

AMIA Spring 2017 Poster



i2b2 ON OMOP CDM

¹Lori C. Phillips MS, ¹Christopher D. Herrick MBA, ^{1,2}Shawn N. Murphy MD PhD
¹Partners HealthCare, Somerville MA, ²Massachusetts General Hospital, Boston MA



Introduction

The legacy i2b2 data model is comprised of a central fact table (*observation_fact*) surrounded by multiple dimension tables (star schema). In the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)¹, rather than a central fact table, we have a collection of them distinguished by domain: procedures, condition, drug, measurement, observation, etc.

In this exercise, we investigate how to modify the i2b2 CRC cell to query against a collection of OMOP fact tables as dictated by domain.

Revisiting i2b2 query building

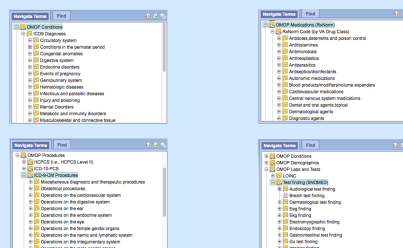
METADATA	
C_HLEVEL	INT NULL
C_FULNAME	VARCHAR(900) NULL
C_NAME	VARCHAR(2000) NULL
C_SYNONYM_CD	CHAR(1) NULL
C_VISUALATTRIBUTES	CHAR(3) NULL
TOTLNUM	INT NULL
C_ACCESSMODE	VARCHAR(450) NULL
C_METADATAXML	TEXT NULL
C_FACTTABLECOLUMN	VARCHAR(50) NULL
C_FACTNAME	VARCHAR(50) NULL
C_COLUMNNAME	VARCHAR(50) NULL
C_COLUMNNAME2	VARCHAR(50) NULL
C_OPERATOR	VARCHAR(10) NULL
C_DBCODE	VARCHAR(900) NULL
C_COMMENT	TEXT NULL
C_TGDTYP	VARCHAR(900) NULL
UPDATE DATE	DATE TIME NULL
DOWNLOAD DATE	DATE TIME NULL
IMPORT DATE	DATE TIME NULL
SOURCE SYSTEM_CD	VARCHAR(50) NULL
NUM_ROWS	INT NULL

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

c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Diabetes mellitus	concept cd	concept dimension	concept path	LRR	11292;Diabetes;Endocrine disorders (260-259);Other endocrine gland diseases

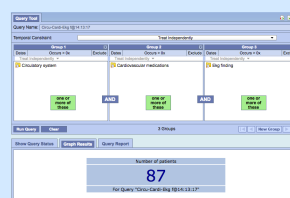
```
select patient_num from
observation_fact where concept_cd
IN
(select
concept_cd from concept_dimension where
concept_path like
'\i2b2\Diagnoses\Endocrine disorders (240-
259)\Other endocrine gland diseases (250-
259)\(250) Diabetes mellitus\%')
```

Step 3: 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.

Step 4: Build OMOP CDM and run queries



Query across 3 domains: condition, medication, measurement.
CDM populated with OMOP's Synthetic Public Use File data (synPUF)² for 1000 patients

Step 1: Create views for CDM tables

	OBSERVATION_FACT	CONDITION
PK	PATIENT_NUM	PERSON_ID
	ENCOUNTER_NUM	VISIT_OCCURRENCE_ID
	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_high	

```
CREATE VIEW dbo.[CONDITION_VIEW]
(patient_num,obsno,num_concept_cd,provider_id,
start_date,modifier_cd,instance_num,end_date,
valtype_cd,trial_char,num_val,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,
'0',
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
```

Step 2: Direct OMOP metadata to views

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

```
select patient_num from [domain_view] where
[c_facttablecolumn] IN
(select [c_facttablecolumn] from [c_tablename]
where [c_columnname] [c_operator]
[c_dimcode])
```

c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Diabetes mellitus	condition	view_concept_05	concept_dimension	concept_path	LTX \ 12b\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 '%\z2b2\Diagnoses\Endocrine
disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus%\')
```

Results

We were able to successfully query against OMOP CDM fact tables and produce meaningful query results.

Query types included

- Multi-panel, multi-domain queries;
- Date constrained queries;
- Occurs > x queries;
- Value constrained queries;
- Temporal queries.

Our ability to run value-constrained queries was limited by the SynPUF data set as it did not contain any value-based fact data. To test this, we created a small set of value-based measurement and observation lab data for 5 patients.

Results were verified by comparing to an i2b2 classic hive populated with the same data set.

Acknowledgements

This work was funded by NIH grant U01HG008685.

References:

1. Reich, C, Ryan, P. OMOP Common Data Model V5.0.1, 4-Apr-2016
http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:common_data_mo2_del
2. LTS Computing LLC: Sample 1000 person CMS SynPUF simulated data in CDMDV5 format: <http://www.ltscomputingllc.com/downloads/>

