Mapping of Critical Care EHR Flowsheet data to the OMOP CDM via SSSOM

Polina Talapova, MD, PhD¹, Andrew Williams, PhD¹, Nicolas Matentzoglu²,
Anna Ostropolets, MD, PhD², Michael Kallfelz, MD³
¹Tufts Clinical and Translational Science Institute, Boston, MA, United States
²Semanticly Ltd, London, United Kingdom
³Observational Health Data Sciences and Informatics, North Bethesda, MD, United States

Background

A significant amount of patient data collected in Intensive and Critical care settings is typically represented in flowsheets, which are EHR-integrated tools that are used to document longitudinal data, assessments, observations, and routine care tasks in a grid-like format.¹ They cover a wide range important unstandardized data sources including questionnaires, vital signs, fluid intake and output, body system assessments (neurological, respiratory, cardiac, renal, gastrointestinal), nutrition, wound care, pain management, and nursing procedures. To facilitate deep phenotyping approaches that require the full richness of healthcare generated data, it is necessary to harmonize this data. One of the validated approaches to achieve this is by converting the data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).² This involves semantically mapping the source terms to standard concepts from the OMOP CDM Standardized Vocabularies. However, the OMOP CDM Standardized Vocabularies, which serve as the mapping targets, still require improvements in terms of maintenance, provenance, precision, and justification of the mappings.

Within the OHDSI community, mappings are typically provided in the form of either a source_to_concept_map (STCM) table or staging tables (such as concept_stage, concept_relationship_stage, concept_ancestor, vocabulary_stage, etc.). Alternatively, wide mapping tables may also be used. However, all of these approaches share a common drawback of lacking up-to-date documentation and mapping metadata. This limitation compromises the precision and usability of the mappings.³

Therefore, in this work, we propose the utilization of the Simple Standard for Sharing Ontological Mappings (SSSOM),⁴⁵⁶ which offers a solution to address these issues by enabling the storage of a comprehensive set of standard metadata elements that depict the mapping reliability.

Methods

To map EHR Flowsheet data collected in critical or intensive care settings to the OMOP Standardized Vocabularies using SSSOM, we followed the next steps. Firstly, we conducted a structural (field-level) mapping from the source data to SSSOM. This mapping was used to define the operational model, as shown in Figure 1.
Subsequently, we performed a structural mapping from SSSOM to the OMOP CDM, as depicted in Figure 2, in order to identify interoperability potential.

To distinguish between Flowsheet Items and Flowsheet Item Values, we utilized the subject_type field. Next, we manually mapped the top 1000 critical care flowsheet items in the SSSOM table format. During this process, we employed interim SSSOM-like fields such as cnt (count), subject_synonym, subject_value, subject_unit, object_code, object_class, and object_vocabulary. These additional fields enhanced the user-friendliness of the table during manual curation. Furthermore, to incorporate the SSSOM mapping table into an OMOP CDM instance, we developed a Python tool, which generates INSERTS into the OMOP Vocabulary staging tables, including concept_stage, concept_relationship_stage, concept_synonym_stage, and vocabulary_stage.
Results

The applied methodology enabled us to retain crucial information regarding the origin, intent, creation, maintenance, and quality of the mappings. This approach allowed us to identify and address mapping issues that may arise when working with OMOP CDM Standardized Vocabularies, as shown in Table 1.

Table 1. SSSOM mapping precision metadata as an indicator of mapping issues

<table>
<thead>
<tr>
<th>SSSOM Field Name</th>
<th>Field Value</th>
<th>Meaning</th>
<th>Issue Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicate_id</td>
<td>skos:exactMatch</td>
<td>One-to-one full equivalent mapping: semantics of one subject (source) is fully covered by an object (target).</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>skos:broadMatch</td>
<td>1. An object is semantically broader than a subject; 2. It is similar</td>
<td>The mapping results in the loss of source term granularity. This implies that in the OMOP CDM, there is no exact equivalent for the source term. To accommodate such cases, an OMOP extension concept may be necessary to capture the specific details that are not fully represented within the standard OMOP vocabulary.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to the “Is a” relationship in the OMOP CDM, which indicates a link to a semantic parent (a more generic and less granular term).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>skos:narrowMatch</td>
<td>1. An object is semantically narrower than a subject; 2. It is similar</td>
<td>The mapping brings additional details that are not mentioned in the source. Alternatively, in the OMOP Vocabulary, there are several similar standard concepts that might require deduplication.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to the “Subsumes” relationship in the OMOP CDM, that indicates a link to a semantic child (a more specific and more granular term).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>skos:closeMatch</td>
<td>Indicates a standard OMOP concept that has a relation to a subject, but</td>
<td>The standard equivalent is absent, and an OMOP extension concept might be required.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nothing is listed above (e.g., to store information about devices related to intervention or manipulation).</td>
<td></td>
</tr>
<tr>
<td>confidence</td>
<td></td>
<td>A score between 0 and 1 to denote the confidence or probability that the match is correct.</td>
<td>“1” denotes full confidence in mapping, while “0” - absence of such a confidence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The higher the confidence, the higher mapping precision.</td>
</tr>
</tbody>
</table>
To ensure the storage of all important mapping metadata in the OMOP CDM, the inclusion of an additional mapping_metadata table is necessary. The table should include the following fields as a bare minimum: predicate_id, confidence, mapping_justification, author_id, author_label, reviewer_id, reviewer_label, mapping_provider, mapping_tool, mapping_tool_version.

Conclusion

The implementation of the OMOP-on-SSSOM approach enables the export of mappings in a computationally accessible format, facilitating their direct integration into ETL pipelines and supporting data exchange between different semantic spaces and data models. By including basic metadata, such as the type of predicate_id and the level of mapping confidence, users can assess the suitability of mappings for different purposes. They can filter for high-confidence mappings and exact predicates to obtain more reliable evidence.

To enhance the transparency of the OMOP Vocabulary, we propose the addition of a mapping_metadata table to the OMOP CDM. This table would capture additional details about the mapping development process and provide insights into the resulting mapping quality.

Moving forward, we plan to evaluate the proof of concept of our methodology by measuring user experience in mapping-related manipulations. Additionally, we will create an SSSOM representation of ICD-10-CM-to-OMOP mappings to assess the impact of mapping corrections on a cohort. Furthermore, we intend to submit a request for the deduplication of Standard Concepts available in OHDSI Athena across Measurement and Observation domains, focusing on those relevant to Critical and Intensive care. These steps will contribute to further refining and validating our approach while expanding its applicability to real-world scenarios.

References


