Always ♥ Your Data:
An Approach to Observational Data Lifecycle Management in the OHDSI Ecosystem

Erica A. Voss, MPH[1,2,3,*], Patrick B. Ryan, PhD[1,2,4], Clair Blacketer, MPH[1,2], Ajit A. Londhe, MPH[1,2], Martijn J. Schuemie, PhD[1,2]

[1] Janssen Research and Development, Raritan, NJ, [2] Observational Health Data Sciences and Informatics (OHDSI), New York, NY, [3] Erasmus University Medical Center, Rotterdam, Netherlands, [4] Columbia University, New York, NY

INTRODUCTION

- Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) must be maintained. This includes:
  - Testing
  - Updating the OMOP Vocabulary
  - Managing raw data schema or CDM version changes
- The CDM enables research and can efficiently leverage diverse raw data sets to produce timely evidence

DESIGN & IMPLEMENT

- **CDM Data Custodians**:
  - design an Extract, Transform, & Load (ETL) process for each raw data source
  - develop a program to convert raw data to the CDM
- **CDM_BUILDER** [1] is an example of a program, it has evolved over time:
  - Initially SQL queries were used
  - Then generic off-the-shelf ETL tools
  - Ultimately, a customized program that standardizes where it can but is flexible enough to handle each data source's nuances

MAINTAIN

- **CDM ETL Coordinators**:
  - decide on changes to CDM_BUILDER
  - important role of driving changes to the timeline
- Monthly changes necessary are decided upon, usually there is a theme:
  - Data Refreshes (i.e. raw data sources have new data)
  - Updating the OMOP Vocabulary
  - Tuning CDM_BUILDER (i.e. time to improving the run time)
  - Adding enhancements / fixing bugs
- Tasks managed in issue tracking tool
- Initial time investment of developing a CDM is high, however over time the energy required to maintain not as significant

Design & Implement — Test — Release — Maintain

CDM Data Custodians — CDM ETL Coordinators

TESTING

- **Rabbit In a Hat (RIAH) Testing Framework** [2] to build unit tests:
  - CDM_BUILDER is run on the test cases and the outputted CDM is evaluated
  - Test cases / CDM_BUILDER augmented until all test cases pass
- OHDSI **Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (ACHILLES)** enables characterization of a CDM
  - Critical components from ACHILLES are used:
    - **Data Density report** shows by domain, year by year number of records
    - **Heel report** was designed by OHDSI to check for data quality issues
    - Additional diagnostics are recommended:
      - **Compare OMOP Vocabularies** - if OMOP Vocabulary is to be updated, old version is compared to the new
      - **Compare "cohorts of interest" in new/old CDM** - the cohorts have conditions, drug exposures, measurements, and procedures frequently used for analysis within an organization

RELEASE

- There are a series of steps handling "cut over":
  - Communication describing changes is shared with the analysts
  - Date / time are set to update the standardized tools
  - Final communication asking them to start using the new sources
- **ACHILLES'** preprocessed summarizations of each CDM allow users databases' characterization and data quality
- **ATLAS** allows an organization to perform typical epidemiologic data analysis work however in a standard, systematic way improving quality and transparency
  - develop concept sets of standard terminology
  - create cohorts of patients
  - generate R code for performing analysis
- **Methods Library** uses CDM to do population-level estimation and patient-level prediction

CONCLUSION

- CDMs must be maintained / updated as the world around them changes
- Efficiencies gained by using CDMs offset the energy spent on development / maintenance
- Processes highlighted in this poster could help improve lifecycle of CDMs

REFERENCES

1. ETL-CDMBuilder. [GitHub] 2018.02.25]; Available from: https://github.com/OHDSI/ETL-CDMBuilder.
2. Schuemie, M.J. Rabbit-In-a-Hat Testing Framework. [Wiki] 2018.02.25]; Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:software:whit0erabbit:test_framework.

OHDSI — Erasmus MC University Medical Center Rotterdam — Janssen PHARMACEUTICAL COMPANIES OF Johnson&Johnson

*Corresponding Author – evoss3@its.jnj.com