Implementation of the OMOP Common Data Model in a Multi-Site Canadian Network

Presenting Author: Lisa Lix, Chair, Algorithms and Harmonized Data Working Group, Health Data Research Network Canada; University of Manitoba, Winnipeg, CANADA; e-mail: lisa.lix@umanitoba.ca

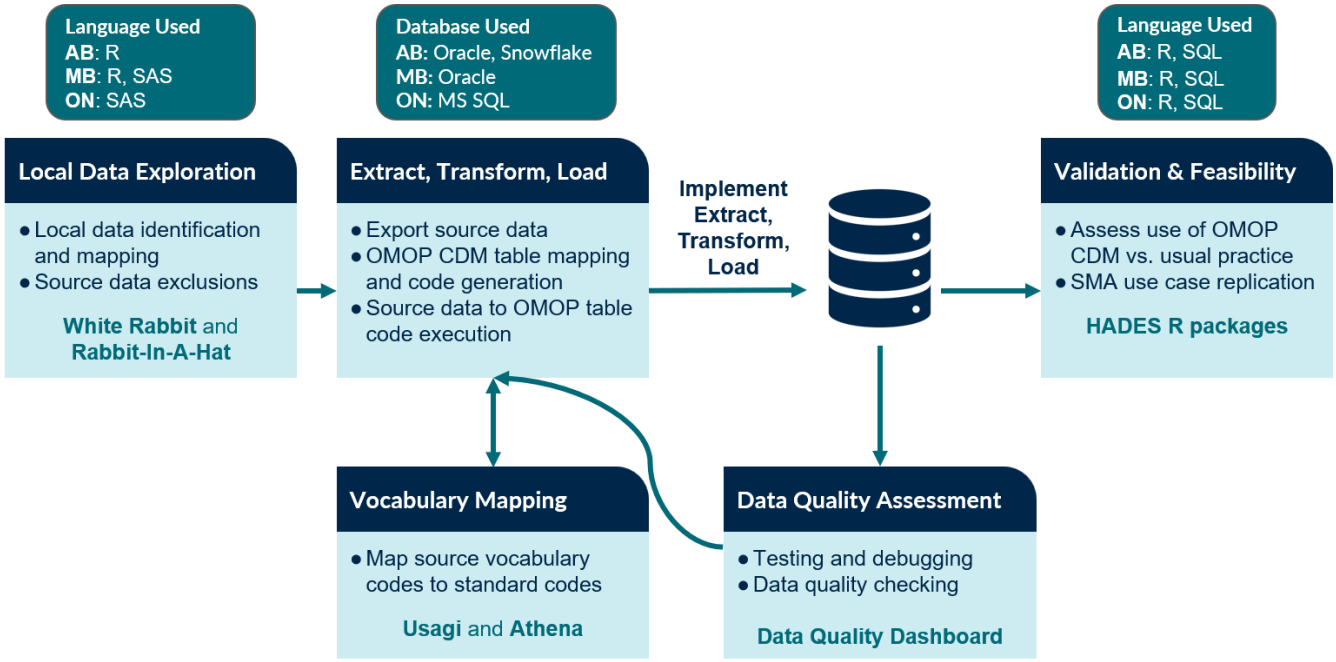
Background

- HDRN Canada began working towards implementing the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) in 2023 across provincial and territorial data centres to enable consistent, interoperable data and support systematic analysis of disparate observational databases across jurisdictions.
- OMOP is only one approach that HDRN Canada is exploring to harmonize data and develop research-ready datasets.

Summary of Progress

- Five data centres participated in HDRN Canada's OMOP CDM implementation efforts to date: Population Data BC (PopData), in BC; Alberta SPOR SUPPORT Unit (AbSPORU), in AB; the Manitoba Centre for Health Policy (MCHP), in MB; ICES in ON; and the Canadian Institute for Health Information (CIHI), a pan-Canadian organization.
- Two HDRN Canada sites (ON and MB) have successfully transformed limited administrative data holdings into the OMOP CDM version 5.4. A third site (AB) is in the process of completing the transformation process but has experienced delays due to local IT infrastructure challenges.
- The two sites with functional OMOP datasets have used OHDSI tools and analytical software packages to map vocabularies and concepts, check data quality, develop cohorts, and run analyses (see **Figure 1** for a workflow overview).

Figure 1. OMOP CDM Pilot Project Workflow and OHDSI Tools, Databases, and Languages Used.



Mapped Data and OMOP Tables

- Ten OMOP CDM tables were created as part of the effort to replicate a multi-regional spinal muscular atrophy (SMA) study. Mapped data include patient demographics, inpatient admissions, outpatient physician visits, and prescription drug dispensing, sourced from site-specific administrative data (see **Table 1**).
- Custom fields have also been created to capture census-based area-level indicators (e.g., rurality, deprivation indices).
- Mapping of Canadian classification systems (e.g., ICD-10-CA, the Canadian version of the Statistical Classification of Diseases and Related Health Problems) to the OMOP standard vocabulary (e.g., SNOMED CT, RxNorm) was initially limited to diagnostic codes and prescription medications related to SMA, diabetes, and contraception.

Table 1. OMOP CDM Tables, mapped data elements and corresponding data sources

OMOP Data Table	Data Source(s)	Mapped Elements
PERSON	Health insurance registry *(contains information on individuals registered to receive provincial health insurance)*	• Health card number • Date of birth • Sex
OBSERVATION_PERIOD	Health insurance registry	• Provincial insurance coverage start and end dates
VISIT_OCCURRENCE	Hospital discharge abstracts *(inpatient acute discharge records)* Physician claims *(for publicly insured medical services)*	• Inpatient admission dates • Inpatient discharge dates • Discharge disposition (destination/patient status) • Outpatient visit dates (from physician claims)
CONDITION_OCCURRENCE	Hospital discharge abstracts	• Diagnosis codes (ICD-10-CA mapped to SNOMED CT)
DRUG_EXPOSURE	Pharmaceutical data *(community/outpatient pharmacy data or public drug program claims)*	• Medication (Drug Identification Numbers (DINs) mapped to RxNorm) • Dispensing date
MEASUREMENT	Hospital discharge abstracts	• Diagnosis codes (ICD-10-CA mapped to SNOMED CT, assigned to OMOP Measurement domain)
OBSERVATION	Hospital discharge abstracts	• Diagnosis codes (ICD-10-CA mapped to SNOMED CT, assigned to OMOP Observation domain)
DEATH	Health insurance registry OR Vital Statistics records	• Date of death
LOCATION	Health insurance registry	Patient's: • Last known postal code • Province of residence
LOCATION_HISTORY (OMOP v 6.0)	Health insurance registry Postal Code Conversion File (PCCF/PCCF+) *(enables the linkage of individual postal codes to standard census geographies and census data)*	One postal code per person per calendar year Custom fields to capture census-based area-level socioeconomic data: • Rurality • Canadian Marginalization Index (CAN-Marg) / Ontario Marginalization Index (ON-Marg) • Canadian Index of Multiple Deprivation (CIMD)

Key Activities

OMOP CDM Pilot Project and SMA Use Case Replication

- Five sites (BC, AB, MB, ON, and CIHI) participated in a limited-scale OMOP implementation sufficient to replicate a multi-regional SMA study.
- The SMA use case was a population-based, descriptive cohort study of pediatric SMA in Canada, examining sociodemographics, comorbidities, treatment patterns, health care use, and costs.
- Weekly meetings supported collaborative data mapping across sites, and external technical support was provided by IQVIA and Odysseus Data Services.
- All participating sites mapped data to the OMOP CDM tables and fields required for the SMA use case; however, only two sites (ON and MB) ran their Extract, Transform, Load (ETL) pipelines to transform and load source data into an OMOP CDM database.
    - Data loaded into the OMOP CDM included demographics, inpatient discharge abstract, outpatient physician claims and drug data for the SMA cohort from 2004-2022 and the entire provincial population from 2022-2023.
- These two sites successfully ran the SMA analysis on the OMOP dataset; results were equivalent to those obtained from the source data.
- The OMOP CDM supported the use case replication but was challenging, and potential efficiency gains were not fully realized. The steep learning curve and time constraints limited analysts' ability to become proficient with the tools.

Vocabulary Mapping

- HDRN Canada is using a combination of automated and manual review processes to support the mapping of Canadian vocabularies to OMOP standard vocabularies, including SNOMED CT and RxNorm.
- In a pilot study, 240 ICD-10-CA codes related to spinal muscular atrophy (SMA) were mapped to SNOMED CT, and three SMA-related Drug Identification Numbers (DINs) were mapped to RxNorm using both automated tools and expert review.
- HDRN Canada has completed a mapping project for diagnostic and drug codes to OMOP standard vocabularies for diabetes, pregnancy, and birth outcomes.

    - 1770 ICD-10-CA 1547 ICD-9-CM, and 1147 ICD-9 diagnostic codes relevant to diabetes and obstetrics were mapped to SNOMED CT.
    - 631 DIN codes relevant to diabetes and contraception were mapped to RxNorm.
- Planned work includes mapping Canadian diagnosis codes used in the Charlson and Elixhauser Comorbidity Indices, along with codes from commonly used algorithms in HDRN Canada's Algorithms Inventory, to OMOP standard vocabularies.

Challenges and Lessons Learned

- **Governance, operational, and resource challenges:** Sites have encountered delays due to internal approval processes and limited staff capacity, which has affected the ability of some sites to complete or participate in OMOP implementation.
    - Some sites require project-specific approvals to create and maintain new databases.
- **Technical issues:** Common issues included software installation failures, configuration problems with relational database systems, and challenges managing OHDSI software dependencies.
- **Vocabulary mapping difficulties:** Challenges included the absence of equivalent standard concepts, source codes with multiple clinical elements, and ambiguity due to multiple mapping options in ATHENA.
    - Support from Odysseus Data Services and CIHI has been instrumental in resolving mapping discrepancies, maintaining consistent and semantically correct mappings, and ensuring that mappings reflect a pan-Canadian lens.
    - These local mapping efforts also have the potential to improve the accuracy and relevance of mappings for the broader OHDSI community.
- **OMOP model limitations:** Some elements of the SMA use case and anticipated future studies cannot be fully represented in the OMOP CDM or may require customizations (e.g., multiple diagnosis fields in a single record, linked hospital encounters). Additionally, US-centric concepts present challenges for transforming Canadian data.