As we finalize the LAISDAR project, we are set to begin the next chapter of OMOP-based federated networks both in Rwanda and around the African continent; Ares, along with other powerful open-source OHDSI tooling, will be central in those efforts.

Title: Implementation of the ARES application to monitor network-wide data quality and mapping coverage for 16 unique OMOP sources across Rwanda

BACKGROUND: Leveraging Artificial Intelligence and Data Science Techniques in Harmonizing, Accessing and Analysing SARS-COV-2/COVID-19 Data in Rwanda, or LAISDAR, aims to establish a nation-wide federated data network based on the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) [1, 2]. The project was initially intended to support research on COVID-19 but given the quantity and quality of electronic health record (EHR) data available at the various participating hospitals, the scope has since widened to other relevant communicable and noncommunicable disease areas. Most Rwandan hospitals have implemented one of two EHR systems for storing electronic medical data: openClinic GA and openMRS [3]. A first step in the project was to define structural and semantic mappings and logic to transform the two source systems to the OMOP CDM. We have since developed an Extract-Transform-Load (ETL) process that can transform data in either source EHR system to the target format and that uses a variable switch to select EHR-specific transformations and mappings. Currently, the network comprises 16 OMOP instances; 14 different hospitals have EHR data in OMOP format, with a combined total of more than 3.5M (approximately 25% of the national population) individuals represented. Additionally, we have transformed national data related to tests results for COVID-19 into an independent OMOP dataset, and we have also transformed COVID-19-related survey data from a 10’000+ participant survey conducted in early 2022 into a separate OMOP instance. Tracking data quality, mapping coverage, and transformation versions across a network of this size is nontrivial; for this task, we have employed a set of new OHDSI tools, Ares [4] and AresIndexer [5], which when combined compile aggregate statistics and information about a set of OMOP data sources and releases for validation and exploration. In this work, we present our experiences using Ares, with the intention of (1) highlighting the power and ease-of-use of the tool, and (2) motivating others facing similar multi-OMOP-source challenges to implement the tool as a plug-and-play solution.

LIMITATIONS AND DISCUSSION: Ares has been tremendously useful in compiling a network-wide overview of pain points, data gaps, and deployment status across this network. Importantly, the AresIndexer relies on common dependencies (e.g. Achilles) used by other OHDSI tooling, so integrating it into the deployment workflow is both simple and computationally efficient. In this case, we created two separate docker images (Ares web application & AresIndexer) that can be deployed quickly and easily on different operating systems and architectures. We plan to make these images publicly available in the coming weeks. Used in combination with a remote orchestration tool like SimpleMDD, Ares completes a powerful feedback loop for ensuring data quality and enables a detailed overview that is critical for defining network-wide studies. The most important functionalities Ares has provided in this project are: (1) a dynamic benchmark that motivates data managers to improve their local data quality and to participate actively in the network, (2) an overview of the diverse data quality issues that stem from subtle differences (e.g. date handling, local mappings) between EHR system configurations, and (3) a detailed tool for evaluating feasibility and potential impact of future federated studies. Given that the Ares tool is still in beta testing, we look forward to updated functionality and capabilities in future releases, and we plan to take an active role in its development moving forward.