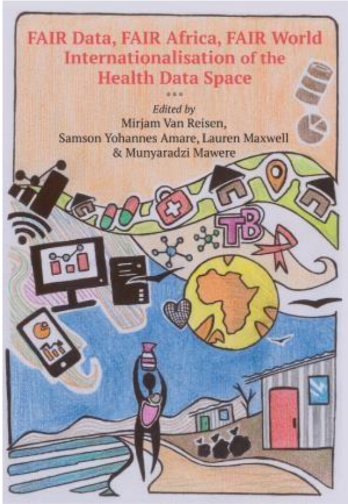
Testing the Cross-Border Africa Health Data Space: Monitoring, Tracking and Prevention of Mother to Child Syphilis Transmission

Samson Yohannes Amare, Liya Mamo Weldu, Araya Abrha Medhanyie, & Mirjam van Reisen

Chapter in:
Fair Data Fair Africa Fair World:
Internationalisation of the Health Data Space



Cite as: Amare, S. Y., Weldu, L. M., Medhanyie, A. A., & Van Reisen, M. (2025). Testing the cross-border Africa health data space: Monitoring, tracking, and prevention of mother-to-child syphilis infectious transmission. https://doi.org/ 10.5281/zenodo.15382985. In M. Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. (pp. 517–552). Bamenda: Langaa. URL: https://www.researchgate.net/publication/391750151_FAIR_Data_FAIR_Africa_FAIR_World_The_Internationalisation_of_the_Health_Data_Space

The About the Authors note can be found here: https://raee.eu/wp-content/uploads/2025/05/About-the-Authors-and-Editors.pdf
The list of figures and tables can be found here: https://raee.eu/wp-content/uploads/2025/05/List-of-Figures-and-Tables.pdf

Contents

Chapter 16

Testing the Cross-Border Africa Health Data Space: Monitoring, Tracking and Prevention of Mother to Child Syphilis Transmission

Samson Yohannes Amare, Liya Mamo Weldu, Araya Abrha Medhanyie, & Mirjam van Reisen[i]

Abstract

Effective use of patient data enhances maternal healthcare by improving diagnosis, treatment, and follow-up care. Detecting syphilis during pregnancy is critical to preventing congenital infections and protecting foetal health. Beyond individual care, structured patient data supports broader applications, including clinical trials, and surveillance, which are essential for public health. This research examined a shift from traditional data reporting, where health data is periodically aggregated and submitted to higher jurisdictions, to a dynamic, granular approach. Instead of relying on statistical summaries, the study proposed a model enabling direct patient-level analysis, linking datasets through semantic relationships within a knowledge graph. The feasibility of this model was demonstrated by enabling healthcare facilities to access data via Query Language and Resource Description Framework (SPARQL) remote queries among themselves, allowing real-time interrogation and analysis. The study highlighted the benefits of cross-facility and cross-country analytics, which provide comparative insights for refining screening and treatment protocols. This approach can support federated querying, federated learning, and multi-layered insight generation for infection surveillance, demonstrated using syphilis. By identifying and tracking congenital syphilis cases, healthcare providers can minimise the risk of mother-to-child transmission.

Keywords: Africa Health Data Space, congenital syphilis, infection prevention, FAIR data, RDF, SPARQL, health information systems

Introduction

Pregnancy is a susceptible period during which maternal infections have significant consequences. Infection during pregnancy is harmful to both the mother and the unborn foetus, leading to complications like miscarriage, premature labour, congenital malformations, and stillbirth (Sappenfield et al., 2013), generating concerns that the novel coronavirus (COVID-19) may have negatively impacted on maternal health outcomes (Ministry of Health and Social Services, 2022; Roberton, 2020). Hence, the monitoring and follow-up of disease outbreaks are a concern that needs urgent attention (Terris, 2003).

Among several infectious diseases, syphilis remains a public health concern of high priority due to its catastrophic consequences if left untreated. Syphilis transmission during pregnancy poses severe risks to both the mother and the baby. The dangers of syphilis transmission during pregnancy include, among other things, stillbirth, and if untreated at birth, congenital syphilis (CS) that can lead to neurological impairment, bone deformities, and organ damage appearing months or years later. Surveillance and monitoring of diagnosed cases is a good strategy, to support the elimination of mother-to-child transmission of the infection that results in CS (Elarrat, 2019).

The World Health Organization (WHO) has provided guidelines on syphilis screening and treatment for pregnant women (WHO, 2017). Countdown to 2030 (CD2030) is a global collaboration that generates evidence and analyses data to track progress and promote accountability for reproductive, maternal, newborn, child, and adolescent health and nutrition, focusing on equity and key drivers such as policy, financing, and health systems. Syphilis screening and treatment during antenatal care (ANC) is a priority indicator in CD2030. Using secondary data from mortality and disease reporting systems, Elarrat (2019) found that there is still underreporting of congenital syphilis related fatal and neonatal mortality.

CS has been designated as a target for global elimination. The Prevention of Mother-to-Child Transmission (PMTCT) programme in the Regional Office for Africa (AFRO) is aligned with WHO's Global Health Sector Strategies (2016–2021) on HIV and sexually transmitted infections (STIs), which set a target of achieving zero new infections among infants by 2020. The programme provides technical assistance and normative guidance for the development, implementation, strengthening, and evaluation of policies, strategies, and programs related to PMTCT. Efforts are directed towards improving access to antenatal services, HIV and syphilis testing and treatment during pregnancy, as well as safe childbirth practices, optimal infant feeding, and the uptake of infant HIV testing, treatment, and retention in care (WHO Africa region, u.d.).

Zhu et al. (2022) developed a predictive model to estimate the incidence of syphilis cases, identifying an increasing trend in China. Data systems and predictive models have been implemented to monitor mother-to-child transmission (MTCT) of infectious diseases, particularly syphilis, facilitating surveillance and outcome prediction. In South Africa, CS is classified under category two of the national Notifiable Medical Conditions (NMC) surveillance platform, enabling the systematic tracking of progress toward its elimination (Kufa et al., 2023; Morifi et al., 2021). Surveillance programmes play a critical role in measuring progress toward the elimination of syphilis. In addition to surveillance, data collected on syphilis can directly contribute to improved health outcomes by informing health workers in obstetric departments and aiding in timely intervention and management.

Access to syphilis testing also remains a challenge in some settings. For instance, in Tigray, Ethiopia, during the war, a shortage of test kits prevented many pregnant women from receiving the recommended routine syphilis screening during antenatal care visits, despite WHO guidelines mandating such testing. This shortage, a result of the two-year siege, highlights the critical need for sustained access to essential diagnostics in conflict-affected regions (Gebreslassie et al., 2024).

Data analysis can play a critical role in this especially if this data is handled under federated FAIR principles (that data is Findable, Accessible, under well-defined conditions, Interoperable, and Reusable), with data held under Ownership, in Locale under Regulatory Compliance (OLR) (Amare et al., 2023; Amare et al., 2024). Initial work on query and visualisation of antenatal care data focusing on syphilis as a use case was published as part of a study by the Value-driven Ownership of Data and Accessibility Network (VODAN) research group, focusing on data from Ayder Hospital in Tigray (Van Reisen et al., 2024). This work builds on the use case by expanding it to wider geographic coverage making cross country queries and cross-facility analysis as well as monitoring and tracking of cases within a facility.

This chapter investigates the possibility of using the proposed architecture of VODAN as a data pipeline for clinical decision support. We investigated whether the FAIR data-based proposed landscape for patient data can be used to track, and monitor mothers during their ANC visits. It investigated how the African Health Data Space can be used for monitoring and surveillance using dashboards and data visiting for syphilis management and beyond during ANC. The study investigated how providing tracking and monitoring functionality could help manage syphilis cases at facility, cross-facility/community, and cross-country levels.

Two research questions were formulated for this research. First, to what extent and how are insights into infectious diseases enabled by data capture and storage in health facilities across-border settings harmonised based on controlled semantic vocabularies? Second, how can syphilis cases that can result in congenital syphilis be identified and closely followed up to prevent mother-to-child transmission?

Relevance

The relevance of this study is to determine an ethically and practically robust framework for patient data handling for decision-making at the point of care, as well as to create insights that improve the services to patients and the health system delivery. The recording and use of patient data within the health facility has not been directly connected to the aggregation of data for surveillance and monitoring, as the individual and aggregate data show variation. Multiple unconnected datasets exist that are not interoperable. Aggregate data which are reported periodically have little granularity to be of value for the decision-making processes within health facilities (Nalugala et al., 2025).

The lack of attention given to the value of the data for better outcomes at the point of care has an impact on the way the data is produced. A lack of value given to the use of data in the facility can harm the quality of the data. The divorce of aggregate data from the provenance of the data diminishes the value of the data for further insights (Amare et al., 2025; Jati et al., 2025). The lack of integration of heterogeneous data hinders the functional use of the data (Callard, 2025).

In the context of infectious diseases that have regional and global implications, a robust data system is needed to monitor progress and limit spread, as was evident during the COVID-19 pandemic. Public health crises such as this require sharing of data and lessons, as well as a systems approach (Ros et al., 2021; Van Reisen et al., 2022). Depending on the different geographical coverage and nature of some of the infectious diseases that have contagious properties, the proposed data landscape has the potential to provide solutions such as early warnings of pandemics or use in routine care to save the lives of patients. The cross-border query and analytics functions can provide insights into what is happening elsewhere in the region and help support decisions and make necessary preparations. This makes the study relevant not only at the facility and community level, but also at broader national and regional levels.

This study contributes to the understanding of the feasibility of new and alternative solutions that build on established norms of data FAIRification in academia (Stocker et al., 2022) to establish federated repositories for the interoperability and reuse of data and to ensure that source data is curated with the right provenance.

Theoretical framework: the Africa Health Data Space as a basis for screening for early prevention of infection

The Africa Health Data Space is designed to improve how health data is collected, stored, and used across the continent. It is a way of organising data production, services and access for intelligence. Imagine it as a three-layered system, like a well-organised library. The Africa Health Data Space website (https://aun.mu.edu.et/ahds/), published by the Africa University Network on FAIR Open Science (AUN-FOS), further elaborates the concept of the Africa Health Data Space.

The concept is pictured in Figure 1.

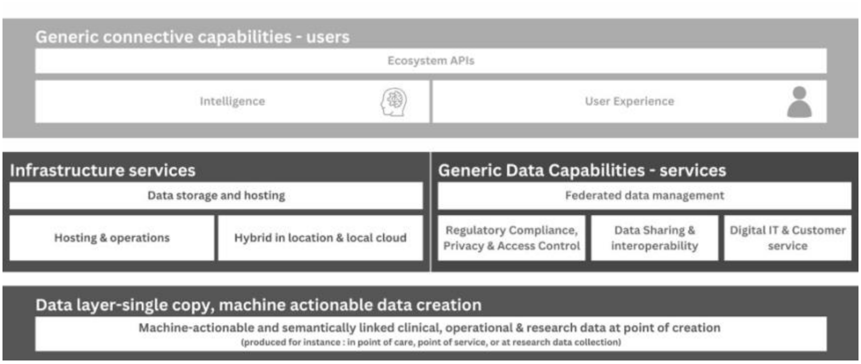


Figure 1. Federated Africa Health Data Space – a proposed landscape

Source: Van Reisen et al., 2024

The Africa Health Data Space proposes three functional layers. The data curation layer, which can also be referred as the foundational layer, is where health data is first created. At hospitals and clinics, patient records and research data are collected in a way that ensures they are FAIR. This ensures that healthcare providers and researchers can easily find and understand the data they need.

The exposure and services layer is referred to as the middle layer. This layer makes sure the data is properly structured and categorised. It provides services that make the data from the lower layer accessible to the intelligence layer, providing the right hosting of software and hardware, interoperability and service support that make data use seamless. To cater for systems that have a legacy approach, this layer can provide standard templates and use of ontologies formulated in triples to make sure that data from different sources can be semantically interoperable and work together smoothly. It also helps update older records, making them easier to use in modern health systems.

The intelligence layer is referred to as the top-level layer. This is where the data is put to good use. Health professionals, policymakers, and researchers can analyse the data to improve healthcare services, detect disease outbreaks, and make informed decisions that benefit communities.

By organising health data in this way, the Africa Health Data Space ensures that patient information is more reliable, accessible, and useful, helping improve healthcare outcomes across Africa (Van Reisen, et al. 2024).

The three-layered architecture shown in Figure 1 is designed to improve how health data is collected, stored, and used. This creates a well-organised system that ensures health information is reliable, easy to find, and beneficial for both healthcare providers and researchers.

In the foundation of the system, health data is first produced and collected. Imagine it as a digital filing system, as in a library, where every patient's medical information is recorded in a structured way. Instead of scattered and inconsistent records, this layer ensures that a single, accurate copy of patient data exists and is managed properly.

To achieve this, the system follows FAIR principles. The curation involves adding metadata as triples (subject-predicate-object), structured through templates and standardised with ontologies that are standardised and machine-actionable. FAIR resources are used to help structure data properly. These tools ensure that health records are labelled in a consistent and meaningful way so that different hospitals, clinics, and researchers can use them without confusion. This means that:

- Data is easy to search and locate when needed.
- Only authorised people (like doctors and researchers) can access the data securely, based on permission controls.
- Different health systems can share and understand the data, preventing information silos, based on prior access provided.
- The data can be reused for research and policymaking, rather than being locked away in individual hospital records.

Older records (legacy data) can also be updated and included in the system, preventing the loss of important historical health information (Amare, et al., 2025).

Once data is collected, it needs to be organised and standardised to ensure that it makes sense and can be used properly. Think of this as the librarians carefully classifying books in a library so that anyone can find and understand them easily and know their broad topics and content. The services layer is crucial because health data often comes from many different sources, such as hospitals, labs, and research studies. Without proper organisation, data can be messy and difficult to use. By ensuring that all data follows the same structure, this layer makes it possible to combine and compare information from different places.

At the top of the system is the intelligence layer, where all the carefully organised data is used and reused to create new insights from it. This is where doctors, researchers, and policymakers can analyse health trends, improve patient care, and make better decisions based on real-world data. Think of it as the library reading room, where the researchers who have been given access, can study the literature. Special access may be required for sensitive archives. The FAIR-data layered Health Data Space follows the same logic, focused on the data itself.

Examples of insights that can be acquired include the following:

- Doctors can track disease outbreaks in real-time, helping them respond faster to prevent the spread (Van Reisen et al., 2024).
- Hospitals can use data to predict patient needs, ensuring they have enough staff and medicine available (Callard, 2024).
- Governments and health organisations can create better policies, ensuring healthcare resources are used efficiently (Nalugala et al., 2025).

Because the data is curated and structured properly in the lower layers, this intelligence layer can provide accurate and meaningful insights rather than just raw, unorganised data.

The Africa Health Data Space ensures that health information is collected, stored, and used in a way that benefits both patients and healthcare providers. By following a FAIR approach and using this three-layered system, the system is expected to:

- Improve patient care by ensuring doctors have the right information at the right time.
- Enhance medical research by making high-quality data available to scientists.

- Strengthen disease surveillance by identifying outbreaks and health trends early.
- Increase efficiency by reducing paperwork and administrative burdens.

This structured approach ensures that health data is not just stored, but actively used to improve healthcare across Africa (VODAN, 2021). These hypotheses need to be further researched. This study is part of this new research agenda.

A particularly important assumption underpinning this approach is that data should reside in and/or be under the control of the health facility and benefit the services provided in the facility. This means that there is particular attention paid to the ownership of the data handler (in the case: the health facility), where the data resides in locale, and data production, interoperability and re-use is performed under regulatory compliance in the place where the data was produced and collected. This is referred to as FAIR under Ownership in Locale under Regulatory Compliance (FAIR-OLR) (Van Reisen et al., 2023).

For the data on the lower layer to be processed, it requires the right infrastructure and a generic data handling capacity and services that make it accessible to the intelligence layer and beyond (Aktau et al., 2025). For the data to be accessible, especially in low-resource settings, it requires the right hosting software and hardware infrastructure (Amare et al., 2025). The software infrastructure includes data visiting systems, data analytics tools, and federated learning platforms (Jati & Lasroha, 2025). The infrastructure can be organised in the form of a local host, local cloud, or cloud infrastructure in the form of a virtual private network that ensures data privacy and security (Amare et al., 2025).

To help run and sustain those services, there must be a support system. This layer indicates what is required to run and sustain the services in the Health Data Space. The top layer leverages the middle layers as an enabler to deliver services, enhance user experience and systematically measuring its effectiveness. Adopting a human-centred approach, this layer ensures that user interactions are optimised based on real-world data insights derived from the foundational data layer.

Additionally, the intelligence layer facilitates seamless integration with external systems, enabling the generation of deeper insights, the development of applications, and the augmentation of existing data through external sources. This interconnected approach enhances data-driven decision-making, ensuring a more comprehensive and effective use of health information.

At the core of these is the need for strong data stewardship to provide technical support, the need to comply with the relevant regulatory framework, and the need to make the data and systems in the facility interoperable internally as well as with external systems (Aktau, 2025; Zhang et al., 2022; Taye et al., 2024; Jati, 2025; Folorunso, et al., 2025).

Methodology

This simulation study was set up as an engineering study in a natural setting, with the strong involvement of stakeholders including doctors and health workers, hospital administration, data clerks, and data stewards (Sanders, 2008). The case-study, carried out in a real-life situation with patient data, looks specifically into syphilis, as a case study on how data insights can serve to improve decision-making at the point of care/service.

VODAN Africa Research Network

The VODAN Africa Research Network is a collaborative initiative focused on developing FAIR data infrastructure to support health research and outbreak response across Africa (https://aun.mu.edu.et/vodan/). The network was established in response to the COVID-19 pandemic to ensure privacy-preserving, locally governed, and machine-actionable health data that can be used for clinical decision-making and scientific research without compromising data sovereignty (VODAN, 2021; Van Reisen et al., 2023).

VODAN Africa promotes a FAIR federated approach, meaning that health data remains in its country of origin while enabling secure access for analysis. This enhances global health collaboration while ensuring ethical data sharing and compliance with local regulations. The network aligns with FAIR and open science principles (VODAN, 2021).

In its first phase, the research network had set up the data system in 8 countries (Amare et al., 2025). To further this effort and make FAIR data production and use at the core, the research group set up a quality data pipeline by reducing the number of countries to 4 and the health facilities to 12. This case-study re-utilises the data produced in this set-up.

Study setting

The simulation study was conducted as part of a VODAN Africa quality data pipeline project conducted in 2023, which collected nine months of antenatal care data. The study was conducted in 4 countries, namely: Ethiopia, Kenya, Uganda, and Nigeria. The study engaged 12 health facilities: Railway, Tunga-Primary Health Care, Minna General Hospital Lapai, Olabisi Onabanjo University Health Center, and Federal University Lokoja University Medical Centre (in Nigeria); Pumwani Hospital, Beacon of Hope Hospital, Lira Hospital, and Ark Specialist Hospital (in Kenya); Kampala International University Teaching Hospital, and Hoima Referral Hospital (in Uganda); and Ayder Comprehensive Specialized Hospital and Mekelle General Hospital (in Ethiopia).

Covering the ANC and outpatient department (OPD) domains, the study used the natural setting of the health facilities and their routinely used health management information system (HMIS) abstract registers (Amare et al., 2025). Syphilis screening is carried out as part of routine ANC follow-up in each of the countries studied to track pregnancy risk. The data was curated to provide a data source for visualisation in the form of a dashboard and as a data repository for query and analytics.

Data collection

Data was extracted and FAIRified using the relevant abstract HMIS registers (usually District Health Information System 2 [DHIS2]) in the health facilities studied. Both ANC and OPD data covering different diseases were collected.

Table 1. Antenatal care data from four countries in Africa

Country	Freq.	Percent	Cum.
Kenya	1,181	15.77	15.77
Nigeria	1,599	21.35	37.12
Tigray	3,808	50.84	87.96
Uganda	902	12.04	100.00
Total	7,490	100.00	

The data in Table 1 shows the ANC data collected as part of the quality data pipeline. It is a summary of the data that was collected across all the ANC visits. The data is from 4,112 women from 12 facilities in 4 countries over 9 months in 2023.

This study focused particularly on the ANC data and on pregnant women who made frequent visits to health facilities for follow-up before delivery. The frequency of these visits varied depending on the complexity of the cases. To track the pregnant women, an Android-based mobile application, KoboCollect, was used on smartphones and tablets.

The HMIS registers were selected as the primary data source due to their routine use in health facilities and the presence of designated personnel, such as health information technology (HIT) professionals in Ethiopia, responsible for data collection. Data collection was conducted under close supervision by data stewards, who were overseen by the VODAN country coordinators. The progress of the study was continuously monitored and discussed in weekly virtual VODAN meetings.

Ethical considerations

Patient data requires ensuring privacy and security in most cases to the subjects. The data handling process in the VODAN research group adheres to European Union's General Data Protection Regulation (GDPR), while comply with the data regulations of each of the sovereign countries. The data is handled expressly under FAIR-OLR principles.

Data use agreements were signed by the respective health facilities and higher jurisdictions, the regional health bureaus and ministers of health offices in the four respective countries. Moreover, a separate data processing agreement was signed that allowed three technical experts from VODAN and partners to inspect the data quality and provenance, analytics, and generation of insights and prospects for modelling and development of privacy-preserving dashboards, with a use-case approved particularly on syphilis risks for pregnant women.

Results

This research was guided by two key questions. The first explored the extent to which insights into infectious diseases, facilitated by data capture and storage in health facilities across border settings, can be harmonised using controlled semantic vocabularies. The second investigated how syphilis cases, with the potential to cause congenital syphilis, can be effectively identified and monitored to prevent mother-to-child transmission. This use case examined three key aspects: (i) how data is curated at the lower layer, (ii) how it is hosted and maintained in the middle layer, and (iii) how data is accessed within federated repositories to enable computations and analytics in the federated manner in the intelligence layer. The goal is to perform data analysis, including query functionalities and data visualisation, to identify mothers at risk of congenital syphilis.

Following the research questions, the first part of the findings outlines the data curation process, detailing the steps involved in FAIRification. The second part focuses on data analytics, demonstrating how the system can track and monitor infections, in the case of syphilis. The technical process of how to query potential congenital syphilis through a query is subsequently presented.

Workflow

To help support syphilis screening and monitoring and decision support, a data system was set up to curate data in the form of FAIR data. This process involved data FAIRification, visiting, and visualisation functionalities.

Data is currently being recorded in most of the health facilities enrolled in this study in patient charts where doctors take patient history directly. To avoid the difficulty of acquiring important indicators, countries have also developed structured abstract registers/forms that extract data from patient charts. These templates are used as a source of data for tally sheets from which the HIT professionals count and aggregate data for HMIS reports.

Analysing the coherence of variables in the data collection forms based on the HMIS registers, the research team identified significant variability. This led to considerable differences in the data elements recorded during ANC visits. The variations were particularly evident across countries, with some data elements documented in the abstract register of one country, but absent in others during routine data capture and registration (Amare et al., 2025). These inconsistencies posed challenges in terms of data analysis. However, syphilis case monitoring was consistently implemented across all facilities and in all four countries, enabling comprehensive analysis and tracking through the dashboard and other monitoring tools.

In this work, we used the FAIRification workflow to curate the data as FAIR-OLR, based on the proposed landscape to make it available for data visiting and visualisation.

Building the data pipeline

The data pipeline and framework for developing use cases within the intelligence layer of the architecture required the curation of data in accordance with FAIR-OLR principles. ANC data extracted from abstract registers underwent a process to enhance its findability and accessibility, enabling its use for modelling, surveillance, and reuse and placement in triple stores in location.

To ensure semantic interoperability, metadata templates were used to represent terms from the registers in the Simple Knowledge Organization System (SKOS), which were subsequently converted into Resource Description Framework (RDF) Turtle format using SKOS Play. Legacy data was incorporated through bulk uploads, where it was parsed and annotated using the same controlled vocabulary in cases data was available in Comma-Separated Values (CSV). The bulk upload was also used to annotate the data after it was collected using mobile phones and downloaded in CSV. The vocabularies were hosted on BioPortal, ensuring their reusability in both bulk data uploads and individualised data collection.

The choice of data collection tools depended on what was found to work best in the local context: KoboCollect and Center for Expanded

Data Annotation and Retrieval (CEDAR) Workbench were used to facilitate standardised data entry and ensure compliance with FAIR data principles.

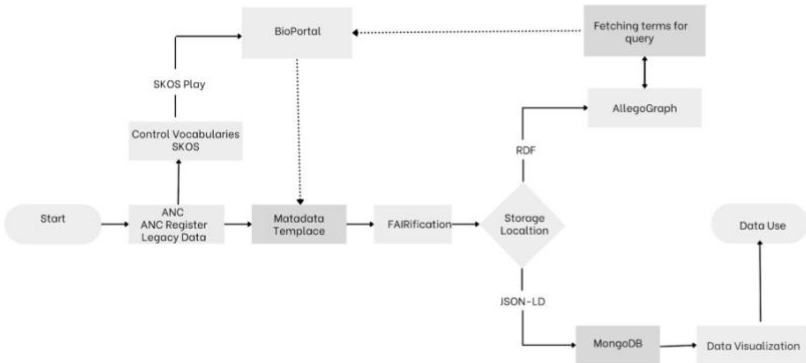


Figure 2. Data curation steps to make it available for analytics and query

Figure 2 shows how data from the source HMIS abstract registers were processed during the FAIRification process and made ready for visualisation and query. The vocabularies were picked from the registers and hosted on BioPortal after curating them in SKOS and later used for metadata template creation. The data was then stored in both document and graph databases to be used for visualisation and query respectively.

Table 2. Reactive and non-reactive cases and non-available tests for syphilis captured from across 12 health facilities in 4 countries from 2023 data

Syphilis test	Ethiopia	Kenya	Nigeria	Uganda	Total
Reactive	8	32	4	3	47
Non-reactive	1,915	1,128	146	653	3,842
Not done	1,885	21	1,450	245	3,601
Total	3,808	1,181	1599	902	7490

Table 2 shows the distribution of syphilis test results from screening during ANC visits. The results were presented as reactive when the test result was positive and nonreactive otherwise. Those who were not tested during screening are presented here indicated as 'Not done'.

As can be seen from the frequency distribution, many of the pregnant women who came for ANC visits did not have a syphilis test. This is contrary to WHO recommendations. This was particularly the case in Ethiopia, followed by Nigeria. The data in Ethiopia is from Tigray, for which test kits were unavailable at the time of data production because of the two-year siege during the war (Niguse et al., 2024; Medhanyie et al., 2024; Taye et al., 2024; Amare et al., 2024; Stocker & Medhanyie, 2024).

The data curated with metadata was then processed in two formats. The data processed for storage in a document database named humongous database (MongoDB) was used to build the visualisation. Internal dashboards that provided access to the data at the facility were designed as well as an aggregate dashboard that stored data from different facilities and different countries to provide broader insights. The data was also processed as RDF triples that were stored in AllegroGraph, which is a graph database that stores the data in the form of knowledge graphs. This data was made available for local and remote queries using SPARQL. During the query, the Internationalised Resource Identifier (IRI)s and field labels for the controlled terms were fetched from the BioPortal ontology repository.

There were challenges related to infrastructure, network availability, and technical capability. Some of the challenges include the difficulty of hosting the triple store and the dashboard locally within the health facilities. To make query and data curation possible, repositories were created for specific countries, to keep the data in-country and as close as possible to the health facility and under its direct control.

Analysis from own community and cross-communities

Once the data was curated as RDF triples and stored in a triple store called AllegroGraph, the metadata created was ready for re-use, provided access permission was given.

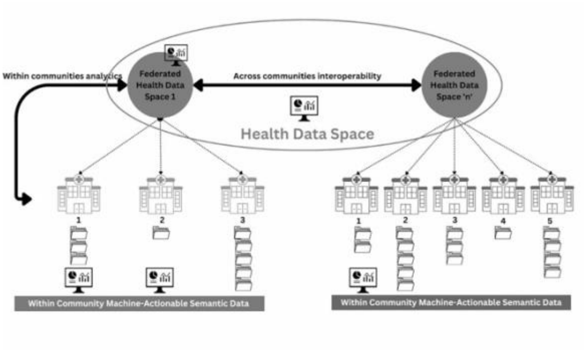


Figure 3. Health data space analytics scopes and possibilities in a community and across the communities

Source: Van Reisen et al., 2022

SPARQL was used for querying the data. To simplify the query process for doctors and nurses, their natural language queries were translated into SPARQL queries. The architectural process of querying across the different data stores in federated set up is pictured in Figure 3.

Figure 3 shows how the proposed data landscape architecture allows the generation of insights and data use at different levels. Starting from facility-level insights, it shows how insights can be generated within a community. These can be sets of health facilities within a single country to provision cross-country analytics and insights based on data that is curated at the point of care/services by the health workers at health facilities.

Data visualisation

To enable data visualisation, both internal and external dashboards were developed. These dashboards displayed key indicators for syphilis, including (i) daily cases, (ii) monthly cases, and (iii) overall cases. Monitoring and tracking were implemented by having the dashboards capture and display data, entered once and reused across all dashboards, as shown in Figure 4. While the data remained within the health facilities, only the aggregate statistics were shared and displayed on the facility dashboards.

Syphilis results were tracked within each of the health facilities. The data was also shown on the aggregate dashboard to show cross-country comparison and follow treads on syphilis prevalence. It showed daily, monthly and overall syphilis reactive and non-reactive, as well as tests that were not being done.

The monitoring and tracking can also happen by making a direct query through data visiting. The subsequent section looks at how the overall monitoring and tracking was done, as well as the sub-question of how the congenital syphilis was tracked, based on the proposed landscape and tools implemented to realise it.

Data visiting approaches

Data visiting is the process of querying data in a federated set up, in which the algorithm visits to the data. To make data visiting possible at different levels of granularity, SPARQL queries were used to compute data at the health facilities, which were stored in the form of RDF. This allows the creation of a knowledge graph, which incorporates the RDF metadata from the different storage facilities.

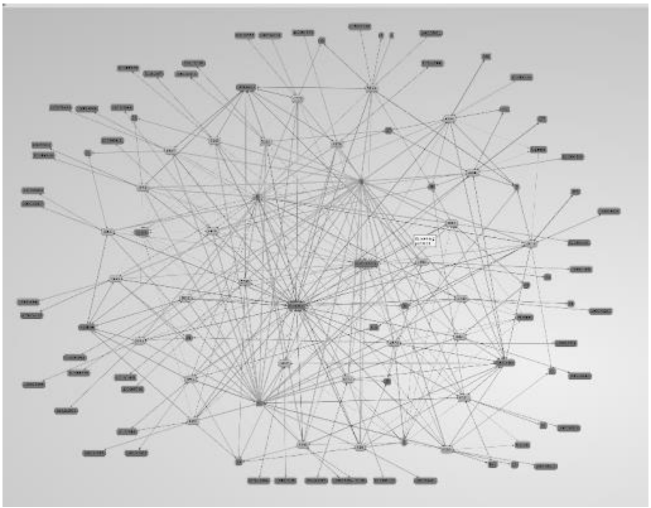


Figure 4. Displaying the knowledge graph across the four countries indicating syphilis test result

Figure 4 shows the result of a query from the graph data, visualised as linked data. The triple store supports both text and visual queries, and the above display represents the result of a query. Users can navigate through the data to inspect the results and explore the relationships within the data.

Federated queries services

Queries in federated format were formulated to visit data in different countries and health facilities. Prior arrangements were made to establish the access provided for this query.

The federated queries formulated using the 'SERVICE' keyword allowed data from multiple sources to be combined by seamlessly running a query and generating results based on distinct sources.

To ease the accessibility of the queries and avoid the need for end users to write complex queries, two approaches were taken. The first approach was to formulate the queries and save them as stored queries to be run when needed within the triple store itself.

```
PREFIX franzOption_authorizationBasic: <franz:username:pwd>

PREFIX franzOption_serviceTimeout: <franz:300> # Increased
timeout to account for network latency issues

SELECT ?s ?p ?o WHERE {
  {
   # Query from a local endpoint
   SELECT ?s ?p ?o WHERE { ?s ?p ?o } LIMIT 100
  }
  UNION
  {
   # Create from a remote endpoint using service keyword
   SERVICE
<http://remoteURL:10035/repositories/RepositoryName/SPARQL>
{
     SELECT ?s ?p ?o WHERE { ?s ?p ?o } LIMIT 100
   }
  }
}
```

The query above can be run from the query browser of the local triple store. It requires authentication of the remote end point and creates a UNION result combined from the local and remote data sources. Using the 'CONSTRUCT' keyword, the visualisation of the results from local and remote data sources was made possible by creating a unified graph as follows:

```
PREFIX franzOption_authorizationBasic: <franz:admin:letmein>
PREFIX franzOption_serviceTimeout: <franz:300>
PREFIX franzOption_logQuery: <franz:yes>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX localage: <http://local-urle>
PREFIX remoteage: <https://remote-url/Age_of_the_Mother>
```

```
PREFIX viz: <http://example.org/viz/>

CONSTRUCT {
  ?person ?p ?o .
  ?person viz:age ?ageValue .
  ?person viz:source ?source .
}
WHERE {
  {
    # Local results with age filter (LIMIT 10)
    SELECT ?person ?p ?o ?ageValue ("local" AS ?source) WHERE {
      ?person ?p ?o .
      ?person localage: ?rawAge .
      BIND(xsd:integer(?rawAge) AS ?ageValue)
      FILTER(?ageValue < 25)
    }
    LIMIT 5
  }
UNION
{
  # Remote results with age filter (LIMIT 10)
  SERVICE
<http://dashboard.vodana.health:10035/repositories/AyderANC/SPARQL> {
    SELECT ?person ?p ?o ?ageValue ("remote" AS ?source) WHERE {
      ?person ?p ?o .
      ?person remoteage: ?rawAge .
      BIND(xsd:integer(?rawAge) AS ?ageValue)
      FILTER(?ageValue < 25)
    }
    LIMIT 5
  }
 }
}
```

The other approach tested was to build a Python script that used the SPARQL Wrapper Library to encode the queries so that they can be run when needed from an external application.

Health facility-based queries: In-facility insights

To enable with the overall tracking and monitoring of the data curated as a knowledge graph, queries could be made to the triple store. As most doctors lack the technical skills required to develop the code, we translated their natural language queries and visualisation requests into SPARQL. These SPARQL queries were saved, allowing authorised users to query and visualise the data for tracking and monitoring purposes.

Figure 5. Saved SPARQL queries that can be run by health workers by selecting one and clicking on execute

Figure 5 show a list of stored SPARQL queries that health workers can click and run. This avoids the need to write code and allow non-technical people to query the data easily.

The data stored in the knowledge graph can be queried to understand the data within the context of the health facility further. Based on local guidelines and treatment protocols, physicians need to make a distinction between cases.

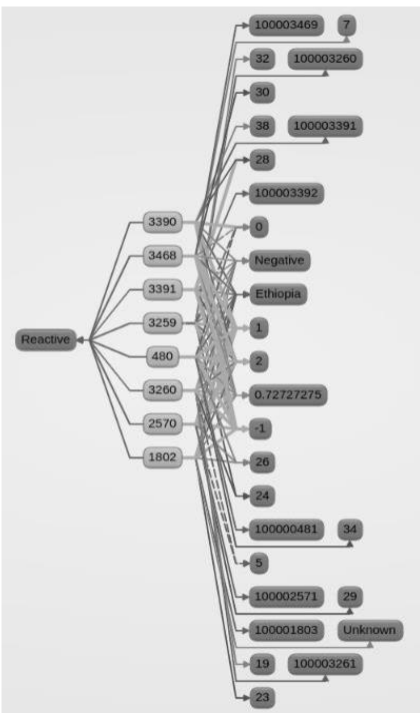


Figure 6. All syphilis reactive

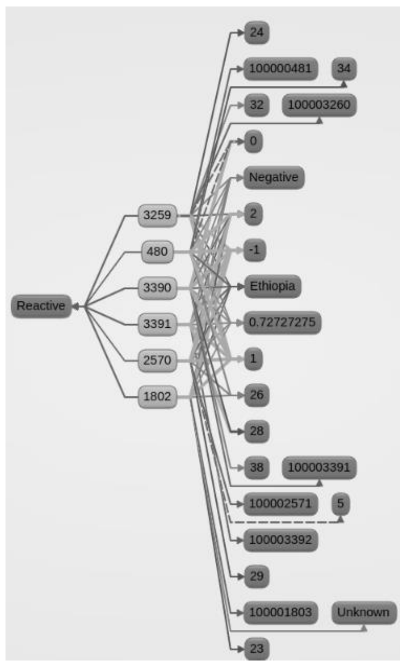


Figure 7. Suspected congenital syphilis

Figure 6 shows all reactive cases at the health facility level and Figure 7 shows those with a risk of congenital syphilis, whose gestational age > 25. This figure is based on selected facilities in Tigray, Ethiopia. This showed the potential to identify syphilis risks in health facilities where test kits for syphilis were not available.

This graph was based on the reasoning that the larger the gestational age of the mother, the higher the chances of transmission of the infection if not properly treated, resulting in congenital syphilis requiring advanced care. As delays in treatment of most of the infections lead to mother-to-child transmission, those mothers who have infections such as syphilis, HIV, hepatitis need to be stratified for advance care. This usually happens during ANC follow-up by assigning them to senior doctors, sending them as a referral out to advanced clinics, and ordering further screening to identify risk of infection transmission.

Community-based queries: Cross facility insights

Health facilities within each country typically follow a tiered structure with linked reporting systems, some adopting hub-and-spoke models. In Ethiopia, for example, the healthcare system is organised into a hierarchy, including tertiary hospitals, general hospitals, primary hospitals, health centres, and health posts, ranging from advanced to lower-level care facilities. Health centres and health posts are collectively referred to as primary health care units (PHCUs). Health posts report to health centres, health centres report to primary hospitals, and both report to other administrative bodies. Physicians can access data from other facilities through remote queries, rather than requesting reports. The data remains within the facilities, but is accessible to the relevant bodies they report to.

```
PREFIX c: <https://dashboard.vodana.health/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?s ?ga ?co WHERE {
  ?s c:ga_in_weeks ?ga;
     c:country ?co;
     c:syphilis_test_result "Reactive".

  # Filter cases where GA > 25 and country is Ethiopia or Nigeria
  FILTER (xsd:integer(?ga) > 25 && (?co = "Ethiopia" || ?co = "Nigeria"))
}
LIMIT 100
```

The SPARQL query above filters those cases whose gestational age is greater than 25 weeks, are syphilis reactive and are from either Ethiopia or Nigeria. Once the query results are returned, one can create a graph as shown below. The graph can be adjusted and displayed in different formats with different colours referring to different variables.

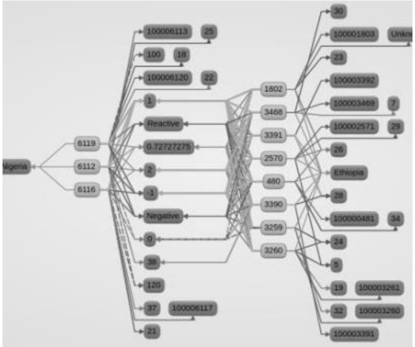


Figure 8. Syphilis reactive cases across Ethiopia and Nigeria

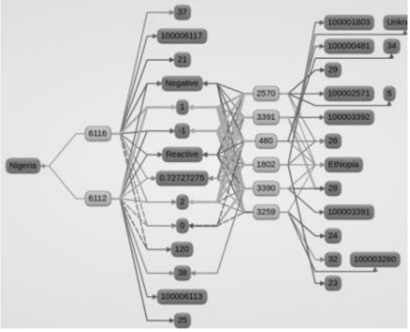


Figure 9. Congenital syphilis suspected cases across Ethiopia and Nigeria

Figure 8 displays all syphilis reactive cases, with suspected congenital syphilis cases shown on the right. It illustrates the monitoring of syphilis reactive and non-reactive cases across the Ethiopia and Nigeria datasets. These figures are similar to Figure 6 and 7, but, in this case, the query is conducted across health facilities in both Ethiopia and Nigeria, forming a cross-border query. The data remains within the health facilities under their control, stored securely in-country, without leaving the facility.

Remote queries: Cross country insights

The 12 health facilities in this study engaged in the quality data pipeline curated their data and stored it in a way that makes cross country query possible using SPARQL. The health facilities, having used the data for facility based and in country decision support, made it available for further insight generation and two-learning. This is achieved with the right prior access and control mechanism and having secured the data was de-identified for data privacy and security. The health facilities can query and compare their status with other facilities across the regions. This has significance for outbreak monitoring and public health surveillance.

Figure 11 shows the result of a remotely conducted query using the federated query service. The federated query allows data to be fetched from both local and a remote endpoint with a SPARQL endpoint. This url shows an endpoint to which a SPARQL can respond to a SPARQL service, (http://dashboard.vodana.health:10035/repositories/AyderANC/SPARQL). The graph shows the query results on the prevalence of infections such as syphilis, HIV, hepatitis and other relevant features on a local and remote end point. The results indicated 1 to be negative, 2 positive and 3 not available.

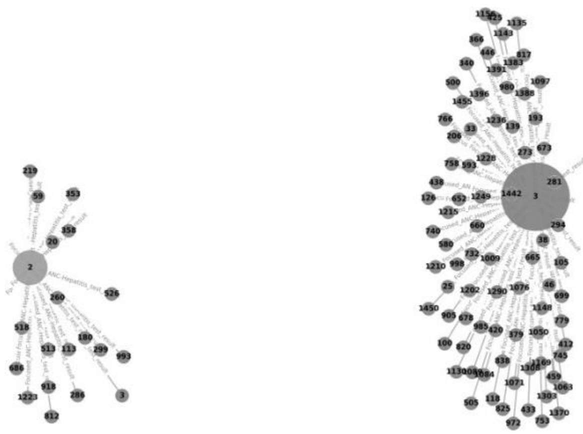


Figure 10. Results of union of local and remote SPARQL endpoints: The results indicate: 2 positive and 3 not available

Figure 10 shows the list of patient anonymised patient identification numbers from local and remote triple store, whose result is either positive or not done cases. As can be seen, the number of cases for which screening has not been done is very significant.

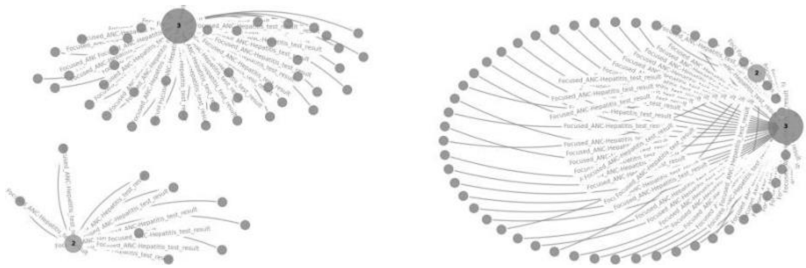


Figure 11. SPARQL query using 'SERVICE' to query distribution and the landscape of the data

Figure 11 visualises the two results. The results indicated as 1 to be negative, 2 positive and 3 not test not done. The results indicated as 1 to be negative, 2 positive and 3 not available.

The ability to make queries and visualise it gives an insight on the distribution and the landscape of the data. This can lead to further investigation and indications of where closer analysis of the data may be useful. It can also act as a starting point for making decisions on focus areas for public health intervention.

Discussion

Typically, health tracking and reporting systems rely on aggregated statistics, limiting the depth of analysis. The study challenges the traditional approach of transmitting healthcare data and reports to higher jurisdictions in the form of monthly or quarterly summaries. Instead, it proposes a model in which data analysis and interrogation occur at a more granular level, allowing for real-time insights and adaptive responses. The approach introduced in this study enables an individual-level data analysis that is dynamically linked to other datasets through semantic linkages and a knowledge graph structure that is accessible through a local and remote query.

The FAIRification of the data and storing it in a federated manner that keeps the right context and provenance of the data with the right metadata, enhances the Findability, Accessibility, Interoperability and Reusability (FAIRness) of the data. Such data is usually not only federated but also AI-ready, as it is not only human readable, but also machine-readable. The health data space thus provides an architecture that makes data from a health facility accessible to a broader community.

The FAIRification process within the Africa Health Data Space framework offers several advantages. Standardisation and use of controlled vocabularies and ontologies make broader interoperability possible. This makes it easier to share across institutions and research networks. It enables data integration across different sources, providing a more comprehensive view of disease patterns. Compliance with regulatory frameworks in each place is possible as data stays in residence, promoting ethical data use while safeguarding patient privacy. It can support longitudinal studies and predictive modelling, helping researchers anticipate disease outbreaks and optimise intervention strategies. This ties very well with the WHO recommendation (2017) on the tracking of syphilis (and other infectious diseases) across countries.

The integration of metadata templates and ontologies enhances the usability of the data. By standardising terminologies and data structures, healthcare providers can ensure consistency across different datasets, facilitating seamless data exchange and collaboration. This structured approach to data curation and linkage provides a robust foundation for building scalable, privacy-preserving digital health solutions.

Traditional extraction, transform, load (ETL) processes manage data by first extracting it such as reading from a legacy database, then transforming the data into to the right format and finally loading it onto destination storage, which is a lot easier with RDF as the triples can just be appended creating the aggregated database very quickly and allowing ease of query if the right modelling is done. Without the need for traditional data ETL activity, one can build datasets that are helpful to track and monitor diseases.

The study showed that the proposed landscape of the Africa Health Data Space has laid a foundation that can act as a basis for building a strong monitoring and tracking system for syphilis, which can also be generalised to other public health concerns. We have demonstrated a pipeline that curates FAIR data at a health facility that enables one-time data entry and creates an interoperable system that allows the creation of insights from the data at different levels. We have demonstrated the relevance of each layer in the architecture of the landscape that realises the different features at all three levels.

Conducted in real-life, in existing health facilities and on real digital patient data, this study found that the procedure that was tested, can be adopted. The spread of 12 health facilities across 4 countries with different context shows the realisation potential of this procedure. The study was conducted under low-resource potential. The readiness of the health facilities to participate in this study, despite the limited support available, shows the adoption potential in real life. This is also true for health facilities in vulnerable settings, including post-war humanitarian settings.

The monitoring feature of the study demonstrated that the outcomes are well aligned with the WHO recommendation for monitoring and tracking of public health concerns such as syphilis, which poses a risk not only to the mother, but also to the foetus. We have also indicated steps on how to identify mother-to-child infection transmission risks using congenital syphilis. Running subsequent queries over the data and narrowing those down to pregnancies that have syphilis, the cases with greater gestational age could be monitored closely and could result in a recommendation for preferential treatment to reduce the risk of mother-child transmission of the infection.

Beyond syphilis surveillance, the implication of this approach extends to a wide range of healthcare applications. By leveraging federated query, learning and analytics, healthcare providers can develop predictive models for maternal and child health, and infectious disease management. This approach aligns with global efforts to strengthen digital health ecosystems, enabling data-driven healthcare innovations that can improve patient outcomes and optimise resource allocation.

Conclusion

This research was driven by two central questions. The first examined how insights into infectious diseases, supported by data capture and storage in health facilities across border settings, can be harmonised using controlled semantic vocabularies. The second focused on how syphilis cases, which have the potential to lead to congenital syphilis, can be effectively identified and monitored to prevent mother-to-child transmission. Using the functionalities in the health data space architecture, this study demonstrates how to leverage the value of FAIR data for improved health outcomes.

A key finding of this study is the positive testing of the data-seeking process by making data queries using SPARQL remote queries to facilitate data visiting. This means that, provided permission is given, health workers in higher level facilities can make queries to lower healthcare facilities to assist the continuum of care across referrals of patients. Health workers in lower-level facilities can also make queries on data, if permission is provided, to generate surveillance insights.

The ability to send algorithms that query the data for computational analysis and return results enables federated learning. This approach allows for direct querying of distributed datasets without the need to physically transport data to a central repository. It provides strong control over access permissions, ensuring that only authorised users can query specific subsets of data within the health facility, with the purpose of the computation clearly defined before access is granted.

Health facilities can retain data sovereignty while enabling real-time, federated analysis. This federated approach enhances data privacy and the efficiency of healthcare interventions by ensuring that patient data remains in its original context while still being accessible for broader analytical purposes. It also means that the data of patients curated in the health facility add to the quality of services for that patient. One of the critical elements of this study was the implementation of knowledge graphs and linked data at health facility level. This makes identification of the susceptible cases of congenital syphilis stand out and allows for timely treatment. These tools enable deeper insights by establishing relationships between those cases that have syphilis with different datasets on syphilis and other infections. The procedure that was tested, operationalises the FAIR-OLR principles.

To ensure availability of the right data and infrastructure, the Africa Health Data Space framework introduces a structured, federated approach to data querying, federated learning, and multi-level insight generation. Unlike conventional centralised health data systems, this model promotes decentralised data management, ensuring that healthcare facilities retain control over their data while still contributing to a larger intelligence network on surveillance and clinical studies in health. This could be extended also to Pharmacovigilance. By adopting this framework, healthcare providers and policymakers can monitor infectious diseases more effectively, allocate resources strategically, and improve public health interventions. The strength of the procedure is that the provenance of the data is real patient data.

The health facilities act as knowledge hubs, as they are host the knowledge, which is only accessible locally, but can also be queried remotely. Knowledge sharing can foster in-facility and across-facility learning. The development of cross-facility and cross-country analytics further strengthens the value of this data-driven approach. By comparing local data with broader benchmarks, healthcare providers can identify trends, refine treatment protocols, and assess health system performance in a global context. This comparative framework allows decision-makers to draw lessons from other facilities and countries. While each medical case is unique, this method fosters knowledge-sharing and adaptive learning, improving patient outcomes and healthcare strategies.

Through semantic reasoning and inference, new patterns and correlations can be uncovered, supporting evidence-based healthcare policies and personalised treatment approaches. Based on the semantic data model, those cases will be automatically linked with others that have a similar pattern. This structured linkage of data nodes enhances the ability of researchers and practitioners to discover hidden associations between diseases, treatments, and patient outcomes, ultimately leading to improved healthcare decision-making.

This study demonstrates how syphilis surveillance can be conducted when data is curated as FAIR. As syphilis remains a significant public health concern, particularly in regions with limited access to diagnostic tools and treatment resources, this approach can provide support. By implementing FAIR principles, health facilities can improve data transparency, interoperability, and reuse, ensuring that syphilis data is available for real-time monitoring and policy decision-making.

Identifying data points that predict syphilis, picked from the digital patient records of pregnant women of patient data held within the health facility, pregnant mothers at risk of infecting their baby were identified. Data visiting across countries with data held in residence was demonstrated. This study illustrates how a federated, knowledge-driven health data system can transform healthcare analytics, moving from aggregated statistics to granular, real-time insights. This approach not only improves infectious disease surveillance and response but also enhances data-driven decision-making at all levels of healthcare delivery. By embedding FAIR principles, knowledge graphs, and federated query mechanisms, this approach ensures that health data is more actionable, accessible, and ethically managed. As the model continues to evolve, it has the potential to revolutionise public health strategies across Africa and beyond, fostering a more resilient and responsive healthcare system.

Acknowledgements

Sincere thanks to the administrators and medical staff, health workers and data stewards in the health facilities who assisted in conducting this study. The study was implemented in low-resource conditions and the research team is so grateful for accepting to participate in this study.

The researchers express their gratitude to Getu Taye and Tesfit Gebremeskel, who advised on the technical engineering organisation of the study.

I would also like to thank the Country Coordinators in Nigeria, Uganda and Kenya for their support to the implementation of the study.

Authors' Contributions

Samson Yohannes Amare is the main researcher of this study He conceptualised the research, wrote the research and implementation plan, set up the approach and conducted all main elements of this study. **Liya Mamo** supported the data management plan and helped instruct the data stewards in the health facilities. **Araya Medhanyie** assisted in the medical decisions underlying the approach and analysis of the study. He also provided comments on the various drafts of this chapter. **Mirjam van Reisen** supervised the overall research and edited the different drafts and provided comments on all drafts.

Ethical Considerations

Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023, for VODAN research group. Data Use Agreements were obtained signed by administrators of Health facilities, Ministries/Bureaus of Health and, the Country Coordinator of the VODAN Research group and the VODAN coordinator.

Tilburg University, Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences REDC#2020/013, June 1, 2020-May 31, 2024, on Social Dynamics of Digital Innovation in remote non-western communities. Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023.

References

Africa University Network on FAIR Open Science (AUN-FOS). (2025). *Africa Health Data Space.* https://aun.mu.edu.et/ahds/

Aktau, A., Amare, S. Y., Van Reisen, M., Taye, G. T., Gebremeskel, T. G., Jati, P. H., & Plug, R. (2025). GO TRAIN: A protocol for metadata creation for the FAIRification of patient data health records. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space.* Bamenda: Langaa

Aktau, A. (2025). The Potential of Adoption of FAIR Guidelines in Digital Healthcare in Kazakhstan. In M. Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space.* Bamenda: Langaa

Amare, S. Y., Taye, G. T., Gebreselassie, T. G., Plug, R., & Van Reisen, M. (2023). Realizing health data interoperability in low connectivity settings: The case of VODAN-Africa. *FAIR Connect, 1*(1), 55-61. IOS Press. https://doi.org/10.3233/FC-221510

Amare, S. Y., Medhanyie, A. A., & Van Reisen, M. (2024). Data visiting in digital black holes: FAIR-based digital health innovation during war. In M. Van Reisen, A. A. Medhanyie, & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 3*. Langaa, Bamenda.

Amare, S.Y., Weldu, L.M., Gebremeskel, T. (2025). Harmonizing Antenatal Care Records Across Four African Countries: A Comparative Analysis and Development of a Standardized Data Model. In M. Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space.* Bamenda: Langaa

Amare, S. Y., Taye, G. T., Plug, R. B. F., Medhanyie, A. A., & Van Reisen, M. (2025). Federating tools for FAIR patient data: Strengthening maternal health and infectious disease surveillance from clinics to global systems. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space.* Bamenda: Langaa

Bezuidenhout, L. (2020). Being fair about the design of FAIR data.

Callard, B. (2025) Complexities of recording and reporting perinatal data: A Case Study from Namibia. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa

Charumbira, F. (2024). Closing remarks. *Speech held at the conference Gi FAIR Go Africa.* https://aun.mu.edu.et/ahds/conferences/

Countdown 2020. Available at: https://www.countdown2030.org/about

Elarrat Canto, S. V., Leite Araújo, M. A., Espinosa Miranda, A., Paulo Cardoso, A. R., & Freitas de Almeida, R. L. (2019). Fetal and infant mortality of congenital syphilis reported to the Health Information System. *PLoS One, 14*(1), e0209906

Folorunso, S. O., Oladipo, F. O., & Van Reisen, M. (2025). A higher education curriculum for cultural competence, representation, and social responsibility in AI and FAIR data practices. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa

Gebreslassie, T. G., Kidanu, G., Weldu, L. M., Amare, S. Y., & Van Reisen, M. (2024). Life in darkness: The communication blockade during the Tigray siege. In M. Van Reisen, A. A. Medhanyie, & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 3*. Langaa, Bamenda.

GO FAIR. (2022). IN-Africa - GO FAIR. GO FAIR. https://www.go-fair.org/implementation-networks/overview/implementation-networks-archive/in-africa/

Jati, P. H., & Van Reisen, M. (2025). Narratives in public agenda-setting for FAIR data and health data management in Africa: Enhancing maternal health and infectious disease outcomes. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa

Jati, P. H., Lasroha, M. S. T., & Van Reisen, M. (2025). Data privacy and interoperability in Indonesia's SATUSEHAT platform. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. (pp. xx–xx). Bamenda: Langaa

Jati, P. H., Amare, S. Y., Kawu, A. A., Nandwa, W., Taye, G. T., & Van Reisen, M. (2025). GO CHANGE: Adoption of FAIR-OLR architectures to support insights from patient health data records in Africa. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. (pp. xx–xx). Bamenda: Langaa

Kahsay, B. G., Van Reisen, M., & Lin, Z. (2025). GO FAIR: Ontology development of health semantics with cultural specificity: Traditional health practices using Tsebel in conflict zones. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa

Kojima, N., & Klausner, J. D. (2018). An update on the global epidemiology of syphilis. *Current Epidemiology Reports, 5*, 24-38.

Kufa, T., Thabane, E., Leshaba, D., Makhubele, M., & Puren, A. (2023). Congenital syphilis annual surveillance report, South Africa 2023. *Public Health Bulletin South Africa.*

Medhanyie, A. A., Wuneh, A. D., Tefera, A. H., Stocker, J., Kidanu, G., Abebe, G. G., & Van Reisen, M. (2024). Genocide through health care violence: The systematic destruction of health facilities in the Tigray

war. In M. Van Reisen, A. A. Medhanyie, & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 3*. Langaa, Bamenda.
Melberg, A., Hailemariam, A., Sisay, T. A., Sisay, M. M., & Moland, K. M. (2019). 'Maternal deaths should simply be 0': Politicization of maternal death reporting and review processes in Ethiopia. *Health Policy and Planning, 34*(7), 492–498. https://doi.org/10.1093/heapol/czz075
Ministry of Health and Social Services. (2022). *National maternal health, death, near miss, stillbirth and neonatal death review committee: Report on the confidential enquiry into maternal deaths, stillbirths, and neonatal deaths for the period of 1 April 2018 to 31 March 2021*. Republic of Namibia, Windhoek.
Morifi, M., Malevu, N., Odayan, S., McCarthy, K., & Kufa, T. (2021). Congenital syphilis case surveillance in South Africa 2017–2019: Experience, challenges, and opportunities. *Journal of Tropical Pediatrics, 67*(4), fmab079.
Nalugala, R., Jati, P. H., Amare, S. Y., Omare, M., Wairimu, J., Kahiro, C., Nandwa, W., Okeyo, S., Kinoti, D., Aktau, A., Mulingwa, A., & Van Reisen, M. (2025). Introducing data sovereignty over patient data: Patient data ownership in residence of health facilities in Kenya. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa
Niguse, S., Tseghay, H. T., & Van Reisen, M. (2024). A reinforcement feedback loop: Medical care services in Ayder Hospital during war. In M. Van Reisen, A. A. Medhanyie, & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 3*. Langaa, Bamenda.
Roberton, T., Carter, E. D., Chou, V. B., Stegmuller, A. R., Jackson, B. D., Tam, Y., Sawadogo-Lewis, T., & Walker, N. (2020). Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: A modelling study. *The Lancet Global Health, 8*(7), e901–e908. https://doi.org/10.1016/S2214-109X(20)30229-1
Ros, F., Kush, R., Friedman, C., Gil Zorzo, E., Rivero Corte, P., Rubin, J. C., & Van Houweling, D. (2021). Addressing the COVID-19 pandemic and future public health challenges through global collaboration and a data-driven systems approach. *Data Intelligence, 5*(1), e10253.
Sanders, L. (2008). An evolving map of design practice and design research. *Interactions, 15*(6), 13–17. https://doi.org/10.1145/1409040.1409043
Sappenfield, E., Jamieson, D. J., & Kourtis, A. P. (2013). Pregnancy and susceptibility to infectious diseases. *Infectious Diseases in Obstetrics and Gynecology, 2013*(1), 752852.
Stocker, J., & Medhanyie, A. A. (2024). Measuring system change: Shifts in the health landscape under the Tigray siege. In M. Van Reisen & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 1* (pp. 407-

438). Langaa, Bamenda. https://www.researchgate.net/publication/385402687_Tigray_War_in_a_Digital_Black_Hole_Book_3
Stocker, M., Stokmans, M., & Van Reisen, M. (2022). Agenda setting on FAIR guidelines in the European Union and the role of expert committees. *Data Intelligence, 4*(4). https://doi.org/10.1162/dint_a_00168
Taye, M., Medhanyie, A. A., & Van Reisen, M. (2024). War-related destruction of the digital health data infrastructure: Discovering features for a resilient digital health information system. In M. Van Reisen, A. A. Medhanyie, & M. Mawere (Eds.), *Tigray. War in a digital black hole, Volume 3* (pp. 439-476). Langaa, Bamenda. https://www.researchgate.net/publication/385554308_Warrelated_Destruction_of_the_Digital_Health_Data_Infrastructure_Discovering_Features_for_a_Resilient_Digital_Health_Information_System
Terris-Prestholt, F., Watson-Jones, D., Mugeye, K., Kumaranayake, L., Ndeki, L., Weiss, H., & Hayes, R. (2003). Is antenatal syphilis screening still cost-effective in sub-Saharan Africa? *Sexually Transmitted Infections, 79*(5), 375-381.
Trivedi, S., Taylor, M., Kamb, M. L., & Chou, D. (2020). Evaluating coverage of maternal syphilis screening and treatment within antenatal care to guide service improvements for prevention of congenital syphilis in Countdown 2030 countries. *Journal of Global Health, 10*(1), 010504. doi: 10.7189/jogh.10.010504
Van Reisen, M., Stokmans, M., Mawere, M., Basajja, M., Ong'ayo, A., Nakazibwe, P., Kirkpatrick, K., Kudakwashe, C. (2020). FAIR practices in Africa. In: *Data Intelligence* 2, 264–275. doi: 10.1162/dint_a_00049 (eds. Mons, B., Jacobsen, A. & Schultes, E.). MIT Press, http://www.mitpressjournals.org/dint
Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihigo, M., Folorunso, S., Schultes, E., Basajja, M., Aktau, A., Yohannes Amare, S., Tadele Taye, G., Hadi Purnama Jati, P., Chindoza, K., Wirtz, M., Ghardallou, M., van Stam, G., Ayele, W., Nalugala, R., Abdullahi, I., Osigwe, O., Graybeal, J., Abrha Medhanyie, A., Abubakar Kawu, A., Liu, F., Wolstencroft, K., Flikkenschild, E., Lin, Y., Stocker, J., & Musen, M. A. (2021). Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics, 2*(2). https://doi.org/10.1002/ggn2.10050
Van Reisen, M., Oladipo, F., Mpezamihigo, M., Plug, R., Basajja, M., Aktau, A., Purnama Jati, P. H., Nalugala, R., Folorunso, S., Amare, Y. S., Abdulahi, I., Afolabi, O. O., Mwesigwa, E., Taye, G. T., Kawu, A., Ghardallou, M., Liang, Y., Osigwe, O., Medhanyie, A. A., & Mawere, M. (2022). Incomplete COVID-19 data: The curation of medical health data by the Virus Outbreak Data Network-Africa. *Data Intelligence, 4*(4). https://doi.org/10.1162/dint_e_00166

Van Reisen, M., Amare, S. Y., Nalugala, R., Taye, G. T., Gebreselassie, T. G., Medhanyie, A. A., Schultes, E., & Mpezamihigo, M. (2023). Federated FAIR principles: Ownership, localisation and regulatory compliance (OLR). *FAIR Connect, 1*(1), 1–7.

Van Reisen, M., Yohannes, S., Plug, R., Tadele, G., Gebremeskel, T., Kawu, A. A., Smits, K., Woldu, L. M., Stocker, J., Heddema, F., Folorunso, S., Kievit, R., & Medhanyie, A. A. (2024). Curation of federated patient data: A proposed landscape for the Africa health data space. In A. L. Imoize, M. S. Obaidat, & H. Song (Eds.), *Federated learning for digital healthcare systems* (Intelligent Data-Centric Systems series). Elsevier. https://doi.org/10.1016/B978-0-443-13897-3.00013-8

Van Stam, G. (2017). *Reflections: A Narrative on Displacement of Technology and Meaning in an African Place* [Doctoral dissertation, Tilburg University]. Tilburg University Research Portal.

VODAN. (2021, January). *Technical requirements document [Version 1]: Phase II implementation project*. VODAN Africa & Asia.

WHO Africa Region. (n.d.). *Prevention of mother-to-child transmission of HIV/AIDS program.* https://www.afro.who.int/node/9110

World Health Organization. (2017). *WHO guideline on syphilis screening and treatment for pregnant women*. Geneva: WHO. https://iris.who.int/bitstream/handle/10665/259003/9789241550093-eng.pdf

World Health Organization. (2011). *Methods for surveillance and monitoring of congenital syphilis elimination within existing systems.* Geneva, Switzerland: Author.

World Health Organization. (2024, May 21). *Syphilis.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/syphilis

Zhang, S., Dos Santos Vieira, B., Bernabé, C. H., Abaza, H., Benis, N., Cámara, A., Cornet, R., Le Cornec, C. M. A., ‘t Hoen, P. A. C., Schaefer, F., van der Velde, K. J., Swertz, M. A., Wilkinson, M. D., Jacobsen, A., & Roos, M. (2022). Towards FAIRification of sensitive and fragmented rare disease patient data: Challenges and solutions in European Reference Network registries. *Orphanet Journal of Rare Diseases, 17*, 436. https://doi.org/10.1186/s13023-022-02558-5

Zhu, Z., Zhu, X., Zhan, Y., Gu, L., Chen, L., & Li, X. (2022). Development and comparison of predictive models for sexually transmitted diseases—AIDS, gonorrhea, and syphilis in China, 2011–2021. *Frontiers in Public Health, 10*, 966813. https://doi.org/10.3389/fpubh.2022.966813

---

[i] Two changes were made tot he original chapter on 7 May 2025 to add ‘simuation’ tot he study design to reflect that the location of the data in location was simulated in the study design.