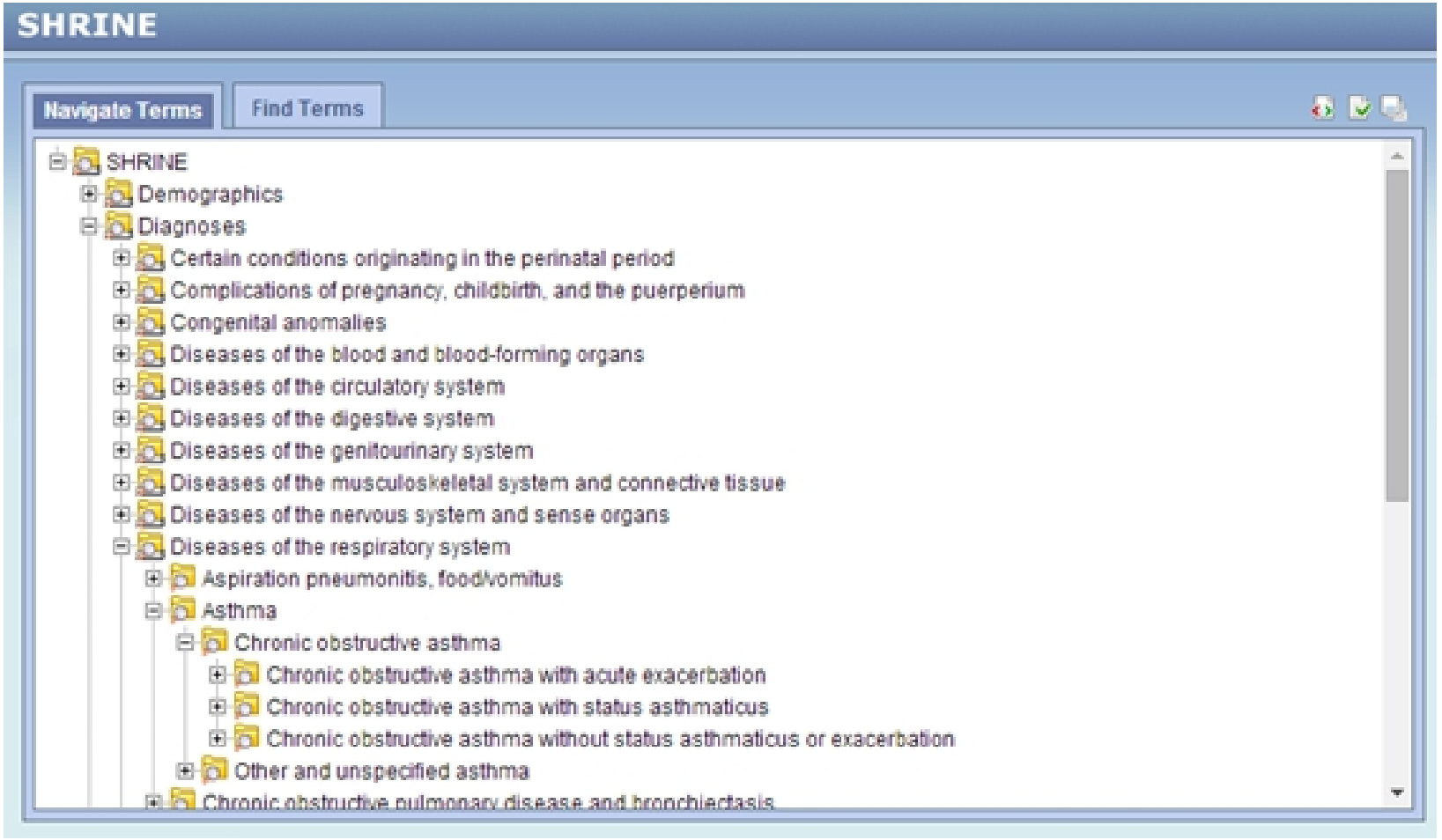
Represents Hierarchies



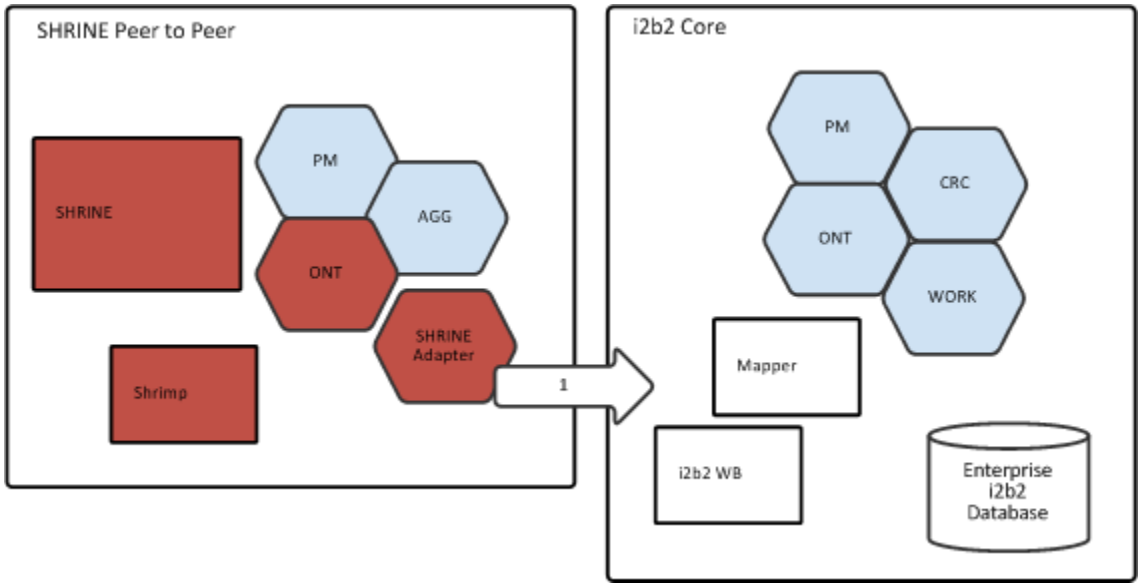
Approach - Networks

- The flexibility of i2b2 becomes a weakness when constructing data networks
  - There is no guarantee that people will choose the same ontologies at every site.
  - These can be roughly approximated using mappings

SHRINE Mappings



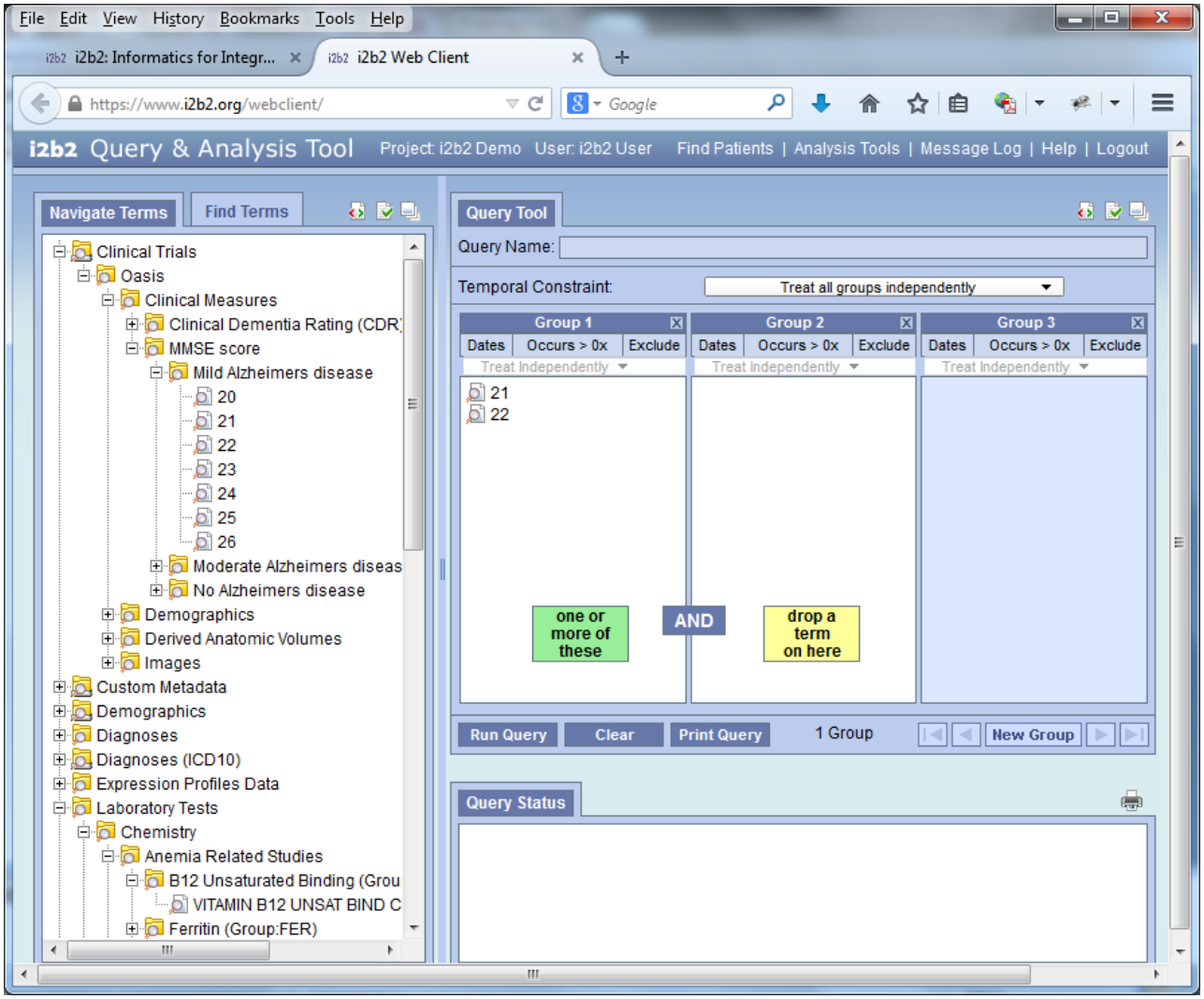
Scope of SHRINE Mappings



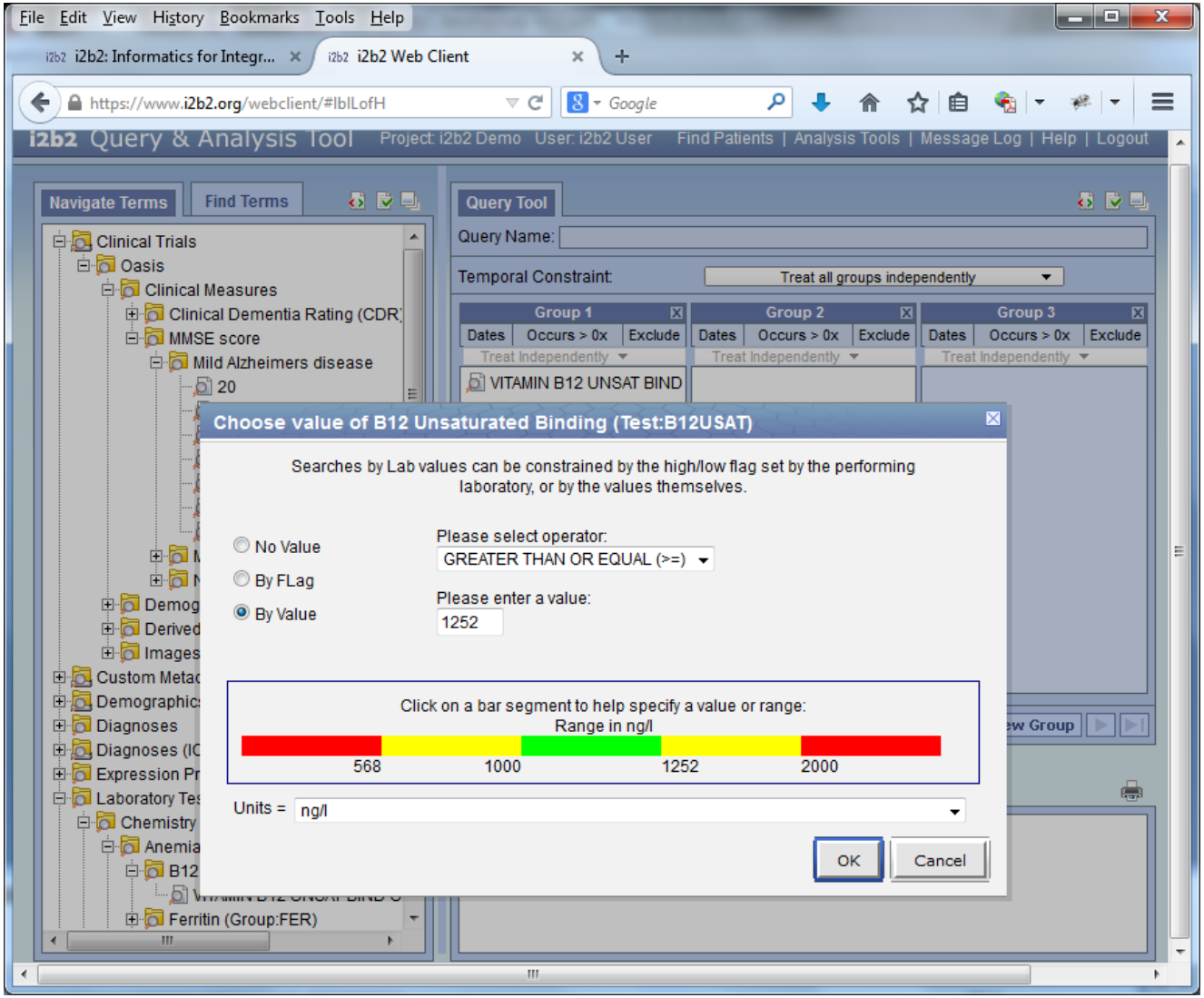
Approach – Modifiers and Values

- Full data representations in i2b2, often needing modifiers and values, are labeled “Information Models”
- These can be managed in i2b2 with relatively minor impact on the local observation codes
- However, the modifiers and values of the basic concepts need to be represented in similar ways to allow simple mappings to be effective.
- Also, for performance, it is useful to represent concept hierarchies in similar ways in the mappings

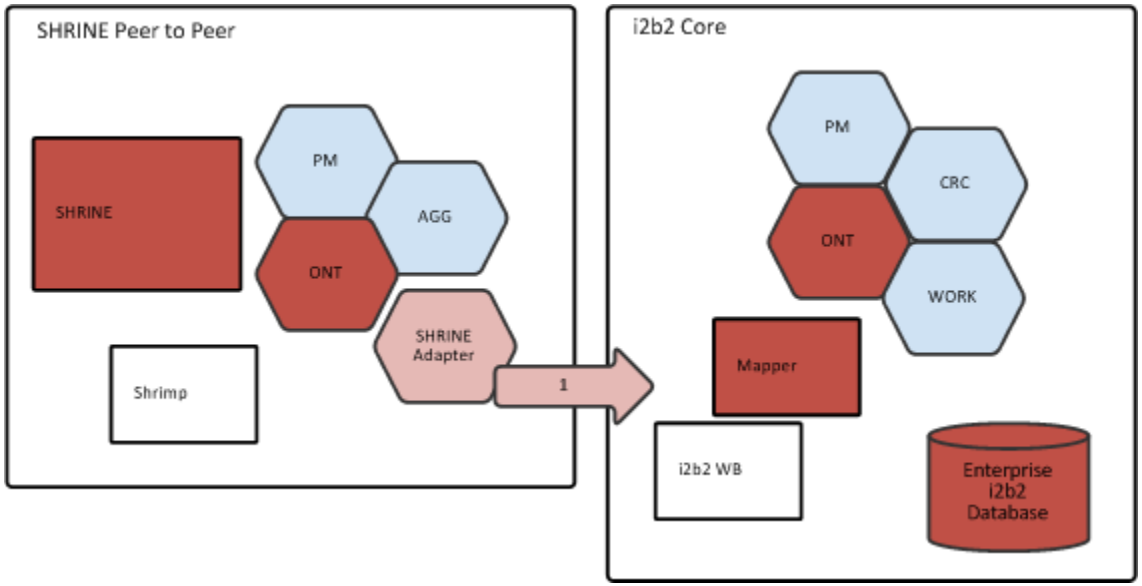
Represent scores as concepts



Represent scores as values



Scope of solution to allow information model equivalencies



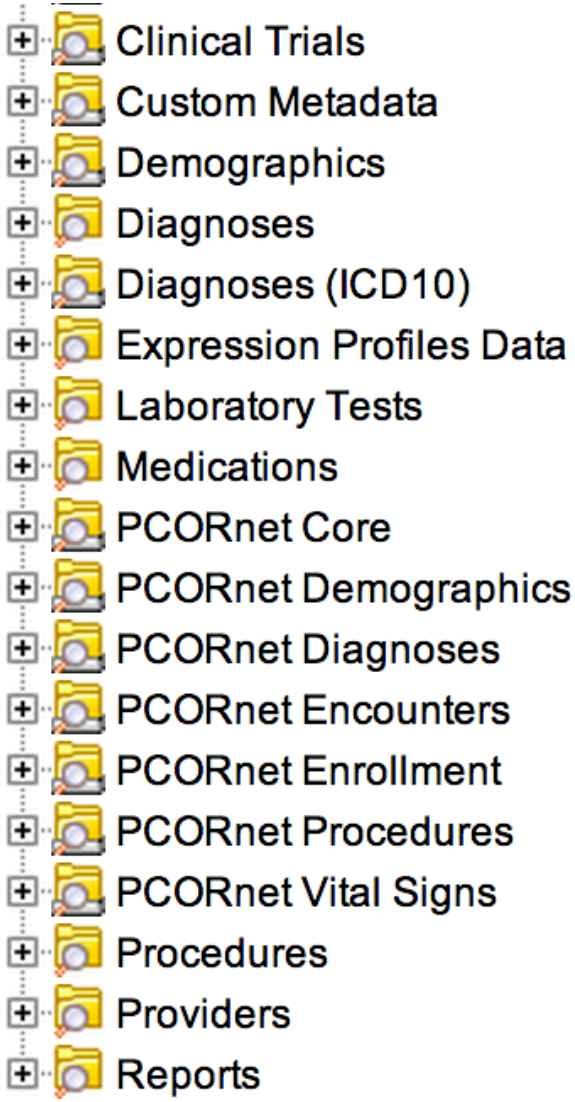
Arrive at consensus through distributed ontologie

- Approach
  - Bring together existing ontologies of the group to understand the best ontology representation of a new information model
  - Develop a consensus ontology and attach to local instance
  - Raw codes in observation-fact table are mapped to “Consensus” ontology
  - SHRINE mapping capabilities used for “finishing touches”

New Information Model Ontology

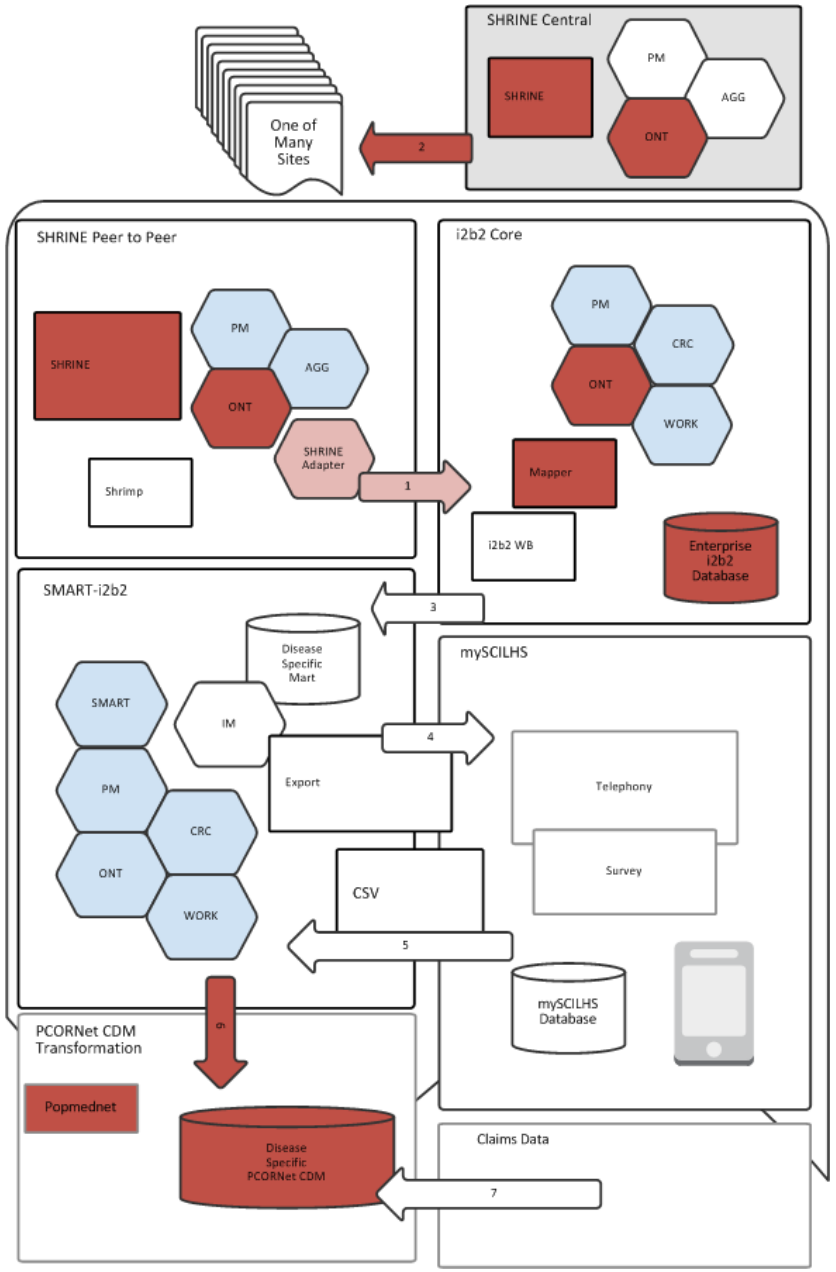
Consensus Ontology can live alongside other ontologies

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



Supporting Full PCORI SCILHS workflow in i2b2

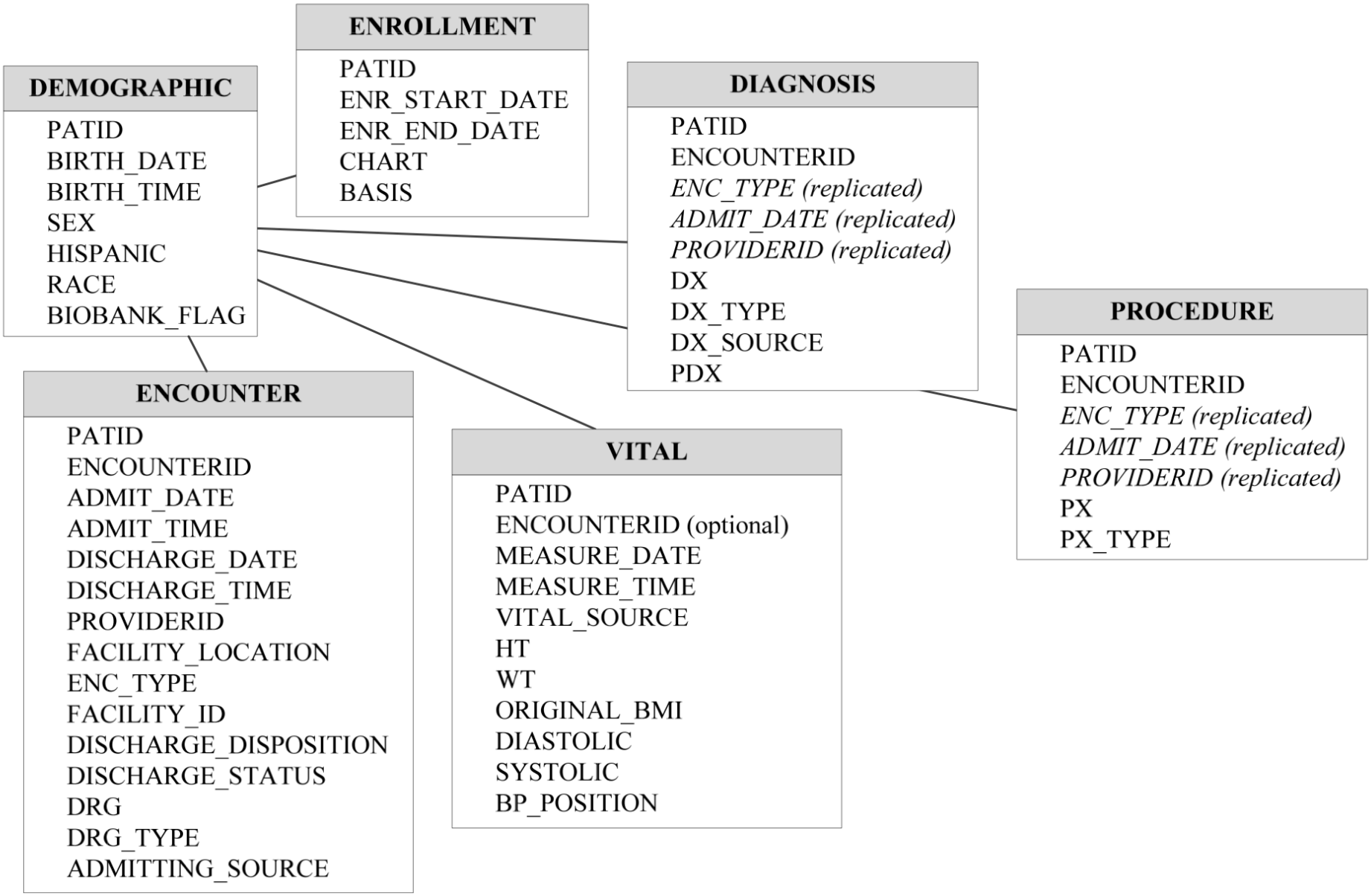
- To support the PCORNet CDM Information Model, we are distributing full SCILHS ontologies that are to be attached to the local SCILHS instance of i2b2.
- The SCILHS ontologies can accommodate the information model with simple mappings
- The SCILHS ontologies will support data transformations to the PCORNet CDM



Contributors

- Ontology Contributors
  - Jeff Klann
  - Dan Connolly
  - Lori Phillips
  - Nathan Wilson
- Release
  - https://www.i2b2.org/webclient/
  - Change the username to pcori
  - Public release early fall

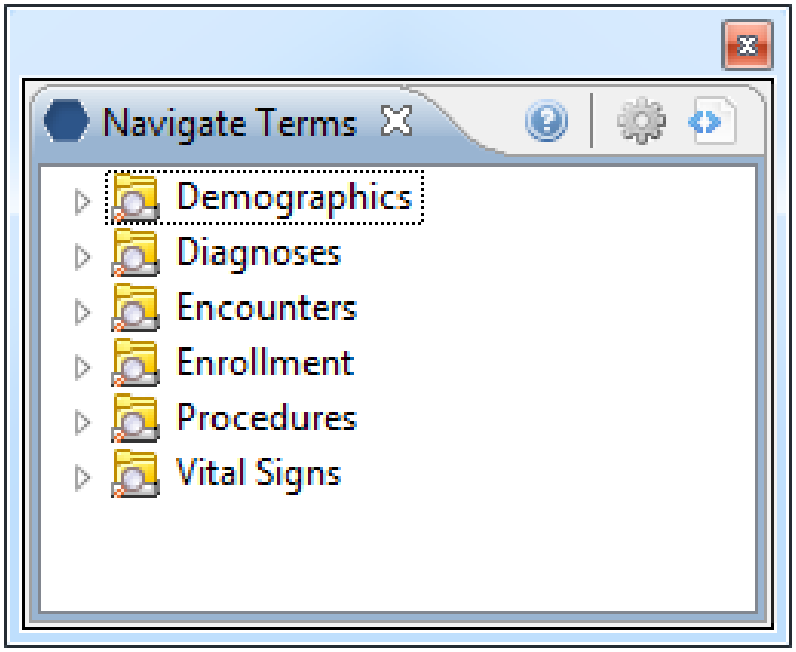
PCORnet CDM


ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
BASIS
DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG
DIAGNOSIS
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX
PROCEDURE
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX
PX_TYPE
ENCOUNTER
PATID
ENCOUNTERID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITY_ID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE
VITAL
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
ORIGINAL_BMI
DIASTOLIC
SYSTOLIC
BP_POSITION

PCORI-centric Use Case

A requirement in PCORI is that all i2b2 instances have to map their ontologies to a central PCORI ontology.

- Local codes need to be integrated into the PCORI hierarchy
- Tools and strategies to assist with this effort



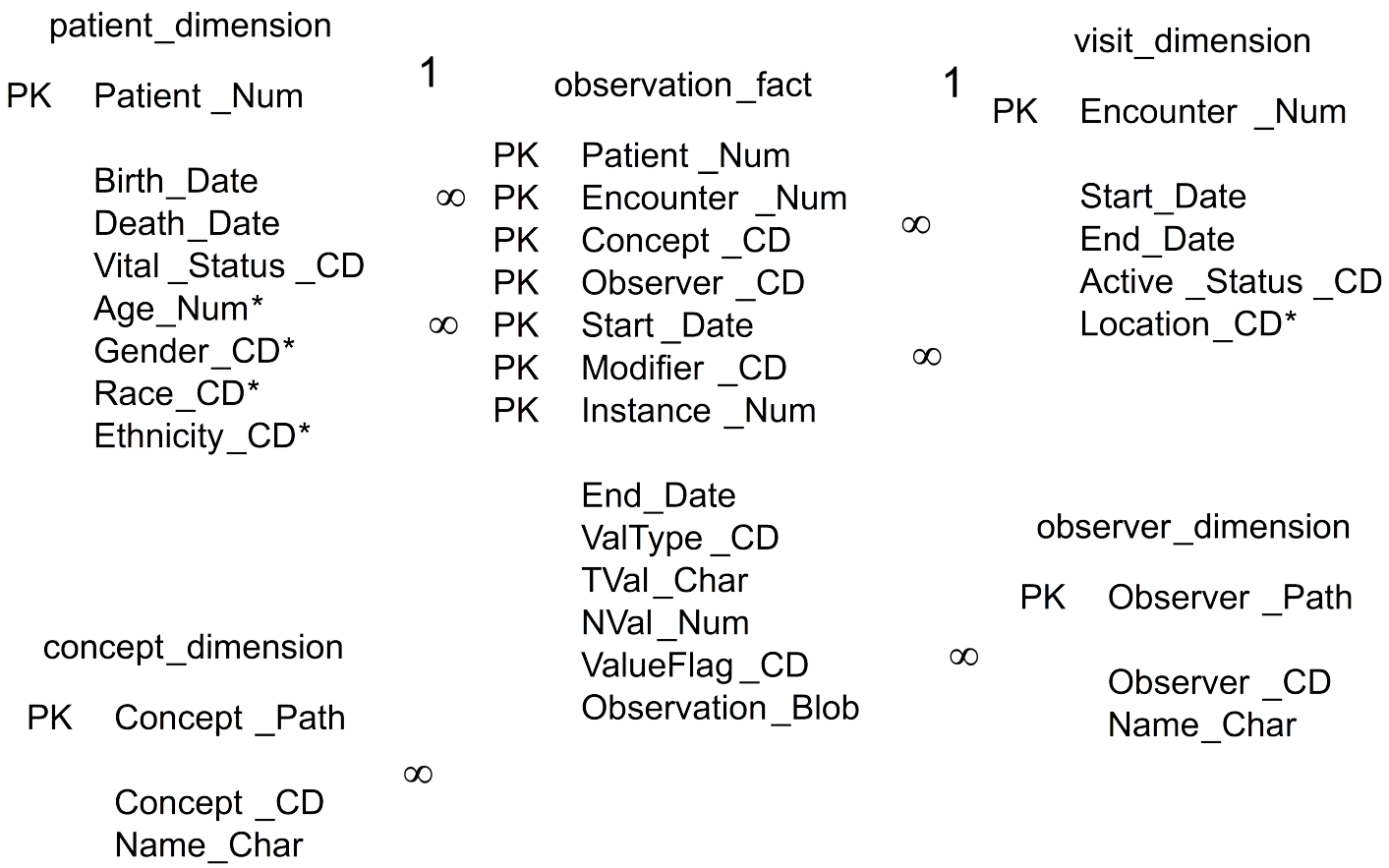
Constructing queries from metadata

METADATA

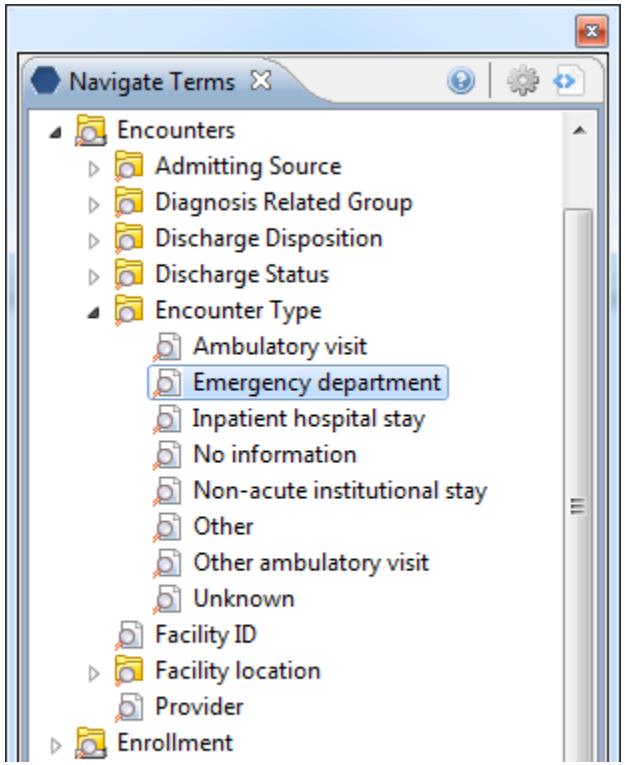
C_HLEVEL	INT NULL
C_FULLNAME	VARCHAR(900) NULL
C_NAME	VARCHAR(2000) NULL
C_SYNONYM_CD	CHAR(1) NULL
C_VISUALATTRIBUTES	CHAR(3) NULL
C_TOTALNUM	INT NULL
C_BASECODE	VARCHAR(450) NULL
C_METADATAXML	TEXT NULL
C_FACTTABLECOLUMN	VARCHAR(50) NULL
C_TABLENAME	VARCHAR(50) NULL
C_COLUMNNAME	VARCHAR(50) NULL
C_COLUMNDATATYPE	VARCHAR(50) NULL
C_OPERATOR	VARCHAR(10) NULL
C_DIMCODE	VARCHAR(900) NULL
C_COMMENT	TEXT NULL
C_TOOLTIP	VARCHAR(900) NULL
UPDATE_DATE	DATETIME NULL
DOWNLOAD_DATE	DATETIME NULL
IMPORT_DATE	DATETIME NULL
SOURCESYSTEM_CD	VARCHAR(50) NULL
VALUETYPE_CD	VARCHAR(50) NULL

```
select patient_num from
   observation_fact where
   [c_facttablecolumnname]
   IN
(select
   [c_facttablecolumnname]
   from [c_tablename]
   where [c_columnname]
   [c_operator] [c_dimcode])
```

i2b2 Star Schema



Visit_dimension types of queries:Encounter types



select patient_num
from visit_dimension
where inout_cd IN
('ED','E')

C_NAME	C_FULLNAME	C_BASECODE	C_FACTTABLECOLUMN	C_TABLENAME	C_COLUMNNAME	C_OPERATOR	C_DIMCODE
Ambulatory visit	\PCORI\ENCOUNTER\ENC_TYPE\AV\	ENCTYPE:AV	patient_num	VISIT_DIMENSION	inout_cd	in	'AV','O'
Emergency department	\PCORI\ENCOUNTER\ENC_TYPE\ED\	ENCTYPE:ED	patient_num	VISIT_DIMENSION	inout_cd	in	'ED','E'
Inpatient hospital stay	\PCORI\ENCOUNTER\ENC_TYPE\IP\	ENCTYPE:IP	patient_num	VISIT_DIMENSION	inout_cd	in	'IP','I'
No information	\PCORI\ENCOUNTER\ENC_TYPE\NI\	ENCTYPE:NI	patient_num	VISIT_DIMENSION	inout_cd	is	NULL
Non-acute institutional stay	\PCORI\ENCOUNTER\ENC_TYPE\IS\	ENCTYPE:IS	patient_num	VISIT_DIMENSION	inout_cd	in	'IS'
Other	\PCORI\ENCOUNTER\ENC_TYPE\OT\	ENCTYPE:OT	patient_num	VISIT_DIMENSION	inout_cd	in	'OT'
Other ambulatory visit	\PCORI\ENCOUNTER\ENC_TYPE\OA\	ENCTYPE:OA	patient_num	VISIT_DIMENSION	inout_cd	in	'OA'
Unknown	\PCORI\ENCOUNTER\ENC_TYPE\UN\	ENCTYPE:UN	patient_num	VISIT_DIMENSION	inout_cd	in	'UN','@'

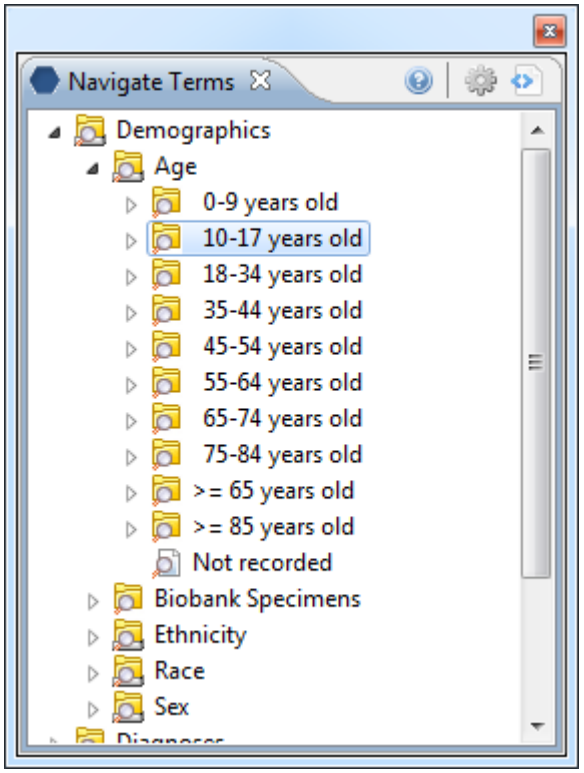
Edit c_dimcode to match your inout_cds.

A word about missing/unknown values

- The PCORnet CDM uses the HL7 conventions of “Null Flavors” (http://hl7.org/implement/standards/fhir/v3/NullFlavor/) as a basis for representing missing or unknown values. Specifically, we will populate these values as follows:

  - ‘No Information’ means that the data field is present in the source system, but the source value is null or blank.
  - ‘Unknown’ means that the data field is present in the source system, but the source value explicitly denotes an unknown value.
  - ‘Other’ means that the data field is present in the source system, but the source value cannot be mapped to the choices presented in the PCORI CDM ontology.

Patient_dimension types of queries:Age calculations

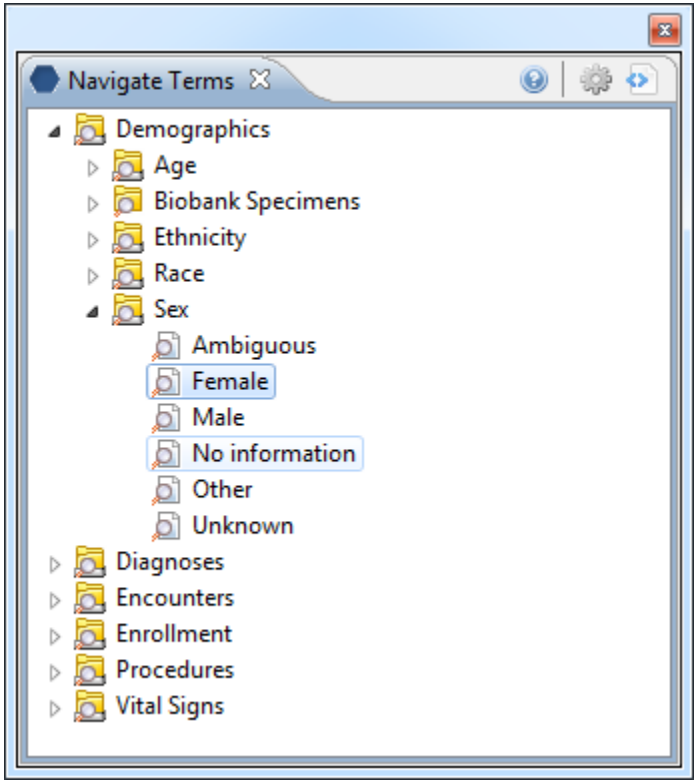



```
select patient_num from
patient_dimension
where birth_date
BETWEEN
sysdate – (365.25*18)
AND sysdate –
(365.25*10)
```

	c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
1	0-9 years old	patient_num	patient_dimension	birth_date	>	getdate() - (365.25*10)
2	10-17 years old	patient_num	patient_dimension	birth_date	BETWEEN	getdate() - (365.25*18) AND getdate() - (365.25*10)

No modifications necessary.

Patient_dimension types of queries:Sex

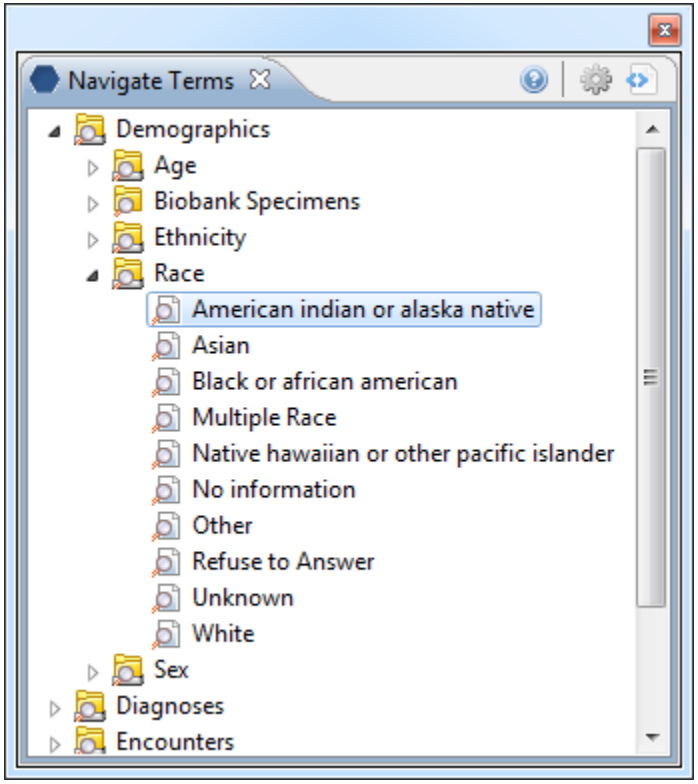



```
select patient_num
from
patient_dimension
where sex_cd IN ('F')
```

C_NAME	C_FULLNAME	C_FACTTABLECOLUMN	C_TABLENAME	C_COLUMNNAME	C_OPERATOR	C_DIMCODE
Ambiguous	\PCORI\DEMOGRAPHIC\SEX\A\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'A'
Female	\PCORI\DEMOGRAPHIC\SEX\F\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'F'
Male	\PCORI\DEMOGRAPHIC\SEX\M\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'M'
No information	\PCORI\DEMOGRAPHIC\SEX\NI\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'NI'
Other	\PCORI\DEMOGRAPHIC\SEX\OT\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'OT'
Unknown	\PCORI\DEMOGRAPHIC\SEX\UN\	PATIENT_NUM	PATIENT_DIMENSION	SEX_CD	IN	'UN'

Edit c_dimcode to match your sex_cds.

Patient_dimension types of queries:Race

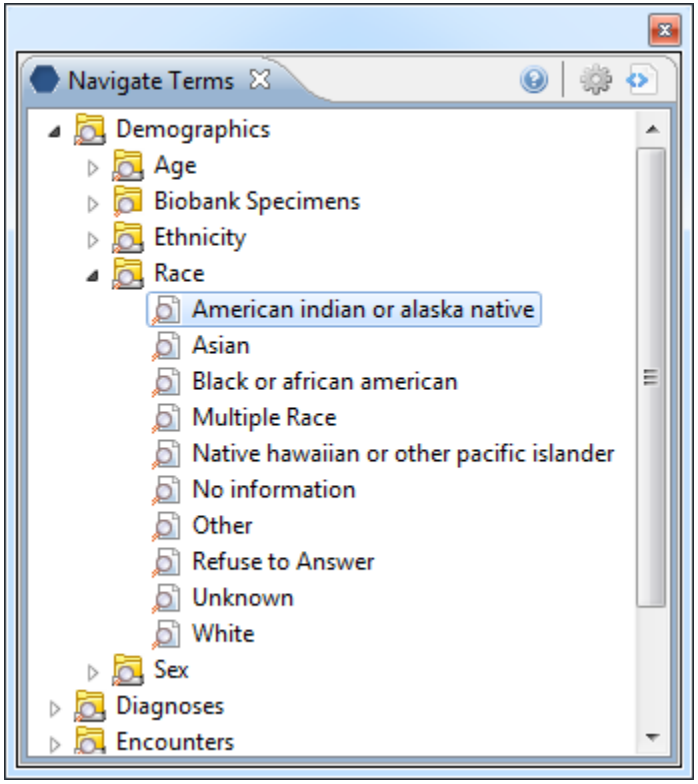



```
select patient_num
from
patient_dimension
where race_cd IN
('01','amer. Indian','i',
'na','nat. am')
```

c_name	c_fullname	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
American indian or alaska native	\PCORI\DEMOGRAPHIC\RACE\1\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'01','amer. Indian','i','na','nat. am'
Asian	\PCORI\DEMOGRAPHIC\RACE\2\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'02','a','asian','or','oriental','in','indian'
Black or african american	\PCORI\DEMOGRAPHIC\RACE\3\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'03','b','black','hib','his/black'
Native hawaiian or other pacific islander	\PCORI\DEMOGRAPHIC\RACE\4\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'04','native_hawaiian','pac. isl'
White	\PCORI\DEMOGRAPHIC\RACE\5\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'05','his/white','hiw','c','w','white','m','mid.eastern'
Multiple Race	\PCORI\DEMOGRAPHIC\RACE\6\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'06'
Refuse to Answer	\PCORI\DEMOGRAPHIC\RACE\7\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'07','r','refused'
No information	\PCORI\DEMOGRAPHIC\RACE\NI\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'NI','ni'
Other	\PCORI\DEMOGRAPHIC\RACE\OT\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'OT','o','d','@','deferred'
Unknown	\PCORI\DEMOGRAPHIC\RACE\UN\	PATIENT_NUM	PATIENT_DIMENSION	RACE_CD	IN	'UN','u','unk'

Edit c_dimcode to match your race_cds.

Patient_dimension types of queries:Ethnicity

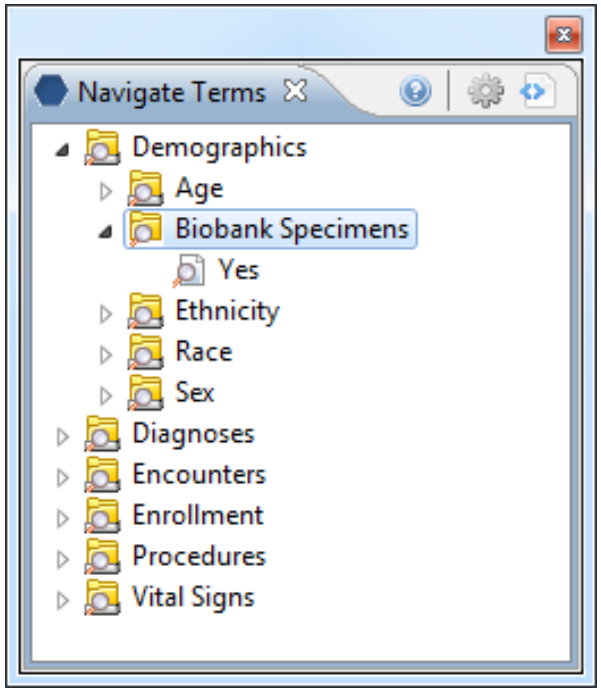



```
select patient_num
from
patient_dimension
where ethnicity_cd IN
('HISPANIC')
```

c_name	c_fullname	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Hispanic	\PCORI\DEMOGRAPHIC\HISPANIC\Hispanic\	patient_num	PATIENT_DIMENSION	ethnicity_cd	IN	'HISPANIC'
Non-Hispanic	\PCORI\DEMOGRAPHIC\HISPANIC\NonHispanic\	patient_num	PATIENT_DIMENSION	ethnicity_cd	IN	'NON-HISPANIC'

Edit c_dimcode to match your ethnicity_cds.

Concept_dimension types of queries:Biobank

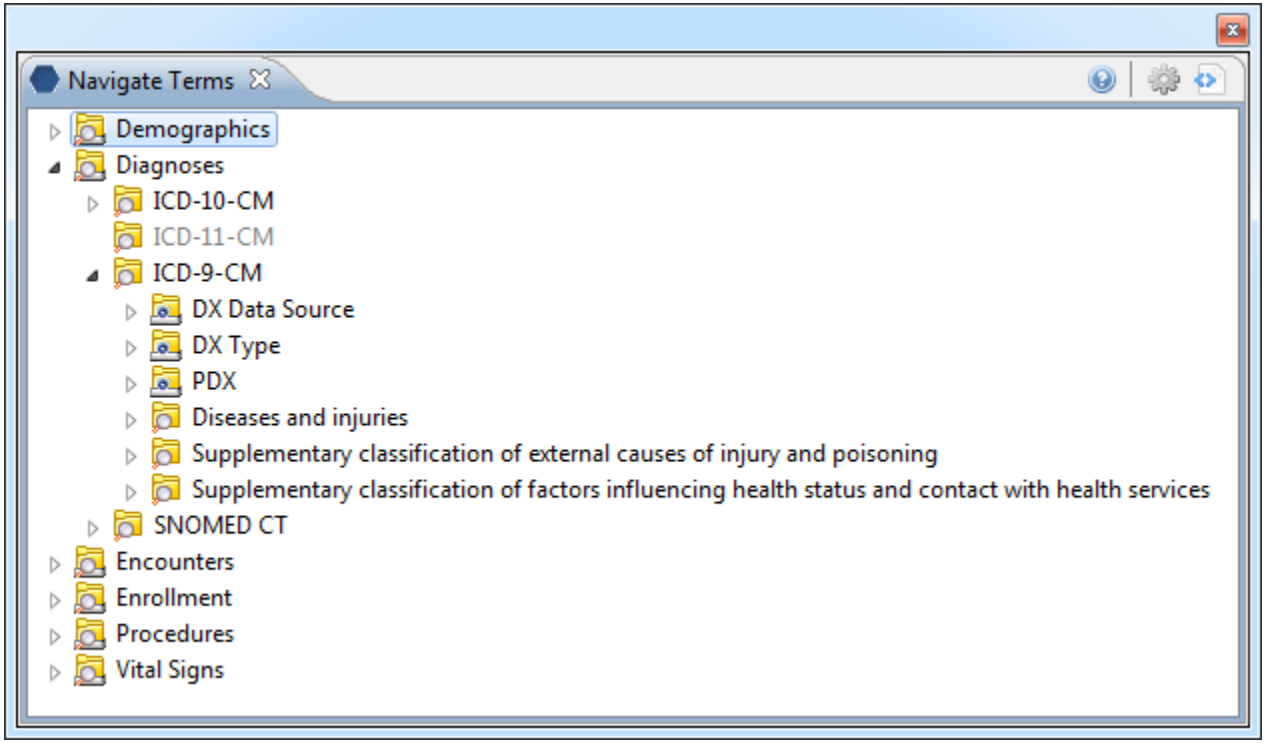



```
select concept_cd from
concept_dimension
where concept_path like
'\PCORI\DEMOGRAPHIC
\BIOBANK_FLAG\Y\%'
```

C_NAME	C_FULLNAME	C_BASECODE	C_FACTTABLECOLUMN	C_TABLENAME	C_COLUMNNAME	C_OPERATOR	C_DIMCODE
Biobank Specimens	\PCORI\DEMOGRAPHIC\BIOBANK_FLAG\	*(null)*	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\BIOBANK_FLAG\
Yes	\PCORI\DEMOGRAPHIC\BIOBANK_FLAG\Y\	BIOBANK_FLAG:Y	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\BIOBANK_FLAG\Y\

Edit c_basecode to match your code for biobank specimens.

Broader concept_dimension based queries



Diagnoses, Procedures typically require a merge of local terms within the tree.

ICD-9 Example

- If your institution uses ICD-9 …
  - Change the c_basecode to match your code format.

PCORI_BASECODE	C_BASECODE
ICD9:250.1	MY_ICD9:250.1

```
update pcornet
set c_basecode = replace(pcori_basecode, ‘ICD9:’, ‘MY_ICD9:’)
where c_basecode is not null
and pcori_basecode like 'ICD9:%'
and c_fullname like ‘\PCORI\DIAGNOSIS\09\%‘;
```

ICD-9 plain_code example

- If your institution uses ICD-9 but in a plain_code format…
  - Change the c_basecode to match your code format.

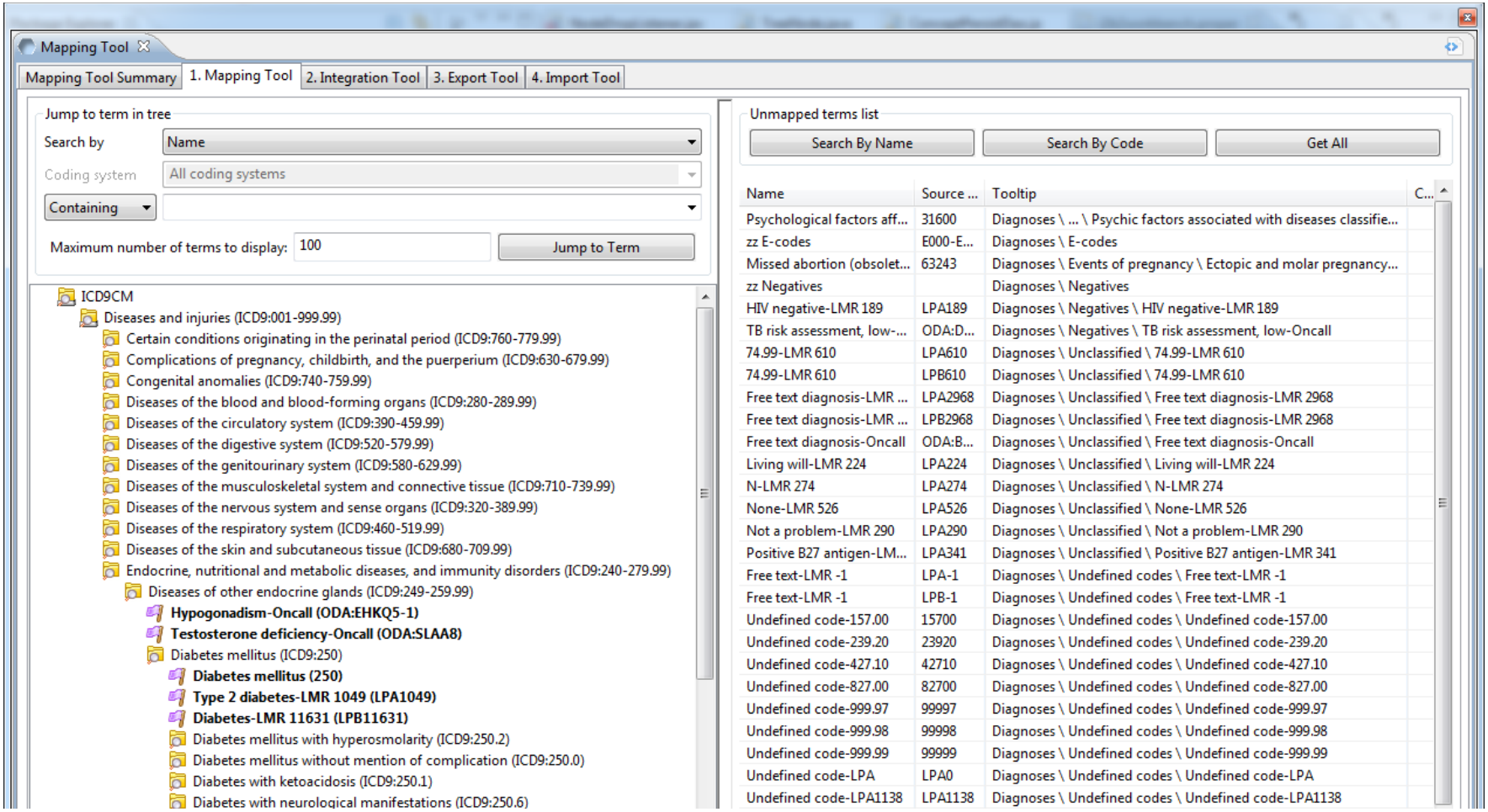
PCORI_BASECODE	C_BASECODE
ICD9:250.1	2501

```
update pcornet
set c_basecode = substring(pcori_basecode, 6, 25)
where pcori_basecode like 'ICD9:%'
and c_fullname like ‘\PCORI\DIAGNOSIS\09\%‘;

update pcornet
set c_basecode = replace(c_basecode, '.', '')
where pcori_basecode like 'ICD9:%'
and c_fullname like ‘\PCORI\DIAGNOSIS\09\%‘;
```

Local diagnoses codes (non-ICD9)

- Local codes need to be mapped/merged.
- Mapper cell / mapping tool plugins exist for workbench.

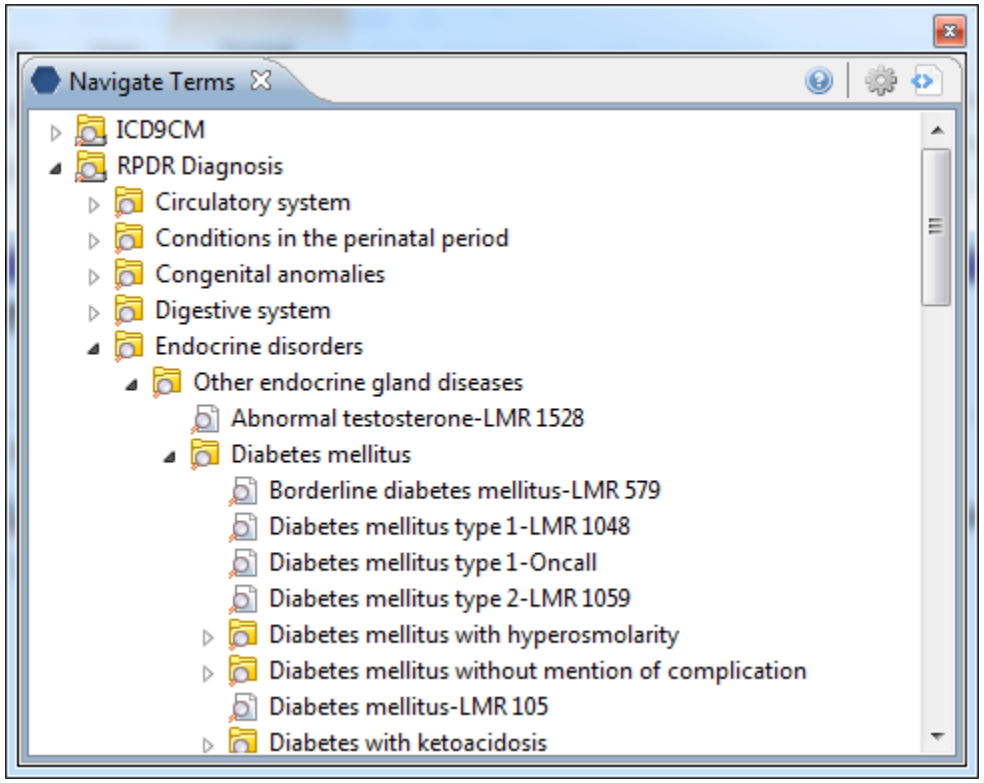


PROJECT_ONT_MAPPING	
SOURCE_CODING_SYSTEM	VARCHAR(50)
SOURCE_BASECODE	VARCHAR(50)
SOURCE_NAME	VARCHAR(2000)
SOURCE_FULLNAME	VARCHAR(700)
SOURCE_TOOLTIP	VARCHAR(700)
SOURCE_TABLE_CD	VARCHAR(25)
SOURCE_KEY	VARCHAR(725)
DESTINATION_CODING_SYSTEM	VARCHAR(50)
DESTINATION_BASECODE	VARCHAR(50)
DESTINATION_NAME	VARCHAR(2000)
DESTINATION_FULLNAME	VARCHAR(700)
DESTINATION_TABLE_CD	VARCHAR(25)
DESTINATION_KEY	VARCHAR(725)
MAPPING_SOURCE	VARCHAR(50)
VARFLAG	VARCHAR(25)
FLAG	INT
STATUS_CD	VARCHAR(25)
UPDATE_DATE	DATETIME
C_TOTALNUM	INT

Source fields are local terms.
Destination fields are PCORI terms.

Automated mappings

- If your local codes already exist within an ICD9 hierarchy..
- Find the c_fullname, c_basecode and c_name of parent of the term and use that as your mapping's destination.

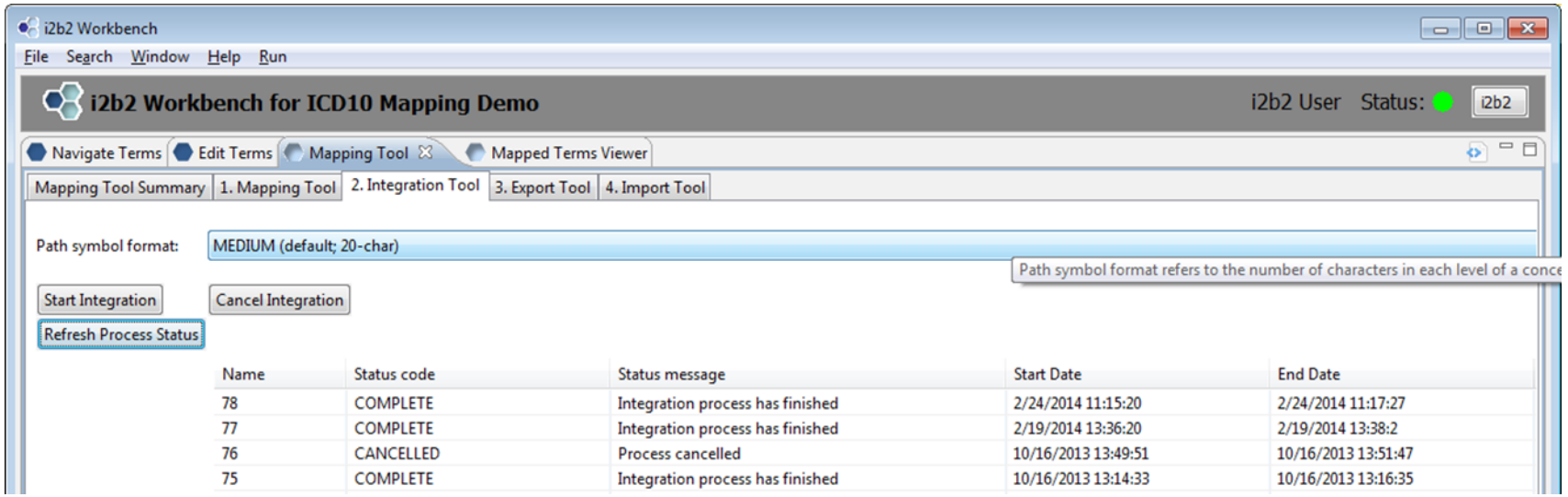


Automated mappings, cont.

```
WITH joined as(
select r2.c_basecode as sourceCode, r1.c_fullname as
destination, r1.c_basecode  as destCode, r1.c_name as
destName from local_ont r1 inner join local_ont r2 on
r1.c_fullname=r2.c_path where r2.c_basecode in (select
source_basecode from project_ont_mapping where
destination_fullname is null))

update project_ont_mapping
set destination_fullname = joined.destination,
    destination_basecode = joined.destCode,
    destination_name = joined.destName
from joined
where joined.sourceCode = source_basecode
```

Integration tool



Select path symbol format (S,M,L)

\i2b2\Diagnoses\Metabolic and immunity disorders (270-279)\

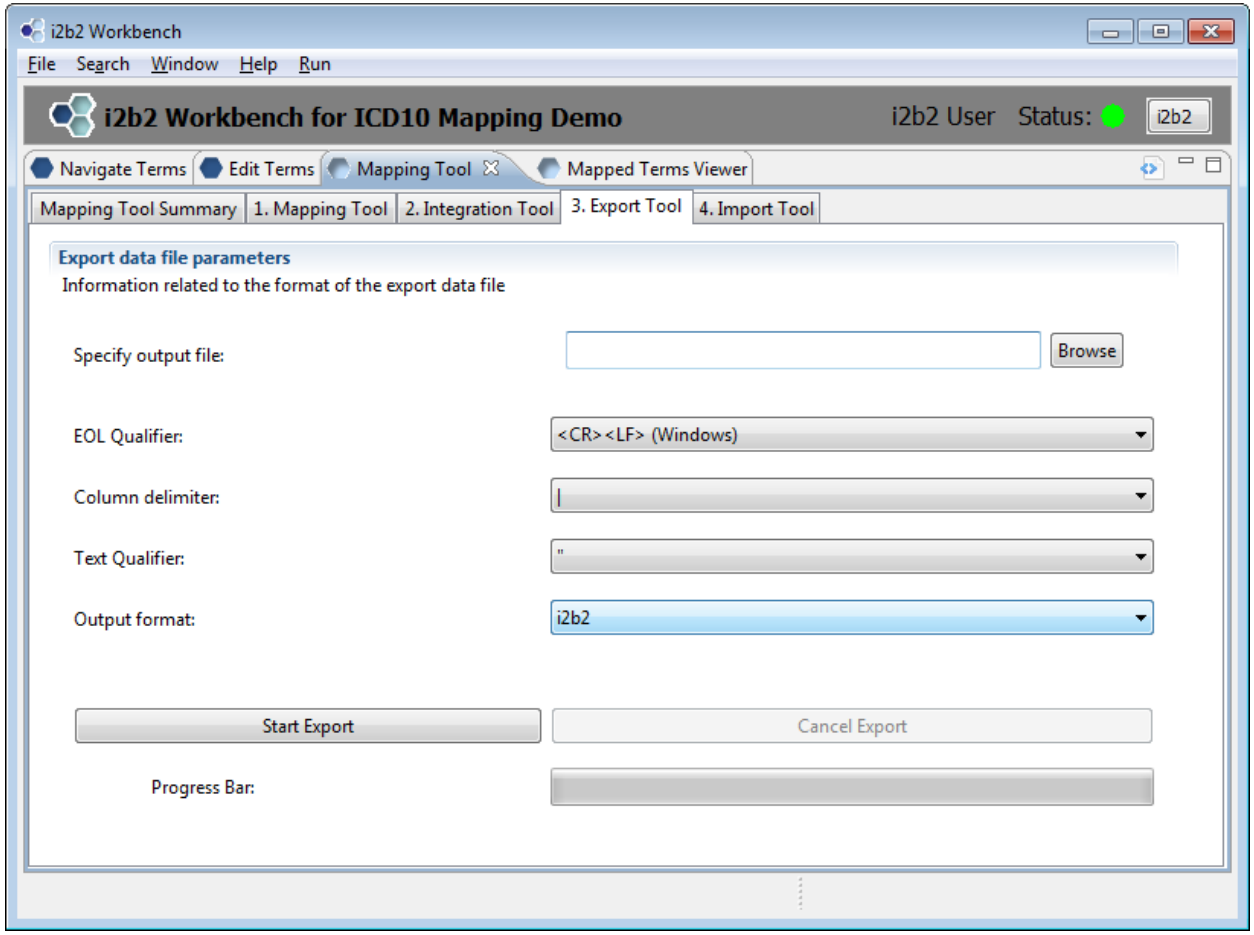
(277) Other and unspecified disor~\(277-0) Cystic fibrosis\

Start integration

Refresh process status until integration is complete.

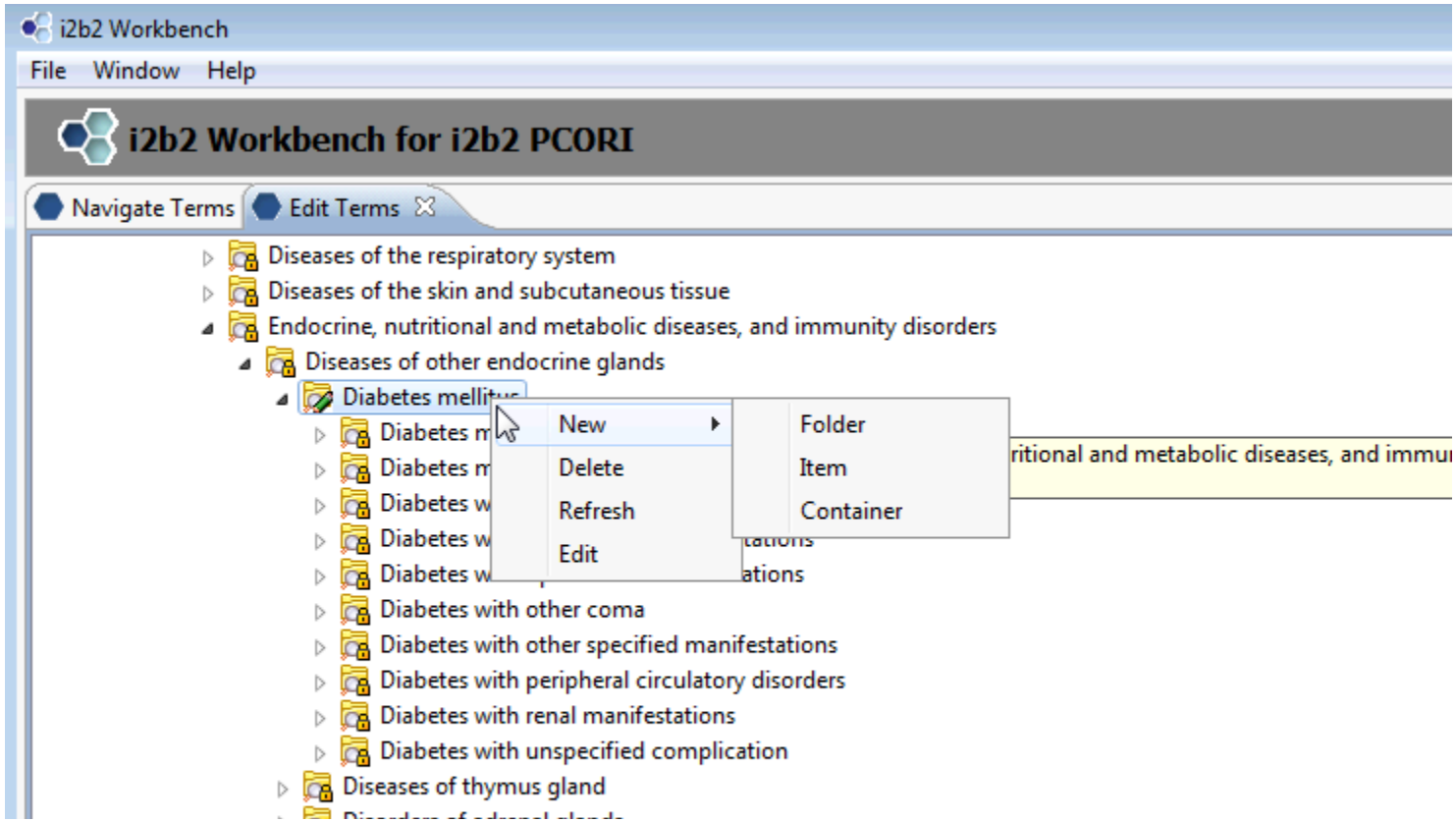
Export tool

- Exports result of integration process as a single, delimited i2b2 metadata file.

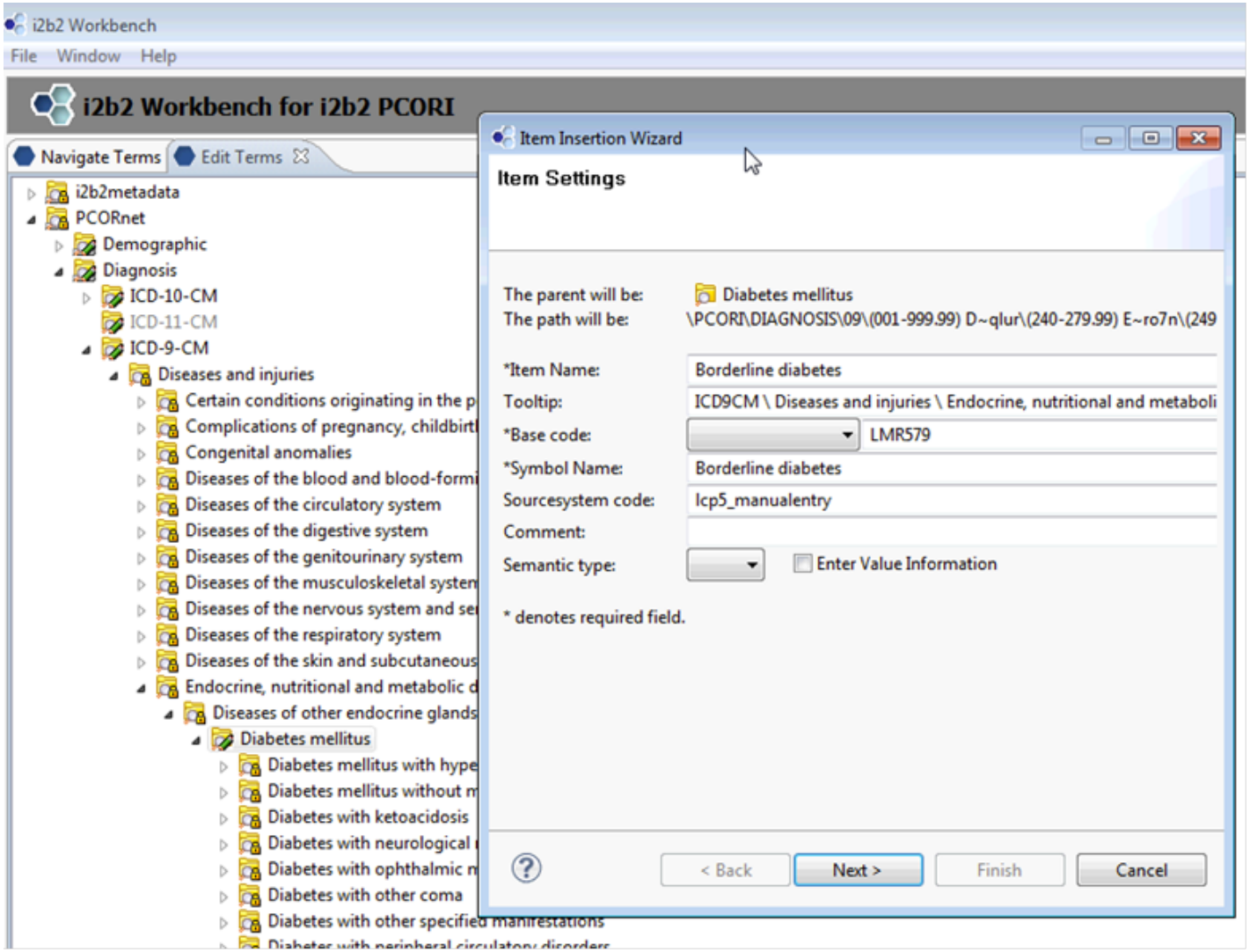


Alternate map/merge method

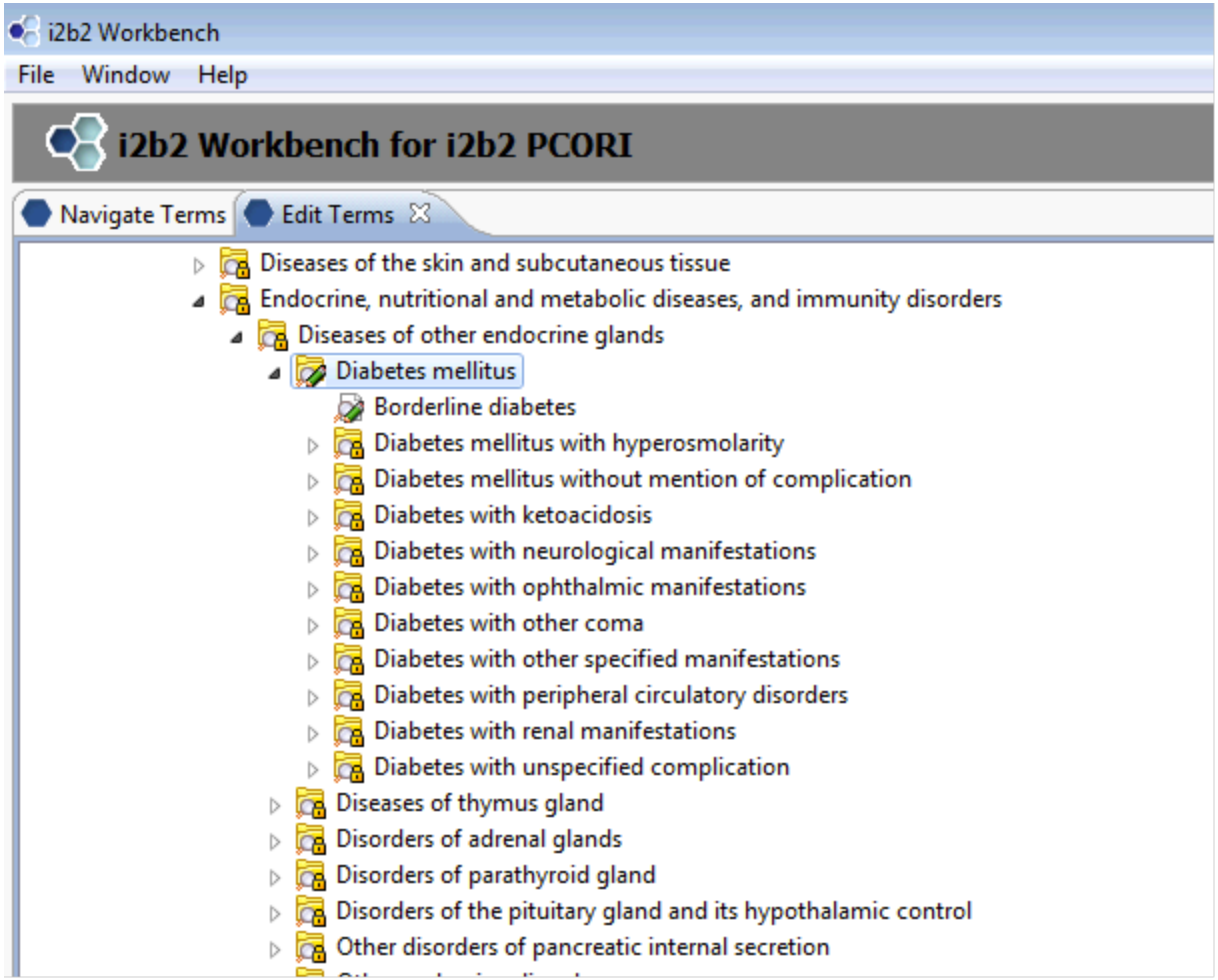
- If you have small number of local codes to merge, use the Edit Terms view.
  - User must be set up with EDITOR role



Edit Terms View



New term added



No need to export file, term is added to table automatically.

SHRINE Adapter Mapping File

<entry>
<key>\\SHRINE\SHRINE\Diagnoses\Endocrine, nutritional, and metabolic diseases and immunity disorders\Cystic fibrosis\</key>
<value>
<local_key>\\i2b2\i2b2\Diagnoses\Metabolic and immunity disorders (270-279)\(277) Other and unspecified disor~\(277-0) Cystic fibrosis\</local_key>
</entry>
<entry>
<entry>
<key>\\SHRINE\SHRINE\Demographics\Gender\Female\</key>
<value>
<local_key>\\i2b2\i2b2\Demographics\Gender\Female</local_key>
</value>
</entry>

Meaning behind <key>

- Concept key == \\c_table_cd\c_fullname\

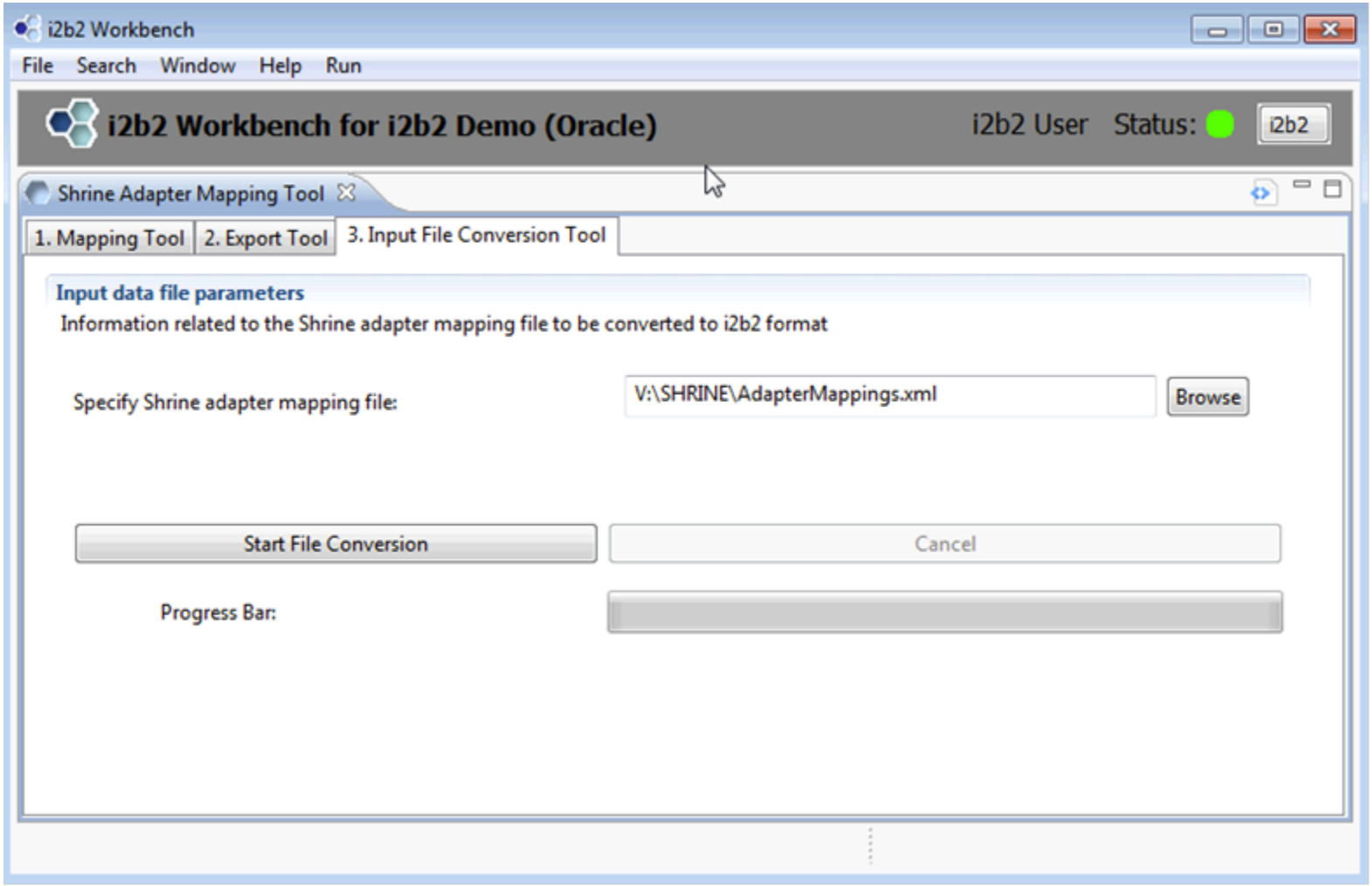
```
<concept>
    <level>1</level>
    <key>\\i2b2_DIAG\i2b2\Diagnoses\</key>
    <name>Diagnoses</name>
     ...
</concept>
```

TABLE_ACCESS

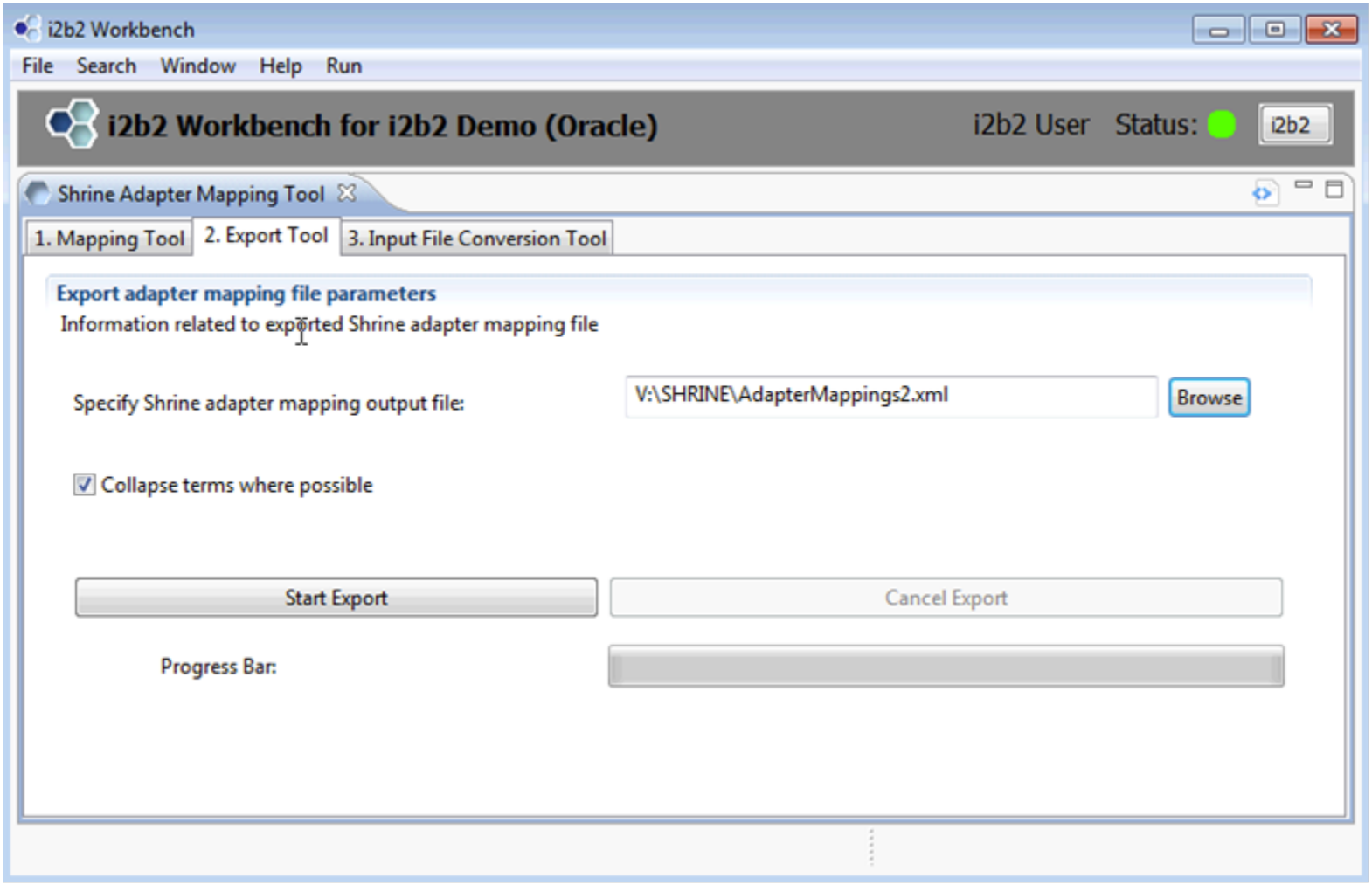
	C_TABLE_CD	C_TABLE_NAME	C_FULLNAME	C_NAME
1	BIRN	BIRN	\BIRN\	Clinical Trials
2	CUST	CUSTOM_META	\Custom Metadata\	Custom Metadata
3	i2b2_DEMO	I2B2_DEMO	\i2b2\Demographics\	Demographics
4	i2b2_DIAG	I2B2_DIAG	\i2b2\Diagnoses\	Diagnoses
5	i2b2_EXPR	I2B2_EXPR	\i2b2\Expression Profiles Data\	Expression Profiles Data
6	i2b2_LABS	I2B2_LABS	\i2b2\Labtests\	Laboratory Tests
7	i2b2_MEDS	I2B2_MEDS	\i2b2\Medications\	Medications
8	i2b2_PROC	I2B2_PROC	\i2b2\Procedures\	Procedures
9	i2b2_PROV	I2B2_PROV	\i2b2\Providers\	Providers
10	i2b2_REP	I2B2_REP	\i2b2\Reports\	Reports
11	i2b2_VISIT	I2B2_VISIT	\i2b2\Visit Details\	Visit Details
12	ICD10_ICD9	ICD10_ICD9	\Diagnoses\	Diagnoses (ICD10)

SHRINE Adapter Mapping Tool

- Version of Mapping Tool for editing Adapter Mapping File.



Export new file



Final output

entry>

<key>\\SHRINE\SHRINE\Diagnoses\Endocrine, nutritional, and metabolic diseases and immunity disorders\Cystic fibrosis\</key>

<value>

<local_key>\\i2b2\i2b2_DIAG\Diagnoses\Metabolic and immunity disorders (270-279)\(277) Other and unspecified disor~\(277-0) Cystic fibrosis\</local_key>

</entry>

<entry>

<entry>

<key>\\SHRINE\SHRINE\Demographics\Gender\Female\</key>

<value>

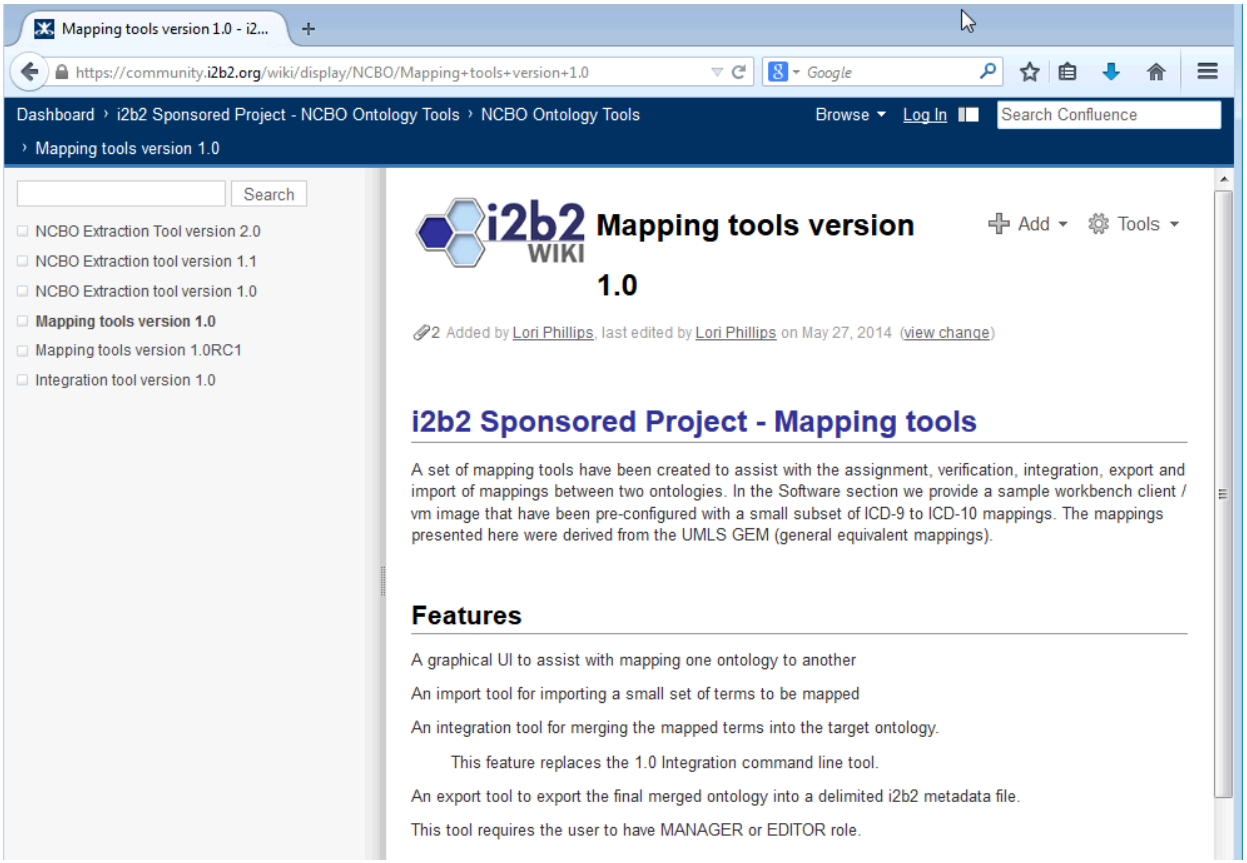
<local_key>\\i2b2\i2b2_DEMO\Demographics\Gender\ Female</local_key>

</value>

</entry>

Where can I find these tools?

- Edit terms view is distributed in standard release of i2b2.
- Mapping tools will be available soon on wiki:
  https://community.i2b2.org/wiki/display/NCBO/Mapping+tools+version+1.0



Fields used to construct queries

c_facttablecolumn	c_tablename	c_columnname	c_columndatatype	c_operator	c_dimcode
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender\Female
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender\Male
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender\Unknown
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender\Unknown\Unknown-@
concept_cd	concept_dimension	concept_path	T	LIKE	\RPDR\Demographics\Gender\Unknown\Unknown-U

Select * from observation_fact where c_facttablecolumn in

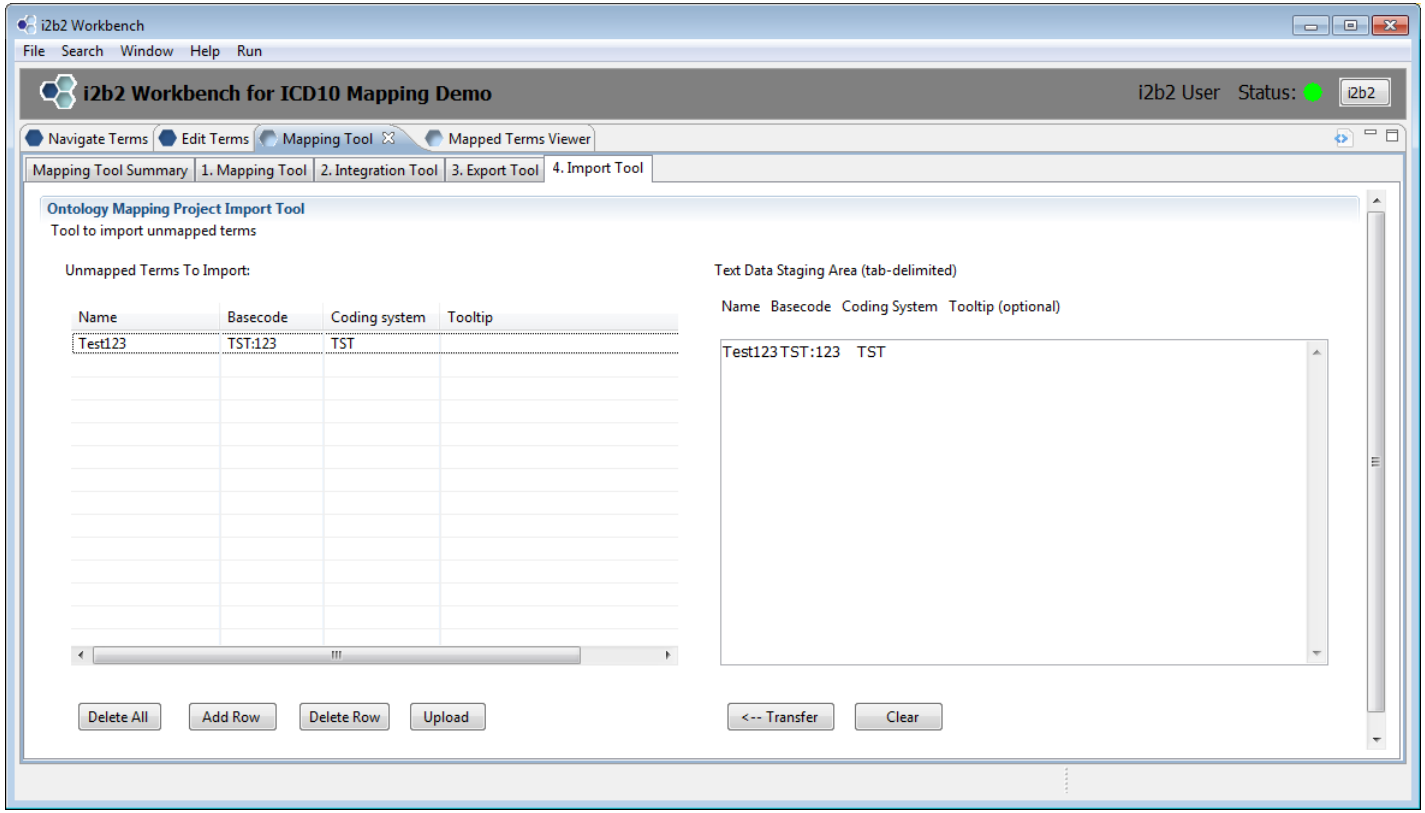
(select c_facttablecolumn from c_tablename where c_columname c_operator ‘c_dimcode %’)

Select * from observation_fact where concept_cd in

(select concept_cd from concept_dimension where concept_path like ’\RPDR\Demographics\Gender\Female\%’)

Import tool

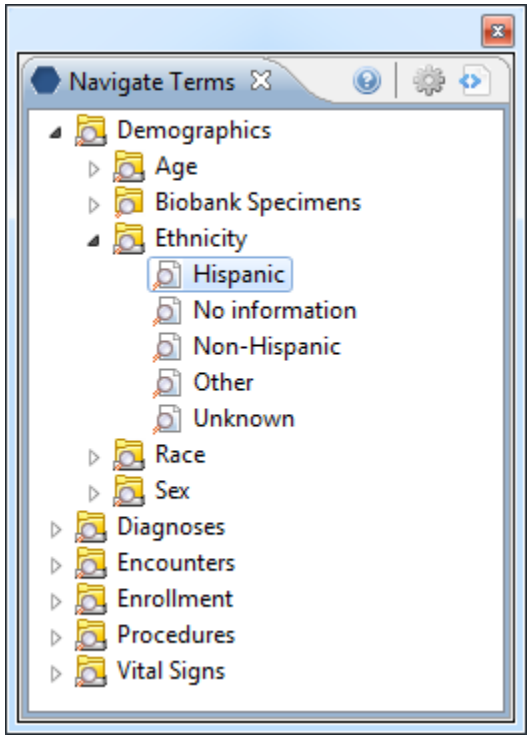
- Intended for small sets of local data
- Enter data into staging area or into table directly.
- Upload to mapping table when complete.
- Entries will appear in unmapped terms table.



patient_dim v concept_dim approaches

- HISTORY: early versions of i2b2 did not support the patient_dimension approach.
  - Demographic data was inserted as entry in observation_fact table and as a result appeared as a tick mark on the timeline.
- Patient_dimension approach does not require an entry in the observation_fact table for that concept.
- Patient_dimension approach will not result in a tick mark in the timeline for that concept.
- Patient_dimension approach is easier to maintain.

Ethnicity, two choices: concept_dimension query




```
select concept_cd from
concept_dimension
where concept_path like
'\PCORI\DEMOGRAPHIC
\HISPANIC\Y\%'
```

C_NAME	C_FULLNAME	C_BASECODE	C_FACTTABLECOLUMN	C_TABLENAME	C_COLUMNNAME	C_OPERATOR	C_DIMCODE
Hispanic	\PCORI\DEMOGRAPHIC\HISPANIC\Y\	ETHNICITY:HISPANIC	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\HISPANIC\Y\
No information	\PCORI\DEMOGRAPHIC\HISPANIC\NI\	ETHNICITY:NI	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\HISPANIC\NI\
Non-Hispanic	\PCORI\DEMOGRAPHIC\HISPANIC\N\	ETHNICITY:NOTHISPANIC	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\HISPANIC\N\
Other	\PCORI\DEMOGRAPHIC\HISPANIC\OT\	ETHNICITY:OT	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\HISPANIC\OT\
Unknown	\PCORI\DEMOGRAPHIC\HISPANIC\UN\	ETHNICITY:UN	concept_cd	CONCEPT_DIMENSION	concept_path	like	\PCORI\DEMOGRAPHIC\HISPANIC\UN\

Edit c_basecode to match your ethnicity codes.

Ethnicity, second choice: patient_dimension query

1. Add Ethnicity_cd column to patient_dimension table

2. Edit entry to a patient_dimension query; modify c_dimcode to match your ethnicity codes.

```
select patient_num from patient_dimension
where ethnicity_cd IN  ('HISPANIC')
```

```
(select [c_facttablecolumnname] from [c_tablename]
  where [c_columnname] [c_operator] [c_dimcode])
```