

Establishing Best Practices for Mapping Standard Assessment Instruments to the OMOP CDM

Nicole M. Gerlanc, PhD¹ and Piper Ranallo, PhD² on behalf of the OHDSI CDM Survey Sub-Working Group

¹National Cancer Institute (NCI), National Institutes of Health (NIH)

²University of Minnesota

Background

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) was designed to standardize observational health data drawn primarily from clinical sources such as electronic health records, claims, and registries.¹ Survey and assessment instrument data, by contrast, are collected directly from patients or their informants, possess a complex question-answer structure, and frequently lack standardized vocabularies or mappings to clinical terminologies such as SNOMED CT or LOINC. As a result, surveys are less frequently mapped to OMOP and lack the conventions, tools, and best practices established for other clinical data types.

The CDM Survey Sub-Working Group was formed in 2024 to address this gap. A 2025 landscape assessment of the OHDSI community confirmed that survey mapping is widespread but inconsistent. All responders who completed a mapping required custom concept identifiers and reported challenges included insufficient standard concepts for social and behavioral content, mismatches in granularity, and a lack of shared guidance.² These findings motivated the Sub-Working Group's 2026 objective of producing a formal best practices proposal for representing survey data in the OMOP CDM. This report describes the development of that proposal and the outcomes of its presentation at the 2026 OHDSI European Symposium in Rotterdam.

Methods

Beginning in January 2026, members of the Sub-working Group representing Survey, Psychiatry, Vocabulary, and CDM communities developed a written best practices proposal through a series of bi-weekly meetings. The proposal establishes a classification framework that distinguishes two primary classes of Standardized Assessment Instruments (SAIs): Standardized Data Collection Instruments (SDCIs), for which the individual question-answer pairs carry the analytic value and Psychometrically Validated Instruments (PVI), for which a validated score is the primary output and is treated analogously to a laboratory result³. The proposal further defines three data representation strategies—Facts, Question/Answer (Q/A) Pairs, and a combined approach that simultaneously preserves the Q/A Pair and a derived clinical fact linked via FACT_RELATIONSHIP—together with guidance on metadata, negative findings, hierarchical structure, vocabulary submission, and complex mapping scenarios such as copyright and negation.

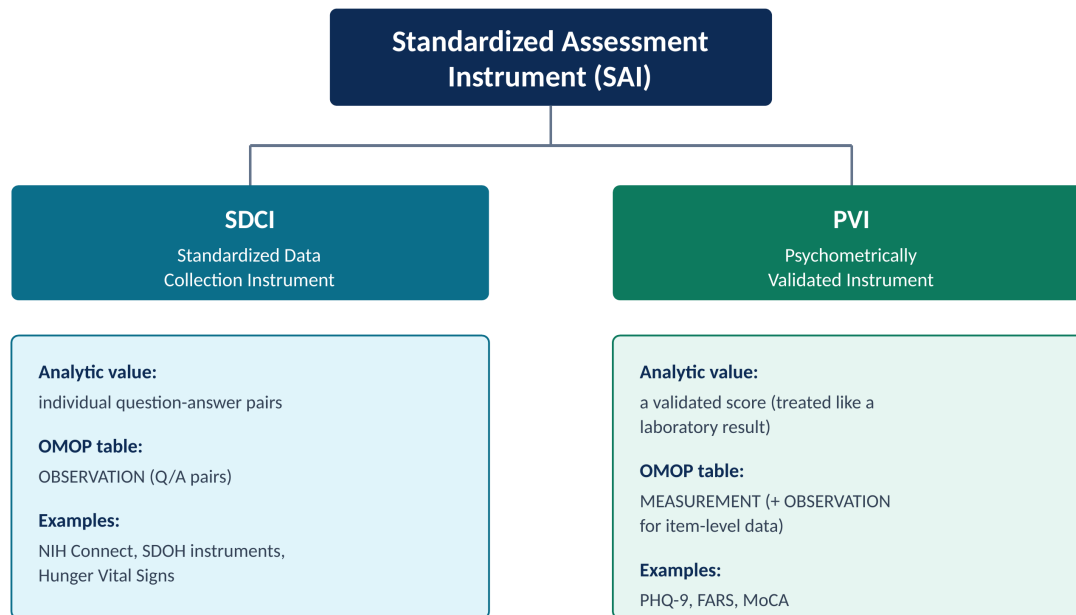


Figure 1. Classification of Standardized Assessment Instruments (SAIs) into the two primary classes, showing the analytic value, OMOP target table, and representative examples for each.

Throughout February and March 2026, the draft was circulated to the broader OHDSI community for asynchronous comment and review sessions were held to refine the classification definitions and the decision framework for choosing a representation strategy. An outline version of the proposal was then presented at a workshop at the OHDSI European Symposium in Rotterdam in April 2026. The session was structured around five objectives: conveying the recommended best practices; reaching consensus on the SAI classification definitions; reaching consensus on the recommended representation strategy; identifying gaps, edge cases, and points of contention; and agreeing on next steps for finalization and dissemination.

Results

The Rotterdam workshop drew a strong turnout despite competing concurrent sessions and generated substantive discussion that expanded the pool of community members actively commenting on the proposal. Discussion centered on the combined representation strategy—the Q/A Pair together with its derived clinical fact—which found some support as the central recommendation, although the discussion is ongoing. Because the term “triplet” already carries a specific meaning in the OHDSI vocabulary concept model (Item → Response → Finding semantic model), the Sub-Working Group subsequently adopted the distinct term “Dual Representation” for the CDM-level strategy to avoid ambiguity.

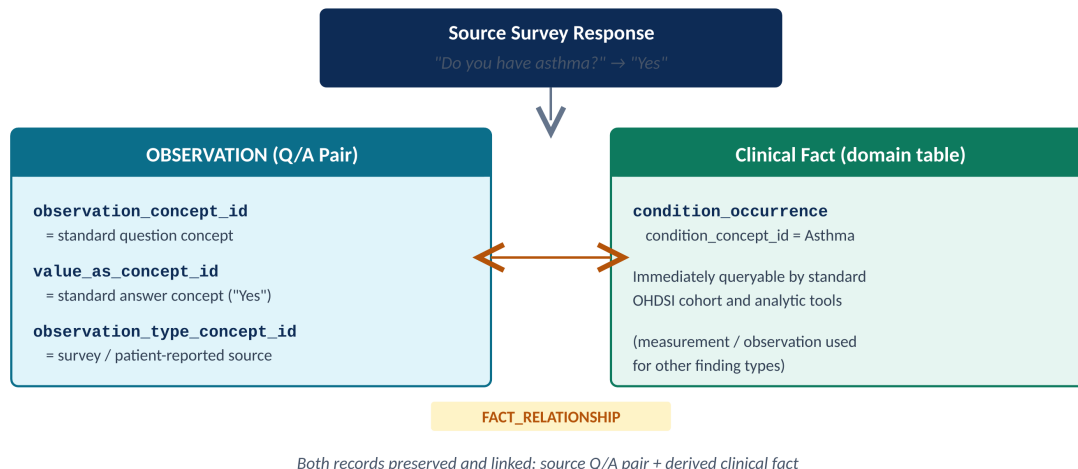


Figure 2. Dual Representation, a survey response stored as both a Q/A Pair in OBSERVATION and a derived clinical fact in a domain table, linked via FACT_RELATIONSHIP, is a remaining open question as a recommended best practice.

Workshop feedback surfaced several issues now being incorporated into the proposal. A significant discussion concerned where instrument-level metadata should reside; rather than further burdening the OBSERVATION table, the group is evaluating whether the dedicated Survey table introduced in CDM version 6.0 could serve as a metadata extension. Participants also debated the balance between encoding instrument detail in the vocabulary (for example, an observable-entity concept such as “frequency of pain over the past two weeks”) versus capturing it in discrete fields of the information model, drawing an analogy to how LOINC codes encapsulate the component, timing, and specimen of a laboratory measurement. The use cases discussed mapped onto all three OHDSI analytic domains—characterization, population-level effect estimation, and patient-level prediction—reinforcing that survey data harmonized under these conventions can be analyzed with much of the existing OHDSI tooling, provided longitudinal and phenotyping requirements are addressed.

Following the Symposium, the Sub-Working Group defined a path to finalization. The revised proposal will be shared and discussed with the THEMIS, CDM, and Vocabulary working groups so that each can review and give comment on the portions that intersect with their domains. A decision meeting is planned to resolve outstanding points by majority where full consensus is not attainable, with the recognition that an agreed foundation is more valuable than indefinite deliberation. The Connect for Cancer Prevention cohort as well as other SAIs used by members of the Sub-Working Group will serve as demonstration projects for applying the best practices to both SDCIs and PVI.

Conclusion

In 2026 the CDM Survey Sub-Working Group moved from problem definition to a concrete, community-reviewed proposal for representing survey and assessment instrument data in the OMOP CDM. Community discussion affirmed not only the two-class SAI taxonomy (SDCIs and PVIs) but also several foundational conventions: explicit negative responses should be preserved as Question/Answer pairs rather than represented as the absence of a record; PVI scores should be stored in the MEASUREMENT table and treated analogously to a laboratory result rather than re-derived from item-level data; temporality should be enforced through the data rather than encoded in the concept; implementers should search existing LOINC, SNOMED CT, and OHDSI vocabularies before creating new concepts and

organize vocabulary submissions by clinical content area to curb the proliferation of near-duplicate concepts; and each administration of a repeated instrument should be stored as a distinct, independently queryable event. Post-coordination of questions and answers emerged as the preferred default where source and target components align. Dual Representation as the recommended strategy for SDCI items and the appropriate home for instrument-level metadata — the parent OBSERVATION panel record versus the CDM v6 SURVEY_CONDUCT table — remain the principal open questions. Through coordinated review with the THEMIS, CDM, and Vocabulary working groups, demonstration projects on both SDCIs and PVI, and a hands-on workshop at the 2026 Global Symposium, the group aims to establish adoptable conventions that let the OHDSI community harmonize survey data alongside other real-world data using familiar analytic tools. We invite the community to review the proposal, contribute use cases, and join the effort to refine and adopt these standards.

References

1. OHDSI. The book of OHDSI. 2021. <https://ohdsi.github.io/TheBookOfOhdsi/>
2. Gerlanc NM, CDM Survey Subgroup. Mapping survey data to OMOP: the current state and available resources. OHDSI Global Symposium Collaborator Showcase; 2025.
3. Ranallo, PA. The evolving landscape of computable representations of standardized assessment instruments in healthcare. Manuscript in preparation. 2026.