



# A Case Study of Translational Research from a Specific Data Model to a Standard Data Model

Martha Michel, MS Ph.D<sup>1</sup>., Rhiannon Croci RN, Dana Ludwig MD<sup>1</sup>, Vijay Rayanker<sup>1</sup>, Bhuwan Karki<sup>1</sup>, Mark Khayter<sup>2</sup>, Don Torok<sup>3</sup>, Doug Berman, MBA<sup>1</sup>

1. Academic Research Systems, UCSF, San Francisco, CA, 2. Ephir, Boston, MA 3. Observational Medical Outcomes Partnership, Boston, MA

## Abstract

Ordinary common data model or OMOP was created to provide researchers and analysts with ways of representing health data with minimal information lost. The pSCANNER project is an implementation of the OMOP model across a network of institutions including the five University of California medical campuses. UCSFs, one of the largest hospitals in the UC system with 722 beds and 763,000 outpatient visits, implementation of OMOP into PSCANNER illustrated important data modeling decisions about data translation. These issues affect the scope of research questions and ultimately the answers that researchers will receive from this resource. Design of the mapping requires a thoughtful approach to translating data from operational EMRs into a research resource. Issues arise that are neither intuitive nor obvious.

## pSCANNER Goal:

To provide infrastructure for an interoperable, learning healthcare system in which data collected for care are reused for patient-centered outcomes research, with full engagement by patients, clinicians, and administrators in the governance of the consortium

## Institutions:

Pscanner is made of 3 networks that place together a diverse group of health settings including :

1. the VA Informatics and Computing Infrastructure (VINCI) with 151 inpatient, 909 ambulatory care clinics, and community based health clinics
2. the University of California Research Exchange UCRESX
3. Scanner a consortium between UCSD, Tennessee VA, and three federally qualified health systems in the Los Angeles area.

## 3 case studies for research:

- CHF (10000 cases)
- Obesity (10000 cases)
- Kawasaki disease (1000 cases)

## OMOP data model

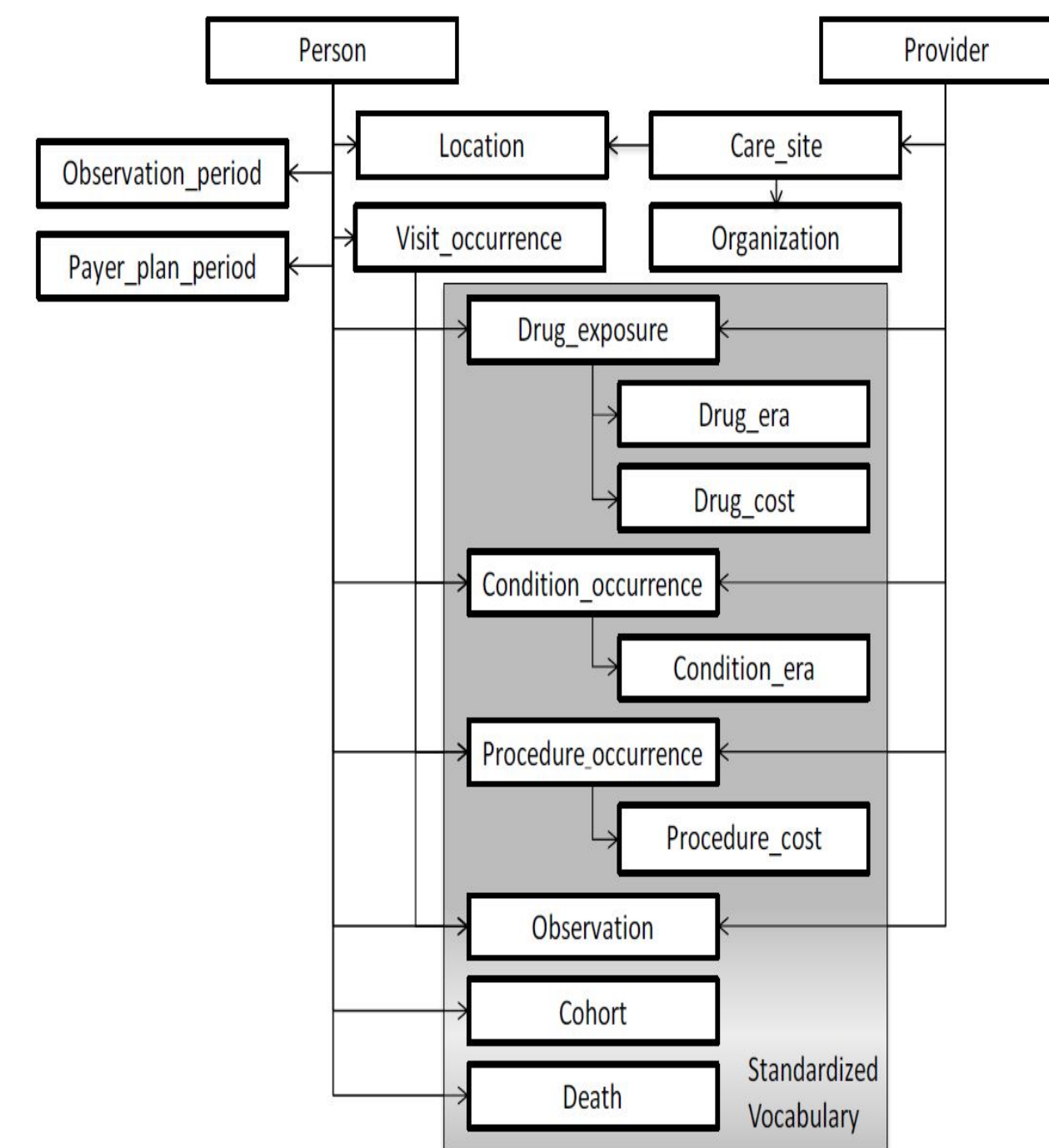


Figure: COMOP common data model 4.0

## Source Data:

Data only includes data from our UCSF EPIC EHR Clarity database. First rollout of EPIC, to Ambulatory clinics only, started April 2010. EPIC Barcode Medication Administration, initiated February 2012. Ambulatory Clinic and Hospital data included in Clarity source data. Legacy system data included in Clarity source data for encounters, labs, procedures, measurements, and notes. Medication order and administration information is not available prior to March 2011. Between 3/2011 and 2/2012, medication information will only be related to Ambulatory visits.

	Present in EHR	Absent in EHR
Present in OMOP	Proceed to harmonization with other institutions ex: ser.Prov_id to Prov_id	Document and note the type of null value for harmonization (enrollment)
Absent in OMOP	Gap in OMOP – propose to governance board	Document and revise research protocol

## Observations:

- Decision to leave out most of flow sheet data for feasibility (large data) but may affect the questions researchers are able to answer which affects the timeline of research projects.
- Decisions not to collect or harmonize data from mother/baby at infancy may affect longitudinal research questions for the 3 case studies;

## Pcornet Summary Results from UCSF

	Demographics	Encounters	Vitals	Diagnosis	Procedures
ALL PATIDs (n)	3,238,111	22,227,229	4,281,037	16,879,204	36,937,395
UNIQUE PATIDs(n)	3,238,111	1,883,021	579,212	461,463	1,218,205

## Conclusion:

To have an effective translation of one data model to another involves careful coordination with a large group of people. The researchers need informaticists experienced in both research and technical implementation. Tools such as shared websites and White Rabbit make the documentation process easier as well as frequent communication between ETL and technical teams with the principal investigators.

## Acknowledgements:

This work is funded by PCORI contract CDRN-1306-04819 and supported by CTSI.