

Transforming Oral Health Standards with the OMOP-CDM

Vidhya Venkateswaran³, Julie Cha², Michael Feolo³, John O' Malley⁴, Jin Xiao⁵, Jennifer Holtzman⁶, Jay Patel⁷, Liz Amos⁸, Aarthi Shanmugavel⁹, Robert Koski¹

Johns Hopkins University¹, Georgia Institute of Technology², National Institute of Dental and Craniofacial Research³, Carequest Institute for Oral Health⁴, University of Rochester⁵, Center for Oral Health⁶, Temple University⁷, National Library of Medicine National Institutes of Health⁸, American Dental Association⁹

Background

Electronic health records are key sources of clinical data that can be used to support large-scale observational research¹. Dentistry has been slow to adopt common data models for observational research due to fragmented practices, complex and ambiguous data structures, and challenges with terminology and data schemas. As a result, few, if any dental datasets have been transformed to a common data model^{2,3} (CDM).

This project will define, implement, and validate mappings from American National Standards Institute/American Dental Association 1084 standard (ANSI/ADA 1084)⁴, which is the Reference Core Data Set for Communication Among Dental and Other Health Information Systems, into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The work will include vocabulary alignment and ETL specifications to the existing OMOP base model while potentially proposing an OMOP extension for dental chart related elements.

A standardized dental data-to-OMOP mapping framework is required to facilitate large scale oral health research. We will pursue a pilot implementation with an OMOP-based database as the finished product. The outcome will be reusable and scalable to allow for integration in large health systems that host both medical and dental data.

Given the growing need to better understand the connections between oral health and systemic disease, researchers require a dental specific CDM implementation. The Observational Health Data Sciences and Informatics (OHDSI) Dentistry Workgroup, the National Institute of Dental and Craniofacial Research (NIDCR), and the American Dental Association (ADA) have partnered to transform the dental standard ANSI/ADA 1084 to the OMOP-CDM to understand the gaps and barriers to conducting observational research on dental data with the OMOP-CDM.

Methods

The OHDSI Dentistry Workgroup, the NIDCR, and the ADA formed a working group to conduct a formal transformation of the ANSI/ADA 1084 standard⁴ to the OMOP-CDM V5.4. The working group is composed of 10 members with clinical dentistry or data science backgrounds. The working group selected ANSI/ADA 1084 because it represents the core set of elements utilized in dental practice and would be

representative of many dental use cases of interest to observational research studies in oral health. The working group's objectives for this study are:

1. Transform the data elements specified by the 1084 standard to the base OMOP-CDM V5.4.
2. Document the data elements that cannot be mapped and identify reasons for inability to map to the OMOP-CDM.
3. Identify potential methods for extension to the OMOP-CDM or modification to the OMOP-CDM vocabularies to accommodate these items.
4. Document the data elements that may be mapped to the OMOP-CDM but have ambiguous or ill-defined possible mapping to the OMOP-CDM.

The WG members were assigned sections of the ANSI/ADA 1084 standard. They were instructed on OMOP-CDM documentation and provided example mappings/transformations to reference. The WG members were instructed to transform their data to the best of their abilities to the OMOP-CDM and report findings back to the group for evaluation and discussion. Participants used OMOP-CDM documentation from GitHub and the open-source Athena tool to accomplish their respective mappings. The findings were documented in a standardized Excel spreadsheet and collected every two weeks. Results were discussed among the group to highlight findings and unique mapping experiences.

Results

This study is ongoing and preliminary results are reported.

The list of elements outlined in the ANSI/ADA 1084 standard was divided into patient-specific general findings and oral/dental findings. The information captured by this list of elements was taken to represent a single patient encounter with a general dentist. Thus, the elements comprise data classes that include document properties (document author, custodian), patient baseline characteristics (family history, social history), and findings from a clinical oral exam. The oral findings include data classes that span dental findings through an odontogram, periodontal measurements, temporomandibular joint findings, and soft tissue findings.

From a preliminary analysis of these data classes, the group found that several of the person and provider centric data fields could be transformed to appropriate OMOP tables, and sample elements could be transformed to appropriate concept IDs. Similarly, the patient baseline characteristics had strong equivalent representations in the CDM, allowing for accurate transformation of these data classes of the clinical record.

Illustrative examples of the table used by participants are shown below (Table 1):

Table 1:

Encounter

1084 Class	1084 Data Field	Sample Data Value	Unique Identifier	Vocabulary	OMOP Table	OMOP Field	Vocabulary
------------	-----------------	-------------------	-------------------	------------	------------	------------	------------

Encounter	Provider	Provider ID	Eg., 123456	NPI	PROVIDER	Provider_id	NPI
Social History	Tobacco Use Duration	X years	-	-	OBSERVATION	Observation _concept id = 36303803 with value _ as _ number and unit _concept _id for years	LOINC

When considering more dental and oral specific findings, we found that tooth level and site level findings in the oral cavity required complex representations when mapped to the person-centric CDM. Each tooth is broken down into 5 surfaces to record dental findings. Accurate capture of disease on each surface in OMOP would require observations for each tooth surface per encounter ($32 * 5 = 160$ total observations for dental caries). Adding additional phenotypes, such as restorations, fractures, or other dental diseases on the same tooth, further complicates the data representation within the CDM. In addition to tooth findings, the standard dental clinical exam also includes periodontal examination (multiple sets of measurements around each tooth), temporomandibular exam, and occlusion which taken together compound the information tied to the oral cavity. A potential alternate representation would be to create a site-specific extension to localize the oral findings to one table which is linked back to the person.

Conclusion

Dental data requires both person-centric and tooth-centric data, which the OMOP-CDM does not currently accommodate easily. Data from the ANSI/ADA 1084 standard that could be mapped was consistent with the person-level data; however, challenges and opportunities remain with mapping tooth level data to the OMOP-CDM. We project that this gap analysis will be completed in the Fall of 2026.

OMOP-CDM handles person-level data, but not the tooth-, surface-, and site-specific detail at the core of a dental care visit, necessary for many dental use cases. Adding site-aware fields to tables will simplify dental representation without complex workarounds. Further, terminology expansions will allow each site to carry its own attributes. A site-aware OMOP model is critical for interoperable, analysis-ready dental data.

References

1. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, Ryan PB. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015;22(3):553–64. doi:10.1093/jamia/ocu023
2. Benoit B, Frédéric B, Jean-Charles D. Current state of dental informatics in the field of health information systems: a scoping review. BMC Oral Health. 2022;22(1):131. doi:10.1186/s12903-022-02163-9

3. Thyvalikakath TP, Duncan WD, Siddiqui Z, LaPradd M, Eckert G, Schleyer T, Rindal DB, Jurkovich M, Shea T, Gilbert GH. Leveraging Electronic Dental Record Data for Clinical Research in the National Dental PBRN Practices. *Appl Clin Inform.* 2020;11(2):305–314. doi:10.1055/s-0040-1709506
4. American Dental Association. (2019). ANSI/ADA Standard No. 1084 Reference Core Data Set for Communication Among Dental and Other Health-D.

Funding Acknowledgements

This work was supported in part by the National Library of Medicine (NLM), National Institutes of Health (NIH). The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Service