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

Jared Houghtaling[a], Emma Gesquiere[a], Lars Halvorsen[a], Marc Twagirumukiza[b], and Charles Ruranga[c]

[a] edenceHealth NV (BE)
[b] University of Gent (BE)
[c] University of Rwanda (RW)

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 test 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.

Methods:

We have provided each hospital with a dedicated Mac Mini (2021, M1 processor, 8-16 GB RAM, 500-1000 GB Disk), all of which are connected to a Mobile Device Management client (SimpleMDM) used to deploy software packages and launch scripts remotely. In support of the project, we have created a suite of Docker images that enable us to support the services shown in **Figure 1.** All hospitals currently have this software suite installed; ETL processes have been executed end-to-end on all Mac Minis, and 11 sources are currently visible in the central Ares application. To support a centralized view on the various OMOP instances, we configured a central server (Ubuntu 20.04.6) managed on the premises of the Ministry of Health in Rwanda, which houses both the COVID case information and survey data in OMOP CDM format, along with the Arachne Central service and an ETL-enrichment API. We have deployed Docker-based versions of both the AresIndexer (for generating the network-wide index) as well as the Ares web application to this location as well. We have established a four-step orchestration for the ETL deployments and subsequent Ares processes via SimpleMDM, in which: (1) a hospital Mac Mini receives a script to launch a Docker-based ETL process, (2) upon completion, a second script initiates Achilles, DQD, and the AresIndexer processes, (3) a third script securely transfers the AresIndexer output to a common location on the central server, and (4) the appearance of the updated directory on the central server triggers AresIndexer on the central server to rebuild the network index, after which all results can be viewed and contextualized on a protected URL.

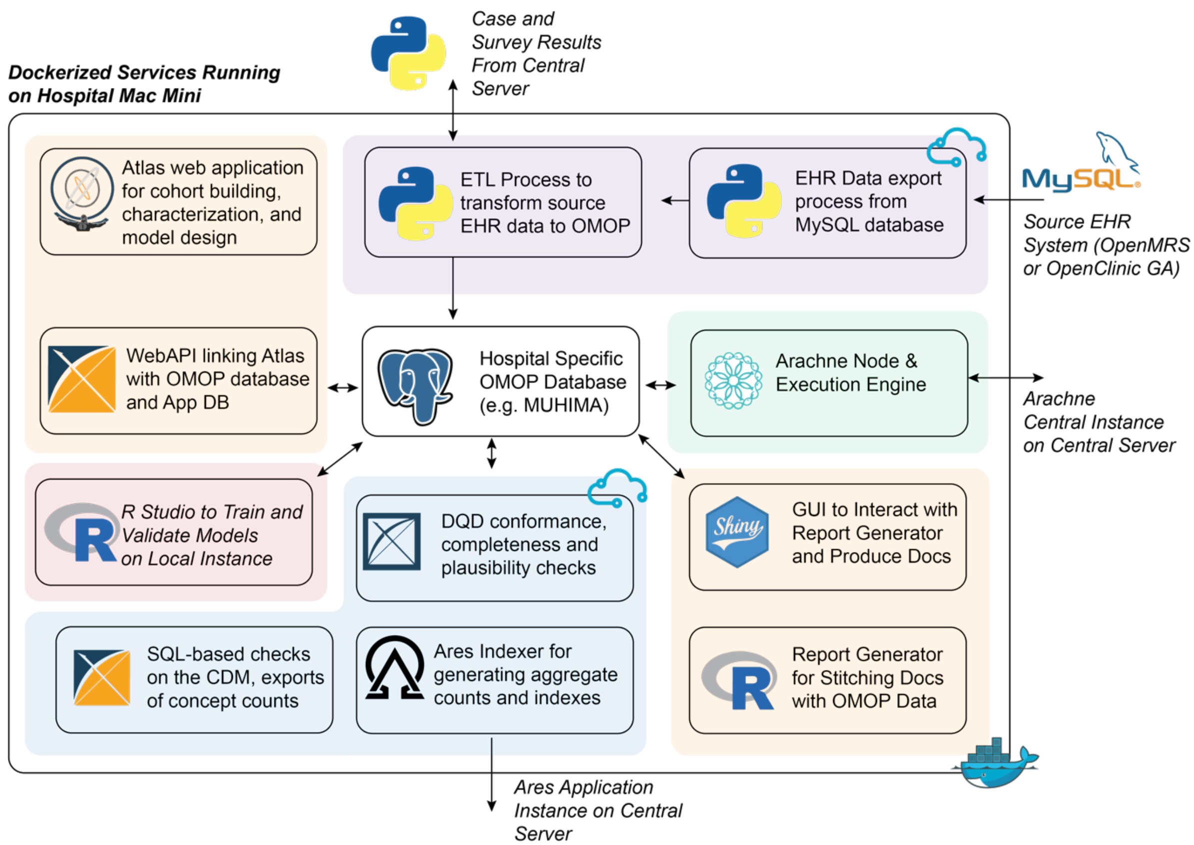

Figure 1. Services deployed (via Docker) on each hospital's Mac Mini machine, and directional connections between the Mac Mini and central server, as well as between the Mac Mini and the local EHR system.

Results:

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 SimpleMDM, 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.

Conclusion:

As we finalize the various implementations for 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 will play a critical role in these efforts, and we would like to acknowledge both Frank DeFalco and the rest of the OHDSI community who have devoted significant time and energy toward developing this invaluable tool.

References:

[1] https://rbc.gov.rw/laisdar/

[2] Nishimwe, A., Ruranga, C., Musanabaganwa, C. et al. "Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analyzing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR Project): study design and rationale." *BMC Medical Informatics and Decision Making 22, 214 (2022).*

[3] Akanbi, Maxwell O., et al. "Use of electronic health records in sub-Saharan Africa: progress and challenges." *Journal of Medicine in the Tropics* 14.1 (2012): 1.

[4] Ohdsi. (n.d.-a). *GitHub - OHDSI/Ares: A Research Exploration System*. GitHub. https://github.com/OHDSI/Ares

[5] Ohdsi. (n.d.). *GitHub - OHDSI/AresIndexer: R package that creates the index and relevant files for an Ares deployment.* GitHub. https://github.com/OHDSI/AresIndexer