Vocabulary Versioning System for OMOP-CDM: Enabling Vocabulary Management Across Studies

Tasmeia Yousaf[1], Olivier Bouissou[2], Elisabeth Ross[3]
[1] Special Advisor, Clinical Datawarehouse
[2] Architect, Clinical Datawarehouse
[3] Special Advisor, Clinical Datawarehouse

Background

Observational research using the OMOP Common Data Model (CDM) depends on standardized vocabularies to ensure consistency across studies. However, different studies may require different vocabulary versions, creating challenges in data integrity, reproducibility, and cross-study comparability. We have developed a version-based system that will keep track of different versions of vocabularies and enable the switch between vocabulary versions across studies. The system will also help to build cohorts using different vocabulary versions in ATLAS in the future.

Methods

We have developed a Vocabulary Versioning System with the following key components:

- A version-controlled vocabulary archive **(**OMOP_ARCHIVE**)** storing multiple vocabulary versions for retrieval and study-specific application.
- A switching mechanism that allows researchers to select and activate the appropriate vocabulary version per study.
- A buffered loading system **(**OMOP_BUFFER**)** that facilitates smooth transitions between source and vocabulary archive.
- Retrieval and archival process ensuring that clinical data is mapped correctly to the active vocabulary version.

Results

- A single database for OMOP CDM along with archive database for vocabulary versions.
- The system enables seamless switching between vocabulary versions, supporting study-specific requirements.
- It enhances historical comparison by preserving vocabulary changes, ensuring study reproducibility**.**
- It guarantees that only one active vocabulary version is applied at a time, maintaining data integrity**.**
- It facilitates quality control by allowing clinical data to be archived alongside specific vocabulary versions.
- Vocabulary versions archive will be used for building cohorts with different OMOP vocabularies in the future.

Conclusion

Our Vocabulary Versioning System provides a transparent and reproducible approach for managing OMOP vocabulary updates. By allowing researchers to easily switch between different vocabulary versions across studies, it enhances study reproducibility, data consistency, and cross-study comparability.