

The Data2Evidence platform: facilitating healthcare research through better data discovery and governance.



Enhancing Clinical Data Management and Utilization with the Data2Evidence Platform

Background: Storing and managing multiple large OMOP datasets in an organization with numerous users poses significant challenges in storage, security, and IT expertise. Moreover, given the large number of data sets, it is often challenging for researchers to navigate and find relevant data. Our software platform, Data2Evidence, is designed to facilitate the streamlined management and analysis of clinical data in the OMOP CDM.

Enabling Data Discovery and Evaluation in Dataset Collections

The system administrators can flexibly define and modify the metadata stored for each dataset. Moreover, our solution allows the administrator to run existing OHDSI tools for data characterization and data quality on demand and store the results per dataset.

Our solution also provides the researcher with a comprehensive overview of the datasets in an easy-to-use web interface allowing them to make informed decisions. This includes information such as the metadata, schema version, data characterization, and data quality reports for each dataset. Users can search across this information and free text to get the relevant dataset.

Visibility	Study ID	Tenant	Taken study code	Schema name	Schema version	Update available	Data model	Actions
👁	345164	Dev Tenant	synpuf5ct	CDMeH0702...	questionnaire V1.0.0	No	omop5-4	Select action
👁	3451310	Dev Tenant	synpuf5ct	CDMeH086g...	questionnaire V1.0.0	Yes	omop5-4	Select action
👁	345164	Dev Tenant	synpuf5ct	CDMeH0702...	questionnaire V1.0.0	No	omop5-4	Select action
👁	3451310	Dev Tenant	synpuf5ct	CDMeH086g...	questionnaire V1.0.0	Yes	omop5-4	Select action

Create data mart - synpuf5ct

New dataset name

Date Filter

Cohort Filter

Table Filter

Snapshot Location

Store snapshots in DB

Store snapshots as parquet file

Cancel Create

Improving Governance and Security through Targeted Data Use

Our platform supports typical and more complex workflows to request and grant access to the dataset. It also integrates with existing identity providers and data governance systems in the organisation.

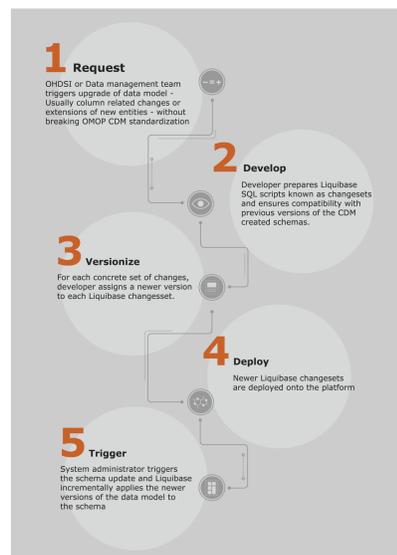
New data subsets (data marts) can be created by filtering by timeframes, patients matching a specific phenotype and specific entities and attributes. The new subset is created from the primary dataset based on the defined filter criteria. The isolation of the new subset begins at the database layer to ensure clear governance and boundaries. This empowers researchers to work with the minimum necessary data while safeguarding information about patients with specific diseases and filtering out sensitive information in data sets.

Easier Data Model Evolution

During the lifecycle of a dataset, it is expected that changes to the existing data model are required with each new version of the OMOP CDM which is not trivial. Also, it might be expected to make these changes only on certain datasets in the platform while keeping the others as is.

To address this challenge, our platform has a robust integrated schema evolution solution. Developers define “changesets” (SQL scripts) and the software maintains a comprehensive record of applied changesets, thereby facilitating the update or rollback of changes to the database.

This ensures a consistent state of the database across multiple deployments including development, testing, and production systems. The flexibility to apply changes to specific datasets also allows users to take sufficient time to amend their analysis scripts based on these changes in the CDM schema.



Karthik Seetharaman¹, Santan Maddi¹, Satish Anbazhagan¹, Afreen Sikandara¹, Brandan Tan¹, Alicia Jing Wen Koh¹, Peter Hoffmann¹

¹D4L data4life Asia Limited

