Patient-Level data export based on ATLAS Cohorts

ATLAS is a crucial tool in healthcare data analysis, but sharing insights while protecting privacy is challenging. To address this, we have developed a method to export an OMOP database with only the subjects and data defined in a Cohort Definition in ATLAS. By extracting a subset of the OMOP database based on identified cohort patients and anonymizing identifiers, we ensure compliance with data protection regulations while facilitating collaborative research efforts.

Methods

We developed an export process that requires cohort definitions and a list of sources. Initially, we retrieve person and concept identifiers from the hospital and Atlas databases, respectively, as filters for data retrieval. Then, using OMOP schema-defined foreign keys, we query relevant references to these identifiers. Anonymized IDs are generated to prevent data collisions and reidentification. Data extraction involves creating a new OMOP schema and joining staging tables with original ones while maintaining anonymized identifiers. Certain columns are masked for GDPR compliance. Ultimately, we achieve a unified and anonymized OMOP instance from all sources.

Results: The utilization of this export process has been successful with cohorts exceeding 15000 patients, integrating data from over 10 distinct sources. This implementation facilitated validation within a targeted cohort and enabled the successful dissemination of patient-level results to stakeholders, while ensuring compliance with privacy and integrity requirements.