

Using OMOP-CDM to Develop Dynamic Disease Registries and Analytic Data Enclaves to Share and Use Real-world Evidence

Erin M. Henninger¹, Selvin Soby¹, Manuel Wahle¹, Boudewijn Aasman¹, Pavel Goriacko², Chandra has Nelapatla¹, Reetam Nath¹, Parsa Mirhaji²

¹ Montefiore Medicine, ² Albert Einstein College of Medicine at Montefiore

Background

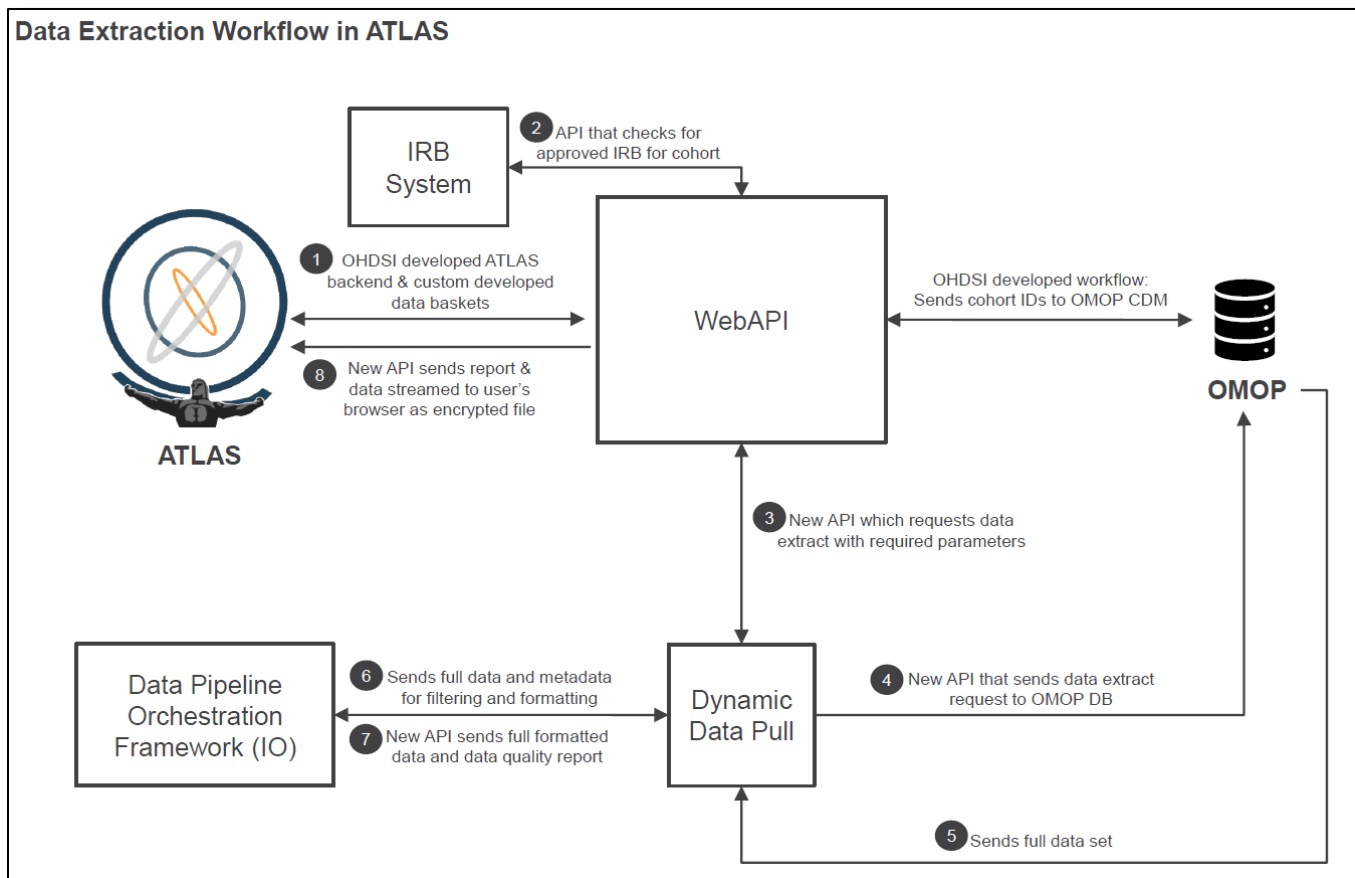
Implementing an OMOP Common Data Model involves identifying data sources, extracting required data, mapping to standardized terminologies, transforming logic to OMOP format, loading to specified data structures, and conducting quality checks.¹ Institutions can implement the OMOP CDM in two ways: create a cohort-specific instance for specific clinical or research use cases (e.g., Covid Cohort Collaboratives), or create an institutional clinical research data warehouse with all available patients. Cohort-specific instances are useful for researchers running various types of analysis, including sophisticated prediction or machine learning pipelines. Developing and maintaining cohort-specific OMOP-CDMs is a resource-intensive process, either involving the creation of separate instances for each research use-case or having to provide tools for researchers which requires the use of their own infrastructure.²

Our project describes and extension of OHDSI's Atlas which enables users to automatically export data pertaining to a specific cohort and specified variables, from a clinical data warehouse into a cohort-specific OMOP data-mart within our university's High-Performance Computing (HPC) Enclave. This allows providing exclusive access to a specific dataset and creating shared data analytic resources across communities of research inside the institution, without exposing information about the overall patient population. The Dagster automation and orchestration framework is used to keep the satellite (mini-OMOP) data-marts always updated and in sync with the main clinical data-warehouse. Investigators can freely extend and incorporate derivatives and new data features into their exclusive enclaves while maintaining an updated access to all data points longitudinally. All data enclaves are centrally managed through linkage between the IRB-WEB API and Atlas for regulatory compliance². This also allows analytic teams to develop their own custom algorithms and innovative analytics while continuously benefiting from existing libraries and community developed tools and services.

The HPC at Albert Einstein College of Medicine offers researchers access to computing resources, storage arrays, and expertise for running scientific applications. These resources are enabled using serial cluster nodes, large memory and parallel systems, and scheduling systems. Importantly, the HPC is compliant to the highest standards so raw identified or deidentified data is kept secure project initiation all the way to the end.

Methods

Using the data extraction framework, we had developed in the past (figure 1), we extended the capabilities to provide an additional full OMOP-CDM format. If our advanced user wants to create a cohort-specific OMOP-CDM, the system first checks for IRB approval and the existence of a previous environment for that cohort². If an environment already exists, the user is given the option to either append to the existing environment or create a new one, granted they have the necessary permission. The creation of a new environment can be achieved through a Docker container with a Postgres database or by utilizing the institution's HPC cluster. To facilitate data extraction, our query engine, Dynamic Data Pull (DDP), sends the configured extraction query from the user's data basket to the Spark cluster, which handles the streaming of the data².



Results

The mini-OMOP enclaves have been successfully deployed and utilized to support multiple departmental and multi-institutional projects that required construction of dynamically updating patient registries and developing and running custom developed analytics regularly. Some examples of ongoing research projects using the mini-OMOP secured enclaves as follows:

1. Radiology and Imaging Registry of Covid Positive Patients (+ DICOM images)
2. Montefiore Liver Transparent Registry
3. Montefiore Sepsis registry
4. Medical Student Computable Phenotyping Projects

Conclusion

By combining ATLAS cohort definition and generation functionality, our scalable Data Extraction and Basket functionality, IRB oversight and audit, and mini-OMOP secured enclaves, we have been able to democratize and provide easy and secure access to real-world evidence across interdisciplinary teams who would like to develop their own analytics without being disrupted by a centrally controlled infrastructure, all while continuously remaining in sync with the centralized clinical data warehouse.

References:

1. Biedermann, P., Ong, R., Davydov, A. A., Orlova, A. V., Solovyev, P., Sun, H., Wetherill, G., Brand, M., & Didden, E. (2021). Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Medical Research Methodology*, 21(1). <https://doi.org/10.1186/s12874-021-01434-3>
2. Mirhaji, P., Soby, S., Henninger, E., Nelapatla, C., Wahle, M., Aasman, B., Belin, E., Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance. Presented at 2022 OHDSI Symposium in Baltimore, MD.