Establishing Interoperability Standards between OMOP CDM v4, v5, and PCORnet CDM v1

Rimma Belenkaya, MS, MA[1], Parsa Mirhaji MD, PHD [1], Mark Khayter [2], Don Torok, MS [2], Ritu Khare, PhD [3] Toan Ong[4], Lisa Schilling[4]

[1]Montefiore Health Center, Bronx, NY (NYC-CDRN); [2] Ephir, Inc., Boston, MA (pSCANNER CDRN); [3] The Children's Hospital of Philadelphia, Philadelphia, PA (PEDSnet CDRN); [4] School of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO

Abstract

The Clinical Data Research Network (CDRN) Interoperability Workgroup, represented by three CDRNs, NYC-CDRN, pSCANNER, and PEDSnet, aim to create interoperability standards between OMOP and PCORnet Common Data Models (CDM) to support data integration for comparative effectiveness research. We describe our approach and progress for OMOP CDM v4, v5, and PCORnet CDM v1.

Introduction

PCORnet, the National Patient-Centered Clinical Research Network, funded by the Patient-Centered Outcomes Research Institute (PCORI), integrates data from 11 heterogeneous networks to enable large-scale comparative effectiveness research **(1)**. While the PCORnet CDM has been evolving, all 11 networks choose to first integrate their source data into more established CDMs, such as i2b2 **(2)** and OMOP **(3)**, and then port these data into PCORnet CDM. Crosswalking from healthcare source systems to OMOP CDM and then to PCORnet CDM poses a substantial challenge. To ensure data harmonization with minimal loss of source granularity, comprehensive CDM interoperability standards are required.

Three OMOP-based CDRNs: NYC-CDRN comprised of 22 organizations and representing over 2.5 million patients**(4)**; pSCANNER integrating three existing networks covering over 21 million patients**(4)**; and PEDSnet which includes eight of the nation's largest children's hospitals and provides service to 4.6 million**(5)** children per year, formed a workgroup to create standards and align approaches for first populating their OMOP CDMs and then migrating the data into PCORnet CDM v1. The multi-disciplinary workgroup created standards for representing PCORnet unique attributes in both OMOP CDM v4 and v5 and the subsequent migration to PCORnet CDM v1.

Methods

Our two-step approach consists of enforcing, to the extent possible, PCORnet requirements on the population of the OMOP CDM from the source so that the second step, conversion from OMOP to PCORnet CDM, conforms to a set of mechanistic transformation rules.

Many of the PCORnet CDM v1 attributes exists in OMOP CDM v4. However, even when common attributes existed, there were differences in the allowable values for these attributes. A simple example is gender. The OMOP vocabulary limits acceptable values to 'Male', 'Female' or 'Unknown' whereas PCORnet also includes "Ambiguous."

An additional difference that affected many otherwise common attributes is the interpretation of unknown values. PCORnet, supports a subset of HL7 delineation of unknown values that is not supported in standard OMOP vocabulary. The HL7 delineation differentiates between 'refused to answer', NULL, unknown and unmapped values. To compensate for this, we extended utilization of PCORnet source concepts as standard concepts in the OMOP vocabulary. Although this is a violation of OMOP CDM conventions, it provides complete coverage of PCORnet required terminology, gives additional insight into source data quality, and does not affect any OMOP CDM based software tools.

The interoperability standards: (1) identify matching domains, attributes and vocabularies between the two CDMs; (2) propose a solution to account for data elements that are missing in the OMOP CDM; (3) add data representation conventions in OMOP CDM that provide closer alignment between the two models.

PCORnet fields that are not available as first class attributes in the OMOP CDM are represented in the Observation table, the generic entity-attribute-value table in the OMOP, and are linked to their respective domains either by foreign keys in the Observation table or via the generic many-to-many Fact_Relationship table. The Fact_Relationship table is also utilized to establish explicit links between unconnected measurements and observations in OMOP CDM that have to be transformed into a single record in PCORnet CDM.

An example of a first class attribute not in the OMOP CDM but required by PCORnet is Hospital Admission Source. The Hospital Admitting Source is added as an observation record for the person with the same date as the Inpatient Visit Occurrence *visit_occurrence_start_date* and the *observation_concept_id* equal to Admission from Establishment (4145666). The *value_as_concept_id* contains the OMOP concept that represents the admission source and the *observation_source_value* holds the code or description used to determine the concept id.

Synchronizing Diastolic and Systolic blood pressure (BP) measurements is an example of using the Fact_Relationship table to establish an explicit link between related measurements. For each pair of BP measurements, two records are created in the FACT_RELATIONSHIP table. The first record contains domain_concept_id_1 and domain_concept_id_2 equal to 21 ('Measurement'), Fact_id_1 and Fact_id_2 equal to the respective measurement_id of diastolic and systolic BP records in the Measurement table coming from the same measurement, and relationship_concept_id equal to 46233682 (' Diastolic to systolic blood pressure measurement'). The reciprocal record is similarly defined with the relationship_concept_id equal to 46233683 ('Systolic to diastolic blood pressure measurement').

Results

In this work, we defined the interoperability standards from source (i.e. electronic health records) to OMOP CDM v4 (Specific to PCORI CDM) and from source to OMOP CDM v5 (Specific to PCORI CDM) while accounting for different data perspectives and peculiarities from at least three different PCORI CDRNs. The main deliverables include a conventions document for populating the OMOP CDM, and an extract-transform-load specifications document for transforming to the PCORnet CDM v1, for both OMOP CDM v4 and OMOP CDM v5. The documents will be publicly available for the OHDSI community. The workgroup was formed in September 2014 and the interoperability standards were established in December 2015 for OMOP CDM v4, and in May 2015 for OMOP CDM v5, through approximately 30 hours of group meetings and several email conversations. As of August 2015, all three participating networks have used these standards to guide implementation of their ETL processes and the PEDSnet CDRN implementation has been successfully audited by the PCORI board for data quality assurance.

The workgroup addressed several challenge. For any major decision where consensus was not established the workgroup members voted. This occurred for decisions such as representation of a missing field (in the source) in the observation table, adjudication of multiple records of biobank flag for a given patient, representation of various flavors of NULL in OMOP (e.g. missing, unknown, and other), linking systolic and diastolic blood pressure readings in OMOP. In many cases, we also resorted to discussions with the outside community, e.g. consultation with OMOP Vocabulary Steward, and clarification from PCORI about certain PCORnet CDM requirements, e.g. "complete data capture," "end date" in the enrollment table. Also, several complex domains required intra-network investigations and discussions for establishing a comprehensive solution based on evidence from the source EHRs, e.g. linking emergency and inpatient visits and determination of diagnosis related groups. In addition, the workgroup ensured that the OMOP concept identifiers specified in the conventions document were accurate and current (i.e. not obsolete).

Conclusion

With rare exceptions, the OMOP CDM supports greater granularity of data representation in both the CDM and vocabulary than PCORnet CDM. These features allow for adequate preservation of source data granularity, and transformation from the more granular OMOP to less granular PCORnet representation is straightforward. Once the OMOP CDM population conventions have been established, the process of creating OMOP-to-PCORnet ETL standards is reduced to describing simple mappings and transformations.

It is possible to use the OMOP CDM as an intermediary data representation when converting various healthcare datasets to PCORnet CDM. However it was necessary to use vocabulary concepts that violated the domain rule established in the OMOP CDM v5 standard. In addition attributes required in PCORnet that do not exist in OMOP can be represented without altering the standard table schema, but require a set of documented conventions to be understood and extractable from the OMOP CDM.

Works Cited

1. Fleurence RL,ea. Launching PCORnet, a national patient-centered clinical research network. Journal of the American Medical Informatics Association 21.4. 2014: p. 578-582.
2. Murphy SN,ea. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association 17.2. 2010: p. 124-130.
3. Stang PE,ea. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Annals of internal medicine 153.9. 2010: p. 600-606.
4. Rainu Kaushal GHea. Changing the research landscape: the New York City Clinical Data Research Network. J Am Med Inform Assoc: 21 (4). 2014 Jul: p. 587-590.
5. Christopher B Forrest PAMea. PEDSnet: a National Pediatric Learning Health System. J Am Med Inform Assoc: 21 (4). 2014 Jul 1: p. 602-606.
6. Ohno-Machado Lea. pSCANNER: patient-centered Scalable National Network for Effectiveness Research. Journal of the American Medical Informatics Association 21.4. 2014: p. 621-626.