Agenda

• OHDSI News: April Olympians by Clair Blacketer
• OHDSI Evidence Network by Clair Blacketer
• Vocabulary Contribution from Korea by Seng Chan You
• Upcoming OHDSI/OMOP Events in APAC
Introducing the April Olympians

CDM & THEMIS Collab-a-thon

Melanie Philofsky & Clair Blacketer
Current State of Conventions

- CDM and THEMIS conventions are housed on the CDM website, CDM github, THEMIS github and OHDSI forum.
- This does not provide users with concise and clear documentation on the how to standardize their data.
- The correct approach is often ambiguous.
Goals of the Collab-a-thon

1. Identify all *currently ratified* CDM and THEMIS conventions for every CDM table and field.
2. Write clear documentation for each THEMIS convention
3. Establish a repository for THEMIS conventions
4. Update the CDM documentation to link to relevant THEMIS repository entries
5. Create CDM documentation related to expansion module efforts around the community
Teamwork Makes the Dream Work

Hunters of Artemis

Writers of Apollo

Builders of Hephaestus
Hunters of Artemis

• On the hunt for ratified conventions
• This group will systematically search OHDSI resources by table and field
• They will then open github issues sharing critical information for each one they track down
Writers of Apollo

• Will pick up each convention the Hunters find
• This group will systematically fill out a template designed to house important information about the convention
• They will then tag the github issues for the next group
 Builders of Hephaestus

• Responsible for building the repository
• Will take the written information given by the writers and incorporate the convention into the repository
• This group will also clean up the CDM website by removing any duplicate information and linking out to the THEMIS repository
Expansion Module Owners

• You will be responsible for either creating documentation for your tables similar to the existing CDM docs or providing information on the tables under construction and how people can participate

• This includes:
  – GIS
  – Medical Imaging
  – Oncology*

*While the Episode and Episode Events tables are in the canonical CDM, there are some outstanding articles and missing information
Please Join Us!

• To sign up for one of the teams, please scan the QR code and indicate your team preference
• The time commitment is as much or as little as you want
• The idea was to create discrete pieces of work so you can pick and choose how much you would like to do
• No technical skills necessary! If you can google you can hunt!
OHDSI Evidence Network

Clair Blacketer

Lead, CDM Workgroup
Lead, Network Data Quality Workgroup
Why are we here?

...to collaboratively generate evidence that promotes better health decisions and better care.
Why are we here?

Network studies are hard!

...to collaboratively generate evidence that promotes better health decisions and better care.
Regulatory Guidelines

Considerations for the Use of Real-World Data

- FDA recognizes that evaluation of relevant data sources or databases is an important step in the design of a study and in evaluating a study’s feasibility. Such evaluations of data sources or databases for feasibility purposes serve as a way for the sponsor and FDA to (1) assess if the data source or database is fit for use to address the research question being posed and (2) estimate the statistical precision of a potential study without evaluating outcomes for treatment arms.

- Sponsors should describe in the study protocol, or as an appendix to the protocol, the data sources evaluated when designing the study, including results from feasibility evaluations or exploratory analyses of those data sources. Sponsors should provide a justification for selecting or excluding relevant data sources from the study. Sponsors should also describe how the choice of the final data sources, study design elements, and analytic approaches aligns with the research question of interest and that the data sources, study design elements, and analytic approaches were not selected to favor particular study findings.
Pillar #2: Standardized data network

- Opportunity: Increase transparency and maturity of OHDSI data network
- Proposed solutions:
  - Create OHDSI data network catalog to encourage network studies across interested partners and promote data quality practices
  - Generate OHDSI network concept prevalence data and make accessible for ATLAS users to enable more generalizable phenotype development
  - Promote database diagnostics by having data partners share limited subset of ACHILLES to allow for users to identify databases that satisfy study criteria

What is Database Diagnostics?

R package that allows us to determine...

...which databases have the elements required to answer a research question...

...using only a set of aggregated summary statistics.
Save our Sisyphus Challenge

Amongst people with psoriasis, does exposure to Risankizumab increase the risk of cerebrovascular events while on treatment relative to other biologic therapies? Lead: Zenas Yiu

Characterization: incidence of progressive multifocal leukoencephalopathy (PML) during Multiple Sclerosis (MS) biologic exposure Lead: Thamir Alshammary

Intravitreal Anti-VEGF and Kidney Failure Lead: Cindy Cai

Is fluoroquinolone use really associated with the development of aortic aneurysms Leads: Jack Janetzki, Jung Ho Kim, Seonji Kim, Jung Ah Lee, Nicole Pratt, Seng Chan You,
## SOS Database Diagnostics Results

### Data Diagnostic Explorer

<table>
<thead>
<tr>
<th>database id</th>
<th>A1: afibercept vs. bevacizumab for blinding diseases with esrd outcome</th>
<th>B1: fluoroquinolone vs. cephalosporin for urinary tract infection and risk of aortic aneurysm</th>
<th>D2: risankizumab vs. tildrakizumab for psoriasis and risk of ischemic stroke</th>
<th>C2: biologics vs disease modifying treatments for multiple sclerosis and risk of PML</th>
</tr>
</thead>
<tbody>
<tr>
<td>truen_mdcd_2359_20230215</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>US_PharMetrics_Plus_20230330</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>JHM_OMOP_20230406</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>truen_ccae_2324_20230201</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>optum_ehr_2247_202211205</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>US_OPENCLAIMS_20230313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Japan_HIS_20220120</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>jmdc_2325_20230126</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>US_Hospital_20230130</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CUIMC_20221214</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>VA-OMOP_20230411</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

[https://data.ohdsi.org/DataDiagnostics/](https://data.ohdsi.org/DataDiagnostics/)
# Inaugural Data Sources of the OHDSI Evidence Network

<table>
<thead>
<tr>
<th>University / Research Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ajou University • Ajou University</td>
</tr>
<tr>
<td>Casa di Cura Igea • Casa di Cura Igea</td>
</tr>
<tr>
<td>Clinical Center of Montenegro • Clinical Center of Montenegro</td>
</tr>
<tr>
<td>Columbia University Medical Center • Columbia Medical Center</td>
</tr>
<tr>
<td>University College London • UK THIN</td>
</tr>
<tr>
<td>IQVIA • Australia EMR</td>
</tr>
<tr>
<td>IQVIA • Disease Analyzer France</td>
</tr>
<tr>
<td>IQVIA • Disease Analyzer Germany</td>
</tr>
<tr>
<td>IQVIA • Japan Claims</td>
</tr>
<tr>
<td>IQVIA • Japan HIS</td>
</tr>
<tr>
<td>IQVIA • Longitudinal Patient Database (LPD) in Belgium</td>
</tr>
<tr>
<td>IQVIA • Longitudinal Patient Database (LPD) in France</td>
</tr>
<tr>
<td>IQVIA • Longitudinal Patient Database (LPD) in Italy</td>
</tr>
<tr>
<td>IQVIA • Longitudinal Patient Database (LPD) in Spain</td>
</tr>
<tr>
<td>IQVIA • OMOP US Hospital Data Master</td>
</tr>
<tr>
<td>IQVIA • Pharmetrics Plus</td>
</tr>
<tr>
<td>IQVIA • UK Medical Research Data EMIS</td>
</tr>
<tr>
<td>IQVIA • UK Medical Research Data THIN</td>
</tr>
<tr>
<td>IQVIA • US Open Claims</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • JMDCC</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Merative®</td>
</tr>
<tr>
<td>Marketscan® Commercial Claims and Encounters</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Merative®</td>
</tr>
<tr>
<td>Marketscan® Medicare Supplement</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Optum’s Clinformatics® Data Mart - Date of Death</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Optum’s Clinformatics® Data Mart - Socio-Economic Status</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Optum’s Longitudinal EHR Repository</td>
</tr>
<tr>
<td>Janssen Research &amp; Development • Premier Healthcare Database</td>
</tr>
<tr>
<td>Johns Hopkins University • Johns Hopkins University</td>
</tr>
<tr>
<td>National University of Singapore • National University of Singapore</td>
</tr>
<tr>
<td>Northeastern • IQVIA Pharmetrics Plus</td>
</tr>
<tr>
<td>Organization Name • Data Source Name</td>
</tr>
<tr>
<td>Taipei Medical University • Taipei Medical University</td>
</tr>
<tr>
<td>Tufts University Medical Center • Tufts University Medical Center</td>
</tr>
<tr>
<td>University of Nebraska Medical Center • University of Nebraska Medical Center</td>
</tr>
<tr>
<td>University of Southern California • Keck Medical Center</td>
</tr>
<tr>
<td>US Department of Veteran’s Affairs • US Department of Veteran’s Affairs</td>
</tr>
<tr>
<td>Yinzhou Bigdata Platform • Yinzhou Bigdata Platform</td>
</tr>
</tbody>
</table>
APAC Community call: Vocabulary Contribution(Korea)

Seng Chan You

March 21, 2024
Topics

• What is **OHDSI Standardized Vocabularies**?

• Introduce the **Korean EDI Vocabulary**

• Create a Semi-automated process: **EDI to OHDSI package**
  
  • The advantages of incorporated EDI

• **Process** for Data Integration in ATHENA
What is OHDSI Standardized Vocabularies?

- **OHDSI Standardized Vocabularies** is a collection of public standard vocabularies used in the OHDSI network.
- It consolidates a system of Vocabularies, Classifications, Domains, and Concepts into a **common format** and stores them in a **set of CDM tables**.
- It enables the **application of standardized large-scale analytical methods** in a federated setting.
What is OHDSI Standardized Vocabularies?

- **Standard Concept** (standard_concept = ‘S’)
  - Official representation of unique clinical entities, recorded with their concept IDs in designated fields.
  - Typically sourced from well-established vocabularies like SNOMED, ensuring clear definitions and comprehensive coverage.

- **Classification Concept** (standard_concept = ‘C’)
  - Have a hierarchical relationship to Standard Concepts.
  - Can be used to query for Standard Concepts using the records of `CONCEPT_ANCESTOR` table.

- **Non-standard Concept** or **Source Concept** (standard_concept = NULL)
  - Entities that are neither Standard or Classification Concepts.
  - Only appear in the `source_concept_id` fields of data tables.
  - Each Non-standard Concept is mapped to one or more Standard Concepts.
Introduction of EDI

- **EDI** (Electronic Data Interchange) is a code system for the reimbursement or claim data in Korea.

- EDI concepts are divided into **drugs, devices, and medical services**.

- EDI is developed and maintained by **HIRA** (Health Insurance Review & Assessment Service), **updated on the 1st of every month**.

- The number of EDI codes by domain announcement on October 1, 2023 is as follows.

<table>
<thead>
<tr>
<th>Number of EDI codes</th>
<th>Drug</th>
<th>Device</th>
<th>Medical Service</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65,637</td>
<td>44,453</td>
<td>457,740</td>
</tr>
</tbody>
</table>
Introduction of EDI

- Despite widespread adoption in Korean EHR systems, limitations still persist.
  - Validity dates are not recorded in the official monthly announcement.
  - There are expired or replaced EDI code and outdated EDI can be assigned to new concepts.
  - EDI has duplicated identifiers due to the lack of a unified encoding system across domains.

OUR GOAL IS ...

To enhance EDI vocabulary for a controlled and standardized vocabulary system

- For this purpose, we incorporated the EDI into OHDSI vocabulary using a semi-automated process.
EDI to OHDSI package

- EDI to OHDSI package is a **semi-automated process** we made.
- This package makes EDI vocabulary as a **Source Concepts** using OHDSI table structure.
- There are **four main steps** to incorporate EDI into OHDSI standardized vocabulary.

1. Improved the classification of EDI domains and separated medical services into procedures and measurements
2. Assigned a unique identifier and validity dates for each EDI concept
3. Built a vertical hierarchy between EDI concepts
4. Added an English definition for each EDI concept using Google Translation

https://github.com/dr-you-group/EDItoOmopPackage
EDI to OHDSI package

1. Improved the classification of EDI domains and separated medical services into procedures and measurements

- EDI concepts are divided into **Drugs**, **Devices**, and **Medical Services**.

- The scope of **Medical Services** is too broad for the OHDSI standardized vocabularies.

- So, we subclassified Medical Services into **Procedures** and **Measurements** to match the OHDSI domains.

https://github.com/dr-you-group/EDItoOmopPackage
2 Assigned three attributes for each EDI concept

Attributes: Valid start date, Valid end date, Invalid reason

- **Valid start date, Valid end date**
  - When an EDI concept is newly registered or deprecated, the term’s date is updated or expired.

- **Invalid reason**
  - If a concept is valid $\rightarrow$ invalid reason : NULL
  - If a concept is replaced by another concept or deleted $\rightarrow$ invalid reason : U or D

<table>
<thead>
<tr>
<th>Concept code</th>
<th>Valid start date</th>
<th>Valid end date</th>
<th>Invalid reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>A29506361</td>
<td>2008-04-01</td>
<td>2099-12-31</td>
<td>NULL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept code</th>
<th>Valid start date</th>
<th>Valid end date</th>
<th>Invalid reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>A29506361</td>
<td>2008-04-01</td>
<td>2010-01-31</td>
<td>U</td>
</tr>
<tr>
<td>670600010</td>
<td>2010-02-01</td>
<td>2099-12-31</td>
<td>NULL</td>
</tr>
</tbody>
</table>
EDI to OHDSI package

3 Built a vertical hierarchy between EDI concepts

- We built a formal vertical hierarchy for EDI concepts as ICD-9 or ICD-10 code system.

- The first five digits of the EDI code in the medical service domain (procedure, measurement) represent the ancestor terms for longer descendent EDI code.

<table>
<thead>
<tr>
<th>Concept code</th>
<th>Concept name</th>
<th>Ancestor concept code</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC281</td>
<td>Whole Body Scan</td>
<td></td>
</tr>
<tr>
<td>HC281006</td>
<td>Whole Body Scan, Nuclear Medicine and other physician reading</td>
<td>HC281</td>
</tr>
<tr>
<td>HC281300</td>
<td>Whole Body Scan, Under 8 years old</td>
<td>HC281</td>
</tr>
<tr>
<td>HC281306</td>
<td>Whole Body Scan, Under 8 years old, read by nuclear medicine physician</td>
<td>HC281</td>
</tr>
<tr>
<td>HC281600</td>
<td>Whole Body Scan, Under 72 months</td>
<td>HC281</td>
</tr>
<tr>
<td>HC281606</td>
<td>Whole Body Scan, Under 72 months, Nuclear Medicine physician reading</td>
<td>HC281</td>
</tr>
</tbody>
</table>
EDI to OHDSI package

4 Added an English definition for each EDI concept using Google Translation

- You have to add an English definition for each EDI term.
- We utilized Google Cloud Translation API for the initial translation.
- Inaccurately translated words underwent review by nurses and were retranslated.

<table>
<thead>
<tr>
<th>Korean definition</th>
<th>Using Google Translation API</th>
<th>Using Google Translation API With glossary</th>
</tr>
</thead>
<tbody>
<tr>
<td>맥산중독-정량그룹1,b형 감염바이러스 [중합효소연쇄반응교잡반응법]</td>
<td>Nucleic acid amplification-quantitative group 1 hepatitis B virus [polymerase chain reaction hybridization method]</td>
<td>Nucleic acid amplification-quantitative group 1_HBV [PCR-Hybridization]</td>
</tr>
<tr>
<td>단기사용담관용튜브·카테터</td>
<td>Short-term use bile duct tube and catheter</td>
<td>Cahteter, bile duct short-term use</td>
</tr>
</tbody>
</table>

https://github.com/dr-you-group/EDItoOmopPackage
### The advantages of Incorporated EDI

#### Terminology evaluation criteria and explanation

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Explanation</th>
<th>EDI vocabulary</th>
<th>EDI in OHDSI vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniqueness and exclusivity of the concept</td>
<td>Concept orientation A concept must be linked with only one term</td>
<td>△</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Non-semantic concept identifiers There must be a unique code representing a concept</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Coverage The domain covered by the terminology system must be consistent and obvious</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Synonyms uniquely identified and mapped to relevant concepts Synonyms, including abbreviations are managed by unique identifiers, and related concepts are mapped</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td>Hierarchies and relationships between concepts</td>
<td>Relation The relation of each concept should be defined</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Multiple hierarchy A concept can have multiple hierarchies</td>
<td>X</td>
<td>△</td>
</tr>
<tr>
<td></td>
<td>Formal definition Having a structure and definition that can be indexed and processed by computer</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Compositional definition Terms can be separated into atomic units and have compositional extensibility</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Management system for vocabulary</td>
<td>Concept permanence Even if the used term is updated, the previously used term should not be deleted</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Version control When terminology is updated, version information, including changes, must be specified</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>Multi-language The terminology system supports multiple languages</td>
<td>△</td>
<td>○</td>
</tr>
</tbody>
</table>

- All criteria except compositionality, indicate that converted EDI demonstrates a higher quality index than the original EDI.
- Users can easily search for related concepts using formal English definitions within Athena.
- Users can activate versioning by storing metadata that indicates the start and end dates for each concept.
Process for Data Integration in ATHENA

- We successfully incorporated 313,431 EDI codes of Korean medical information in 2019.

- Now, our objective is to expand not only the incorporation of EDI vocabularies but also add mapping with Standard Concepts.

<table>
<thead>
<tr>
<th></th>
<th>Non-standard Concepts</th>
<th>Standard Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>EDI</td>
<td>RxNorm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RxNorm Extension</td>
</tr>
<tr>
<td>Device</td>
<td>EDI</td>
<td>SNOMED</td>
</tr>
<tr>
<td>Procedure</td>
<td>EDI</td>
<td>1) SNOMED</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) LOINC</td>
</tr>
<tr>
<td>Measurement</td>
<td>EDI</td>
<td>1) LOINC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) SNOMED</td>
</tr>
</tbody>
</table>
Process for Data Integration in ATHENA

- For the integration, you need to adhere to the **Community Contribution pipeline**.
- This process is required for quality assurance and control process
- **We are still in the process** working with the Vocabulary Team, so consider this just for reference.
- If you are interested in loading the vocabulary, please contact the Vocabulary Team.

Contact the Vocabulary Team
or
Vocabulary WG meeting.

https://github.com/OHDSI/Vocabulary-v5.0/wiki/Community-contribution
Process for Data Integration in ATHENA

- **Prerequisites**
  - PostgreSQL database (plpython3u, plpgsql extensions)
  - Preparation Schemas: sources, dev_xyz, devv5
  - Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

- **Process**
  1. Run load_stage.sql in the dev_xyz schema
  2. QA/QC part 1
  3. Generic update
  4. QA/QC part 2: semi-automatic process
Process for Data Integration in ATHENA

- **Prerequisites**
  - PostgreSQL database (plpython3u, plpgsql extensions)
  - Preparation Schemas: sources, dev_xyz, devv5
  - Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

- **Process**
  1. Run load_stage.sql in the dev_xyz schema
  2. QA/QC part 1
  3. Generic update
  4. QA/QC part 2: semi-automatic process
Process for Data Integration in ATHENA

• Prerequisites
  • PostgreSQL database (plpython3u, plpgsql extensions)
  • Preparation Schemas: sources, dev_xyz, devv5
  • Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

```sql
-- Install the extensions
CREATE EXTENSION plpython3u;
CREATE EXTENSION plpgsql;

-- Check the list of extensions
SELECT * FROM pg_extension;
```
Process for Data Integration in ATHENA

• Prerequisites
  • PostgreSQL database (plpython3u, plpgsql extensions)
  • Preparation Schemas: sources, dev_xyz, devv5
  • Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

Schema sources
  • Put your source vocabulary

Schema dev_xyz
  • Set as Working directory
  • Run DevV5_DDL.sql to create empty tables
  • Put copy of vocabularies downloaded from Athena

Schema devv5
  • Reference Schema of Working directory
  • Run DevV5_DDL.sql to create empty tables
  • Put copy of vocabularies downloaded from Athena

https://github.com/OHDSI/Vocabulary-v5.0/wiki/Community-contribution
Process for Data Integration in ATHENA

- **Prerequisites**
  - PostgreSQL database (plpython3u, plpgsql extensions)
  - Preparation Schemas: sources, dev_xyz, devv5
  - Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

https://athena.ohdsi.org/search-terms/start
Process for Data Integration in ATHENA

- **Prerequisites**
  - PostgreSQL database (plpython3u, plpgsql extensions)
  - Preparation Schemas: sources, dev_xyz, devv5
  - Copies of tables, fully indexed (downloaded from Athena and put into Schema dev_xyz, devv5)

④ Run `--Main DDL in postgresQL`

⑤ Upload the corresponding tables from the vocabulary bundle in Athena

⑥ Create PKs, external keys, indexes and checks

https://github.com/OHDSI/Vocabulary-v5.0/blob/bd54f495932acbe529d4534eae25195aa6455759/working/DevV5_DDL.sql
Process for Data Integration in ATHENA

• **Prerequisites**
  • PostgreSQL database (plpython3u, plpgsql extensions)
  • Preparation Schemas: *sources, dev_xyz, devv5*
  • Copies of tables, fully indexed (downloaded from Athena and put into Schema *dev_xyz, devv5*)

• **Process**
  1. Run `load_stage.sql` in the *dev_xyz* schema
  2. QA/QC part 1
  3. Generic update
  4. QA/QC part 2: semi-automatic process
Process for Data Integration in ATHENA

**Prerequisites**
- PostgreSQL database (plpython3u, plpgsql extensions)
- Preparation Schemas: `sources`, `dev_xyz`, `devv5`
- Copies of tables, fully indexed (downloaded from Athena and put into Schema `dev_xyz`, `devv5`)

**Process**
1. Run `load_stage.sql` in the `dev_xyz` schema
2. QA/QC part 1
3. Generic update
4. QA/QC part 2: semi-automatic process

[GitHub Link](https://github.com/OHDSI/Vocabulary-v5.0/wiki/Community-contribution)
Process for Data Integration in ATHENA

- **Process**
  1. Run `load_stage.sql` in the `dev_xyz` schema
     - If `load_stage` references `devv5` or `sources`, replace them with the names of your schema
     - Run function `Vocabulary_pack.SetLatestUpdate.sql`

```sql
DO $$
BEGIN
  PERFORM VOCABULARY_PACK.SetLatestUpdate(
    pVocabularyName => 'vocabulary_id of your vocabulary as in Vocabulary table',
    pVocabularyDate => 'date of new vocabulary version',
    pVocabularyVersion => 'name of the vocabulary version, if none use date of the version',
    pVocabularyDevSchema => 'name of your development schema'
  );
END $$;
```

[GitHub Link](https://github.com/OHDSI/Vocabulary-v5.0/tree/master/working/packages/vocabulary_pack)
Process for Data Integration in ATHENA

- **Process**
  1. Run `load_stage.sql` in the `dev_xyz` schema
     - If `load_stage` references `devv5` or `sources`, replace them with the names of your schema
     - Run function `Vocabulary_pack.SetLatestUpdate.sql`
     - Run supporting functions to stage tables

Usually use these support functions:

```python
VOCABULARY_PACK.ProcessManualRelationships()
VOCABULARY_PACK.AddFreshMAPSTO()
VOCABULARY_PACK.DeprecateWrongMAPSTO()
VOCABULARY_PACK.DeleteAmbiguousMAPSTO()
```

- You should download function queries you need
- Run query
- If you've done it correctly, it should be contained within the `vocabulary_pack` schema functions
Process for Data Integration in ATHENA

- **Process**

  2. **QA/QC part 1**

  - As a result of previous step, you will populate stage tables
    
    \(\text{concept\_stage, concept\_relationship\_stage, concept\_synonym\_stage, etc}\)
  
  - Run `qa_tests.check_stage_tables()` in `create_qa_tests.sql`

```sql
CREATE OR REPLACE FUNCTION qa_tests.check_stage_tables ()
RETURNS TABLE (error_text TEXT, rows_count BIGINT)
AS $BODY$
BEGIN

RETURN QUERY
SELECT reason, COUNT(*) FROM (--concept_relationship_stage
    SELECT
      CASE WHEN v1.vocabulary_id IS NOT NULL AND ...
)

https://github.com/OHDSI/Vocabulary-v5.0/blob/master/working/packages/QA_TESTS/create_qa_tests.sql#L623
```

This function performs conformance checks and should return no errors.

- If the concept is valid, check `valid\_end\_date = 12/31/2099`
- Field length does not exceed limits in the standard DDL
  - There are no duplicates
  - Vocabulary_id exist in `VOCABULARY` table
Process for Data Integration in ATHENA

- **Process**
  3. **Generic update**
     - This function integrates the content of the _stage_ tables into basic tables and assigns concept_ids.
     - If you execute this function, you can inspect _CONCEPT_ and _CONCEPT_RELATIONSHIP_.

```sql
CREATE OR REPLACE FUNCTION devv5.GenericUpdate ( )
RETURNS void AS
$BODY$
BEGIN

--1. Prerequisites:
--1.1 Check vocabulary table, at least one vocabulary must
    PERFORM FROM vocabulary WHERE latest_update IS NOT NULL LT
    IF NOT FOUND THEN
        RAISE EXCEPTION 'At least one vocabulary must have
        USING HINT = 'Forgot to execute SetVar'
    END IF;

--1.2 Check stage tables for incorrect rows
DO $$

If you need to modify your scripts or stage tables, you will need to clean stage tables and revert basic tables.

You can easily erase all changes using _fast_recreate_schema.sql_.
```
Process for Data Integration in ATHENA

- Process

4. QA/QC part 2: semi-automatic process
   - Execute `get_checks` to check the compliance of resulting basic tables to the OMOP rules
   - Check high-level statistics
   - Run `manual_checks_after_generic` to review manually

* This query is very heavy
It may take a long time to execute.
Process for Data Integration in ATHENA

- **Process**
  
  4. **QA/QC part 2: semi-automatic process**
     
     - Execute `get_checks` to check the compliance of resulting basic tables to the OMOP rules
     - Check high-level statistics
     - Run `manual_checks_after_generic` to review manually

You can use other functions from the folder, `Vocabulary-v5.0/working/packages/QA_TESTS`

https://github.com/OHDSI/Vocabulary-v5.0/tree/master/working/packages/QA_TESTS
• Process
  
  4. QA/QC part 2: semi-automatic process
    • Execute `get_checks` to check the compliance of resulting basic tables to the OMOP rules
    • Check high-level statistics
    • Run `manual_checks_after_generic` to review manually

https://github.com/OHDSI/Vocabulary-v5.0/blob/master/working/manual_checks_after_generic_update.sql
Upcoming OHDSI/OMOP Events in APAC

• 2 events coming up in April in Japan and Thailand
• Japan: Half-day event with lectures, discussions and a hands-on session
• Thailand: Full-day event with lectures and an introductory tutorial
OHDSI/OMOP Event in Japan

- **Date:** April 17, 2024
- **Venue:** National Cancer Center Hospital, Tsukiji, Tokyo
- **Tentative Agenda**

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00 – 13:20</td>
<td>OMOP and OHDSI Japan Introduction</td>
<td>Tatsuo Hiramatsu</td>
</tr>
<tr>
<td>13:20 – 13:30</td>
<td>Generating RWD/RWE in OHDSI APAC using the OMOP CDM</td>
<td>Mui Van Zandt</td>
</tr>
<tr>
<td>13:30 – 14:00</td>
<td>Ongoing initiatives in Japan</td>
<td>OHDSI Japan Collaborators</td>
</tr>
<tr>
<td>14:00 – 15:00</td>
<td>Why federated (network) studies within a country?: the OHDSI UK Studyathon experience</td>
<td>Dani Prieto-Alhambra</td>
</tr>
<tr>
<td>15:00 – 15:45</td>
<td>Discussion between Dani and biostatisticians in Japan</td>
<td></td>
</tr>
<tr>
<td>15:45 – 16:00</td>
<td>Open Q&amp;A with audience</td>
<td></td>
</tr>
<tr>
<td>16:00 – 16:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>16:30 – 19:30</td>
<td>Hands-on session: Replication of <em>Association of Ticagrelor vs Clopidogrel With Net Adverse Clinical Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention</em></td>
<td>Seng Chan You</td>
</tr>
</tbody>
</table>

Registrations open at [https://odjpn.doorkeeper.jp/events/171041](https://odjpn.doorkeeper.jp/events/171041)!
### OHDSI/OMOP Event in Thailand

- **Date:** April 24, 2024
- **Venue:** Eastin Grand Hotel Phayathai, Bangkok
- **Tentative Agenda**

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:10 – 09:30</td>
<td>Trends in RWD/RWE and data standardization</td>
<td>Mui Van Zandt</td>
</tr>
<tr>
<td>09:30 – 09:50</td>
<td>European OMOP initiatives</td>
<td>Sarah Seager</td>
</tr>
<tr>
<td>09:50 – 10:10</td>
<td>Local perspectives and considerations on RWD/RWE</td>
<td>KOL from Thailand</td>
</tr>
<tr>
<td>10:10 – 10:30</td>
<td>Lessons learned adopting OHDSI/OMOP in Thailand</td>
<td>Natthawut Adulyanukosol</td>
</tr>
<tr>
<td>10:30 – 10:50</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:50 – 11:50</td>
<td>OHDSI/OMOP Introduction</td>
<td>Sarah Seager</td>
</tr>
<tr>
<td>11:50 – 13:30</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>13:30 – 14:45</td>
<td>OMOP CDM and Vocabulary + Vocabulary mapping exercises</td>
<td>Mui Van Zandt/Gyeol Song</td>
</tr>
<tr>
<td>14:45 – 15:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>15:00 – 16:30</td>
<td>OMOP Conversion Process + ETL exercises</td>
<td>Mui Van Zandt/Gyeol Song</td>
</tr>
<tr>
<td>16:30 – 17:00</td>
<td>Closing &amp; Networking</td>
<td></td>
</tr>
</tbody>
</table>
Thank you!