Name:	ZS Associates
Affiliation:	ZS Associates
Email:	harshal.soni@zs.com; ajinkya.patale@zs.com
Presentation type (select one):	Lightening Talk

Demonstrate Reusability and Effectiveness of Using an Intermediate Layer for OMOP CDM Conversion

ZS Associates, Evanston

Abstract

The OMOP CDM data model helps standardize healthcare data (EMR, claims, registries, etc.) and makes it easier to analyze outcomes at a large scale. To be able to use OMOP CDM data for analysis, we have to convert all the data sets into a CDM format, which is currently a hefty process that involves substantial technical and domain expertise as it contains many complex operations. Through this document, we have tried to simplify the traditional conversion process by introducing an intermediate layer between source data set and the OMOP CDM model.

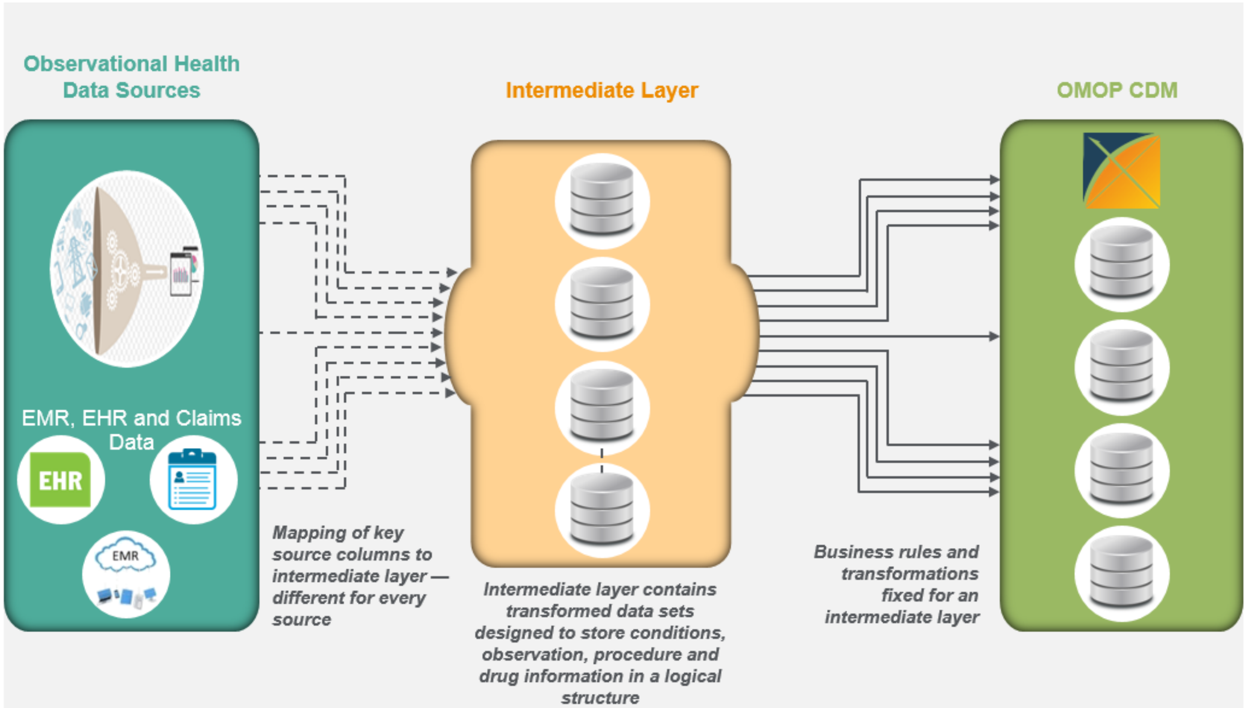
Introduction

We have converted multiple claims and EMR data sources into CDM format per our client's needs and analytics requirements. While converting these data sources in various technologies (RDBMS, Hadoop, Redshift), we realized that to develop this conversion process in a tool, we must present a strong methodology that allows us to make our code or product modular and highly scalable, and attain maximum reusability. The introduction of an intermediate layer in OMOP conversion is an example of this. It allows us to make our codes modular and highly scalable and gives us a chance to reuse an existing mapping for CDM conversion.

Benefit of an Intermediate Layer in OMOP Conversion

It helps to reduce the overall OMOP conversion time for RWE data sources. Enables self-service for making minor adjustments or fixes to the conversion rules to support specific client needs. Enables investment in a scalable solution that can address future business needs in a cost-effective manner.

How Does an Intermediate Layer Achieve Reusability and Flexibility in OMOP Conversion?



Fundamentally, source-to-intermediate layer transformation focuses on data and terminology transformation. It means all the columns from multiple tables are converted to single format and named per the OMOP terminology. Additionally, data scattered across multiple tables is converged into a single intermediate layer. In the intermediate-to-target layer transformation, we perform some additional transformations, generate unique IDs for OMOP tables and populate these IDs in other OMOP tables as foreign keys. We also combine the vocabulary with the target data to give it clinical relevance.

Thus, intermediate-to-target layer mappings are generic in nature and can be reused for new data sources. For every new data source, we only need to map the source data set to the intermediate layer. ZS validated the concept of an intermediate layer by executing conversion of multiple data sources using the intermediate layer and reusing the *intermediate-to-target layer* mappings for those data sources. We have found that through this measure, the effort estimates for converting a data set from an intermediate layer dropped to almost 50%. This approach also tends to provide more flexibility and modularity over the complete solution and supports high-volume data sets well.

Conclusion

ZS has reduced almost 50% of its CDM conversion effort by using the intermediate layer. We hope to share our findings with the community to make CDM conversion more efficient for everyone.