LAISDAR – A federated data network to support COVID-19 research in Rwanda

Lars Halvorsen[a], Freija Descamps[a], Jared Houghtaling[a], Benjamin Burke[a], Francine Birungi[b], Clarisse Musanabaganwa[c], Jean Baptiste Byiringiro[c], Stefan Jansen[b], Celestin Twizere[d], Kizito Nkurikiyeyezu[d], Charles Ruranga[e], Aurore Nishimwe[b], Regine Mugeni[f], Jean N Utumatwishima[f], Damas Kabakambira[g], Sabin Nsanzimana[c], Marc Twagirumukiza[h]

[a] edenceHealth NV, Belgium
[b] College of medicine and health sciences, University of Rwanda, Kigali, Rwanda
[c] Rwanda Biomedical Center, Ministry of Health, Kigali, Rwanda
[d] Center of Excellence in Biomedical Engineering and eHealth, University of Rwanda, Kigali, Rwanda
[e] Center of Excellence in Data Sciences, University of Rwanda, Kigali, Rwanda
[f] Rwamagana District hospital, East province, Rwanda
[g] Kigali Central Teaching Hospital (CHUK), Internal medicine Department, Kigali, Rwanda
[h] Faculty of medicine and Health sciences, Ghent University, Ghent, Belgium

Background

COVID-19 was first confirmed in Rwanda on March 14th 2020, and confirmed cases have since been increasing across the nation, initially due to international movement, and more recently through community spread. As of August 30th, 2021[1], 2.4 million people have been tested in the country, leading to 87.131 confirmed positive cases, and total number of deaths of 1083 on a total Rwandan population of 11 million. However, because testing has been limited to those patients with known COVID-positive contacts, the true number of positives cases (including mild and asymptomatic infections) is assumed to be higher.

By now, there is a vast amount of SARS-COV-2/COVID-19-related data collected through COVID-19 testing, surveys, routine care, and other sources. This data has the potential to substantially improve our understanding of the disease, monitor its progress and generate evidence to guide prevention measures. In addition, it can lend crucial insights into outcomes, which can enable efficient preventive or treatment measures.

It is common, both in Rwanda and elsewhere, for health-related (and specifically, COVID-19) data to be assembled in different formats and scattered across multiple institutions including hospitals, clinics and testing sites. Analyzing those fragmented COVID-19 datasets individually reduces the potential of uncovering important evidence because of limited sample size. On the other hand, combining these datasets together is challenging because (1) they have different data structures and (2) data owners may fear breaches of data privacy. Therefore, an innovative approach to analyze data from multiple data sources was sought.

A consortium made up of multiple Rwandan and Belgian institutions has been established[2], managed by University of Rwanda (UR), under the project titled: "*Leveraging Artificial Intelligence and Data Science Techniques in Harmonizing, Sharing, Accessing and Analysing SARS-COV-2/COVID-19 Data in Rwanda (LAISDAR)*". The LAISDAR project received funding from Canada's International Development Research Centre (IDRC)[3] as part of the Global South AI4COVID program[4].

The project objective is to leverage the federated hospital data sets, extended with data from centralized COVID-19 test results and survey data, to support Rwandan government needs in monitoring and predicting the COVID-19 burden, including on hospital admissions and overall infection rates. The impact

of various public health measures on the pandemic evolution, social-economic situation, and mental health are also key study objectives.

Methods

The LAISDAR project will be a federated data network, based on the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), as well as on open-source Observational Health Data Sciences and Informatics (OHDSI) tools for data analytics and network integration, and R Studio and Python.

The network will include several hospitals, whose EHR (Electronic Health Records) data will be harmonized to OMOP CDM, and enriched both with COVID-19 test results and COVID-19 survey results from a national database.

Figure 1. LAISDAR participants

There are 2 different open source EHR systems used by the participating hospitals; OpenMRS[5] and OpenClinic GA[6]. Therefore, two different ETL (Extract Transform Load) processes will be implemented, with the objective of having as few local adaptations per hospital as possible.

Enrichment of EHR data is part of the ETL process, whereby available COVID-19 test and survey results will be retrieved from a central repository over a secure interface. One critical challenge with this step is consolidating individual patients; different person identifiers (national ID, mobile number, name, address), or combinations thereof, are used across different systems. We envision generating unique

identifiers based on the available keys to facilitate reliable and reproducible matching of records from the different source systems.

The integration of the sites with a central hub will be accomplished by using the open-source version of Arachne[7], which provides a platform for performing network studies, integrating OHDSI standards and tools.

Software deployments at the participating hospitals will rely on Docker-based containerization; this approach ensures consistent and reproducible installation across the different sites. For most participating hospitals, a pre-configured Mac Mini will be provided with the complete LAISDAR Dockerized software suite.

Training for the end users/researchers will be provided as a combination of user documentation, online training, and live training sessions. All technical deliverables will be documented, ETL source code provided, and handover to and training for local technical personnel envisioned to ensure continuity beyond the project lifetime.
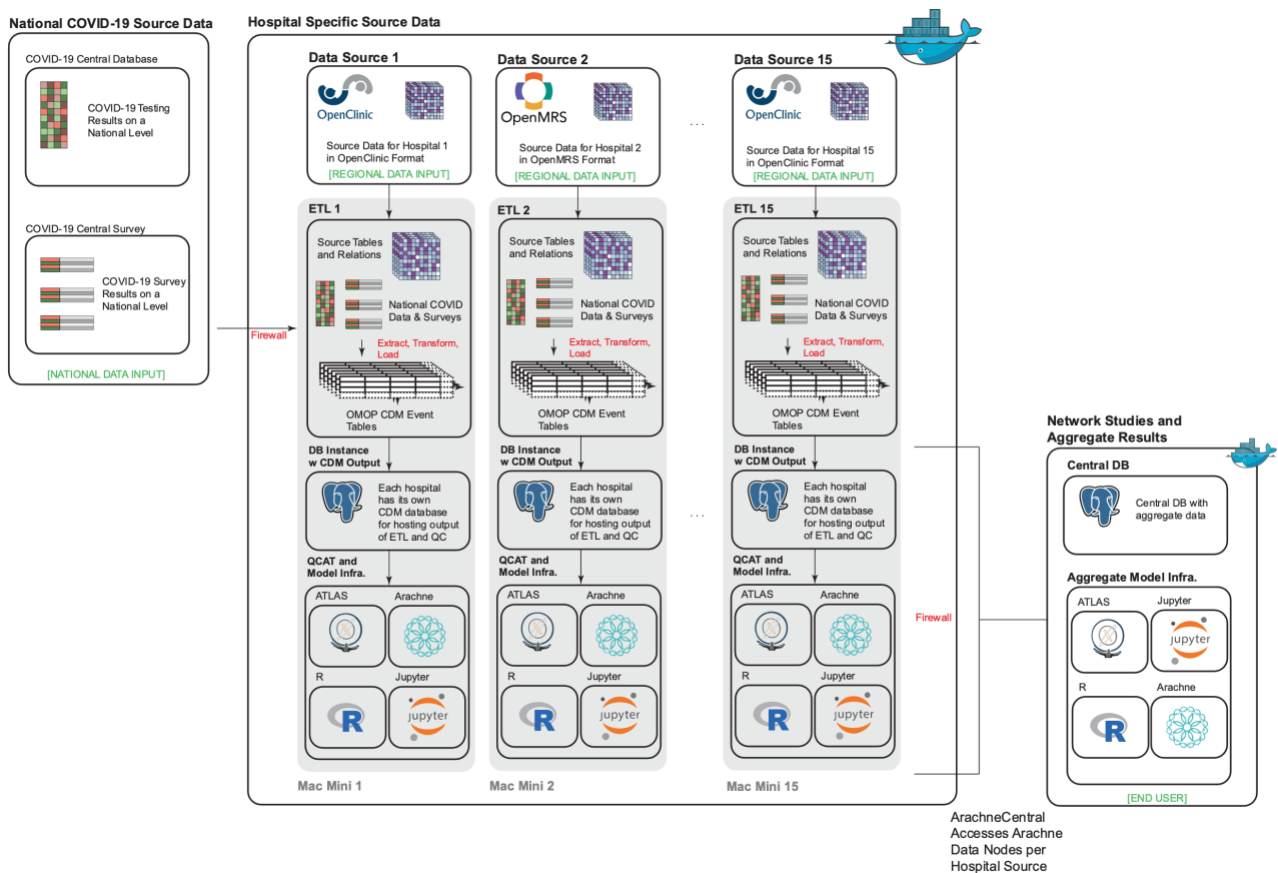

Figure 2. Overall LAISDAR architecture

Results

An initial proof-of-concept (POC) implementation was set up and tested, which included the central LAISDAR instance and 2 data nodes – one on a Mac Mini and one on an AWS EC2 instance. The data nodes were set up with docker containers providing the following services: a (PostgreSQL) OMOP CDM database, Atlas/webAPI, Achilles and Arachne (connected to the Arachne Central instance).

The central server was set up with Arachne Central, where the Data Catalogue was configured, and studies were created and executed for testing the integration at the data nodes.

The objective of the POC was to test the integration layer (Arachne), as well as demonstrating the overall process flow for network studies; these objectives were met.

The next phase of the development is well underway, which includes the completion of the ETL implementations, and the integration with the central COVID-19 test and survey results.

The first phase of the project will include 15 hospitals, with the intention of including additional hospitals in a later phase of the project.

Conclusion

The project is the first on the African continent implementing data harmonization on COVID-19. The design and implementation of an OMOP CDM based federated data network for COVID-19 related studies in Rwanda will provide researchers in Rwanda and elsewhere with the tools and data access needed to better track the disease, predict outcomes, and plan appropriate responses.

The chosen architecture lends itself to expansion to additional hospitals or other data sources, should there be a need.

Building the LAISDAR infrastructure on the open-source OMOP CDM data model and utilizing OHDSI tools and other open-source tools facilitates easy involvement of new partners. In addition, these choices provide opportunities for participation in other OHDSI based network studies around the world.

References/Citations

1. Rwanda Biomedical Centre, Coronavirus Disease COVID-19.
   https://www.rbc.gov.rw/index.php?id=707
2. University of Rwanda (UR), Rwanda, Rwanda Biomedical Center (RBC), Rwanda, University of Ghent (UGent), Belgium, and the Regional Alliance for Sustainable Development (RASD), Rwanda.
   http://laisdar.rbc.gov.rw
3. International Development Research Centre (IDRC), Canada. https://www.idrc.ca/en
4. Global South AI4COVID Program. https://covidsouth.ai
5. OpenMRS: open-source platform to build customized EMR system. https://openmrs.org
6. OpenClinic GA: open-source integrated hospital information management system.
   https://sourceforge.net/projects/open-clinic/
7. Arachne: https://github.com/OHDSI/ArachneNodeAPI