OMOP-2-OPMI: Ontologization of OMOP CDM using OPMI to support clinical data interoperability and analysis

Long Nguyen Minh Tran, Yongqun “Oliver” He

University of Michigan
Ann Arbor, MI 48109 USA
OMOP CDM

- OMOP: Observational Medical Outcomes Partnership
  - Developed by OHDSI
  - >1 billion patient records
- OMOP Common Data Model (CDM): allows for integration of different observational databases
  - Current version: CDM v5.4
- Used by National COVID Cohort Collaborative (N3C, https://ncats.nih.gov/n3c) and many other scenarios

https://ohdsi.github.io/CommonDataModel/
http://ohdsi.github.io/CommonDataModel/cdm54.html
Bottlenecks of OMOP CDM

- **Weak semantics**: OMOP CDM provides database schema, which is powerful but lacks robust semantic relations among terms.
- **Poor interoperability** among CDMs
  - Other clinical CDMs exist: pcornet, cdisc, etc.
  - Difficult to support data standardization among databases with different CDMs.

- **Ontology can be a solution** to solve these bottlenecks
  - Open Biomedical Ontology (OBO) such as OPMI is our solution.
Our strategy: OMOP-2-OPMI

• OMOP2OBO is good but not enough

  https://github.com/callahantiff/OMOP2OBO

  ○ Maps to 8 OBO ontologies, e.g., Human Phenotype Ont. (HP), and MONDO disease Ont., etc.
  ○ Does not cover higher level CDM structure of OMOP
  ○ Does not cover contents of >10 clinical data tables and their relations.
  ○ Many CDM terms are not yet available in OBO ontologies → so new development is needed.

• OPMI:

  ○ Ontology of Precision Medicine and Investigation
  ○ An OBO library ontology
  ○ Used for KPMP (Kidney Precision Medicine Project)

  Goal: use OPMI to map and analyze OMOP CDM.

http://ohdsi.github.io/CommonDataModel/cdm60.htm
OMOP-2-OPMI Mapping Strategy & Workflow

- **Source:** OMOP CDM v5.4:
  - [https://ohdsi.github.io/CommonDataModel/cdm54.html](https://ohdsi.github.io/CommonDataModel/cdm54.html)

- **Mapping strategy:**
  - Use Ontobee to find OMOP CDM terms from OBO ontologies
  - If existed in OBO ontologies, import using Ontofox
  - If new, create new OPMI terms and annotation.
OMOP-2-OPMI: Mapping OMOP CDM to OPMI

- OMOP-2-OPMI ontologizes all terms of OMOP CDM.
- Available on GitHub: https://raw.githubusercontent.com/OPMI/opmi/master/src/ontology/omop2opmi.owl

- Aligned with BFO upper level ontology
Simplified high level OMOP-2-OPMI ontology design pattern (ODP)

- ‘Person’ usually refers to Patients in OMOP
  - centric to OMOP CDM
  - mapped to NCBITaxon:human
- ‘Person’ participates in:
  - 5 medical occurrences
  - Observation process
- ‘Person’ can be a target of:
  - Measurement
- ‘Specimen’ is mapped to OBI:specimen

Figure 2
CDM terms from OMOP tables mapped to OPMI

- **Statistics:**
  - **165 terms** from **15 OMOP tables** mapped
  - **46 newly generated terms**

- **Current mapping:**
  - Clinical data tables
  - Health system data tables

- **Not yet included:**
  - Non-clinically relevant tables (e.g., Standardized Metadata, Standardized Vocabularies, Standardized Derived Elements, etc.)

### Table 1. terms from 10 representative OMOP tables

<table>
<thead>
<tr>
<th>Selected OMOP Tables</th>
<th>Mapped OMOP terms</th>
<th>Mapped Ontology Term Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>12/19* person ID (OPMI_0000470), gender (PATO_0001894), year of birth (OPMI_0000473), race (NCIT_C17049)</td>
<td></td>
</tr>
<tr>
<td>PROVIDER</td>
<td>9/13 case provider (OPMI_0000163), National Provider Identifier (OPMI_0000503), DEA identifier (OPMI_0000504)</td>
<td></td>
</tr>
<tr>
<td>SPECIMEN</td>
<td>6/15 specimen ID (OLI_0001616), date of specimen collection (OBIB_0000714), anatomical structure (UBERCN_0000061)</td>
<td></td>
</tr>
<tr>
<td>VISIT OCCURRENCE</td>
<td>26/17 visit occurrence (OPMI_0000482), visit start date (OPMI_0000487), preceding visit occurrence (OPMI_0000492)</td>
<td></td>
</tr>
<tr>
<td>PROCEDURE OCCURRENCE</td>
<td>13/16 procedure (NCIT_C16218), procedure start date (OPMI_0000508), procedure end date (OPMI_0000510)</td>
<td></td>
</tr>
<tr>
<td>DRUG EXPOSURE</td>
<td>18/23 drug exposure (OPMI_0000572), drug product (DRON_00000005) drug exposure start time (OPMI_0000565)</td>
<td></td>
</tr>
<tr>
<td>CONDITION OCCURRENCE</td>
<td>38/16 condition occurrence (OPMI_0000527), medical condition status (OPMI_0000583), admission diagnosis status (OPMI_0000542)</td>
<td></td>
</tr>
<tr>
<td>DEVICE EXPOSURE</td>
<td>7/15 device exposure (OPMI_0000554), device (OLI_0000948), device exposure start date (OPMI_0000562)</td>
<td></td>
</tr>
<tr>
<td>MEASUREMENT</td>
<td>11/20 clinical measurement identifier (OPMI_0000582), measurement time (OPMI_0000579), measurement unit label (TAC_0000003)</td>
<td></td>
</tr>
<tr>
<td>OBSERVATION PERIOD</td>
<td>5/6 observation period start date (OPMI_0000577), observation period end date (OPMI_0000578),</td>
<td></td>
</tr>
</tbody>
</table>
OMOP CDM term mapping by element types

- Many terms not yet covered
- OMOP element types mostly missing:
  - source concept id
  - source value
- Examples of mapping or no mapping:
  - measurement_concept_id mapped to ‘clinical measurement’ (CMO_0000000)
  - measurement_source_concept_id (no mapping)
  - measurement_source_value (no mapping)

<table>
<thead>
<tr>
<th>types</th>
<th>OMOP terms</th>
<th>OMOP mapped</th>
<th>percent mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>_type_concept_id</td>
<td>11</td>
<td>11</td>
<td>100.00%</td>
</tr>
<tr>
<td>name_id</td>
<td>34</td>
<td>30</td>
<td>88.24%</td>
</tr>
<tr>
<td>_date</td>
<td>34</td>
<td>27</td>
<td>79.41%</td>
</tr>
<tr>
<td>_concept_id</td>
<td>30</td>
<td>18</td>
<td>60.00%</td>
</tr>
<tr>
<td>_concept_name</td>
<td>30</td>
<td>16</td>
<td>53.33%</td>
</tr>
<tr>
<td>source_concept_id</td>
<td>17</td>
<td>1</td>
<td>5.88%</td>
</tr>
<tr>
<td>_source_value</td>
<td>34</td>
<td>1</td>
<td>2.94%</td>
</tr>
<tr>
<td>Total</td>
<td>179</td>
<td>93</td>
<td>51.96%</td>
</tr>
</tbody>
</table>

Table 2
In OMOP, many terms denote events occurring *over a period of time*.

These defined as BFO: **process** terms
- medical condition occurrence
- medical exposure
- device exposure
- drug exposure
- medical procedure occurrence
- medical visit**
- (note: red labels OMOP tables)

All are subclasses of **OPMI: medical occurrence**.
Temporal elements in 6 OMOP tables: Person, Visit, Device/Drug Exposure, Procedure/Condition Occurrence

OPMI representation:
- Direct mapping:
  - visit_start_date → ‘visit start date’ (OPMI)
  - Separate date and time
    - date: which day
    - datetime: day, hours, minutes, …
- BFO: ‘zero (or one)-dimensional temporal region’

Figure 5
OMOP-2-OPMI Use Case 1:

Ontology-level data standardization

- Rooted in the nature of ontology:
  - Standard **representation & definitions** of mapped OMOP CDM terms
  - Computer-understandable logic **axioms** among mapped terms
- Furthermore, it’s **interoperable** ontology:
  - Also used on other ontologies, e.g., Coronavirus Infectious Disease Ontology
  - Support knowledge/data sharing and integration
- Future: extend to represent other CDMs (e.g., **PCORnet** and **CDISC**) → further support data integration
Use Case 2: Adverse event (AE) modeling & analysis

OMOP CDM does not have ‘adverse event’ (AE) concept. However, OMOP-2-OPMI can be semantically extended to model AE.

Ontology of Adverse Event (OAE) AE modeling:

OMOP CDM does not have ‘adverse event’ (AE) concept. However, OMOP-2-OPMI can be semantically extended to model AE.
Example usage of OMOP-2-OPMI modeling:

Modeling of adverse event (AE) using OMOP CDM/data

OMOP-2-OPMI modeling of adverse event (AE):

OMOP CDM:
- Do not represent AE directly.

OMOP-2-OPMI modeling:
- Use OAE-based AE modeling (see earlier slide)
- Transfer OMOP CDM data for AE representation, e.g.,
  - Procedure is a ‘medical intervention’
  - A new condition afterwards is an AE

Citation: He Y, Ong E, Zheng J, Wan L, Schaub J, Kretzler M. Ontological representation of OMOP CDM using the OBO framework. 2018 OHDSI Symposium, Oct 12, 2018, Bethesda, MD, USA.
"Heart surgery AKI AE" using OMOP data (from IQVIA)

(Q: How to detect & find patterns of patients with Acute Kidney Injury (AKI) AE after heart surgery?)

Algorithm: Define “heart surgery AKI AE” based on OMOP-2-OPMI modeling (earlier slide)

30 days before surgery (no AKI) → Heart surgery on a patient within 14 days after surgery (AKI appearance)

Symptoms (Conditions) found at 30 days before heart surgery:

Phenotypic abnormality
- Abnormalities of metabolism/homeostasis
  - Diabetes mellitus
    - Type II diabetes mellitus
  - Hyperlipidemia
- Abnormality of the cardiovascular system
  - Abnormal cardiovascular system physiology
    - Angina pectoris
    - Arrhythmia
    - Congestive heart failure
    - Hypertension
    - Myocardial infarction
  - Abnormal systemic arterial morphology
    - Arterial stenosis
    - Coronary artery atherosclerosis
    - Peripheral arterial stenosis
    - Arteriosclerosis
    - Coronary artery atherosclerosis
  - Abnormality of the respiratory system
    - Dyspnea
  - Abnormality of the urinary system
  - Abnormality of the kidney
  - Abnormal renal physiology
    - Nephropathy
- Pain
  - Chest pain

Operation on heart (SNOMED: 4275564) and its subclasses

- 15,548 patients in the AKI AE cohort.
- Sex effect:
  - Male: 72%;
  - Female: 28%
- Age effect:
  - Age > 55: 78.5%

Finding: Phenotypes including **Type II diabetes** & **Nephropathy** are often observed before heart surgery-associated AKI adverse event.

Ref: He et al, ICBO-2019
Use Case 3: OMOP-2-OPMI-based COVID-19 clinical data standardization, modeling, and analysis

- National COVID Cohort Collaborative (N3C) with >5 mill. COVID cases
  - Analyze vaccine & drug AEs using the AE model (earlier slide)
  - AKI AE can still be a focus since AKI closely relates to COVID.
  - Effects of clinical variables (e.g., age, gender, comorbidities) can be modeled with OMOP-2-OPMI and analyzed using machine learning (ML) methods.

- More powerful by co-using the Coronavirus Infectious Disease Ontology (CIDO)
  - CIDO represents various COVID-19 knowledge and metadata, e.g., COVID-19 viral variants, vaccines, drugs, etc.
Summary and Discussion

- **OMOP CDM** is an open data standard in observational data integration and analysis
- **OPMI** is an ontology of precision medicine and investigation
- We further developed **OPMI** to map **OMOP** on **CDM** level

**Use cases:**
1. Ontology-level data standardization
2. Adverse event modeling and analysis
3. COVID-19 clinical data standardization, modeling, and analysis

- Other CDMs (e.g., PCORnet, CDISC)?
- More COVID-19 related AE studies using N3C data?
- More OMOP terms/concepts and relations?
Acknowledgements

University of Michigan
• Long Tran
• Yongqun “Oliver” He

IQVIA, UK
• Anastasios Siapos
• Christian Reich
• Sarah Seager

• Asiyah Yu Lin (NIH)
• Andrew Williams (Tufts University)

Funding:
• NIH NIAID grants: 1UH2AI132931 & NIH NIAID 1U24AI171008
• NIH-NIDDK grant: 1U2CDK114886
• Michigan Medicine–Peking University Health Sciences Center Joint Institute for Clinical & Translational Research (U072807).