

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

Aurore Nishimwe (✉ auroreshimwa@yahoo.fr)

University of Rwanda

Charles Ruranga

University of Rwanda

Clarisse Musanabaganwa

Ministry of Health

Regine Mugeni

Rwamagana Provincial hospital

Muhammed Semakula

Ministry of Health

Joseph Nzabanita

University of Rwanda

Ignace Kabano

University of Rwanda

Annie Uwimana

University of Rwanda

Jean N Utumatwishima

Rwamagana Provincial hospital

Jean Damascene Kabakambira

The University Teaching Hospital of Kigali (CHUK)

Annette Uwineza

The University Teaching Hospital of Kigali (CHUK)

Lars Halvorsen

edenceHealth NV

Freija Descamps

edenceHealth NV

Jared Houghtaling

edenceHealth NV

Benjamin Burke

edenceHealth NV

Odile Bahati

Regional Alliance of Sustainable Development

Clement Bizimana

Regional Alliance of Sustainable Development

Stefan Jansen

University of Rwanda

Celestin Twizere

University of Rwanda

Kizito Nkurikiyeyezu

University of Rwanda

Francine Birungi

University of Rwanda

Sabin Nsanzimana

Ministry of Health

Marc Twagirumukiza

Ghent University

Research Article

Keywords: Artificial Intelligence, Machine Learning, Data Science, SARS-COV-2/COVID-19, Rwanda

Posted Date: March 30th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1418826/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Since the outbreak of COVID-19 pandemic in Rwanda, a vast amount of SARS-CoV-2/COVID-19-related data have been collected including COVID-19 testing and hospital routine care data. Unfortunately, those data are fragmented in silos with different data structures or formats and cannot be used to improve understanding of the disease, monitor its progress, and generate evidence to guide prevention measures. The objective of this project is to leverage the artificial intelligence (AI) and data science techniques in harmonizing datasets to support Rwandan government needs in monitoring and predicting the COVID-19 burden, including the hospital admissions and overall infection rates.

Methods: The project will gather the existing data including hospital electronic health records (EHRs), the COVID-19 testing data and will link with longitudinal data from community surveys. The open-source tools from Observational Health Data Sciences and Informatics (OHDSI) will be used to harmonize hospital EHRs through the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The project will also leverage other OHDSI tools for data analytics and network integration, as well as R Studio and Python. The network will include up to 15 health facilities in Rwanda, whose EHR data will be harmonized to OMOP CDM.

Expected results: This study will yield a technical infrastructure where the 15 participating hospitals and health centres will have EHR data in OMOP CDM format on a local Mac Mini (“data node”), together with a set of OHDSI open-source tools. A central server, or portal, will contain a data catalogue of participating sites, as well as the OHDSI tools that are used to define and manage distributed studies. The central server will also integrate the information from the national Covid-19 registry, as well as the results of the community surveys. The ultimate project outcome is the dynamic prediction modelling for COVID-19 pandemic in Rwanda.

Discussion: The project is the first on the African continent leveraging AI and implementation of an OMOP CDM based federated data network for data harmonization. Such infrastructure is scalable for other pandemics monitoring, outcomes predictions, and tailored response planning.

Introduction

In December 2019, a critical respiratory disease from an unknown cause was identified in Wuhan, China [1]. Thereafter, the causative pathogen was discovered as a novel coronavirus and was named the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1–3]. Since then, the SARS-CoV-2 virus has quickly spread across China and worldwide [1, 4–6]. Since the first released reports of the confirmed cases of the coronavirus disease of 2019 (COVID-19) in Wuhan, China, the whole world has witnessed severe unprecedented mortality and morbidity due to this disease resulting in serious public health emergencies that defined the disease as a global pandemic [7–9]. Despite the delayed severity of the pandemic in Africa, a strong warning was given to the continent due to existing socio-economic and health-related challenges [10–12]. The first patient diagnosed with COVID-19 was detected in Rwanda in

March 2020 [13]. After detection of the first case, different public health measures have been implemented by the Rwandan government from total lockdowns, inter-district lockdowns, localized lockdowns and others [14, 15]. The classic public health measures to prevent COVID-19 were also emphasized such as the mandatory wearing of masks, social distancing and regular handwashing [16, 17]. Rwanda has also joined the rest of the world to secure vaccines against COVID-19 [18]. Despite all efforts, Rwanda has exceeded 80,000 cases of COVID-19 with more than 1000 deaths and numbers keep increasing [13].

In addition, with the apparent unpredictability of the increase or decrease of COVID-19 cases, both the decision-makers and the general population live in an uncertain situation [19]. Accurate short-term forecasting of COVID-19 spread plays an essential role in improving the management of the overcrowding problem in hospitals and enables appropriate optimization of the available resources [20]. This forecasting effort helps also to reduce the burden of COVID-19 in terms of planning and adjustment of public health measures. This study, the first to our best knowledge in Rwanda, that has proposed the building of data hubs which will later be used to design COVID-19 prediction models using Artificial intelligence (AI) and Machine Learning (ML) techniques. Recently deep learning methods have gained particular attention in time-series modeling and analysis because of their outstanding generalization capability and superior nonlinear approximation [20–23].

In Rwanda, the use of AI and data science techniques is motivated by a large implementation of electronic medical record (EMR) systems in the health care facilities which makes data accessible. However, the data is fragmented, as health facilities are using different EMRs, and the COVID-19 data were not systematically collected. To effectively visualize and re-use the data, there's a need to create a centralized common data model using already collected data for both COVID-19 and other medical conditions. In addition to understanding the spread of COVID-19 data, prospective understanding of respect of public health measures is also mandatory. For example, in the study by Barak et al. (2020), they emphasized that obtaining information on symptoms dynamics is of great importance to control the complications of the disease in the population [24]. In the current study, not only symptoms will be assessed longitudinally, but also any reinforcement of COVID-19 measures will be captured. The Rwanda Artificial Intelligence (AI) project will leverage AI and other Data Science (DS) techniques to create a scalable framework for inventorying, harmonizing and federating the accumulated data from COVID-19 patients and converting it to a standardized data format so that it can be used as part of wider studies on the disease. The harmonized data will consist of COVID-19 diagnosed/serotyped patient data and non-infected individuals from electronic health records (EHRs) of different hospitals and databases of testing centers (positive and negative results). In the second phase of the project, we will collect new data longitudinally. Those newly collected longitudinal data will be enriched with the patient-reported outcome (PROs) and will be in a similar standardized model by design.

The project outcome is to leverage all federated data with Machine Learning (ML) and other mathematical methods to drive evidence. This evidence will fulfil Rwanda's needs and priorities in predicting and monitoring the burden of COVID-19 pandemic in the Rwandan community, on hospital

admissions related to COVID-19 and overall COVID-19 infection rates. The generated evidence will also monitor the impact of different public health measures on the COVID-19 pandemic evolution in Rwanda. This project will also add new knowledge to AI-guided prediction models to control epidemics in Sub-Saharan Africa.

Materials And Methods

Study setting

The study sites will include 13 hospitals (Kigali University teaching hospital, Butare University teaching hospital, Ngoma regional hospital, Ruhengeri provincial hospital, Muhima district hospital, Kibagabaga district hospital, Nyamata district hospital, Nyagatare district hospital, Kinihira district hospital, Kigeme district hospital, Kirehe district hospital, Gisenyi district hospital and Gihundwe district hospital); two health centers (Remera health center and Nyamata health center); and once centralized dataset gathering 22 COVID 19 test centers (Kanyinya, Rwankeri, gatenga, Kicukiro, ASPEK-Ngoma, Kigali Transit Centre, Rugerero, Kabgayi and Rusizi). These study sites have been selected to include all four provinces of the country for prediction and generalizability purposes.

Inclusion and exclusion criteria

The included health facilities were selected based on the availability of EMRs at hospitals. For the community surveys, all participants aged 18 years and above will be eligible to enter the study. Participants will only be excluded if they have no eligible household phone contacts. According to the national regulations, all participants will provide consent to participate in the study. The consent will be electronically signed and embedded into a mobile application used for surveys.

Study design

The LAISDAR project is 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. As demonstrated in Fig. 1, 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, COVID-19 survey results from a national database, and the results of the community surveys. 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 (Amazon Elastic Compute Cloud) instance.

There are 2 different open source EHR systems used by the participating hospitals; OpenMRS [25] and OpenClinic GA [26]. Therefore, two different ETL (Extract Transform Load) processes will be implemented in order to have 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 a 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, 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.

Conceptual Framework for hospital EHRs harmonization

The conceptual framework is presented in Fig. 2. After inventorying existing datasets this project will set up an infrastructure for data harmonization where novel techniques will be developed including the data access and data analysis interface where we will mix the existing methods and innovative techniques. Within this conceptual framework it's planned to include: a data catalogue describing the different data sources, the Arachne central hub, a central OHDSI Atlas instance, a central database, as well as R Studio and Jupyter.

Data collection, analysis and management

The study will involve four main steps with regards to data collection, analysis and management:

Step 1: Data gathering /collection

That includes the inventory of the existing data from the first 24 months of the COVID- 19 pandemic in Rwanda (the first case was identified in march 2020) and the 20 weeks' data collection (through community surveys via telephone calls).

It's anticipated to have different formats of data sources ranging from Covid-19 related data registered in Excel documents, via data sources containing Minimum Clinical Data (MCD) in DHIS2 [25] and other systems, to more granular Electronic Medical Record (EMR) data in Open Clinic GA [26], OpenMRS [25] and other EMR systems. We will start by mapping full hospital patients' records, focusing on 15 health facilities located in regions with a high number of COVID-19 patients and completing with other isolated datasets.

The new data collection (community surveys) will use standardized protocol and questionnaires and will be done according a longitudinal approach. The participants will be randomly sampled following the

sampling frame used for the recent Rwanda Demographic Health Survey (based on the fourth Rwanda Population and Housing Census RPHC) provided by the National Institute of Statistics of Rwanda (NISR) [27]. This sampling frame is a complete list of districts covering the whole country. The data collection is done through be-weekly phone call, 6 phases planned (starting by December 2021), involving 30 well trained data collectors, one per district supervised by 10 investigators. The minimum sample size required was estimated at 107 people per each of 30 administrative district in Rwanda. To anticipate on consent refusal and drop outs, we doubled this number making 214 participants in each district. If the participant has a medical file in participating hospitals, or in other COVID-19 testing dataset, the datasets will be linked with possibilities of other linkage data request in future.

The sample will proportionally include males and females based on the number of inhabitants. Each participant will receive a mobile fee connection and internet bundle each week to allow data collection. To mitigate the expected gap of the gender digital divide but also of selected persons without a mobile phone anymore, the consortium established mitigation measures, including but not limited to, leveraging the community healthcare workers (CHWs). Each village in Rwanda has a CHW who participates in various ministry of health (MoH) programs and they have all received the mobile phones from MoH. If we select a respondent without a mobile phone we will liaise with the nearest CHW to reach out to him.

The questionnaires (which will be translated into 3 languages, Kinyarwanda, English and French) include 10 modules (at least 8 of them has to be fulfilled by the project): 1) Demographics; 2) Face mask use; 3) Hand hygiene; 4) Respect of social distancing measures and risk minimization measures; 5) Recent risk situations exposures and COVID-19 measures. On the outcome side, the collected data will include 6) Coronavirus like-Signs and symptoms; 7) Mental health indicators (based on General anxiety disorder-GAD); 8) Social-economic impact (based on loss of income, or categories) and 9) Covid-19 test results [28].

Gender considerations.

The SARS-CoV-2 virus does not discriminate. In order to respond effectively to the crisis, we need a whole-society approach to understand its differential impact on women and men. Supporting gender analysis and sex-disaggregated data is an integral part of this project. The gender COVID-19 related data are still scarce and a little is known in Rwanda on the topic. Therefore, collecting related data will be a key activity to bridge the gap and contribute to best gender driven policies locally and in the region. Specifically, we articulated the gender considerations in this project in a multi-lever approach including: (1) Fostering gender balance throughout in research teams, in order to close the gaps in the participation of women, and (2) Integrating the gender dimension in research and innovation (R&I) content that will help improve the scientific quality and societal relevance of the produced knowledge, technology and/or innovation. Gender has been integrated as a transversal theme and not a vertical aspect. Gender facets are found in all COVID-19 consequences including morbidity in general and mental health problems in particular and socio-economic outcome. Social and cultural factors related to gender such as specific considerations for some collected data elements will be addressed as well, eg. reproductive health data, the usage of

gender-sensitive research questions and gender-impartial language. Moreover, the sampling will pay special emphasis to gender proportional balance while collecting new COVID-19 data and gender key output/aspects will be driven from data analysis.

Step 2: Infrastructure for data harmonizing (developing novel techniques)

For data harmonization, the custom-designed ETL scripts will be developed per data source to extract, transform and load the source data to an OMOP CDM database instance. In the early stages, when the hospital EHRs are not yet harmonized, we will also use synthetic data approaches to help automate harmonization processes. The data owner-side infrastructure will include the OMOP CDM database instance, the Arachne client, the OHDSI Atlas [29] analytical tool, R Studio [29], and Jupyter [29]. The data harmonization process converts the observational data from the format of the source data system to the OMOP CDM supported by the OHDSI organization.

Step 3: Infrastructure for data access, query, and data analysis (Mixing existing methods and innovative techniques)

The central platform data access, query, and data analysis, will handle the participating data sources. The central site will use Arachne for the central portal with the data and study catalogues, but also a PostgreSQL database, an OHDSI Atlas analytical platform instance and an R Studio instance. Additional tools can be added like a Jupyter server instance. As a standard, the database will include an OMOP CDM schema, and additional schema(s) to support the central data catalogue. At the beginning of the harmonization process, as the data from hospitals EHRs will be not yet available, this project will use synthetic data to help automate harmonization processes and training models, specially we will use the OHDSI community's available mock-up data (like Synthea) to train different algorithms /models before we use them on real data.

The Arachne central server setup will allow central management of network studies, with tight integration with the OHDSI tools such as Atlas. The Arachne data catalogue will incorporate the Achilles output from each participating site; the Achilles tool generates a profile of the participating sites' data on an aggregate level, which will allow a central view of the descriptive statistics for each site. The R Studio and Jupyter instances will allow the development and testing of R scripts as part of a study design, or to analyze data collected from data source sites as part of studies.

Step 4: Data analysis and interpretation (Mixing existing methods and innovative techniques)

The federated datasets are challenging to analyze with traditional statistical methods, because they are, like other real-world-data (RWD), 1) collected without any intention for being used in research; 2) incomplete and not cleaned and 3) collected in a sporadic way, not pure longitudinal approach so no way to derive cohort-like data from them.

The current project will leverage AI techniques including Machine learning techniques and data mining that bring an added value in discovering hidden patterns or relationships between data points. In this project, like in other similar works [20], we will first evaluate the performances of different deep learning methods, including the hybrid convolutional neural networks-Long short-term memory (LSTM-CNN), the hybrid gated recurrent unit-convolutional neural networks (GAN-GRU), GAN, CNN, LSTM, and Restricted Boltzmann Machine (RBM), as well as baseline machine learning methods, namely logistic regression (LR) and support vector regression (SVR) [20]

Additionally, we will evaluate the added value of the dual mode system consisting of (1) a continuous-time version of the Gated Recurrent Unit (GRU), building upon the recent Neural Ordinary Differential Equations (ODE), and (2) a Bayesian update network that processes the sporadic observations (GRU-Bayes). With this new approach the GRU-ODE [30], is responsible for learning the continuous dynamics of the latent process that generates the observations and GRU-Bayes, responsible for dealing with incoming observations and updating the conditional current estimate of the latent process [30]. Those two steps and modules are similar to the propagation and update steps of a Kalman filter. With GRU-ODE, we expect to evaluate the capacity to project in time the hidden process $h(t)$ and hence indirectly future observations. GRU-Bayes performs the update of the hidden state conditioned on new observations.

After this evaluation a final method will be implemented. It's hypothesized however that the hybrid models (i.e., LSTM-CNN and GAN-GRU) will potentially improve the forecasting accuracy of COVID-19 future trends, based on previous works [31].

In prediction models, the sequential reproduction number $R(t)$ will be estimated using the Bayesian approach on the Extended SEIR compartmental model. The Bayes rule is used to update the beliefs about the true $R(t)$ based on model predictions and new cases that have been reported each day.

Model definition

We will use an extension of the SEIR model (Fig. 3) inspired by the previous works [32]. This model splits the population into different categories, i.e. susceptible, exposed, infected and removed. The latter two categories are further broken down into super mild, mild, heavy and critical for the infected part of the population, whereas the removed population indicates the immune and dead fraction. A super mild infection refers to the category of asymptomatic people who are infected but are unaware of their own infection. Recent figures from Chinese scientists put this number at 86% of all infections [32].

Transitioning between different fractions of the population is indicated by the arrows and its rates are expressed by parameters in the model. The two most important parameters in such a model are: 1) the incubation period and 2) rate of virus spread. Other parameters include the odds of having a super mild, mild, heavy or critical infection. For each type of infection, there is an infectious period, etc. All parameters except one were gathered from the available literature on coronavirus. The parameter that remained to be calibrated is 'beta', which determines the rate of transitioning individuals from susceptible to exposed. Beta can be interpreted as the degree of social interaction or the amount of exposure to the

virus. It is this parameter that is targeted when governments impose restrictions on their citizens. We will, therefore, focus on this parameter. Finally, a documented mathematical model will be discussed at a later stage, at the beginning of the project implementation.

Study limitations

We acknowledge the limitations imposed by a federating network approach used by LAISDAR project, where the data remains at each participating hospital's site, rather than the patient-level data being pooled in a central location. There are different approaches to accommodate a federated approach for the ML methods described here; the main challenge relates to training any model across all the considered remote data, as it must be done in a federated manner.

Study expected results, outcomes and impact

Technical infrastructures:

An initial proof-of-concept (POC) implementation was set up and tested early in the project, 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 using Docker containers providing the following services: a (PostgreSQL) OMOP CDM database, Atlas/web API, 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 to demonstrate 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, to include additional hospitals in a later phase of the project.

Capacity building through training.

This project will mainly contribute to research and capacity building through training staff before and during the project both at UR and at participating hospitals. The planned training includes: 1) data mapping infrastructure; 2) training on surveys instruments and 3) training on sensitive patient data handling, data harmonization, interoperability and medical terminology: A team from Ghent University (Belgium) will train the Rwandan research team on OHDSI OMOP CDM mappings including terminology and coding.

Clinical, epidemiological, mental and socio-economic outcomes results:

This project will yield prediction models for the burden of COVID-19 in the community but also the potential impact on hospital admissions or overall infection rates and the impact of various public health measures on 1) the pandemic evolution in the country; 2) on the social-economic situation, 3) and on the mental health (stratified by gender and other vulnerable groups). As intermediate results, the community survey will be analyzed separately on all scopes including descriptive statistics of socio-economic impact, epidemiology, mental and clinical outcomes. For socio-economic outcome, the variables to be analyzed are related to the effect of covid-19 on livelihood with a focus on its effect on basic needs (food, medical, care, school fees and transport), income, employment and saving. A logistic model will be formulated and used to analyze the socio-economic characteristics of people who have been experiencing some economic difficulties due to the COVID-19 situation.

On epidemiological aspects we will investigate the prognosis factors associated with clinical outcome of COVID-19 burden in Rwanda, and the drivers of COVID burden in Rwanda.

Regarding the gender and mental health multiple axes of research are planned including 1) the longitudinal study on stigmatization and associated factors during the COVID-19 pandemic in Rwanda; 2) Behavioral/ Gender based violence outcome of COVID-19 in Rwanda; 3) longitudinal study on mental health wellbeing and associated factors during the COVID-19; and others.

Finally, a cultural analysis is planned to investigate how Rwandans deal with the COVID-19 pandemic and the related control measures.

Conclusion

The project is the first on the African continent to implement 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.

Abbreviations

AI: Artificial Intelligence; LAISDAR: Leveraging Artificial Intelligence and Data Science Techniques in Harmonizing, Sharing, Accessing and Analyzing SARS-COV-2/COVID-19; CHUK: The University Teaching Hospital of Kigali; SARS-COV-2: Severe Acute Respiratory Syndrome Coronavirus 2; COVID-19: Coronavirus Disease of 2019; OHDSI: Observational Health Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership; CDM: Common Data Model; EHR: Electronic Health

Records, EMR: Electronic Medical Records, ETL: Extract transform load ; DS: Data Science, ML: Machine Learning; PROs: with patient reported outcome; CHWs: Community Health Workers; MoH: Ministry of Health; NISR: National Institute of Statistics, Rwanda; R&I: Research and Innovation; RWD: Real World Data; UR: University of Rwanda; SEIR: Susceptible-Exposed-Infected-Removed; API: Application Programming Interface; GAD: General anxiety disorder; POC: Proof of concept; MCD: Minimum Clinical Data; WHO: World Health Organization.

Declarations

Authors' contributions

All authors were involved in the conception and design of the study. AN and MT initiated the draft manuscript. All authors contributed and reviewed the manuscript until the final version. All authors read and approved the final manuscript.

Acknowledgments

The LAISDAR project is funded by the Canada's International Development Research Centre (IDRC) Grant 109587 and the Swedish International Development Cooperation Agency (Sida), under the Global South AI4COVID Program.

The part of community surveys is funded locally by the National Council for Science and Technology (NCST) Rwanda through the research grant NCST-NRIF/COVID-19/002/2020 for the project titled "Longitudinal datasets hub for predicting and monitoring COVID-19 evolution in the community and mitigation measures outcomes in Rwanda (PREDICT project)".

Competing interests

The authors report no conflict of interest.

Availability of data and materials

This article is a research proposal. The dataset generated for some of the phases of the study have not yet been analysed but will be made available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethical considerations

This study has been approved by the Rwanda National Ethics Committee Rwanda (No.112/RNEC/2021), the University of Rwanda, College of Medicine and Health Sciences' Institutional Review Board, and the ethics committee of the Kigali University Teaching Hospital (EC/CHUK/080/2021). Consents will be

obtained from each study participant. Confidentiality of the participants will be maintained at all times. No identifying information will be stored by the research team, reducing the risk of breaches of confidentiality. Questionnaires will be number-coded thereby keeping the identity of the participants anonymous. Given the focus on sensitive clinical data, it is important to govern data adequately and ensure appropriate management of the data. Therefore, the consortium will specifically address the data governance matters, from the sources of data, their integration and use ensuring suitable privacy protection and information governance. No patient data will be shared even anonymized. As the IDRC embraces the principle of sharing research data and encourages researchers to make their data openly available, the researchers will be able to access data where they get only the aggregated data (no data download). Each individual will need to register and request access to the whole or a part of data available from the common analytical interface. He or She will sign data access agreement, limited by the research project duration. Research findings will also be made accessible to research participants through a login credentials, but also findings will be disseminated to participants through various media. This study will be carried out in accordance with relevant guidelines and regulations in the Ethical Declarations.

References

1. She J, Jiang J, Ye L, Hu L, Bai C, Song Y. 2019 novel coronavirus of pneumonia in Wuhan, China: emerging attack and management strategies. *Clin Transl Med.* 2020;9(1):1–7.
2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270–3.
3. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis. *J Infect.* 2020;80(6):656–65.
4. Zhou G, Chen S, Chen Z. Back to the spring of 2020: facts and hope of COVID-19 outbreak. Vol. 14, *Frontiers of Medicine.* Springer; 2020. p. 113–6.
5. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *BioRxiv.* 2020;
6. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;588(7836):E6–E6.
7. Kumar A, Singh R, Kaur J, Pandey S, Sharma V, Thakur L, et al. Wuhan to world: the COVID-19 pandemic. *Front Cell Infect Microbiol.* 2021;11:242.
8. Açıkgöz Ö, Günay A. The early impact of the Covid-19 pandemic on the global and Turkish economy. *Turkish J Med Sci.* 2020;50(SI-1):520–6.
9. Hitt MA, Holmes Jr RM, Arregle J-L. The (COVID-19) pandemic and the new world (dis) order. *J World Bus.* 2021;56(4):101210.
10. Lone SA, Ahmad A. COVID-19 pandemic—an African perspective. *Emerg Microbes Infect.* 2020;9(1):1300–8.

11. Hager E, Odetokun IA, Bolarinwa O, Zainab A, Okechukwu O, Al-Mustapha AI. Knowledge, attitude, and perceptions towards the 2019 Coronavirus Pandemic: A bi-national survey in Africa. *PLoS One*. 2020;15(7):e0236918.
12. Acter T, Uddin N, Das J, Akhter A, Choudhury TR, Kim S. Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency. *Sci Total Environ*. 2020;730:138996.
13. Rwanda_Biomed_Center. Rwanda COVID Response 2021 [Internet]. 2021. Available from: <https://www.rbc.gov.rw/index.php?id=707>
14. Ngamije J, Yadufashije C. COVID-19 pandemic in Rwanda: An overview of prevention strategies. *Asian Pac J Trop Med*. 2020;13(8):333.
15. Karim N, Jing L, Lee JA, Kharel R, Lubetkin D, Clancy CM, et al. Lessons learned from Rwanda: innovative strategies for prevention and containment of COVID-19. *Ann Glob Heal*. 2021;87(1).
16. Nkeshimana M, Igiraneza D, Turatsinze D, Niyonsenga O, Abimana D, Iradukunda C, et al. Experience of Rwanda on COVID-19 Case Management: From Uncertainties to the Era of Neutralizing Monoclonal Antibodies. *Int J Environ Res Public Health*. 2022;19(3):1023.
17. Condo J, Uwizihiwe JP, Nsanzimana S. Learn from Rwanda's success in tackling COVID-19. *Nature*. 2020;581(7809):384–5.
18. Loembé MM, Nkengasong JN. COVID-19 vaccine access in Africa: Global distribution, vaccine platforms, and challenges ahead. *Immunity*. 2021;54(7):1353–62.
19. Musanabaganwa C, Cubaka V, Mpabuka E, Semakula M, Nahayo E, Hedt-Gauthier BL, et al. One hundred thirty-three observed COVID-19 deaths in 10 months: unpacking lower than predicted mortality in Rwanda. *BMJ Glob Heal*. 2021;6(2):e004547.
20. Dairi A, Harrou F, Zeroual A, Hittawe MM, Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. *J Biomed Inform*. 2021;118:103791.
21. Sudat SEK, Robinson SC, Mudiganti S, Mani A, Pressman AR. Mind the clinical-analytic gap: electronic health records and COVID-19 pandemic response. *J Biomed Inform*. 2021;116:103715.
22. Li WT, Ma J, Shende N, Castaneda G, Chakladar J, Tsai JC, et al. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC Med Inform Decis Mak*. 2020;20(1):1–13.
23. Sun C, Hong S, Song M, Li H, Wang Z. Predicting COVID-19 disease progression and patient outcomes based on temporal deep learning. *BMC Med Inform Decis Mak*. 2021;21(1):1–16.
24. Mizrahi B, Shilo S, Rossman H, Kalkstein N, Marcus K, Barer Y, et al. Longitudinal symptom dynamics of COVID-19 infection. *Nat Commun*. 2020;11(1):1–10.
25. Collaborative O, Network. OpenMRS: open-source platform to build customized EMR system. 2016; Available from: <https://openmrs.org/>
26. Verbeke F. OpenClinic GA: open-source integrated hospital information management system [Internet]. 2016. Available from: <https://sourceforge.net/projects/open-clinic/>

27. [Rwanda] NloSo RN, [Rwanda] MoHM II. Rwanda Demographic and Health Survey 2019-20 Key Indicators Report. [Internet]. 2020. Available from: <https://dhsprogram.com/pubs/pdf/PR124/PR124.pdf>
28. Laisdar Project Investigators. Laisdar website. 2020; Available from: <https://laisdar.rbc.gov.rw/>. Accessed on 14th June 2020.
29. OHDSI: Observational Health Data Sciences and Informatics. ArachneNodeAPI. 2020; Available from: <https://github.com/OHDSI/ArachneNodeAPI>
30. De Brouwer E, Simm J, Arany A, Moreau Y. Gru-ode-bayes: Continuous modeling of sporadically-observed time series. *Adv Neural Inf Process Syst.* 2019;32.
31. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. American Physiological Society Bethesda, MD; 2020.
32. Alleman TW, Vergeynst J, De Visscher L, Rollier M, Torfs E, Nopens I, et al. Assessing the effects of non-pharmaceutical interventions on SARS-CoV-2 transmission in Belgium by means of an extended SEIQRD model and public mobility data. *Epidemics.* 2021;37:100505.
33. Alleman T, Torfs E, Nopens I. Covid-19: from model prediction to model predictive control. Unpubl Preprint, URL https://biomath.ugent.be/sites/default/files/2020-04/Alleman_etal_v2.pdf [Google Sch. 2020;

Figures

Figure 1

LAISDAR participants

Figure 2

Overall LAISDAR architecture for data harmonization

