Adopting the OMOP CDM to enable standardized analytics

Christian Reich, Erica Voss, Mui van Zandt
Adopting the OMOP Common Data Model (CDM) to Enable Standardized Analytics

Christian Reich
Erica A. Voss
Mui van Zandt
OMOP Common Data Model & OMOP Vocabularies
Traditional RWE Approach

"What treatment pathway is used in treatment of blood pressure in different healthcare settings?"

Analytical method: Treatment Pathway

Application to data

One study – one script model (often in SAS!)

- Reliant on partner capabilities
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert users
Solution: Systematic and Standardized Research

OHDSI Tools

OMOP CDM

Treatment Pathway
Mortality
Population-level estimation
Safety Signals

North America Southeast Asia
China
Europe UK Japan India
So Africa Switzerland Italy Israel

Standardized data
Analytics can be remote

North America Southeast Asia China
Europe UK Japan India
So Africa Switzerland Italy Israel
Analytics can run behind firewall
How do you create data that you query without having access?
The OHDSI approach: Strict Standardization

- **Data**
  - **Structure**: tables, fields, data types
  - **Content**: vocabulary to codify clinical domains
  - **Semantics**: conventions about meaning

- **Research**
  - **Cohort definition**: algorithms for identifying the set of patients who meet a collection of criteria for a given interval of time
  - **Covariate construction**: logic to define variables available for use in statistical analysis
  - **Analysis**: collection of decisions and procedures required to produce aggregate summary statistics from patient-level data
  - **Results reporting**: series of aggregate summary statistics presented in tabular and graphical form

OMOP CDM
OMOP Standardized Vocabularies
OMOP CDM – One model, multiple use cases

- Drug safety surveillance
- Vaccine safety surveillance
- Comparative effectiveness
  - Health economics
  - Quality of care
  - Clinical research

Standardized clinical data
- Person
  - Observation_period
  - Specimen
  - Death
  - Visit_occurrence
  - Procedure_occurrence
  - Drug_exposure
  - Device_exposure
  - Condition_occurrence
  - Measurement
  - Note
  - Observation
  - Fact_relationship

Standardized health system data
- Location
- Care_site
- Provider
- Payer_plan_period
- Visit_cost
- Procedure_cost
- Drug_cost
- Device_cost
- Condition_era
- Drug_era
- Dose_era
- Cohort
- Cohort_attribute
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- CDM_source
- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

Standardized derived elements
Standardized meta-data

Standardized health economics

Standardized health system data

OMOP CDM

– One model, multiple use cases

Drug safety surveillance
Vaccine safety surveillance
Comparative effectiveness
Health economics
Quality of care
Clinical research
<table>
<thead>
<tr>
<th>Vocabulary ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>All vocabularies stacked up in one table</td>
<td>Concepts related through semantic links, forming hierarchy</td>
</tr>
</tbody>
</table>

4,809,472 total
Vocabularies: Goals

✓ **Domains:** Every Standard Concept belongs to the right Domain

• **No duplicates:** For every entity exists one Standard Concept

• **Comprehensive:** For every Domain exists a complete finite set of Concepts covering all possible entities in this domain

• **Hierarchy:** All Concepts are connected through a comprehensive hierarchy

• **Mapping:** For every existing code in a vocabulary there is a map to a Standard Concept or a map to 0
Standardized Vocabularies: Conditions

Classifications

SNOMED-CT

Source codes

ICD-10-CM  Read  Oxmis  ICD-9-CM  SNOMED  ICD-10-GM  ....

Cohort

Standard vocabulary

SNOMED-CT

Higher-level classifications (Level 2 and up)

Low-level concepts (Level 1)

Top-level classification

System organ class (Level 5)

High-level group terms (Level 4)

High-level terms (Level 3)

Preferred terms (Level 2)

Low-level terms (Level 1)
Why a Hierarchy?
Diseases are not not “parallel”

- Rheumatic Disease
  - Spondyloarthropathy
    - Ankylosing Spondylitis
    - Reactive Spondylitis (Reiter’s Syndrome)
    - Spondylites associated with IBD
    - Isolated Acute Anterior Uveitis
    - Juvenile Idiopathic Arthritis
    - Undifferentiaed Spondylo-arthropaty
  - Bechterew’s Disease
Why a Drug Hierarchy?
Available data might have ingredients, brand names, doses...

5 ML Deferoxamine 100 MG/ML Injectable Solution [Desferal] Box of 4 by Novartis Pharma

- **Ingredient**: Deferoxamine
- **Drug Strength**: 100 mg/ml
- **Dose Form**: Injectable Solution
- **Brand Name**: Desferal
- **Qty Factor**: 5 ml
- **Box size**: Box of 4
- **Supplier**: Novartis Pharma
Granularity through Hierarchies

• Find (count, create cohort of) patients who ...
  
  – Had acute ST segment elevation myocardial infarction involving left anterior descending coronary artery
  – Had myocardial infarction
  – Had cardiovascular disease
  – Have family history of malignant disease
  
  – Took diphtheria and tetanus toxoids, adsorbed for pediatric use
  – Took diphtheria toxoid vaccine, inactivated
  – Took toxoid vaccine
  – Took vaccine
  – Took drug by FDA-approved indication
  – Took off-label drug
  
  – Underwent hysterectomy
  – Underwent abdominal surgery
  
  – Had HbA1C > 7 consistently in last 6 months
  – Smoked longer than 10 packyears
  – Received IUD
  – Has follow-up of 2 years after event
  – Had treatment in excess of $100k per year
<table>
<thead>
<tr>
<th>Product</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotine Chewing Gum [Nicorette]</td>
<td>France, Germany, UK</td>
</tr>
<tr>
<td>Nicotine Oral Lozenge [Nicorette]</td>
<td>France, Germany, UK</td>
</tr>
<tr>
<td>Nicotine Cartridge [Nicorette]</td>
<td>Germany</td>
</tr>
<tr>
<td>Nicotine Dry Powder Inhaler [Nicorette]</td>
<td>France</td>
</tr>
<tr>
<td>Nicotine Inhalant Solution [Nicorette]</td>
<td>UK</td>
</tr>
<tr>
<td>Nicotine Metered Dose Inhaler [Nicorette]</td>
<td>Germany</td>
</tr>
<tr>
<td>Nicotine Nasal Spray [Nicorette]</td>
<td>Germany, UK</td>
</tr>
<tr>
<td>Nicotine Oral Solution [Nicorette]</td>
<td>Germany</td>
</tr>
<tr>
<td>Nicotine Oral Spray [Nicorette]</td>
<td>France, Germany</td>
</tr>
<tr>
<td>Nicotine Sublingual Tablet [Nicorette]</td>
<td>France, Germany, UK</td>
</tr>
<tr>
<td>Nicotine Topical Solution [Nicorette]</td>
<td>Germany</td>
</tr>
<tr>
<td>Nicotine Transdermal System [Nicorette]</td>
<td>France, Germany, UK</td>
</tr>
</tbody>
</table>
Solution: RxNorm Extension

- Make RxNorm for the World
- Build up with relationships to existing RxNorm drugs and attributes
- Same logic, structure and attributes
Drug Markets overlap only partially

- RxNorm: 201,672 concepts
- AMIS: 65,258 concepts
- BDPM: 19,428 concepts
- DPD: 46,771 concepts
- dm+d: 150,188 concepts
“10 ML Nicotine 1 MG/ML Nasal Spray [Nicorette] Box of 5 by Johnson & Johnson”
Resources to get help

http://forums.ohdsi.org/

https://github.com/ohdsi
Teams to help you

Vocabulary Team

Alina  Polina  Anna  Oleg  Eduard  Vlad  Alex  Denis  Polina  Eldar  Dima

CDM WG

Clair
Extract, Transform, & Load (ETL) & OHDSI ETL Tools
ETL Process

Data experts and CDM experts together design the ETL.

People with medical knowledge create the code mappings.

All are involved in quality control.

A technical person implements the ETL.

Documentation

OHDSI Tools

White Rabbit
Rabbit In a Hat
Usagi
White Rabbit
ACHILLES
Designing an ETL

**White Rabbit**
- performs a scan of the source data, providing detailed information on the tables, fields, and values that appear in a field

**Rabbit In a Hat**
- Uses White Rabbit scan to provide a graphical user interface to help build an ETL document
- Does not generate code
<table>
<thead>
<tr>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>AA</th>
<th>AB</th>
<th>AC</th>
<th>AD</th>
<th>AF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>610</td>
<td>white</td>
<td></td>
<td>Frequency: 835</td>
<td>irish</td>
<td></td>
<td>M</td>
<td>562</td>
</tr>
<tr>
<td>M</td>
<td>344</td>
<td>hispanic</td>
<td></td>
<td>Frequency: 112</td>
<td>italien</td>
<td></td>
<td>F</td>
<td>557</td>
</tr>
<tr>
<td>S</td>
<td>156</td>
<td>black</td>
<td></td>
<td>Frequency: 82</td>
<td>english</td>
<td></td>
<td>U</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>asian</td>
<td></td>
<td>Frequency: 70</td>
<td>puerto_rican</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>native</td>
<td>List trunc</td>
<td>Frequency: 20</td>
<td>french</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Scan your Data
Rabbit in a Hat

Map Source Tables to CDM Tables

- Source tables:
  - allergies.csv
  - conditions.csv
  - encounters.csv
  - immunizations.csv
  - medications.csv
  - observations.csv
  - patients.csv
  - procedures.csv

- CDMV5.3.1 tables:
  - condition_occurrence
  - drug_exposure
  - measurement
  - observation
  - observation_period
  - person
  - procedure_occurrence
  - visit_occurrence
Rabbit in a Hat

Source

- id
- birthdate
- race
- ethnicity
- gender

CDMV5.3.1

- *gender_concept_id
- *year_of_birth
- month_of_birth
- day_of_birth
- birth_datetime
- *race_concept_id
- *ethnicity_concept_id
- person_source_value
- value

Table to Table Mapping Details
### Rabbit in a Hat

#### Table to Table Mapping Details

<table>
<thead>
<tr>
<th>Source Field</th>
<th>Destination Field</th>
<th>Logic</th>
<th>Comment field</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>gender_concept_id</td>
<td>When gender = ‘M’ then set gender_concept_id to 8507, when gender = ‘F’ then set to 8532</td>
<td>Drop any rows with missing/unknown gender.</td>
</tr>
<tr>
<td>birthdate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- *gender_concept_id*
**Usagi**

Map Source Codes to Standard Concepts

<table>
<thead>
<tr>
<th>Source code</th>
<th>Source term</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB7</td>
<td>No Illness</td>
<td>500000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target concepts</th>
<th>Source code</th>
<th>Source term</th>
<th>Frequency</th>
<th>Concept ID</th>
<th>Concept name</th>
<th>Domain</th>
<th>Concept class</th>
<th>SNOMED</th>
<th>Concept code</th>
<th>Standard concept</th>
<th>Parents</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>4192174</td>
<td>Illness</td>
<td>Condition</td>
<td>Clinical Finding</td>
<td>SNOMED</td>
<td>39104002</td>
<td>S</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Search**

Use source term as query

Query:

Results

<table>
<thead>
<tr>
<th>Score</th>
<th>Term</th>
<th>Concept ID</th>
<th>Concept name</th>
<th>Domain</th>
<th>Concept class</th>
<th>SNOMED</th>
<th>Concept code</th>
<th>Standard concept</th>
<th>Parents</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.82</td>
<td>Illness</td>
<td>4192174</td>
<td>Illness</td>
<td>Condition</td>
<td>Clinical Finding</td>
<td>SNOMED</td>
<td>39104002</td>
<td>S</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0.80</td>
<td>Mental Illness</td>
<td>4214703</td>
<td>Mental Illness</td>
<td>Observation</td>
<td>Qualifier Value</td>
<td>SNOMED</td>
<td>39815005</td>
<td>S</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.78</td>
<td>Viral Illness</td>
<td>4040928</td>
<td>Viral illness</td>
<td>Observation</td>
<td>Qualifier Value</td>
<td>SNOMED</td>
<td>34014006</td>
<td>S</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>0.77</td>
<td>Mass Illness</td>
<td>45089695</td>
<td>Mass factor value</td>
<td>Value</td>
<td>Assure</td>
<td>LONG</td>
<td>LA1095-0</td>
<td>S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.75</td>
<td>Stillness</td>
<td>40922058</td>
<td>Stillness</td>
<td>Observation</td>
<td></td>
<td>SNOMED</td>
<td>247802008</td>
<td>S</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Rabbit in a Hat

Helps Build Test Cases
Review your Data with High Level Reporting
Tools Demo

• Visit Clair during the Collaborator Showcase

Rabbit in a Hat

Usagi
Purpose of THEMIS

Developing rules and regulations to combat the inconsistent representations of the same data sources across Observational Health Data Sciences and Informatics’ (OHDSI) and inability to gather reliable and scalable evidence.
THEMIS Roadmap

Q1-Q2 2018
Developed 1st set of conventions and documentation

Q1-Q2 2019
- Develop QA/QC to incorporate THEMIS conventions including revisions to Achilles, WhiteRabbit etc..
- Kickoff Certification Process
- Invite 2-3 ETL partners & companies for pilot
- Release certification process

Q3-Q4 2018
Release 1.0 Announcement at OHDSI Symposium

Q3-Q4 2019
- Announce pilot success story and value proposition of consistent/single ETL process for CDM
- Role out at OHDSI Symposia (Europe, China, US)

Q1-Q2 2019
- Drive adoption
- Dedicated web page under ohdsi.org
- Develop tutorial in CDM advanced model
- Submit abstracts to relevant conventions

Q1-Q4 2020
Release 2.0
March 2018 Hackathon

**THEMIS ISSUES**

- **Not Yet Started, 40**
- **Discussed, 3**
- **Accepted, 46**

### Duplicate Drugs on Same Day

**What**

What should someone do if duplicate drugs are reported on the same day? Most likely an idea that can be expanded to procedures.

**Notes**

http://forums.ohdsi.org/t/duplicate-drugs-themis-wg3/4101

**RECOMMENDATION**

If a patient has multiple records on the same day for the same drug or procedures the ETI should not dedupe them unless there is probable reason to believe the item is a true data duplicate.

**Action**

Work with Clair to have posted on the DRUG_EXPOSURE & PROCEDURE_OCCURRENCE page under the CDM Wiki. [https://github.com/OHDSI/CommonDataModel/wiki](https://github.com/OHDSI/CommonDataModel/wiki)

### Missing Visit End Dates/Pharmacy considered visits

<table>
<thead>
<tr>
<th>THEMIS Topic</th>
<th>Resolution</th>
</tr>
</thead>
</table>
| Missing Visit End Dates       | Visit end dates should be required. Use the best information to infer a visit end date. Examples include:  
|                               | • A claims billing end date instead of a discharge date  
|                               | • An encounter end date instead of a visit end date                        |
| Prescription events have visits | concept_ids have been proposed by Gowtham  
|                               | [http://forums.ohdsi.org/Pharmacy-Claims-Drug-exposure-how-to-identify-the-dispensing-billing-pharmacy/379513](http://forums.ohdsi.org/Pharmacy-Claims-Drug-exposure-how-to-identify-the-dispensing-billing-pharmacy/379513) and will be allowed in the CDM. |
Release V1.0.0 Highlights:
- Establishing source concepts for local variables
- Handling overlapping Payer Plan periods
- Establishing conventions for:
  - Multiple death codes
  - Person inclusion
  - Masking items related to a person

Next Release Activities:
- Cleaning up Place_of_Service
- Handling provider specialty, clinical titles, multiple addresses, multiple providers and 1-to-many NPIs
- Mother to baby and father to baby linkages
- Achilles Heel error for spotting non-sensical units
Purpose: To create a certification program, based off the rules and regulations THEMIS has decided on

Who Should Attend: SQL Developers, JAVA Developers, Software Developers

Day 1:
• What are the gaps between THEMIS rules and Achilles
• Create a program to check for these rules
  – Adjust Achilles to take these rules into consideration
  – Take code of Achilles and create new program
  – Start from Scratch

Day 2:
• Code it
• Mostly developers at work
• Analysts/ non technical team work to organize
• Whatever work does not get done strategize the next steps to keep the ball rolling

Where/When: May 14th & 15th at the University of Colorado (Aurora)
Join Us

To learn more about any of these topics, come see us at the tutorials:

• Saturday – OMOP Vocabulary/CDM
• Sunday – OMOP CDM ETL
Q&A Session
All speakers
Moderator: Patrick Ryan
13:00 OHDSI Collaborator Showcase: Demos and Posters
14:00 Lightning Talks