MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Centre for Health Analytics

@kidsanalytics

Transforming Data from an EMR into the OMOP Common Data Model: Goals, Approach and Challenges

Ezra Fass

Senior Project Manager

Centre for Health Analytics

OMOP Transformation: Focus Points

- Centre for Health Analytics
- What is OMOP
- OMOP in Australia
- Centre for Health Analytics objectives with OMOP
- Proof of Concept and Emergency Department Pilot
- Next Steps
- Keeping Data Safe

Centre for Health Analytics

Unleashing the power of data to improve health

healthanalytics.org.au

Melbourne Children's

A world leader in child and adolescent health

The Royal Children's Hospital Melbourne

murdoch children's research institute

The University of Melbourne

Supported by The Royal Children's Hospital Foundation

Data services

Data systems

Workforce

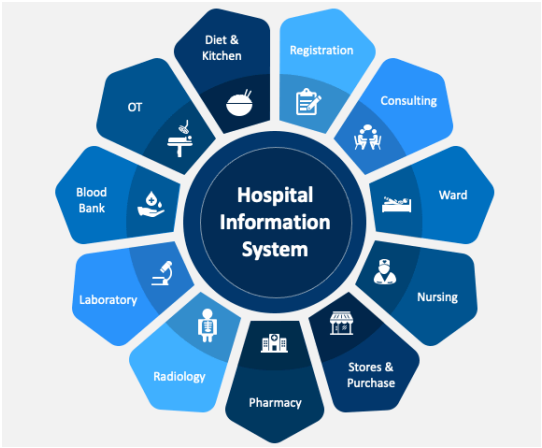
Interoperability

Research

Engagement

Data Systems: Background

- Currently, data across the Melbourne Children's Campus remains largely **siloed** within in transactional data systems.
- Pathways to **access** this data are **complex**, and the data often require manual extraction by technical personnel
- Once extracted, further **specialist** insight is necessary to **interpret** the data to make it useful.



Data Systems: Goal

- Deliver a **data platform** (Health Informatics Virtual Environment - HIVE) for the campus that enables the increased use of quality data, and provides the Campus with secure, replicable, “self-service” pathways to access, analyse and visualise data.
- Starting Point Data Source: RCH EMR (Epic)

OMOP Overview

- The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics.

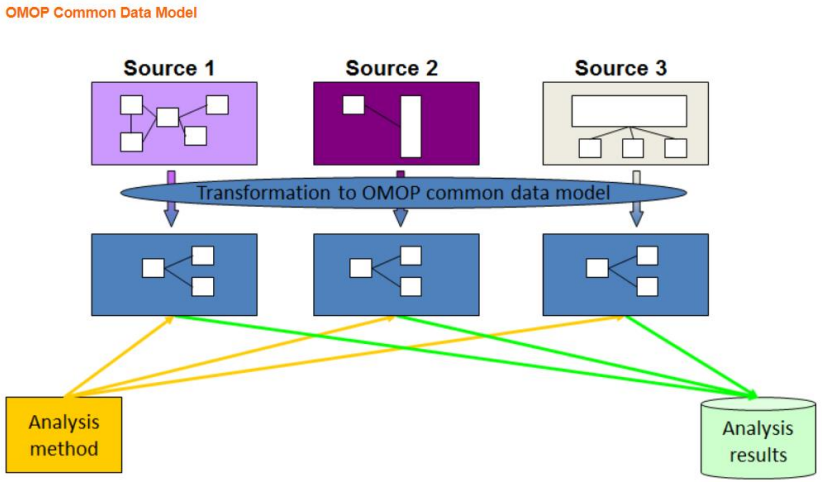
- To realise those goals, the program maintains and promotes the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The OMOP CDM is an open community data standard, designed to standardise the structure and content of observational health data and to enable efficient analyses that can produce reliable evidence.

https://www.ohdsi.org/
https://ohdsi.github.io/CommonDataModel/

OMOP: Common Data Model

OMOP CDM Combines

- Standard Format
- Common vocabulary
- Allows for Shared and Standardised Analysis



What is the OMOP Common Data Model (CDM)?

The OMOP Common Data Model allows for the systematic analysis of disparate observational databases. The concept behind this approach is to transform data contained within those databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes), and then perform systematic analyses using a library of standard analytic routines that have been written based on the common format.

OMOP: Common Data Model

Model helps to harmonise data from different types of observational data / systems, e.g.:

- EMR
- Patient Administration data
- Clinical Registries
- Study Data
- Etc.

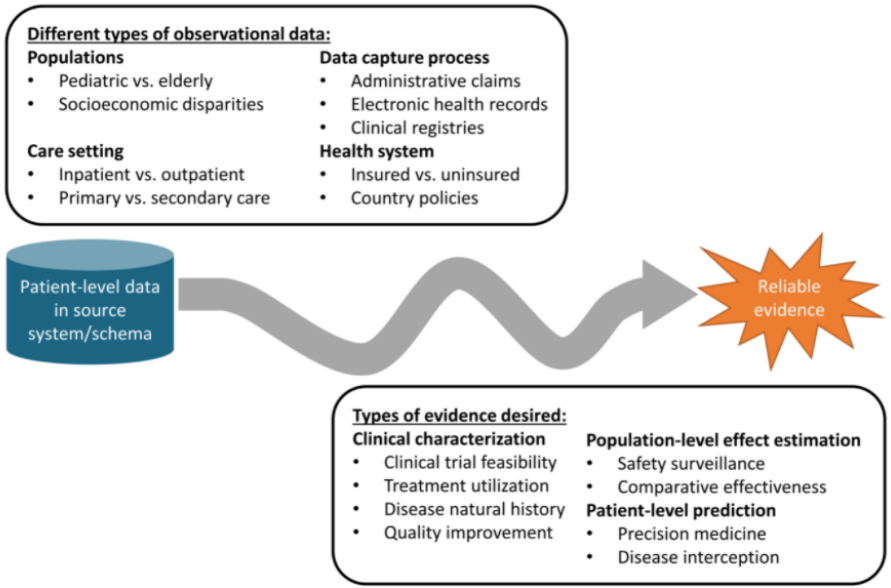


Figure 1.1: The journey from data to evidence

OMOP: Common Data Model

EPIC = 1000s of tables

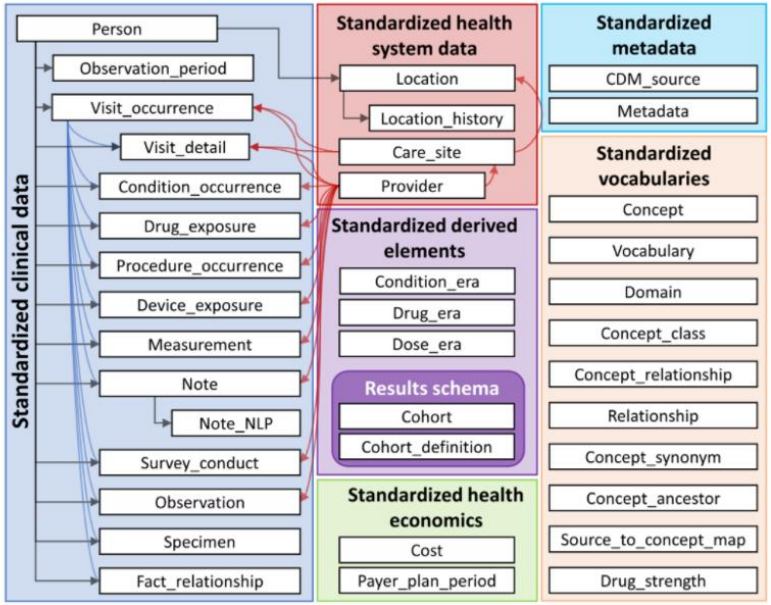
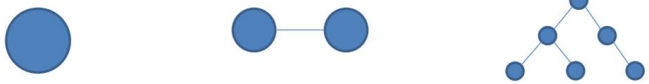


Figure 4.1: Overview of all tables in the CDM version 6.0. Note that not all relationships between tables are shown.

OMOP: Common Data Model

Structure of OMOP Vocabulary



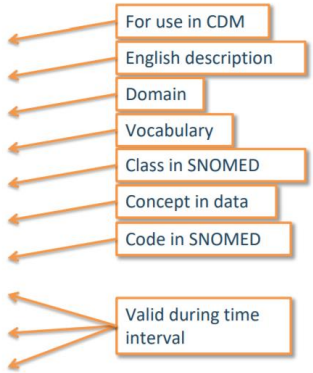
All content: concepts in **concept**

Direct relationships between concepts in **concept_relationship**

Multi-step hierarchical relationships pre-processed into **concept_ancestor**

What's in a Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



OMOP: Common Data Model

Standard Format and Common vocabularies Leads to:

- Standardised and Shareable Analysis

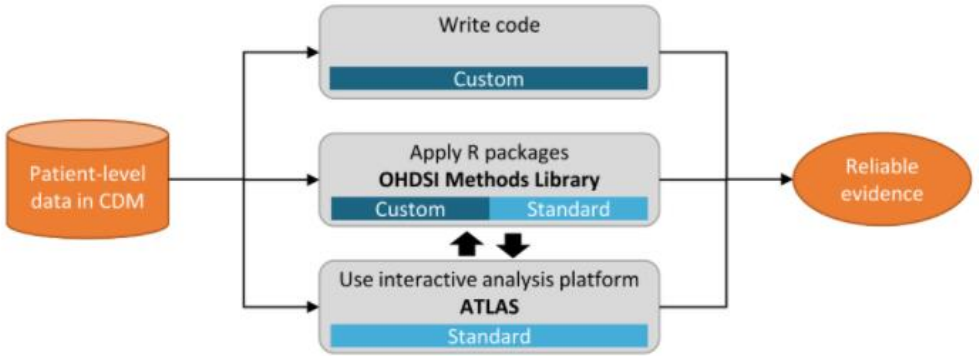


Figure 8.1: Different ways to implement an analysis against data in the CDM.

OMOP: Common Data Model

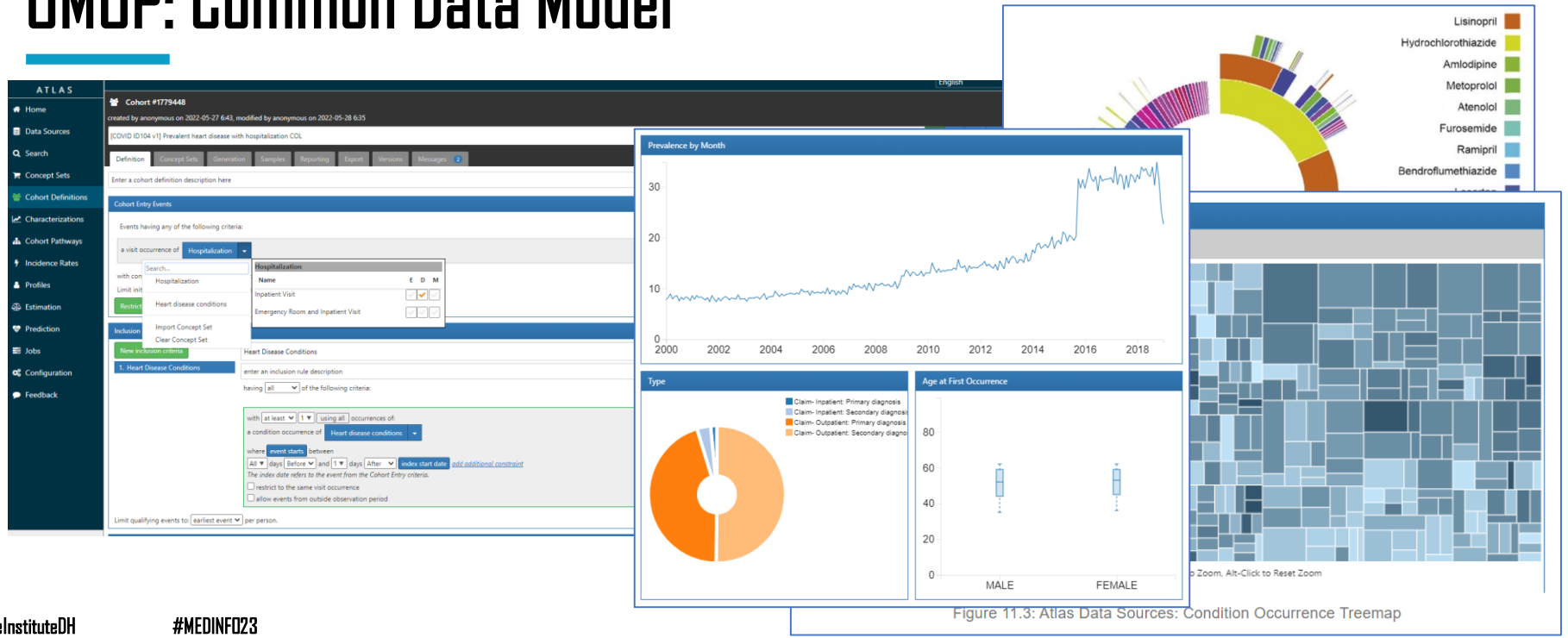


Figure 11.3: Atlas Data Sources: Condition Occurrence Treemap

OMOP: Common Data Model

OHDSI analytical methods

- Characterization
  - What proportion of patients…?
  - What is the distribution of values for lab…?
  - What are the HbA1c levels for patients with…?
  - What are the trends over time in…?
  - What are other drugs that these patients are using?
- Population-level estimation
  - What is the effect of…?
  - What if I do intervention…?
  - Which treatment works better?
- Patient-level prediction
  - What is the chance that this patient will…?
  - Who are candidates for…?

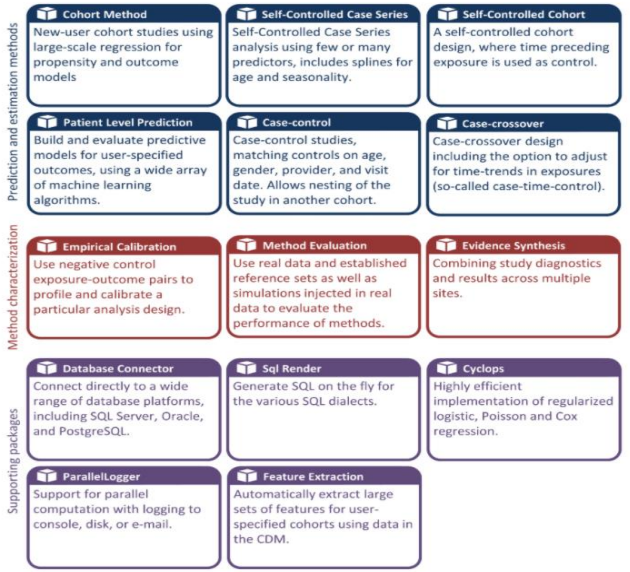


Figure 8.4: Packages in the OHDSI Methods Library.

OMOP Adoption

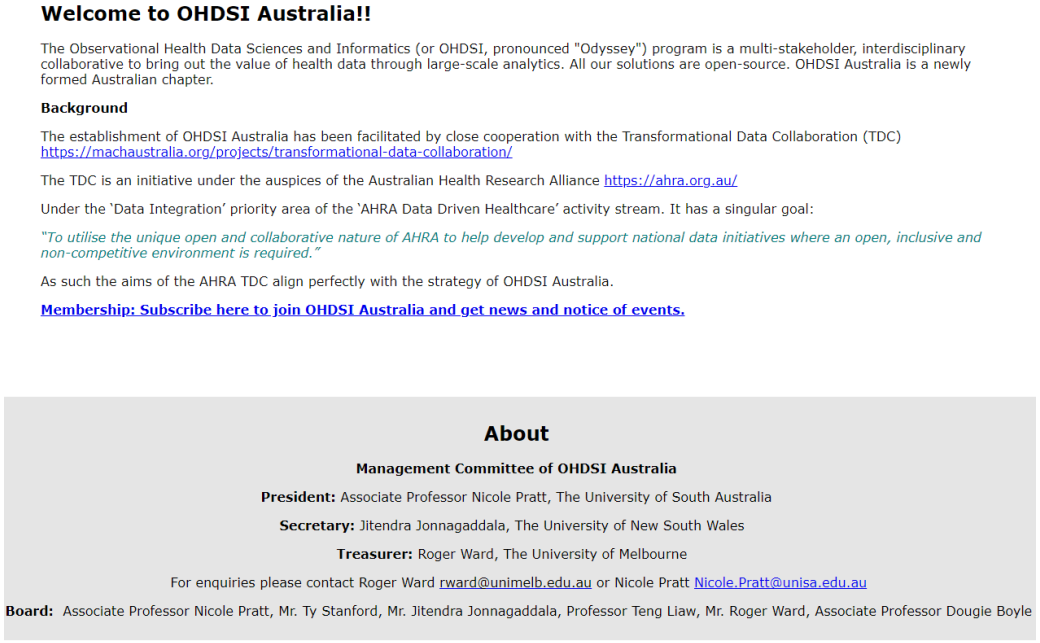
- Global Community
- Expanding
- Active

Figure 1.2: Map of OHDSI collaborators as of August, 2019

As of August, 2019, OHDSI has also established a data network of over 100 different healthcare databases from over 20 countries, collectively capturing over one billion patient records by applying a distributed network approach using an open-community data standard it maintains, the OMOP CDM. A distributed network means that patient-level data are not required to be shared between individuals or organizations. Instead, research questions are asked by individuals within the community in the form of a study protocol and accompanied by analysis code that generates evidence as a set of aggregated summary statistics, and only these summary statistics are shared amongst the partners who opt to collaborate in the study. With the OHDSI distributed network, each data partner retains full autonomy over the use of their patient-level data, and continues to observe the data governance policies within their respective institutions.

OMOP Australia

- https://ohdsi-australia.org/
- OHDSI Australia is part of the OHDSI Asia-Pacific (APAC) community, comprising seven regional chapters (Australia, China, India, Japan, Singapore, South Korea, Taiwan)



Welcome to OHDSI Australia!!

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source. OHDSI Australia is a newly formed Australian chapter.

Background

The establishment of OHDSI Australia has been facilitated by close cooperation with the Transformational Data Collaboration (TDC) https://machaustralia.org/projects/transformational-data-collaboration/

The TDC is an initiative under the auspices of the Australian Health Research Alliance https://ahra.org.au/

Under the 'Data Integration' priority area of the 'AHRA Data Driven Healthcare' activity stream. It has a singular goal:

"To utilise the unique open and collaborative nature of AHRA to help develop and support national data initiatives where an open, inclusive and non-competitive environment is required."

As such the aims of the AHRA TDC align perfectly with the strategy of OHDSI Australia.

Membership: Subscribe here to join OHDSI Australia and get news and notice of events.

About

Management Committee of OHDSI Australia

President: Associate Professor Nicole Pratt, The University of South Australia

Secretary: Jitendra Jonnagaddala, The University of New South Wales

Treasurer: Roger Ward, The University of Melbourne

For enquiries please contact Roger Ward rward@unimelb.edu.au or Nicole Pratt Nicole.Pratt@unisa.edu.au

Board: Associate Professor Nicole Pratt, Mr. Ty Stanford, Mr. Jitendra Jonnagaddala, Professor Teng Liaw, Mr. Roger Ward, Associate Professor Dougie Boyle

OMOP Australia

Data Partner	Type of Data	Number of unique Patient	Provenance	Progress	Data Platform	Contact
Pharmaceutical Benefits Scheme 10% extract	Administrative Claims	2.5M	National	Complete 2021	PostgreSQL	Nicole Pratt
Primary Care GP data (Patron)	Primary care EMR	2.2M	Victoria	Complete	MSSQL	Dougie Boyle
AU-ePBRN (Australian Electronic practice based research network)	EHR	1.1M	New South Wales	Complete		Jittendra
Sydney Local Health District (LHD)	Hospital EHR	1M	New South Wales	Complete	PostgreSQL	Angus Ritchie
Royal Melbourne Hospital and Western Health. Hospital Admissions	Hospital EMR	685k	Victoria	In progress (First draft complete)	MSQL	Dougie Boyle
University of Queensland - Queensland Health	Hospital EHR (Cerner)	>5M	Queensland	In progress	MSSQL	Roger Ward
Austin Health	Hospital EHR (Cerner)	500k (approx.)	Victoria	In progress (early)	MSSQL	Roger Ward
Department of Veterans Affairs	Administrative Claims	approx 180,000 (current treatment population)	National	In progress	PostgreSQL	Nicole Pratt
South Western Sydney LHD	Cancer EMR	80,000 (current population)	NSW	In progress. New sites being onboarded 2023.	PostgreSQL	Georgie Kennedy
AOA National Joint Replacement Registry	Registry		National	Planning	PostgreSQL	Nicole Pratt
Sydney Children's Hospital	Hospital EHR		NSW	Planning		
The Royal Children's Hospital	Hospital EHR		Victoria	Planning		

Table Courtesy of Professor Nicole Pratt, University of South Australia

Centre for Health Analytics Goals and OMOP

The Centre is looking to utilise OMOP to address:

- Campus-wide data democratisation, accessibility, and delivering tools for analysis and visualisation
- More intuitive and standard data representation
- Accommodate a broad range of end-user skill-sets (“clickers” vs. “coders”)
- More Easily Comply with Security and Privacy Considerations

OMOP Implementation approach

- Executed a Proof of Concept on a pre-existing dataset (completed March ‘22)
- With learnings and outcomes of POC moved on to a larger Pilot dataset within the EMR (ongoing)

Proof of Concept Scope and Purpose

Proof of Concept Scope

- Transform a defined clinical dataset already extracted from Epic into the OMOP Common Data Model format

Proof of Concept Purpose

- Understand process of mapping (questions raised, consultations required)
- Measure quality and completeness of data mapping (what gets lost/obfuscated)
- Compare what analyses are possible based on pre and post-mapped data

Proof of Concept – Transformation Results

Clinical Domain	%Transformed	Issues, Decisions, Gaps, Lessons
Demographics	99%	• Source data didn't include ethnic or racial information – would need to address as part of a broader mapping effort
Visits	100%	• Followed SME guidance as to how to classify Inpatient / Outpatient visits (even though this might not align with OMOP convention)
Conditions (Diagnoses)	74%	• EPIC internally maps diagnoses to ICD10, SNOMED. • Achieved 74% through pure automated term matching based on ICD10 (did not attempt to further manually match). • Future efforts should map directly to SNOMED (and not use ICD10 as intermediary)
Procedures	85%	• Epic does not internally map procedures to any standardized terminology • Manual terminology mapping required clinical SME support • OMOP does not explicitly group procedures into surgeries, surgical panels • Some EPIC procedure codes validly map to more than 1 SNOMED code
Medications	99%	• EPIC internally maps to ATC code. • Mapped to active ingredient based on ATC Code (may be able to map to clinical drug format with more specificity in a broader effort)

Proof of Concept – Lessons

- Transformation of clinical data as represented by the source dataset into the OMOP CDM was successful, although certain domains required extensive manual subject matter expert input and review.

- The resulting data model preserved most important aspects of the source data transformed, however certain gaps were identified (specifically in relation to procedure records, and their relationship to surgeries, panels, billing, etc).

- **Knowledge of where and how data is represented within EPIC is Key**

Pilot Transformation of ED Dataset

- Transforming Emergency Department dataset, all encounters since 2016 (~500k)
- Ethics covers transformation and validation only (not data distribution for broader use)
- For Vocabulary Mapping, Focusing on Most Prevalent terms first (Covering 80% of records in the dataset)

Pilot Transformation of ED Dataset

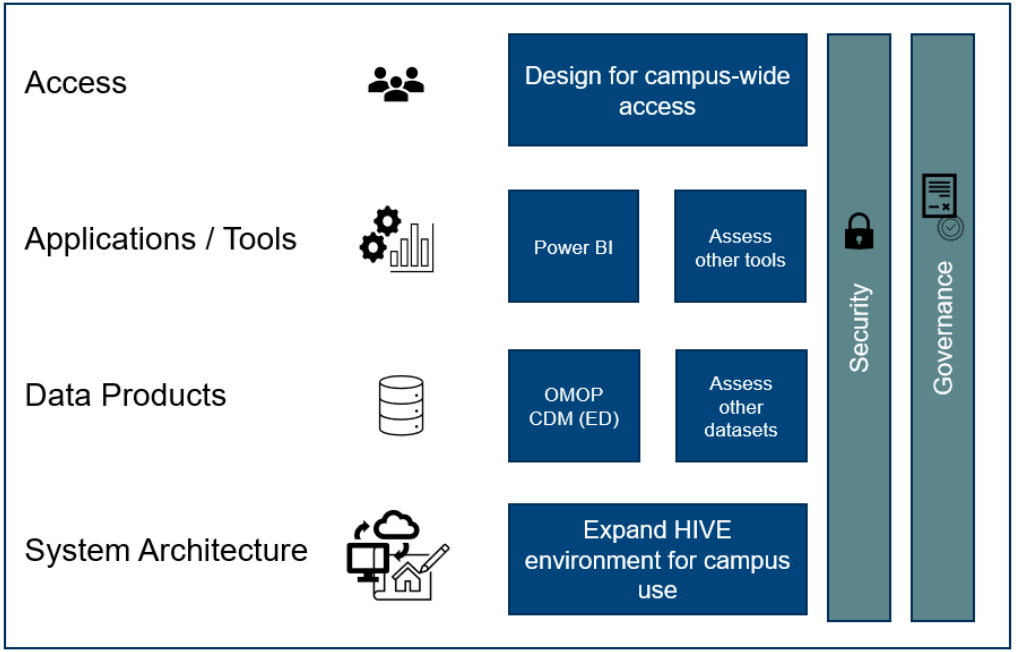
OMOP Domains	EMR Data
Location	The physical location of the point of service/ED
Care Site	The department, care area or room where the care was administered
Provider	The person who provided the care while the patient was in ED
Person	The patient in the ED
Death	Details of death of a patient who has had an ED encounter
Visit Occurrence	The encounter record for the patient in ED
Visit Detail	The intra ED movement for the patient
Condition Occurrence	The reason for visit and the diagnosis for the ED encounter
Procedure Occurrence, Measurement	All procedure orders related to the ED encounter while the patient was within ED
Drug Exposure	All medication orders related to ED encounter while the patient was within ED
Observation	Other items (Allergies, certain diagnoses)

ED Transformation Lessons So Far

- The process may take more time that you expect.
- Much of the Complexity arises out of how and where data is represented in the source system
  - (E.g., splitting off Emergency visits, certain patient movements captured in 2 different tables, Med admins vs. Prescriptions)
- Clinical SME knowledge, and EMR database expertise is required
- The Vocabulary Mapping process can be time consuming and requires clinical input

Next Steps

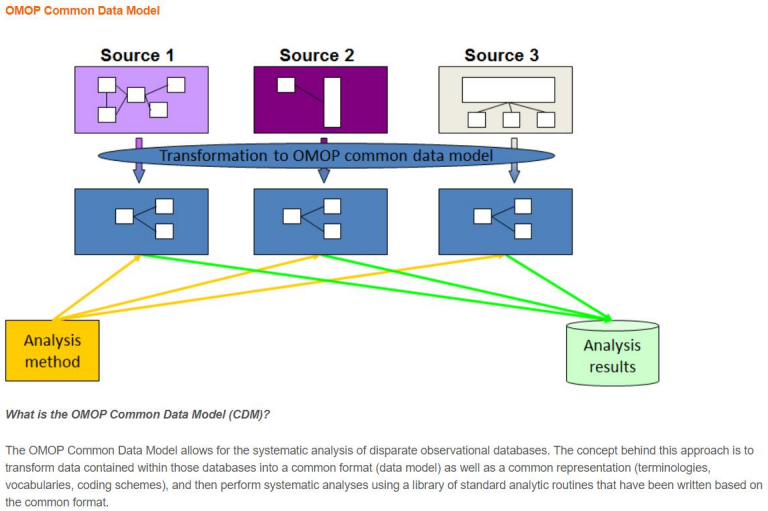
- Immediate focus: Complete Validation
- Design and Implement Platform making data accessible for Campus Use
- Privacy, Ethics and Governance Framework



Keeping Data Safe

- OMOP conceived as a method to minimise data movement

- Instead of Pooling data, craft common analysis that can be run at "the source" and share results



Keeping Data Safe

Centre for Health Analytics to use OMOP as one aspect of providing a secure area for Campus users access to reliable data, minimise one-off data extraction and storage, support analysis sharing and reuse