ETL in the Time of Data Privacy

Kristin Kostka, MPH
Talita Duarte-Salles, PhD
Who We Are

Kristin Kostka, MPH
IQVIA (United States)
Data steward: IQVIA data
Co-lead Project Charybdis

Talita Duarte-Salles, PhD
IDIAPJGol (Barcelona, Spain)
Data steward: SIDIAP, HM Hospitales
Co-lead Project Charybdis
I’ve been seeing all this cool COVID work that OHDSI’s doing... how can I get my data into OMOP?
Process to become an OMOP site

Commit to OHDSI

Build ETL

Finish ETL / Release CDM

Run DQ Queries

Run Study Package

Send Results to OHDSI SFTP
Joining the OMOP Global Data Network

Becoming an OHDSI site involves a variety of decisions across the following dimensions (in no particular order):

Who writes ETL scripts and executes the conversion?
Who builds and maintains the infrastructure?
Who leads OHDSI / OMOP training and support?
Who installs the OHDSI Tools (e.g. ATLAS)?
Who certifies the CDM data quality?
How do receive, review and conduct network studies?
Sites have different rules about patient data.

**In Europe:**
General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA)

*Implemented 25 May 2018*

**In the US:**
The Health Insurance Portability and Accountability Act (HIPAA) of 1996 establishes national standards to protect individuals' medical records and other personal health information
What happens if you need help during developing site extract-transform-load processes but underlying data cannot be shared?
OMOP Conversion Process Flow

Data experts & CDM experts together design the ETL

Medical experts create the code mappings

All are involved in quality control

A technical person implements the ETL

Analysis

Tools:
- White Rabbit
- Rabbit In a Hat
- Usagi

Quality Control
- Internal Quality Checks
- Achilles
- Data Quality Dashboard

Development
- Jenkins
- Code Repository
Take advantage of your OHDSI toolkit.
OMOP Conversion Process Flow

1. Data experts & CDM experts together design the ETL
2. Medical experts create the code mappings
3. All are involved in quality control
4. A technical person implements the ETL

Tools:
- White Rabbit
- Rabbit In a Hat
- Usagi
- Internal Quality Checks
- Achilles
- Data Quality Dashboard
- Jenkins
- Code Repository
WhiteRabbit for ETL design

WhiteRabbit is a software tool to help prepare for ETLs (Extraction, Transformation, Loading) of longitudinal healthcare databases into the OMOP Common Data Model (CDM). The source data can be in comma-separated text files, or in a database (MySQL, SQL Server, ORACLE, PostgreSQL); the CDM will be in a database (MySQL, SQL Server, PostgreSQL).

The main function of WhiteRabbit is to perform a scan of the source data, providing detailed information on the tables, fields, and values that appear in a field. This scan will generate a report that can be used as a reference when designing the ETL, for instance when using the Rabbit-In-a-Hat tool. Rabbit-In-a-Hat uses the scan document and displays source data information through a graphical user interface to allow a user to connect source data structure to the CDM data structure. The function of Rabbit-In-a-Hat is to generate documentation for the ETL process, not generate code to create an ETL.

Download WhiteRabbit: https://github.com/OHDSI/WhiteRabbit
Leveraging the power of WhiteRabbit
Scan reports are our friends
Don’t be afraid to ask for help!
The power of collaboration in action

Talita runs WhiteRabbit to generate the mapping document + ScanReport on her new Madrid data.

Along the way she realizes...

¡Necesito ayuda!

Patrick offers to help... except Patrick no habla español

Talita can do that! She translates the Scan Report into English and shares her artifacts to Patrick.

Patrick uses this to mock-up a DB and helps Talita write ETL for her database. He shares back the code she can deploy in her environment.

How does this work? This is aggregate results profiling the source data. No data moves.
More OHDSI Resources on your ETL journey
You can always phone a friend!

Clair Blacketer, MPH, PMP
https://www.ohdsi.org/who-we-are/collaborators/clair-blacketer/

Erica Voss, MPH
https://www.ohdsi.org/who-we-are/collaborators/erica-voss/
We can help you prioritize data elements!

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health pre-COVID19</td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td>Conditions</td>
</tr>
<tr>
<td></td>
<td>Drugs</td>
</tr>
<tr>
<td>COVID19 presentation</td>
<td>Health service utilization</td>
</tr>
<tr>
<td></td>
<td>Recent health behaviour</td>
</tr>
<tr>
<td></td>
<td>Measurements/tests/lab</td>
</tr>
<tr>
<td></td>
<td>Conditions</td>
</tr>
<tr>
<td></td>
<td>Drugs</td>
</tr>
<tr>
<td>Test time</td>
<td>COVID19 test presence</td>
</tr>
<tr>
<td></td>
<td>COVID19 test type</td>
</tr>
<tr>
<td></td>
<td>COVID19 test results</td>
</tr>
<tr>
<td>Admission for COVID19</td>
<td>Inpatient services</td>
</tr>
<tr>
<td></td>
<td>Complications (eg AKI, sepsis)</td>
</tr>
<tr>
<td></td>
<td>ICU</td>
</tr>
<tr>
<td></td>
<td>Procedures (eg tracheostomy)</td>
</tr>
<tr>
<td></td>
<td>Devices (eg ECMO)</td>
</tr>
<tr>
<td>ICU Admission detail</td>
<td>Date of ICU admission</td>
</tr>
<tr>
<td></td>
<td>CPAP</td>
</tr>
<tr>
<td></td>
<td>Intubation/mechanical vent</td>
</tr>
<tr>
<td></td>
<td>Procedures (eg tracheostomy)</td>
</tr>
<tr>
<td></td>
<td>Devices (eg ECMO)</td>
</tr>
<tr>
<td>Mortality</td>
<td>Date of death</td>
</tr>
<tr>
<td></td>
<td>Cause of death</td>
</tr>
</tbody>
</table>
Join the Journey

http://ohdsi.org