Patient 360 – A Strategic Application for the OMOP Common Data Model

Rakesh Babu, PharmD
Atlantic Health System, Morristown, NJ

Background

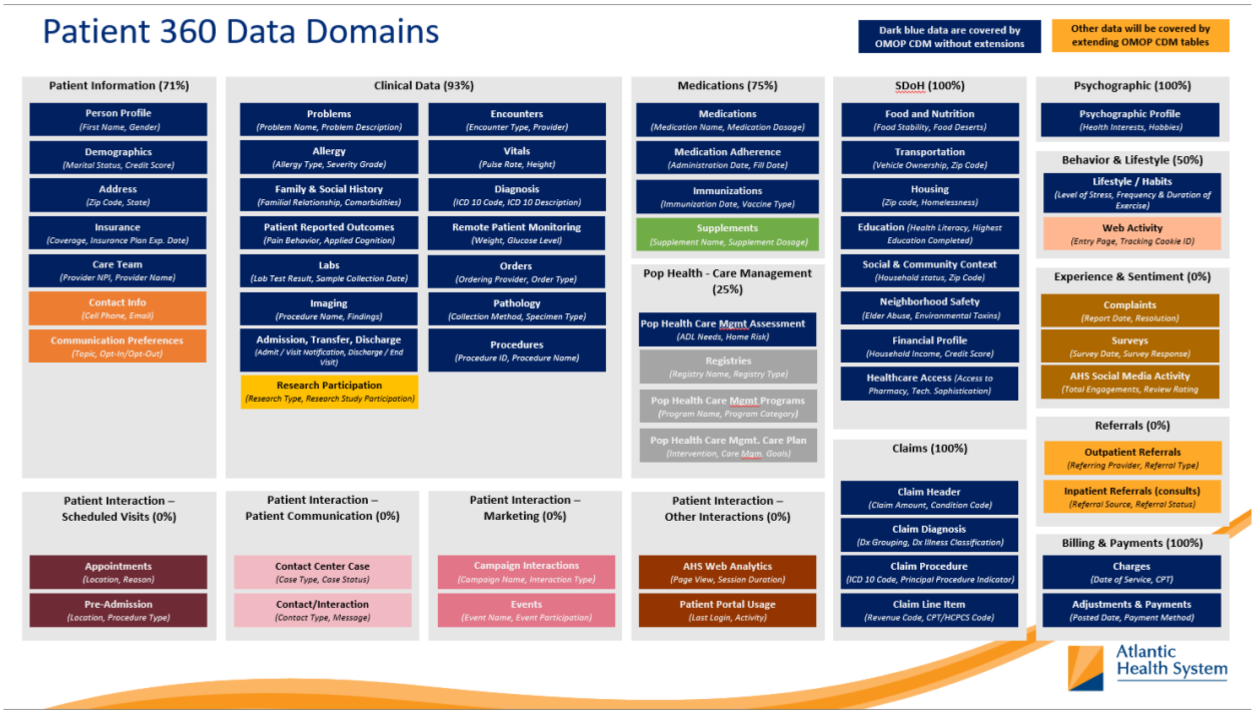
The OMOP Common Data Model (OMOP CDM) provides a standard method of representing healthcare data for the purpose of retrospective outcomes research[1]. The design of the OMOP CDM meets the requirements that data be stored in a standard format, that data be labeled in a standard fashion, and that data be of good quality. However, when using OMOP CDM data for purposes other than retrospective research, additional requirements may be identified.

Atlantic Health System (AHS), a not-for-profit health system based in New Jersey, identified a need to establish several datasets to support strategy and planning analyses. The most complicated of these datasets is Patient 360 (P360), a patient-centered wholistic analytic dataset intended to capture all patient data stored throughout the enterprise. AHS evaluated using the OMOP CDM as a base dataset for P360, and identified several additional requirements: That the dataset be updated daily; that the dataset be able to accommodate multiple overlapping data sources, and that the dataset be extensible to support data not currently in scope for the OMOP CDM.

In this project, we review the novel architecture adopted by AHS to build P360 using the OMOP CDM as a base.

Methods

Prior to this project, AHS engaged in a 3 month process to identify the required scope for P360. The process involved interviewing numerous stakeholders and leaders throughout the organization to collect use cases for P360. These use cases were analyzed to group the data requirements into a set of 15 data domains, summarized in figure 1. During this process, we also identified the requirements for which data sources need to be integrated into P360 and how current the data in P360 needs to be.



Patient 360 Data Domains

Dark blue data are covered by OMOP CDM without extensions

Other data will be covered by extending OMOP CDM tables

Patient Information (71%)
- Person Profile (First Name, Gender)
- Demographics (Marital Status, Credit Score)
- Address (Zip Code, State)
- Insurance (Coverage, Insurance Plan Exp. Date)
- Care Team (Provider NPI, Provider Name)
- Contact Info (Cell Phone, Email)
- Communication Preferences (Topic, Opt-In/Opt-Out)

Clinical Data (93%)
- Problems (Problem Name, Problem Description)
- Encounters (Encounter Type, Provider)
- Allergy (Allergy Type, Severity Grade)
- Vitals (Pulse Rate, Height)
- Family & Social History (Familial Relationship, Comorbidities)
- Diagnosis (ICD 10 Code, ICD 10 Description)
- Patient Reported Outcomes (Pain Behavior, Applied Cognition)
- Remote Patient Monitoring (Weight, Glucose Level)
- Labs (Lab Test Result, Sample Collection Date)
- Orders (Ordering Provider, Order Type)
- Imaging (Procedure Name, Findings)
- Pathology (Collection Method, Specimen Type)
- Admission, Transfer, Discharge (Admit / Visit Notification, Discharge / End Visit)
- Procedures (Procedure ID, Procedure Name)
- Research Participation (Research Type, Research Study Participation)

Medications (75%)
- Medications (Medication Name, Medication Dosage)
- Medication Adherence (Administration Date, Fill Date)
- Immunizations (Immunization Date, Vaccine Type)
- Supplements (Supplement Name, Supplement Dosage)

Pop Health - Care Management (25%)
- Pop Health Care Mgmt Assessment (ADL Needs, Home Risk)
- Registries (Registry Name, Registry Type)
- Pop Health Care Mgmt Programs (Program Name, Program Category)
- Pop Health Care Mgmt. Care Plan (Intervention, Care Mgm. Goals)

SDoH (100%)
- Food and Nutrition (Food Stability, Food Deserts)
- Transportation (Vehicle Ownership, Zip Code)
- Housing (Zip code, Homelessness)
- Education (Health Literacy, Highest Education Completed)
- Social & Community Context (Household status, Zip Code)
- Neighborhood Safety (Elder Abuse, Environmental Toxins)
- Financial Profile (Household Income, Credit Score)
- Healthcare Access (Access to Pharmacy, Tech. Sophistication)

Claims (100%)
- Claim Header (Claim Amount, Condition Code)
- Claim Diagnosis (Dx Grouping, Dx Illness Classification)
- Claim Procedure (ICD 10 Code, Principal Procedure Indicator)
- Claim Line Item (Revenue Code, CPT/HCPCS Code)

Psychographic (100%)
- Psychographic Profile (Health Interests, Hobbies)

Behavior & Lifestyle (50%)
- Lifestyle / Habits (Level of Stress, Frequency & Duration of Exercise)
- Web Activity (Entry Page, Tracking Cookie ID)

Experience & Sentiment (0%)
- Complaints (Report Date, Resolution)
- Surveys (Survey Date, Survey Response)
- AHS Social Media Activity (Total Engagements, Review Rating

Referrals (0%)
- Outpatient Referrals (Referring Provider, Referral Type)
- Inpatient Referrals (consults) (Referral Source, Referral Status)

Billing & Payments (100%)
- Charges (Date of Service, CPT)
- Adjustments & Payments (Posted Date, Payment Method)

Patient Interaction – Scheduled Visits (0%)
- Appointments (Location, Reason)
- Pre-Admission (Location, Procedure Type)

Patient Interaction – Patient Communication (0%)
- Contact Center Case (Case Type, Case Status)
- Contact/Interaction (Contact Type, Message)

Patient Interaction – Marketing (0%)
- Campaign Interactions (Campaign Name, Interaction Type)
- Events (Event Name, Event Participation)

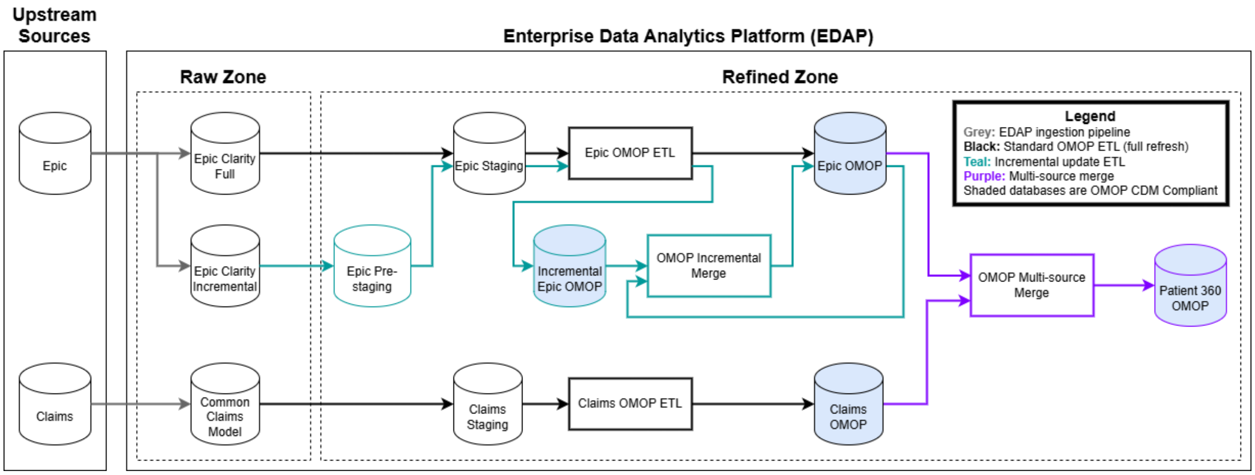
Patient Interaction – Other Interactions (0%)
- AHS Web Analytics (Page View, Session Duration)
- Patient Portal Usage (Last Login, Activity)

Atlantic Health System

We then engaged in a process to design an architecture to support the requirements for Patient 360. At the outset of this process, we decided to limit the scope of the engagement to 2 datasets: the electronic medical record (Epic), and a post-adjudicated insurance claims dataset (Claims). At this time, we decided that we would attempt to create multiple complete OMOP CDM conversions, then merge them together into one master dataset. We determined that we needed an incremental update process due to the size of the CDM's. When scoping the initial implementation, we decided to avoid populating the notes tables and the visit details table for time and cost reasons. We determined that 3 additional routines not generally implemented as part of an OMOP CDM conversion would be necessary: a pre-stage process to identify records that would be in-scope to build an incremental dataset; an upsert process to update and insert records as part of the incremental dataset; and a merge process to combine information from multiple complete OMOP CDM's.

The architecture is designed to work with the preexisting orchestration engine in use at AHS (Apache Airflow). Airflow runners are configured to fetch the SQL files developed as part of this project and run them in the correct order, with the supplied database targets depending on whether the desired action is to fully refresh the dataset, to incrementally update the dataset, or to merge multiple OMOP CDM's together into one master dataset.

Results



The final data flow diagram for the project is shown in Figure 2. There are 4 data flows in this diagram: A grey line that indicates a standard data lake ingestion process that is out of scope for this project; a black line that indicates a standard OMOP CDM Conversion ETL process; a teal line that indicates an incremental update process; and a purple line that indicates a CDM Merge process.

The grey line indicates that both the Epic and Claims sources are loaded into the Data Lake Raw Zone using existing ETL packages. For Epic data, a Change Data Capture process loads daily incremental files into a separate schema. The black lines for Epic and Claims detail the same complete conversion process; from the Raw zone, required tables are copied into a staging schema, then are passed into an ETL package that generates an OMOP CDM-compliant dataset. The teal line shows the incremental load process. The Change Data Capture files are first scanned to find all the new clinical events that transpired since the last CDM update; records related to those facts are placed into a pre-staging schema. Based on those facts, all records from the source that are required to generate an internally consistent OMOP CDM are also placed into the pre-staging schema. These records are then put through the same ETL package that is used for the complete conversion process, and saved into an incremental schema. An incremental merge process then upserts those records into the existing OMOP CDM schema. Finally, the purple line describes the multi-source merge process. All the facts and all the dimensions from each of the source CDM schemas are copied into the final schema.

There were several challenges we needed to resolve in order to make this process possible. The first relates to the incremental upsert. Most OMOP CDM's use a nonsemantic serial number for the primary keys for each table (e.g. person.person_id). However, this approach would make it impossible to identify which records need to be updated when running the incremental upsert process. To solve this problem, we used the Redshift built-in hash function FarmFingerprint64 to generate stable identifiers for all primary keys.

A similar challenge awaited us when attempting to deduplicate patients; because Claims and Epic use different person identifiers, the hashed person_id fields still wouldn't match. To resolve this, we used a person matching service to identify the same person across datasets.

There are several opportunities to continue to improve this ETL process. We do not currently have a process to handle hard or soft deletes as part of the incremental update process. Currently, we run periodic full refreshes to remove any deleted records that have not been cleared from the dataset. At the current time, it is possible for the same clinical fact to be represented in the Epic and Claims datasets, and therefore be represented in the dataset twice. This would make counting the number of occurrences of an event (e.g. admissions) challenging for analysts. In the future, we are planning to pursue a project to identify and resolve such duplication. Finally, we more data sources that we would like to integrate with P360.

Conclusion

The OMOP CDM is a powerful tool for retrospective research, and can be extended to support strategy and planning needs. We have presented a novel approach to adding incremental update and multiple source integration capabilities to existing standard OMOP CDM conversions. This approach retains the considerable benefits provided by the OHDSI community, by ensuring that the datasets produced are OMOP CDM compliant. Further research is required into deduplicating clinical events across multiple sources.

References

1. Observational Health Data Sciences and Informatics, *The Book of OHDSI*.