"OMOP Anywhere": Daily Updates from EHR Data Leveraging Epic's Native Tools

Mujeeb A Basit[1,2], Mereeja Varghese[1], Aamirah Vadsariya[1], Bhavini Nayee[1,2], Margaret Langley[1,2], Ashley Huynh[1,2], Jennifer Cai[2], Donglu Xie[2], Cindy Kao[2], Eric Nguyen[2], Todd Boutte[2], Shiby Antony[2], Tammye Garrett[2], Christoph U Lehmann[1], Duwayne L Willett[1,2]

[1] University of Texas Southwestern Clinical Informatics Center, [2] University of Texas Southwestern Information Resources

Background

What if any healthcare organization with an Epic EHR could populate their data in OMOP CDM format as part of their existing daily Epic extract-transform-load (ETL) run?

Transforming electronic health record (EHR) data into the OMOP Common Data Model (CDM) enables a health system to participate in valuable Observational Health Data Sciences and Informatics (OHDSI) network studies[1]. Local research and quality improvement initiatives also benefit by leveraging rich OHDSI vocabularies and phenotypes.

However, the requirement to develop and operate a de novo ETL system for transforming Epic EHR data into the OMOP CDM poses a barrier to broader implementation. To address this barrier, shareable ETL solutions have been developed to facilitate wider adoption among health systems on an Epic EHR. These solutions, often designed for and by research units separate from the health system's Epic team, involve either sharing SQL code for implementing a local ETL system[2,3] or hosting a shareable ETL system in the cloud and/or a container[4].

Our Clinical and Translational Science Award (CTSA) hub includes five diverse healthcare organizations on separate Epic EHR instances. Cross-institutional research suffers from significant delays in obtaining and combining data. Yet creating a common research data warehouse with data from all sites proved untenable. Accordingly, we proposed adopting the OMOP CDM for conducting federated research, enabling combined analyses while retaining source data at each site. We aspired to simplify cross-site replication of (1) Epic-to-OMOP transformation logic, (2) terminology mapping, and (3) importantly, the ETL system itself. Specifically, we sought to assess feasibility of leveraging each site's existing Epic Caboodle ETL framework as a shareable method for populating an OMOP CDM.

Methods

Participants and Design

Our CTSA hub includes UT Southwestern (UTSW), Parkland Health, Children's Health, Texas Health Resources, and Scottish Rite[5].

Aligning with CTSA aims, we adopted "design for dissemination" as a guiding philosophy. Accordingly, we sourced data primarily from the Epic Caboodle data warehouse, enabling re-use of Epic ETL logic and of any existing mappings of local source Epic data to standard Caboodle fields. We endeavored to minimize the amount of site-specific local code mapping required, mapping to standard terminologies within source EHR records wherever feasible.

Additional design principles included adhering to Epic's custom Caboodle development standards, segregating any added Caboodle objects to a custom schema (OmopX), and leaving all standard Epic-released ETL code untouched. In this design, OMOP CDM data physically resides within the main Caboodle database, readily linkable to other Health System data for ad hoc queries. Presentation of data to OHDSI tools such as Data Quality Dashboard (DQD) and ATLAS is done through views housed in a companion

OMOP database on the Caboodle server, recasting table and field names to OMOP CDM naming standards. (Figure 1).

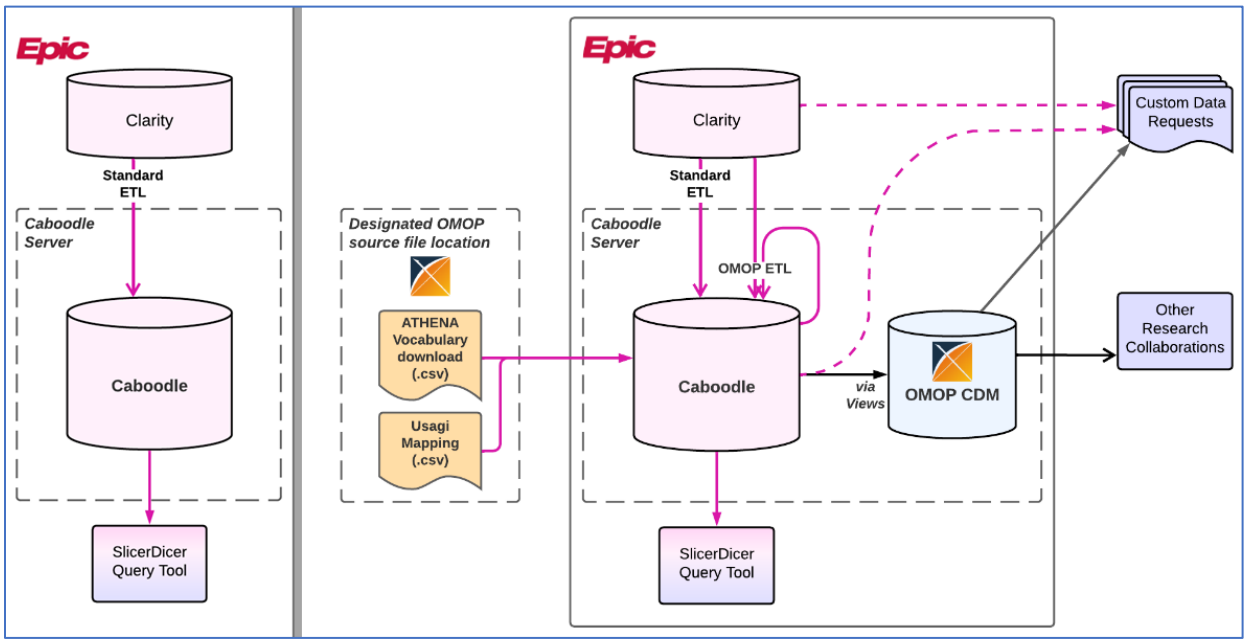

Figure 1: (Left) Pre-OMOP Caboodle ETL data flow; (Right) Additional Caboodle ETL data flows added to populate the OMOP CDM daily with a full set of the health system's EHR data, ready for use with OHDSI tools. Custom data requests can include additional attributes from the EHR found in Caboodle and/or Clarity, when desired to supplement fields in the OMOP CDM.

Development Steps

We reviewed and reused existing ETL logic from well-established Epic-to-OMOP transformations, and applied OHDSI's tools to map Caboodle source tables and fields to the OMOP CDM[6]. We wrote stored procedures to transform Caboodle data to OMOP tables, then developed custom Caboodle Data Model Component (DMC) packages to wrap each stored procedure. Code and database scripts were stored in a version control system. OMOP DMCs were migrated to test and production Caboodle environments, evaluating data contents with DQD checks. Caboodle DMC migration to another health system mirrors the migration process between local environments. Certain fields still require mapping local EHR codes to standard concepts, for which Caboodle queries are provided to expedite discovery.

Software

For initial code development, we employed OHDSI tools (Rabbit in a Hat, Usagi, and DQD), Epic Caboodle as the EHR and ETL infrastructure, Microsoft Team Foundation Server for code source control, and ServiceNow Agile 2.0 for agile project management. Subsequent organizations require only their existing Epic software stack, plus OHDSI's DQD and ATLAS for interacting with the OMOP CDM.

Measures and Outcomes

Our planned main outcomes of interest included:

- The number of OMOP CDM tables implemented as shareable Caboodle DMCs.
- OMOP CDM table sizes.
- Time required for ETL processing.
- Data quality measures and resulting improvements in source data quality.
- Ultimately, time and effort required to implement at an additional Epic site.

Results

We implemented 25 OMOP CDM tables as Caboodle DMCs (10 Vocabulary tables, 3 Health System tables, and 12 Clinical tables). Figure 2 displays the OMOP tables in the Caboodle data dictionary, with an Entity Relationship Diagram of one clinical table.

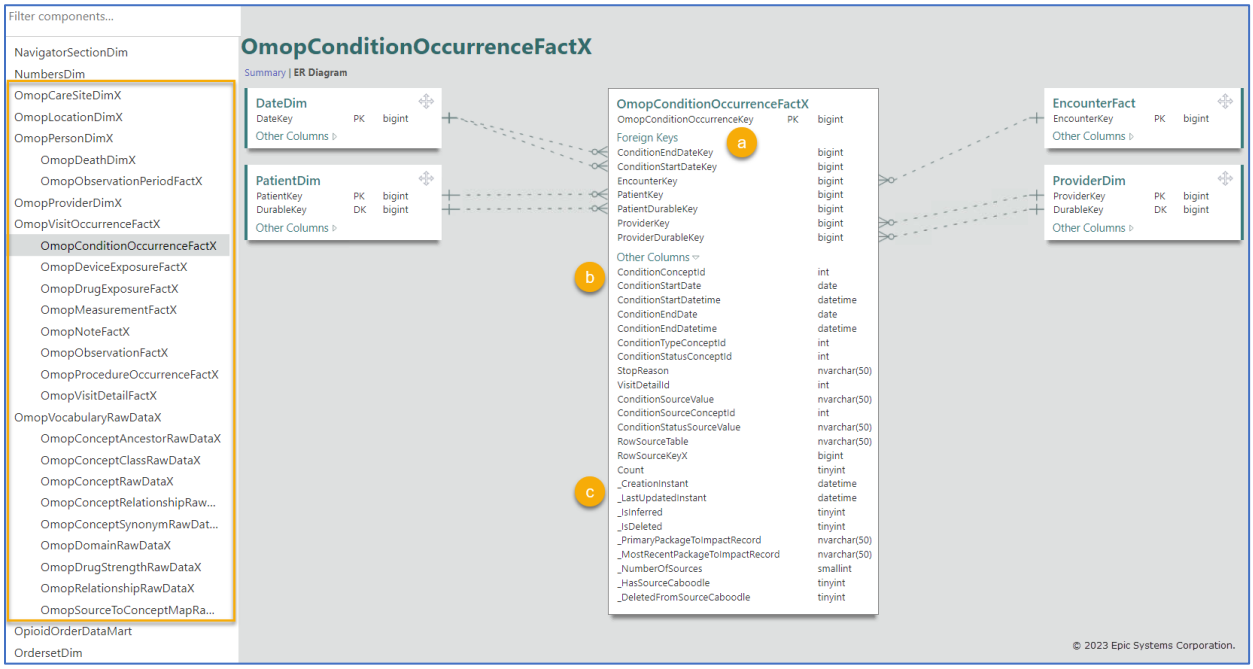

Figure 2: OMOP CDM Tables in the Caboodle Data Dictionary. (Left) List of tables. (Right) An Entity-Relationship Diagram automatically generated by the Caboodle ETL Framework, showing (a) Foreign Key linkages to existing Caboodle tables enabling more detailed queries for custom data requests; (b) OMOP CDM fields (renamed to CDM standard name formats in the corresponding OMOP reporting view); (c) additional metadata columns automatically created and populated by the Caboodle ETL framework, such as timestamps when a given row was initially created and last updated.

Processing time for large tables was optimized by existing Caboodle ETL efficiencies. Despite UTSW's CONDITION_OCCURRENCE table containing >200 million rows, daily processing occurs in under one hour. Diagnoses entered in Epic are available for querying in our production OMOP CDM before 8 am the following day, accurately mapped to OHDSI concepts (Figure 3).

Our efforts identified and addressed several source EHR data quality issues, resulting in improved data accuracy not only in OMOP queries but for all consumers of EHR data.

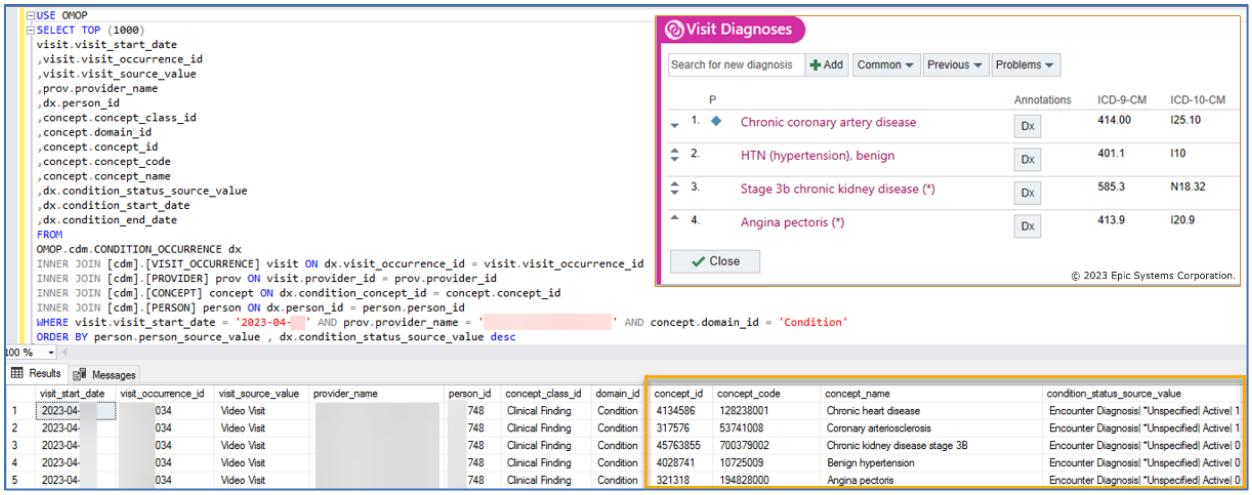

Figure 3: (Upper right): Encounter diagnoses (clinical terms) entered in a 4/2023 Video Visit encounter in Epic. (Rest of screen): SQL query results from the Production OMOP CDM database at 8 am the following day, confirming presence of the 4 encounter diagnoses in CONDITION_OCCURRENCE, accurately mapped to the corresponding 5 OHDSI standard SNOMED CT concepts.

Note: post-coordination of the two SNOMED CT concepts for the source "Chronic coronary artery disease" clinical term is maintained via two added columns (RowSourceTable and RowSourceKeyX) as shown in Figure 2.[7]

Conclusions

Using the Caboodle EDW as the primary data source for Epic data transformation leverages local source field mapping previously accomplished when setting up Epic-released DMCs, and promotes data quality via Epic-provided exception-handling logic.

The Caboodle ETL infrastructure offers valuable capabilities. It delivers a full CDM dataset updated daily, includes ETL job monitoring, allows import of supplemental Clarity fields without modifying Epic-released DMCs, handles patient merges gracefully, leverages existing backup and disaster recovery, enables linking to other EHR data, and aligns with Epic's ongoing investment in reporting technology. With CDM data in Caboodle, optionally Epic's SlicerDicer can be applied for OMOP data exploration (not shown).

The optimal approach to using shareable ETL code for Epic EHR data varies. For OMOP projects with direct access to the Health System's Epic team, employing one's existing Caboodle ETL framework with shared OMOP DMCs can simplify CDM implementation and yield the above benefits. For projects without direct access to the Epic team, utilizing a separate pre-packaged or cloud ETL framework can streamline participation in OHDSI network studies.

For health systems using the Epic EHR, shareable ETL approaches can expedite implementing OMOP by lowering activation effort and expense, in turn promoting broader population representation within the OHDSI network.

References

1. Schuemie MJ, Ryan PB, Pratt N, Chen R, You SC, Krumholz HM, Madigan D, Hripcsak G, Suchard MA. Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study. J Am Med Inform Assoc. 2020 Aug 1;27(8):1268-1277. Doi: 10.1093/jamia/ocaa124. PMID: 32827027; PMCID: PMC7481033.
2. Carlson R. OMOP: The Spectrum Code Reboot [Internet]. Epic UserWeb. Epic Systems; 2022 [cited 2023 Jun 12]. Available from: https://userweb.epic.com/Thread/119330/OMOP-The-Spectrum-Code-Reboot/ (Epic UserWeb account required)

3. The Johns Hopkins University OMOP Team. Introduction — The JHM Epic to OMOP ETL Guideline [Internet]. Pm.jh.edu. 2021 [cited 2023 Jun 12]. Available from: https://pm.jh.edu/omop_etl/html/intro.html
4. Heavner S, Llano T, Wang Z, Schito M, Stone H, Dasher P, Russel T, Kumar V, Saeks B, Cooke B, Kashyap R, Robinson M, Nagy P. Lowering the OMOP ETL Barrier for Clinical Registries – OHDSI [Internet]. OHDSI 2022 Collaborator Showcase. 2022 [cited 2023 Jun 12]. Available from: https://www.ohdsi.org/2022showcase-40/
5. Clinical and Translational Science Award (CTSA) Program [Internet]. www.utsouthwestern.edu. UT Southwestern Medical Center; [cited 2023 Jun 12]. Available from: https://www.utsouthwestern.edu/research/ctsa/
6. Blacketer C, Voss E. Chapter 6 Extract Transform Load | The Book of OHDSI. In: The Book of OHDSI [Internet]. OHDSI; [cited 2023 Jun 12]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html
7. Philofsky M, Olakpe U. How to customize your OMOP CDM to meet internal needs while staying true to OHDSI's mission. OHDSI 2021 Collaborator Showcase. [Online] September 2021. https://www.ohdsi.org/2021-global-symposium-showcase-18/