

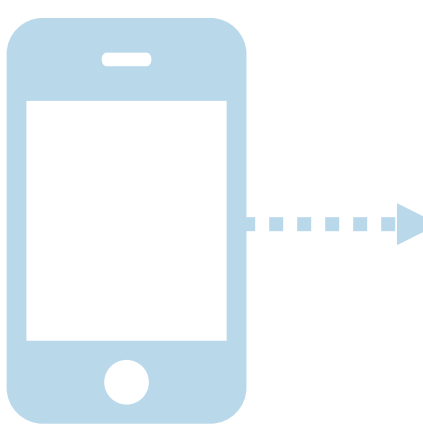


Approaching Data Model Interoperability between OMOP and PCORnet

Shweta P Chavan, Hanieh Razzaghi, Charlie Bailey, Ritu Khare, Childrens Hospital Of Philadelphia, Philadelphia, PA



Take a picture to
download the full paper



Background

PEDSnet, a network of seven of the nation's largest children's hospitals, contains clinical data on 6.2 million children. The PEDSnet Common Data Model (CDM) is an evolving specification, based on the structure of the OMOP CDM, but expanded to accommodate requirements of both the PCORnet CDM and the primary research cohorts established in PEDSnet. To participate in PCORnet, each partner network maps data to the conventions specified in the PCORnet CDM. The PEDSnet Data Coordinating Center (DCC) has developed a seamless process to transform data from OMOP to PCORnet.

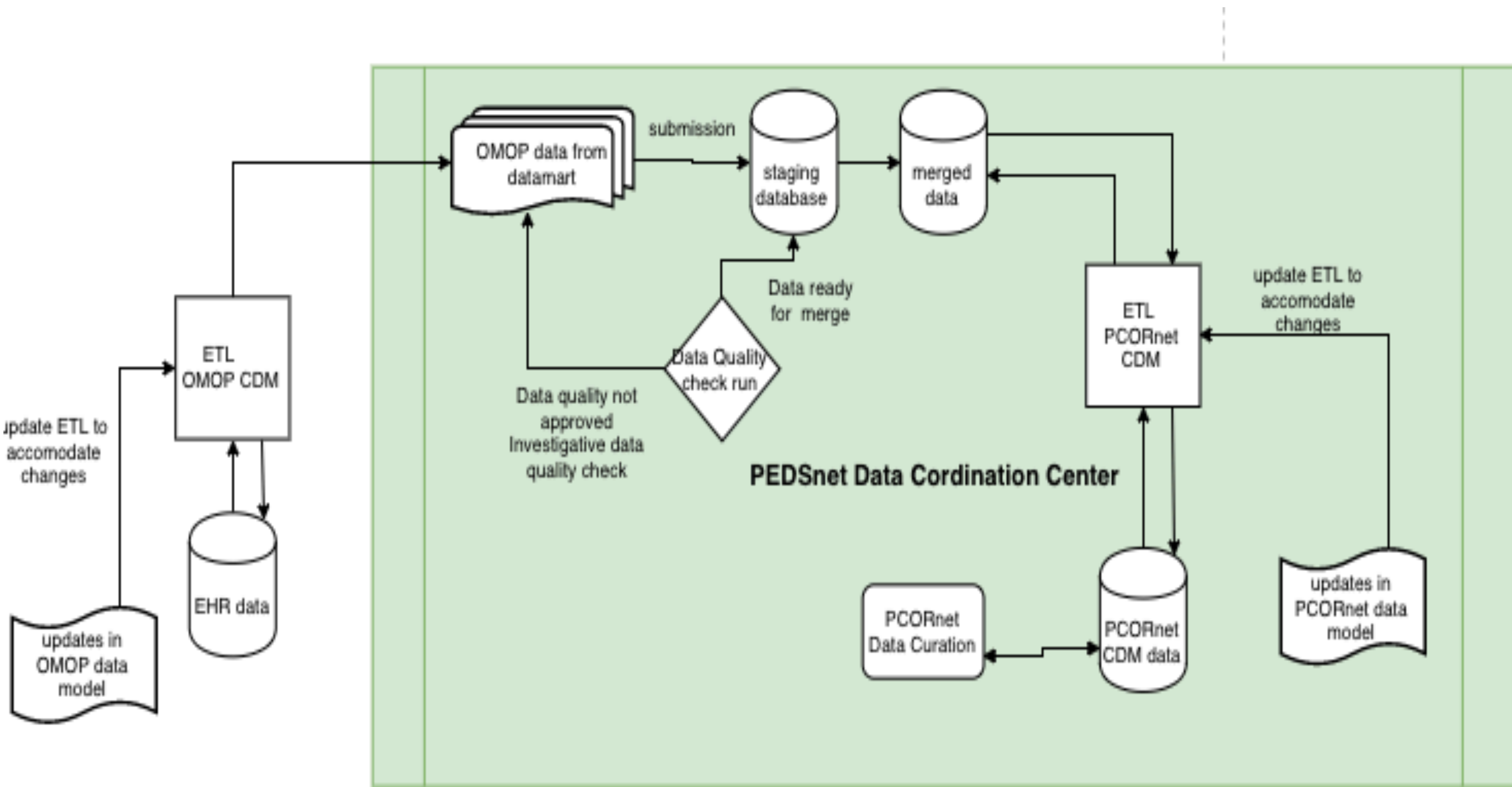


Figure 1: PEDSnet pipeline

Methods

OMOP CDM is the broader CDM. This helps captures lots of information regarding the person. Both PEDSnet and PCORnet data are refreshed quarterly, allowing PEDSnet the ability to adopt the changes and update the ETL process accordingly. PEDSnet has developed a Python CLI app to automate the ETL process for PCORnet transformation. PEDSnet transforms data from OMOP to the PCORnet CDM with minimal loss of source data and is maintained securely behind PEDSnet firewalls to support cross-network querying. Quarterly in the PCORnet network are implemented in the ETL of OMOP to PCORnet. The PCORnet instance of PEDSnet data has been approved by PCORnet for every data cycle. Below is the data cycle overview for PEDSnet. Python and Sqlalchemy was used to develop the ETL application. This ddl is called in the python tool to create the table structures for the CDM. Once the DDL is creates command line python tool is used to upload the valueset map. The ETL scripts are based on SQL scripts. As PEDSnet uses PostgreSQL for the database, the ETL scripts are in SQL format that are called using the python tool.

Results

PCORnet has specified values for representing several of their data variables. We created a crosswalk file that translates PCORnet ontologies to OMOP vocabulary. Figure below shows table structure of valueset:

source_concept_class	source_concept_id	source_description	target_concept	pcornet_description
gender	8532	Female	F	Female
gender
specimen source	4183746	Tissue specimen obtained from gallbladder by biopsy	TISSUE	Tissue
specimen source

Some concepts in OMOP have direct mapping to the PCORnet ontology like gender, race, dose unit, etc., while other like specimen source, facility type needed to develop a text mining tool to standardize the values and manual testing to create the crosswalk.

There are some commonalities in OMOP and PCORnet CDM. Some of the table in PCORnet CDM like demographic, death, procedure, and enrollment can be directly extracted from OMOP person, death, procedure_occurrence, observation_period tables respectively. There exist several domains where the PCORnet and PEDSnet CDMs are structured differently and the transformation must be designed to accommodate this. For example, PCORnet has three different tables to represent drug data; one for drug data, prescribing drug data and administered drug data, while OMOP CDM captures drug data in one table drug_exposure using drug_type_concept_id to specify the setting. Transforming the drug_exposure to PCORnet CDM dispensing, prescribing and medadmin requires splitting of drug_exposure table in drug_type_concept_id domain.

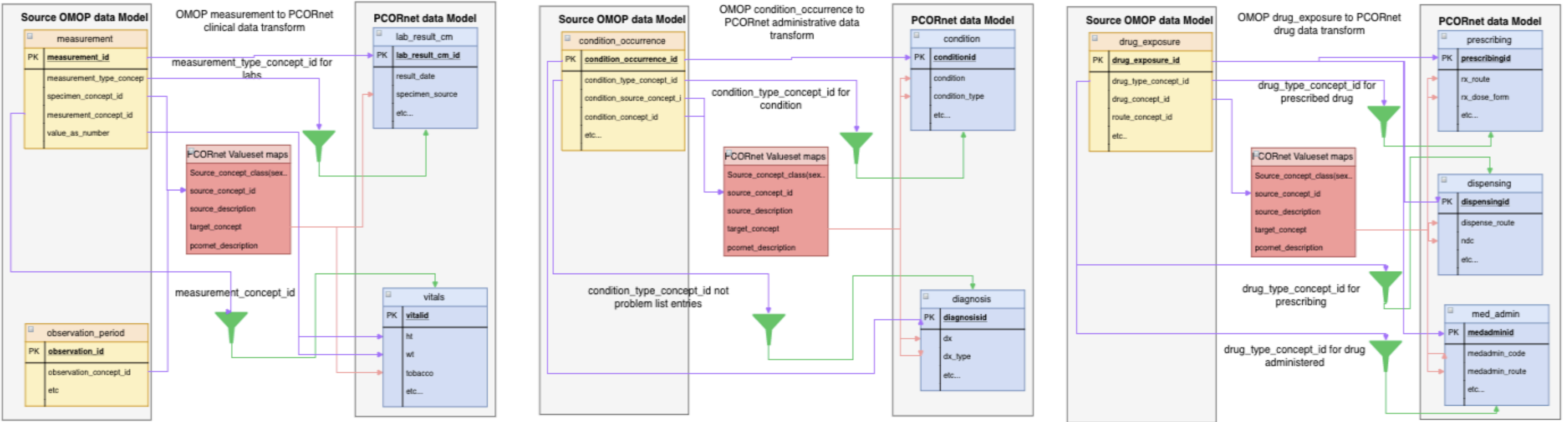


Figure 3: Transformation of tables

Python CLI app

Python ETL application and changes in the DDL are maintained on GIT hub repo and extracted to create the DDL. The tool efficiently transforms data in short period of time.

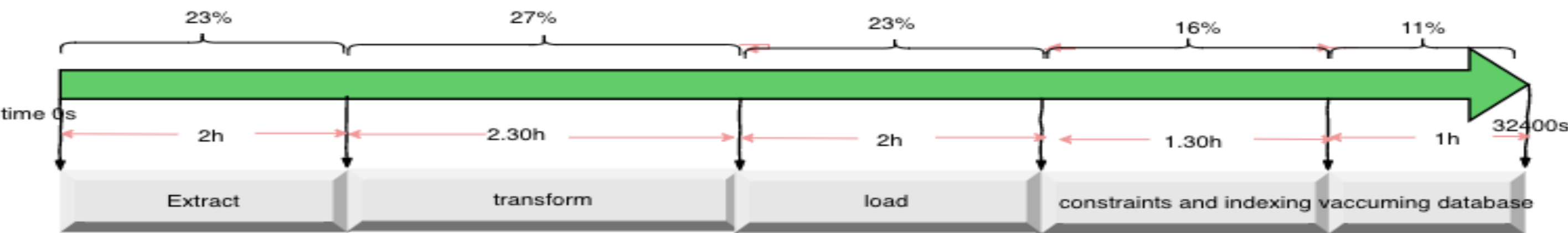


Figure 4: PCORnet ETL Tool process times

Conclusions

PCORnet CDM being less granular, we can extract data from OMOP CDM with minimal loss of the source data. We have developed an efficient process to create interoperability between the OMOP and PCORnet CDMs.