Design and Implementation of FAIR principles in digital health data in Uganda, Africa

Basajja, M.

Chapter 7

Privacy-Preserving Health Data Analytics in Uganda: A Comparative Study of DHIS2 and FAIR VODAN-Africa Architecture

This chapter is partially based on the following publication:

Basajja, M. and Paul-Chima, U. O. (2025) 'Comparative Effectiveness of DHIS2 and FAIR Data Approaches for Privacy-Preserving Health Data Analytics in Uganda: A Systematic Review', ClinicoEconomics and Outcomes Research, 17, pp. 849–864. doi: 10.2147/CEOR.S560265.

Abstract

The increasing use of digital technologies in health systems offers immense potential to improve the reuse of clinical and health data. In Uganda, the implementation of DHIS2 (District Health Information Software 2) and the adherence to FAIR(Findable, Accessible, Interoperable and Reusable) data principles represent significant steps toward improving health analytics. This chapter examines the comparative strengths and limitations of DHIS2 and FAIR data in ensuring data privacy while enabling accessibility and reusability for effective health analytics. Through a detailed critical analysis of their functionalities, data governance frameworks, and application contexts, the study identifies best practices and challenges to ensure privacy-preserving analytics. The chapter elaborates on the integration of VODAN-Africa architecture to enhance the findability, accessibility, integration, and reuse of clinical and research data in various healthcare facilities in Uganda.

Keywords: metadata templates, community standards, digital health data, Outpatient Register, FAIR principles, FAIR data, data life cycle, FAIR data management, data retrieval, data workflows, semantic web standards, ontology, Bio Portal

The source code for the scripts in this chapter can be found here: `https://github.com/mariambasajja/CH7-THESIS-CODE`

7.1 Introduction

The effective use of information at various levels within the health system is fundamental for the planning and delivery of health services. This includes its role in influencing policy decisions, programming actions, and ensuring evidence-based practices [138]. The District Health Information System version 2 (DHIS2), a flexible open source electronic information system, has been adopted in more than 60 countries, including Uganda, to manage and visualize routine health data [139]. Bhattacharya et al. demonstrated the feasibility of using DHIS2 for the local, real-time monitoring of maternal and newborn health services in low-resource settings. The Health Management Information System print-based forms and registers are compiled and submitted to the DHIS2 system for analysis. However, after submission, data analyses are not easily accessible at the point of care due to several factors. The DHIS2 system is often managed at a central or regional level, away from local health facilities, creating a physical disconnect. Many point-of-care locations, especially in rural or resource-constrained areas, may lack reliable Internet access to retrieve analyzed data from the DHIS2 system. The implementation and effective use of DHIS2 remain hindered by significant challenges, including limited interoperability, concerns about data privacy, and a lack of seamless integration with other systems. These issues underscore the need for innovative approaches, such as incorporating FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, to enhance data management and ensure privacy-preserving health analytics.

Using the VODAN-Africa architecture, the data was transformed into machine-actionable data that conforms to the FAIR principles. This supports machine-actionable data and increases health data interoperability, thus enabling efficient incorporation in other systems, analysis workflows, and future applications, including third-party applications that interface with DHIS2.

The primary challenge lies in the current state of health information management in Uganda and many African nations. The Health Management Information System (HMIS) exists primarily as paper forms compiled in registers [144], with some facilities implementing basic digital systems. These pre-printed paper formats, provided by the Ministry of Health to all health facilities, face multiple challenges: poor data quality, late submissions, incomplete capture of data from private health providers, inadequate data segregation, limited political support, and insufficient funding [141]. Additionally, limited access to computers and the reliable Internet impedes the full implementation of digital systems in areas limited by resources. The current health information system workflow consists of the steps, shown in Figure 7.1.

A Health Information System (HIS) follows a hierarchical data flow starting from primary data capture at healthcare facilities and progressing toward consolidated national-level reporting. At the initial stage, data is recorded manually using pre-printed paper forms at the point of care. This represents the foundational level of the system, where raw patient- or service-level information is collected directly within healthcare facilities. Following collection, the data is processed through two parallel pathways: data summarization and data aggregation. In the summarization pathway, raw entries are manually reviewed and converted into simplified indicators such as total counts of patients, cases, or treatments. In parallel, the aggregation pathway involves organizing individual records into structured datasets based on relevant dimensions such as time period, disease category, or healthcare facility. The outputs from both pathways are subsequently integrated into the District Health Information Software 2 (DHIS2) platform, which serves as the central digital repository for health information reporting. Within DHIS2, the data is digitized, standardized, and stored in a structured format, enabling consistent re-

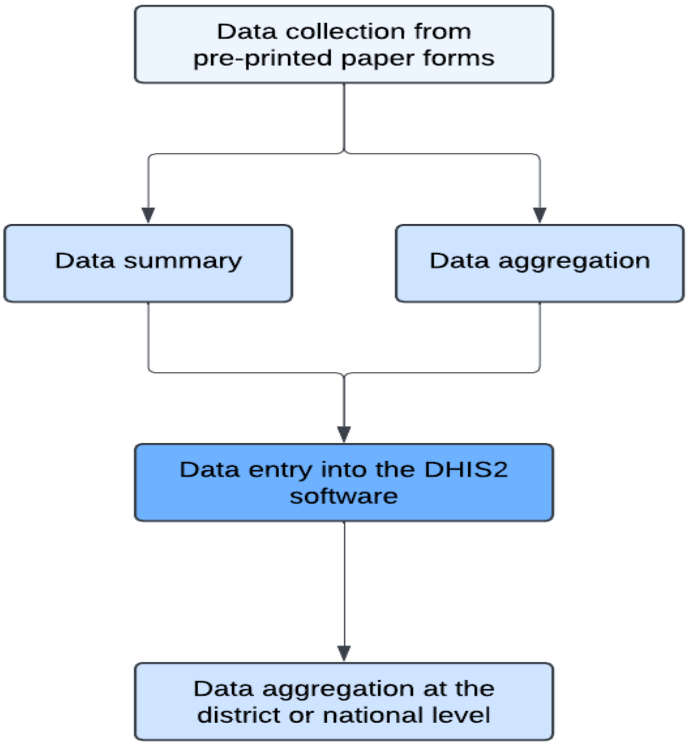

Figure 7.1: Workflow diagram illustrating the Health Information System data management process

porting across the health system. Finally, the integrated data is further aggregated at higher administrative levels, including district and national tiers. This hierarchical consolidation supports health authorities and policymakers in analyzing trends, monitoring disease patterns, and making informed, evidence-based decisions for healthcare planning and intervention.

In a review of the utilization of DHIS2 data for decision making in Ghana, Odei-Lartey et al. [143] identified critical barriers including the absence of guide manuals, infrequent training, and infrastructure limitations. Despite these challenges, their study demonstrated instances where DHIS2 data contributed to action-oriented decisions, emphasizing the system's potential when properly supported and implemented. The terminology used by Uganda's Health Management Information System(HMIS) registers, organized as tables with column headings, represents progress toward interoperability. However, the lack of standardized representation in semantic web formats such as SKOS, OWL, and RDF / XML limits the reusability of data and the interoperability with external datasets. While existing terms and codes facilitate local data usage, integration with broader semantic web standards requires not only adopting these formats, but also dedicated resources to create and maintain the actual ontologies and vocabularies within them.

A critical consideration in health data management is the preservation of privacy and access under well-defined terms whenever needed. Health facilities routinely generate large amounts of clinical data that have potential uses beyond patient diagnosis [147], including research, plan-

ning, decision-making, and improvement in service delivery. However, balancing data utility with patient privacy remains a pressing challenge. Although the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles promote openness and reuse, their implementation must address potential risks to patient confidentiality.
Issues of handling, sharing, and retrieving patient information in healthcare facilities are a challenge, especially in developing countries like Uganda [276]. Technology has advanced massively in the field of health, with digital health technologies promising to improve patient experiences, better resource use, and more effective policies for better health. However, these advancements come with a significant challenge [284]. The basic challenge of creating platforms for handling such big data results from the required balance between privacy protection for patients and making the related data easily accessible and analyzable [271]. This balance becomes even more important as health systems in developing nations turn to digital tools as a means to fix significant health challenges [285]. It also includes, but is not limited to, disease, surveillance, maternal and child health, as well as epidemics.

Digital health tools like DHIS2 (District Health Information Software 2) and the adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles are pivotal in addressing these challenges. DHIS is an open-source health information management system implemented in Uganda for data collection, compilation, and reporting [310]. On the other hand, FAIR principles set a general practice applicable internationally as a data management practice focusing on the reusability of data, which is understood as machine readability and interoperability without compromising data confidentiality [269]. While both conceptually and operationally closely related, these approaches have different focus areas, methods of execution, and methods for maintaining user anonymity and privacy [308].

The VODAN-Africa project has taken steps to address these challenges by incorporating tools from the *Center for Expanded Data Annotation and Retrieval (CEDAR)* [136] to enhance metadata management and make health data FAIR-compliant (as explained in Chapter 6). The project has undertaken the conversion of Uganda's HMIS Registers to a standards-based knowledge representation [146], deploying a localized CEDAR-based system integrated with visualization and query tools [148].
This study aims to analyze the intersection of the FAIR principles with the implementation of DHIS2 in Uganda and broader Africa, focusing on examining how the FAIR principles can enhance the data management capabilities of DHIS2 while maintaining robust privacy preservation for health analytics and evaluating the practical challenges and opportunities to implement these principles within existing health information systems.

7.1.1 Motivation

The motivation for this research stems from the growing recognition of data as a strategic asset in healthcare delivery and public health decision-making. In Uganda, a country grappling with resource constraints and a high disease burden, robust health data systems can dramatically improve health outcomes. However, the sensitive nature of patient data necessitates a dual focus on usability and confidentiality [281]. Data breaches or mismanagement could undermine public trust, exacerbate health inequalities, and lead to serious ethical and legal ramifications [275].
Moreover, while DHIS2 has been successfully deployed in Uganda, there is limited evidence

on how well it safeguards patient privacy in alignment with global data stewardship standards like FAIR principles [271]. By critically examining these two approaches, this research seeks to address a knowledge gap and provide actionable insights for policymakers, health administrators, and technologists in Uganda and similar countries.

7.1.2 Research Significance

This research holds substantial significance for several stakeholders:

- It highlights best practices for managing patient data securely while enabling operational efficiency in public health interventions.
- The findings will inform regulatory frameworks for data privacy, aligning with Uganda's Data Protection and Privacy Act (2019).
- By comparing DHIS2 and FAIR principles, this study identifies opportunities for technical innovation and system integration.

7.1.3 Research Objectives

The primary aim of this research is to evaluate the applicability of DHIS2 and FAIR data principles in fostering privacy-preserving health analytics for patient data in Uganda.

Specific objectives include:

- To explore and discover ways of facilitating the secure sharing of clinical health data in a Findable, Accessible, Interoperable, and Reusable way under well-defined terms using the VODAN-Africa architecture for data privacy-preserving analytics in Uganda.
- To evaluate and compare privacy-preserving health data analytics in aggregated, FAIR-ified DHIS2 datasets retrieved from the AllegroGraph database, which functions as a repository for linked data, specifically storing information in the RDF triple format [175].
- To analyze its mechanisms, such as role-based access control and encryption, for safeguarding patient data.
- To examine how these principles enhance data sharing and interoperability while maintaining privacy.
- To highlight barriers to implementing both systems effectively, including technical, legal, and organizational factors.
- To recommend strategies that leverage the strengths of both DHIS2 and FAIR principles to achieve optimal outcomes for Uganda's healthcare system.

7.1.4 Research Questions

This chapter explores the possibility of improving the analysis of health data at the point of care to propagate, integrate, and maintain it through FAIR data.

Research Questions

Main Research Question: Can implementing FAIR Data principles enhance the privacy-preserving accessibility of various types of health data and support the aggregation and analysis of these data?

Sub-Research Questions:

Question 1:
How does VODAN-Africa's architecture ensure FAIR principles while preserving data privacy?

Question 2:
How can VODAN-Africa dashboards influence healthcare decisions across African countries based on FAIR data analytics?

7.2 Related Work

7.2.1 Overview of DHIS2

DHIS2 (District Health Information Software 2) is a widely used open-source software for health information management. It has been developed by the Health Information Systems Program (HISP) and provides a set of tools for data collection, validation, analysis, and visualization, which allows effective monitoring and reporting of the health system. In Uganda, DHIS2 plays an important role in health data compilation at the national and sub-national levels [264].

Implementation of DHIS2 in Uganda

DHIS2 implementation in Uganda started in 2009, encouraged by the government due to constraints with the paper-based health data management system. Before migrating to the DHIS2 system, the Ministry of Health (MoH) used EpiInfo, an Access-based system, and it proved to be especially challenging in the process of recording health facility data and aggregation on the national level [252]. These manual operations created problems such as time-consuming reporting, wrong information being documented, inaccuracies, and poor management of health information [291].
To overcome these challenges, the MoH, together with HISP Uganda and other partners, introduced and adopted the use of DHIS as the national electronic Health Management Information System (eHMIS), commonly known as DHIS2. Therefore, it was instrumental in realizing the reporting of health data in digital format, making it possible to enter the data on real real-time basis, enhancing the data quality, and making the data easily accessible and usable for the decision-making processes. This flexibility made it possible to apply them in addressing the country's particular health information requirements, including dimensions of the different health programs and indicators [254].

Impact & Significance

The strengthening of routine data collection and utilization in Uganda has immensely benefited from the implementation of DHIS2 [267]. In the health sector, it has enhanced the general health status and disease monitoring and thus informed decision-making [262]. The system can receive and analyze large amounts of data through real-time entry, which has improved the way health programs are supervised and resources are distributed and allocated.

Overall, Uganda's experience with DHIS2 underscores the platform's versatility and effectiveness in strengthening information systems, promoting transparency, and enhancing service delivery in the healthcare sector [267].

7.2.2 FAIR Data Principles

The FAIR data principles (Findability, Accessibility, Interoperability, and Reusability) provide a global framework to optimize the usability of data. In Uganda, the adoption of these principles is still growing more and more, especially within the fields of digital health and research data management [278].
The initiative of the Uganda National Council for Science and Technology (UNCST) and its international counterparts to advance the principles of FAIR data among Ugandan researchers is particularly noteworthy [263]. This initiative attempts to enhance the exchange and interoperability of data across multiple scientific domains to create a more collaborative scientific community.
Moreover, during the celebrations of Open Data Day 2020 in Uganda, stakeholders noted the aspect of FAIR data in enhancing the accountability of creating transparency and supporting monitoring, evaluation, and learning practices [268]. From the discussions, it was evident that government and district planning committees need to appreciate the importance of opening up more key data to assist in informed decision-making.
However, challenges related to the implementation of the FAIR principles in Uganda are encountered, such as low awareness, poor infrastructure support, and the absence of broad and harmonized data management policies. Solving these issues is essential to unlock the full potential of FAIR data to improve the quality of both research and public health [271].

Privacy Challenges in Uganda

Uganda's health sector encounters several significant challenges in managing patient data, particularly concerning privacy and data protection:

- Inadequate Data Protection Laws: Even though Uganda passed the Data Protection and Privacy Act in 2019, the enforcement currently remains limited and weak. A legal analysis conducted in 2015 pointed out that while Ugandans enjoy the protection of their right to privacy in the Constitution, several laws require one to disclose his/her health information in some or other circumstances, thereby eroding his/her privacy rights [264].

- Resource Constraints: The health sector experiences challenges of scarcity of healthcare human capital and restricted infrastructure to put in place elaborate data privacy systems. These constraints limit the ability to provide an effective data governance framework and to meet data privacy requirements [292].

- Cultural and Ethical Considerations: The urge to respect traditional cultural ways when managing personal health information, on the one hand, and the modern world's desire to keep health information private, on the other hand, complicates the process of managing health data. These socio-cultural beliefs and practices can also determine perceptions of privacy in Ugandan society, affecting the acceptance and implementation of data protection measures. In addition, where the patient has AIDS or any other diseases that are socially unacceptable, patient information must be handled with a lot of care to avoid stigmatization of the patient [287].

Addressing these challenges requires a multifaceted approach, including strengthening legal frameworks, investing in capacity building, and fostering a cultural shift towards recognizing the importance of data privacy in healthcare.

7.3 Methodology

This chapter compares privacy-preserving health analytics between datasets generated from the existing data architecture within health facilities (DHIS2) and FAIR data generated after implementing the VODAN-Africa architecture. The chapter elaborates on the integration of VODAN-Africa and FIP tools to enhance the findability, accessibility, integration, and reuse of clinical and research data across various healthcare facilities in Uganda. The focus is on how these tools streamline the exchange of health data while ensuring patient privacy.

The development of the FAIR Implementation Plans (FIPs) was a collaborative effort, as detailed in the Paper "Basajja, M.: Comparing DHIS2 and FAIR data for privacy-preserving health analytics of patient data in Uganda" and further explained in Chapter 7 of this thesis. These FIPs were designed before deployment and specifically tailored to meet the needs of Ugandan health facilities. They provided a structured approach to digitizing and transforming routine Health Management Information System (HMIS) data collection registers into machine-actionable FAIR data using the VODAN-Africa toolset.

During deployment, clinical and research data were merged into the VODAN-Africa architecture following a set of procedures that prioritized the protection of patient privacy. This work is described in greater detail in Section 7.4 of this thesis. The methodology also details the processes for accessing, using, and analyzing the digitized data, ensuring compatibility and smooth operation across platforms, as elaborated in the data section of this chapter.

Data analysis was conducted on the digitized datasets aligned with FAIR principles. The study focused on how the collected data could be accessed, tested, and integrated for cross-continental comparisons and analytics. Specific methodologies for data collection, testing, and integration are described in the Data Section 7.4 of this chapter, where I will explicitly explain which data sets were used, how they were accessed, and the processes followed to ensure that research information flows smoothly between platforms. The results demonstrate the ability of this architecture to analyze and support decision-making across distributed datasets while maintaining privacy-preserving standards.

7.3.1 Data Privacy and Security Considerations and Identification of the Key Statistics of the Selected Datasets

The GDPR Law, introduced in 2018 [149], ensures that organizations and other bodies operating within the EU or collecting data from users within the EU must comply with certain policies detailed in this law. We are also dealing with data from Uganda; we must ensure compliance with local data protection and privacy laws, namely the Data Protection and Privacy Act introduced in 2019 [150, 151]. However, about this project, if we ensure compliance with GDPR along with the Ugandan Data Protection and Privacy Act, the security of health data information is enforced. The GDPR and Ugandan Data Protection and Privacy Act were highly considered and implemented during the implementation of the VODAN-Africa architecture within the different health facilities. Any sensitive data that could be traced back to a single person was excluded. The data subject decides whether the data should not be exposed. The following set of layers was identified as necessary for the VODAN-Africa community:

1. Data processing agreement among all data processing partners.

2. FAIR Equivalency tool based on the regulatory framework in the different contexts.

3. Agreement about machine-actionable templates based on the three Health Management Information System registers, outpatient, Maternal Child Care/Antenatal.

4. An agreement on the data repositories and security to keep the data secure within the residence where the data is stored.

5. An agreement on the data pipelines used for the dashboards providing real-time aggregate data analytics at the facility.

6. An agreement on the data pipelines used for the dashboards providing real-time aggregate data analytics at the community level.

7.3.2 Data Collection

For this preliminary study, four datasets were identified for analysis. These included data from the HMIS Outpatient Register 002 and the Antenatal Care Register (ANC), both of which were produced locally by VODAN-Africa and Asia health facilities. Additionally, two Point of Care datasets were included: one dataset containing data on HIV, schistosomiasis, and HPV for the period 2013-2022 (9 years) from the Ministry of Health (MOH) Uganda and three districts—Lira, Entebbe, and Hoima; and another dataset containing data on schistosomiasis cases and deaths from different districts in Uganda.

The data from the HMIS Outpatient Register 002 and ANC were originally collected on paper forms and submitted to the DHIS2 system daily, weekly, or monthly, as per the standard reporting requirements. Agreements were reached between health facilities and several key individuals involved in the initiative, such as the general coordinator, the executive director of VODAN Africa, and the national Country Coordinators for the use of these data. Specifically, the facilities provided the paper-based data to VODAN-Africa for digitization and integration into the VODAN-Africa system. For this study, anonymized data were used, with the necessary permissions obtained from the relevant Ugandan authorities.

To ensure that the data was accessible and reusable at the point of care, the VODAN-Africa architecture was employed to digitize and "FAIRify" the data. This process involved converting the paper-based registers into machine-actionable formats using locally installed CEDAR platforms within selected health facilities (as described in detail in Chapter 6 section 6.7.3).

This digitization and transformation process was done through template creation and data production, ensuring the data adhered to the FAIR principles (Findable, Accessible, Interoperable, and Reusable).

Data were stored in the AllegroGraph triple store and queries were performed using the internal and community dashboards. These platforms allowed for easy retrieval and analysis of the data, facilitating improved access to point-of-care data. Integration of these data into the VODAN-Africa system demonstrates its potential to improve healthcare outcomes by enabling access to real-time data at the point of care.

In terms of privacy preservation, strict data protection measures were followed. The anonymized data was digitized, ensuring that personally identifiable information was not included. Additionally, data was only made accessible to authorized users within healthcare facilities, with encryption and other privacy-enhancing technologies employed to safeguard sensitive information.

Regarding data availability, these datasets remain part of the VODAN-Africa ecosystem and continue to be accessible for ongoing research and health analytics. In line with VODAN-Africa's decentralized architecture, the data are stored within individual institutions' local repositories, where they remain under the control of the respective health facilities. These datasets can be accessed and reused for future research through secure, distributed queries, provided that the necessary ethical and privacy guidelines are adhered to.

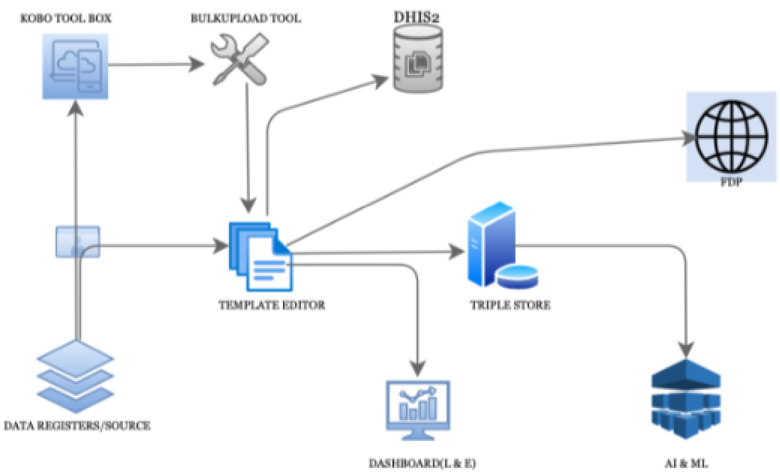

Figure 7.2: VODAN Africa internal and aggregate dashboard HIS integration scheme

Table 7.1 presents the data size produced within the health facilities, illustrating daily patient data collected from sample hospitals in Uganda.

Table 7.1: Data size produced within the health facilities showing daily patient data from sample hospitals in Uganda

S.no.	Health facilities	Country	Daily patient data from previous survey	ANC (Monthly)	OPD (Monthly)	COVID (Monthly)
1	KIU Teaching Hospital	Uganda	130	685	1672	0
2	Ark Specialist Hospital	Uganda	±50	1	0	0
3	Hoima referral Hospital	Uganda	±170	62	118	0
4	Lira Referral Hospital	Uganda	1	1	0	0
5	Ayder Referral Center	Ethiopia	20	750	900	
6	Tukur Anbesa Specialized Hospital	Ethiopia	116	116	508	NA
7	Pumwani	Kenya	±150	595	5203	6
8	Beacon of Hope	Kenya	62	54	1905	22
9	Railway clinic	Nigeria	+150	14	384	0
10	GH Lapai	Nigeria	±130	132	1442	0
11	OOU	Nigeria	50	5	1656	15
12	FULL	Nigeria	100	23	819	NA
Total Uganda			909	1838		

7.3.3 Application of VODAN-Africa and Asia Health Management Information System Architecture

VODAN Africa developed an architecture to record clinical health data, making it accessible both to humans and machines in a distributed system that is locally governed and linked. This architecture supports data analytics directly at individual health facilities by enabling data visits for analytics. It includes the CEDAR system as a one-time data input editor, outpatient instances, ANC instances, and bulk upload of datasets to the CEDAR platform. The CEDAR platform, integrated within the VODAN-Africa architecture, facilitates the organization and storage of health data. The platform works with RDF and JSON-LD structured data. From these formats, template IDs (for Outpatient, ANC, and point of care datasets) and API keys are retrieved to configure the dashboard. This allows for analytics at the point of care using the internal health facility dashboard and the VODAN-Africa aggregate dashboard for real-time statistics. Data are stored in the AllegroGraph triple store, a specialized database designed to store and retrieve data as RDF triples through the running of dynamic SPARQL queries. In addition, it supports the import and export of data in the health facility HIS, such as DHIS2, in the form of reports.

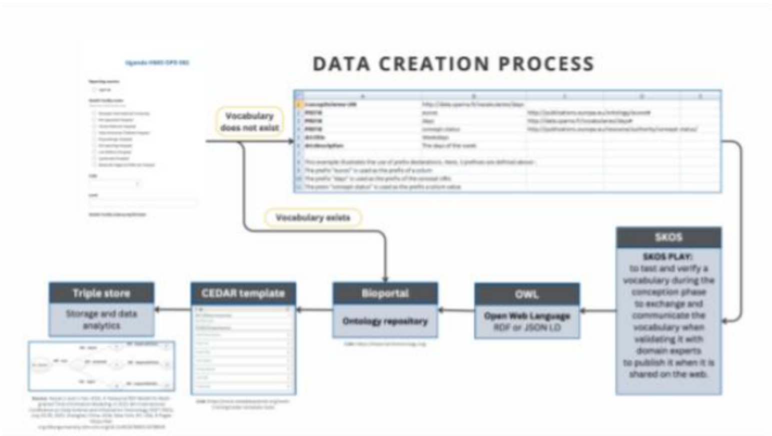

Figure 7.3: Data creation process using the VODAN Africa architecture

Why CEDAR?

The CEDAR is a flexible and agile workbench integrated with the bio portal to produce machine-readable data templates [290]. It was established in 2014 to create a computational ecosystem for the development, evaluation, use, and refinement of biomedical metadata. It was adapted to local data production and for the data capture of repositories controlled by the facility that produces the data by the VODAN-Africa technical team. Its approach is based on the use of metadata templates, which include controlled terms and semantic properties for specific data elements. The data were transformed into Resource Description Framework and JSON-LD format based on BioPortal ontologies (see Table 7.2). It also supports sharing and managing resources both locally and on the web. The CEDAR platform provides an API to support export/import and analysis of metadata, compatible with other tools, i.e., can be used for editing, creating, sharing, and accessing data.

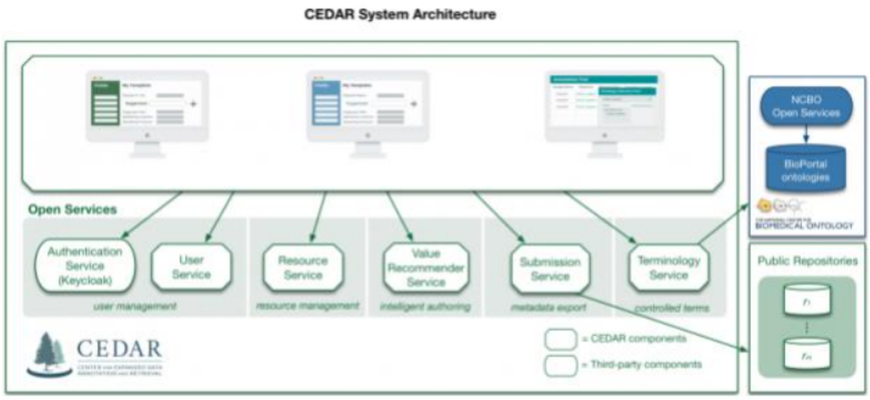

Figure 7.4: Cedar system architecture [1]

Allegro Graph Triple Store

FAIRification by design was successfully realized within the health facilities by digitizing and creating metadata templates based on the Health Management Information System paper registers and the DHIS2 system for aggregated data entry in residence. The data was generated and produced using the CEDAR system integrated into the locally deployed VODAN-Africa architecture within the different health facilities of Uganda. The CEDAR platform supports the creation of customized templates. The templates were then populated with patient data and automatically assigned unique identifiers(F1). The metadata templates are stored in JSON-LD format in MongoDB and the form of RDF triples in the Allegro graph triple store, respectively [152]. The populated templates can also be made accessible via the web under restricted measures, thereby fulfilling the Findable and Accessible aspect of the FAIR principles.

7.3.4 Standardization of the VODAN-Africa Controlled Vocabulary Terms and Semantic Properties

The VODAN-Africa controlled vocabularies and semantic properties were created using the Simple Knowledge Organization System (SKOS) framework, which is a standard for representing knowledge organization systems. These vocabularies are made available in multiple formats: CSV, RDF, and SKOS, to cater to different usage needs and tools.

The Cedar metadata templates (here) were used and examined iteratively to conform to the requirements of Resource Description Framework (RDF) SKOS conversion. Data were produced using the templates to ensure that all appropriate fields were adequately captured, as well as to ensure that the structure inherent in the templates allowed portability.

The conversion of tabular data for the controlled vocabulary terms and properties into SKOS RDF files was initially performed using the SKOS Play Convert tool [140]. However, it is important to note that fully automating the conversion process was not feasible due to the complexity of certain relationships between terms. Before conversion, several steps were taken to model the data to ensure their alignment with the controlled vocabulary. These steps involved assigning the controlled terms and semantic properties on the bioportal, such as ['mapping

[1]Source: `https://more.metadatacenter.org/tools-training/cedar-api`

disease codes to standardized ICD-10 terminology'], to ensure consistency and coherence in the semantic structure, accounting for both quantitative and qualitative associations. This process accounted for both quantitative and qualitative associations, and the conversion scripts used for term mapping can be found on (here). These scripts facilitated the automation of the mapping process for commonly used terms, though some terms required manual intervention to ensure accuracy.

Despite these efforts, the SKOS Play Convert tool was able to automate only the initial stage of the conversion. Some operational associations between terms—particularly those involving more complex or nuanced relationships—were not accurately interpreted by the tool's algorithms and required manual intervention. These issues were addressed through a process of semantic control, where expert review and refinement were necessary to ensure that the relationships between terms were correctly represented.

This limitation highlights an area for potential development in future iterations of the VODAN-Africa system. While some progress has been made in automating data conversion and semantic handling, further advancements are needed to enhance the accuracy and efficiency of the system, particularly in handling complex semantic relationships. Furthermore, flexibility became an issue due to the unfocused error handling of the tool: when input data was supplied in an incorrect format for conversion, the conversion stopped at some point, and incorrectly formatted data could lead to conversion failures. To address these challenges and improve the templates and workflows, the following enhancements were proposed:

- Enhanced Template Design: Validation rules and dropdown menus were incorporated into the templates to minimize user errors during data entry.

- Automated Validation Scripts: Conversion scripts (here) were introduced to validate data completeness and consistency before submission to the SKOS Play Convert tool.

- Hybrid Conversion Approach: A combination of automated conversion and manual review was adopted to mitigate the limitations of the SKOS Play Convert tool and ensure the semantic relationships and properties were accurately represented.

- Continuous Feedback Loop: User and stakeholders feedback was collected regularly to refine templates and address emerging requirements.

The current limitations primarily involve ensuring that non-technical users can effectively use the templates without introducing errors that complicate the conversion process. These enhancements require correcting the existing drawbacks that are related to the stability, simplicity, and applicability of the templates, providing for their wider utilization and constant repeated usage.

In addition, different versions of the VODAN-Africa vocabularies are managed using the GitHub version control system. This ensures that the latest versions of the vocabularies are available for download and use in various formats, including CSV and RDF, providing flexibility for different applications and integration scenarios.

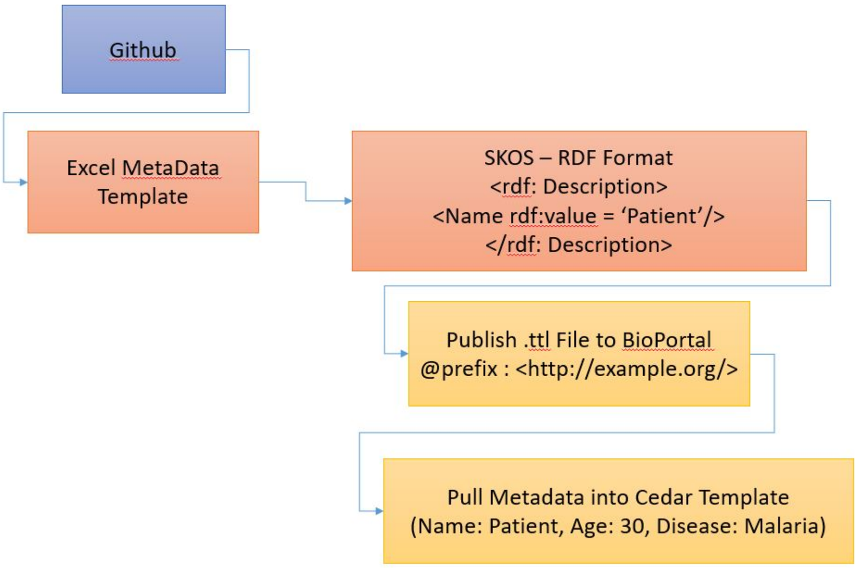

Figure 7.5: Workflow for the standardization of VODAN Africa controlled vocabularies and semantic metadata using RDF and Bioprtal

Semantic interoperability is increased through importing and mapping concepts from semantic resources such as SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms), LOINC (Logical Observation Identifier Names and Codes) and ICD (International Classification of Diseases), and reusing definitions, for example aligning patient demographics and health indicators from these templates with VODAN-Africa data models. Within the MVP of VODAN-Africa, the reuse of suitable existing vocabulary terms and properties was implemented to avoid duplication and increase interoperability within machine-actionable semantic data templates in RDF and JSON-LD linked data formats. The templates contain fields for which each is assigned a controlled vocabulary term or semantic property stored in the BioPortal system. The controlled vocabulary terms and semantic properties created are used to provide values for the metadata attributes in these templates, representing community-endorsed standards and controlled terms from the VODAN-Africa ontologies hosted in Bio Portal (see Table 7.2). For instance, the term 'Hypertension' from SNOMED CT was used to harmonize disease classification across datasets, ensuring semantic alignment with the VODAN-Africa ontology.

S.N	Description	Link
1	VODANA's General controlled vocabulary ontology and semantic properties that are used within Outpatient Department Registration and Antenatal Care Templates within the facilities.	`https://bioportal.bioontology.org/ontologies/VODANA-GENERAL`
2	VODANADISEASES is an ontology of the diseases whose data is collected within the VODANA facilities, especially.	`https://bioportal.vodana.health/ontologies/VODANADISEASES`
3	MCHVODANATERMS is an Ontology containing the controlled terms and semantic properties used in the Maternal Child Care generic template used within different facilities in VODANA	`https://bioportal.vodana.health/ontologies/MCHVODANATERMS`
4	VODANAMFLCODE is the ontology containing the health facility names along with the codes used within the VODANA community.	`https://bioportal.bioontology.org/ontologies/VODANAMFLCODE`

S/N	Description	Link
5	VODANAUGANDA is an Ontology containing the controlled terms and semantic properties used in the OPD registers from different countries used within different facilities in VODANA	`https://bioportal.bioontology.org/ontologies/VODANAUGANDA`

Table 7.2: Overview of the Ontologies implemented within VODAN-Africa framework to facilitate standardized data representation and interoperability.

The ontologies in Table 7.2 were selected because they provide controlled vocabularies and semantic properties essential to ensure data interoperability and adherence to FAIR principles in the following contexts:

- VODANA-GENERAL: This controlled vocabulary offers a foundational framework and common properties used across various VODANA templates, including the Outpatient Department Registration and Antenatal Care templates.
- VODANADISEASES: Focuses specifically on diseases, enabling uniform data collection and reporting related to disease data within VODANA facilities, especially in Uganda.
- MCHVODANATERMS: Provides terms and properties tailored for Maternal Child Health care templates, ensuring consistency across facilities.
- VODANAMFLCODE: This ontology links health facility names with their standardized codes, promoting consistency when referring to facilities within the VODANA community.
- COVIDTERMS: A specialized ontology developed for managing COVID-19-related data, which is critical for pandemic response and monitoring.

Interrelation of Ontologies

These ontologies are interconnected, with the VODANA-GENERAL ontology serving as the overarching framework that supports and standardizes the specific vocabularies within the other ontologies. For example:

- VODANADISEASES and COVIDTERMS inherit the properties of VODANA-GENERAL, but apply them to disease-specific contexts.
- MCHVODANA-TERMS integrates terms from VODANA-GENERAL while focusing on maternal and child health.

- VODANAMFLCODE ensures that the health facility codes used in the other ontologies are consistent and interoperable.

Five main source ontologies were used to constrain metadata fields in the CEDAR template, ensuring that users can enter attribute values correctly. Constraining metadata fields eliminates the possibility of entering syntactically different values or responses, such as misspellings or inconsistent formats. The four VODANA FAIR Vocabularies incorporated into the templates for HMIS OPD 002 REGISTER include VODANA-GENERAL, VODANAUGANDA, VODANA-COVID, VODANADiseases, and VODANAMFLCODE (VODANA Health Facility List).

For example, when entering patient attributes like "Health Facility," users must select from a predefined list provided by VODANAMFLCODE, ensuring uniformity (e.g., "Mulago National Referral Hospital" instead of variations like "Mulago Hosp." or "Mulago NRH"). This approach standardizes entries and supports interoperability in the datasets.

Interoperability of the VODAN-Africa Minimal Viable Product with HMIS systems

Within the VODAN-Africa minimal viable product, syntactic interoperability with point-to-point integration is increased through the support to compute and send aggregated monthly reports to hospital HMIS systems, as illustrated in Figure 7.6. The VODAN-Africa MVP is in compliance with the interoperability component of FAIR in that it guaranties that it integrates easily with other Health Management Information Systems (HMIS), including the DHIS2 system. It is done with the help of standardized data element report skeleton templates which use shared controlled vocabularies and semantic properties. These templates promote the flow of data between systems by responding to the concepts of interoperability described in the FAIR framework. The term interoperability here is understood as the capacity or capability of various systems to communicate with each other in a manner that is significant, advantageous, and without the necessity of putting much effort into it, in terms of manual labor. The MVP of VODAN-Africa has ensured that additional HMIS systems can easily understand and interpreted data through standardized data formats, such as JSON, which facilitates better data circulation and the usefulness of access to information by health systems. The aggregated indicators are then posted to DHIS2, which provides a web API to accept the data values. A separate enterprise service bus for authentication, management of interoperability services, logging, and mediation is envisioned as a microservice to ensure security and add flexibility during interoperability.

Challenges Encountered During VODAN-Africa MVP Deployment and Data Production.

The CEDAR platform was the main data entry tool used within the different VODAN-Africa health facilities. However, data acquisition posed significant challenges and massive data collection is vital for in-depth analysis. During the deployment and data production within the VODAN-Africa health facilities, several issues were identified and addressed through one-on-one interviews with data management personnel in Uganda, Kenya, Ethiopia and Nigeria. The CEDAR platform faced storage shortages and performance issues, leading to the use of alternative data collection tools. The locally installed Kobo toolkit was employed in some facilities, especially in Nigeria.

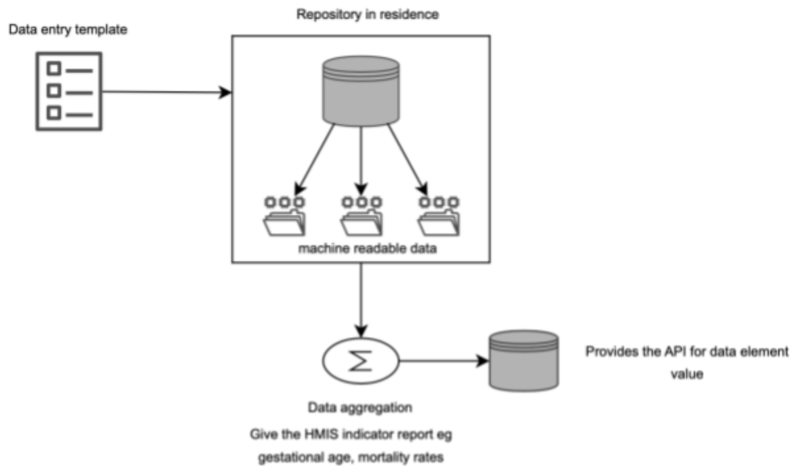

Figure 7.6: Interoperability workflow between VODAN-Afica and HMIS systems for machine-readable data aggregation and indicator reporting

Converting paper-format data into FAIR data required significant manpower. The health facilities in Uganda continued to report to the Ministry of Health (MOH) using more than 20 different paper registers, requiring the assignment of additional personnel to enter local health data. To FAIRify data collected over the past nine months, the bulk upload tool was adopted. Challenges such as internet connectivity and electricity scarcity in some regions of health facilities were mitigated by providing resources to purchase mobile data bundles for data production. Technical issues, such as resetting the admin password, were promptly resolved by the technical team.

To address the inadequate skillset between data management personnel and data stewards, comprehensive capacity-building initiatives were implemented. These included targeted training sessions and the distribution of shared materials aimed at improving competencies in data management and stewardship. Participants were trained on using the bulk upload tool, adhering to FAIR principles, and troubleshooting common issues. These initiatives not only enhanced the technical proficiency of the personnel, but also ensured the sustainability of FAIR data practices in health facilities.

7.4 Data Analytics Using the Local Health Facility Dashboard and VODAN-Africa Aggregate Dashboard

The usability of the dashboard is divided into two components:

- The aggregate dashboard of the VODAN-Africa community.
- The local dashboard (within the health facility).

Data Analytics Using the Local Health Facility Dashboard and VODAN-Africa Aggregate Dashboard

The aggregate dashboard provides a summary of participating countries and the total number of records or instances of the Outpatient 002, COVID-19, and Antenatal Care registers. On the other hand, health facilities can visualize their data through the local dashboard. For example, a health facility can view regular statistics such as the number of patients registered per day, trends in antenatal visits over a specific period, or the distribution of COVID-19 cases within their facility. These insights help health facilities monitor their operations and make data-driven decisions.
The ownership of the data is determined by where the data or metadata is stored or shared. The domain of the local dashboard is attached to the local host. The database is set up through verification, the API key address is set up, admin, standard, user, and password are set up for security, and the update configuration to completed. The local dashboard automatically fetches data from the Cedar platform. The high frequency is attributed to the collection of real-time data that will be reliable in obtaining realistic results. A provenance template on the health facility metadata (template 0) with fields: Country, region, latitude, longitude, address, facility type: government or private, email address, phone number, and facility size is populated and submitted. The VODAN-Africa community dashboard manually updates when the local dashboard is updated and hosted on a public domain [2]

No	Statistics	Aggregate Frequency	Description
1	Patient visits (real-time)	Daily and Monthly	Aggregate of the number of patient visits
2	Patient gender percentage (real-time)	Monthly	Male patient and Female patient percentage
3	Frequency of Diagnosed Diseases (real-time)	Daily and Monthly	Aggregate of the selected diseases grouped by diseases
4	COVID-19 patient visits (real-time)	Daily and Monthly	Aggregate of the number of patient visits for COVID-19
5	Patient tested for COVID-19 (real-time)	Daily and Monthly	Aggregate of the number of patients tested for COVID-19
7	Patient tested positive on COVID-19 (real-time)	Daily and Monthly	Aggregate of the number of patients tested COVID-19 positive
8	Patient death from COVID-19 (real-time)	Daily and Monthly	Aggregate of the number of patients died because of COVID-19
9	ANC Patient Age (real-time)	Monthly	Aggregate of the number of patients categorized by Age (by age in the interval)
10	ANC Patient visits (real-time)	Daily and Monthly	Aggregate of the number of patient visits from ANC form
11	ANC Syphilis test result (real-time)	Daily and Monthly	Aggregate of the number of patient tested for Syphilis categorized by result from ANC form
12	ANC HIV test result (real-time)	Daily and Monthly	Aggregate of the number of patient tested for HIV categorized by result from ANC form
13	Gestational Age (real-time)	Monthly	Aggregate of the number of patient categorized by gestational age in weeks (by weeks in interval) from ANC form

Figure 7.7: Standardized fields selected for the key statistics for the VODAN-Africa local and community dashboards

7.4.1 Key statistics Data Sources for the VODAN-Africa Local and Community Dashboards

[2]Source: https://dashboard.vodana.health/

Template name	Attributes and terms selected
Template 0	Country, region, latitude, longitude, address, status: government or private, facility type: email address, phone number, number of patients, facility size :S, M, L
Outpatient register	Health facility name, attendance frequency, date of visit, diagnosis frequency (daily and monthly), sex (monthly gender composition)
Antenatal childcare register	Serial number, age (age groups: 18-25, 26-35, 36+), attendance frequency, syphilis & Hep Test Results with syphilis, gestational age (1-10, 11-20, 21-30, 31+ weeks)
COVID-19	Health facility name, count of number of visits, date of visit, patients tested for COVID (daily and monthly), mortality rate/death rates (daily and monthly)

Table 7.3: Attributes and terms selected for different templates

Use of FAIR data at the point of care Health Data Analysis was done with the data retrieved from the health facility AllegroGraph database using dynamic SPARQL queries. This included analysing the outcomes of grouping the Ugandan Outpatient Registry Data by aggregation, patient demographic analysis of the Ugandan Outpatient Registry Data, examining the temporal and geographical trends in Uganda, as well as comparing visualisations using FAIR Data.

7.4.2 Data Analytics on the AllegroGraph Repositories

Grouping the Ugandan Outpatient Registry Data by Aggregation

The SPARQL query below is structured to retrieve the number of diagnoses that are recorded under each health facility from the data in the OPDUganda repository. By designing the query to count occurrences of each diagnosis per facility, a clear summary of how often different diagnoses are documented across various health facilities is obtained. The results are ordered in descending order by the count of each diagnosis to highlight the most frequently occurring diagnoses first.

SPARQL Query 1

```
SELECT
?healthFacilityName ?diagnosis (COUNT(?diagnosis) AS ?diagnosisCount)
WHERE {
  ?s <http://dashboard.vodana.health/Diagnosis> ?diagnosis.
  ?s <http://dashboard.vodana.health/HealthFacilityName> ?healthFacilityName.
}
GROUP BY ?healthFacilityName ?diagnosis
ORDER BY DESC(?diagnosisCount)
```

healthFacilityName	diagnosis	diagnosisCount
"KIU teaching Hospital"	"UTI"	"35"
"KIU teaching Hospital"	"Spondylolysis lumbar region"	"25"
"KIU teaching Hospital"	"Peptic ulcer disease"	"23"
"KIU teaching Hospital"	"Dental carries"	"22"
"KIU teaching Hospital"	"Peripheral neuropathy"	"21"
"KIU teaching Hospital"	"Gastritis"	"19"
"KIU teaching Hospital"	"Fibromyalgia"	"18"
"KIU teaching Hospital"	"Irreversible pulpitis"	"18"
"KIU teaching Hospital"	"Spondylolysis"	"17"
"KIU teaching Hospital"	"Malaria"	"16"
"KIU teaching Hospital"	"Dental abscess"	"15"
"KIU teaching Hospital"	"Dog bite"	"14"
"KIU teaching Hospital"	"Acute nasopharyngitis"	"14"
"Hoima Referral Hospital"	"MALARIA"	"14"

Figure 7.8: Output produced from running the SPARQL query to retrieve the number of patients per diagnosis in each health facility from the OPDUganda Repository

The results from running this query can be analyzed to understand how the occurrence of various diagnoses is distributed across the various Ugandan health facilities in the dataset. This information can be used by stakeholders in healthcare management to efficiently allocate resources and implement targeted interventions to hospitals that have a significantly large number of cases compared to the resources that are available for managing the diseases. Consequently, this would potentially improve the effectiveness of disease management in these institutions.

Patient Demographic Analysis in the Outpatient Registry Uganda Data

Patient Demographics Aggregation

The results from running the SPARQL query below can be used to analyze the information on the demographics of patients with certain diagnoses. The query retrieves the number of diagnoses from the OPDUganda repository. The diagnoses are then grouped by age groups of under 18 years, 18 to 25 years, 26 to 35 years, 36 to 45 years, 46 to 55 years, and over 55 years. The diagnoses are also grouped by gender. Finally, the query orders the results by the number of diagnoses in each group in descending order. The pattern of the query is that the subject is a diagnosis that was made, and the query retrieves the age and gender attributes of the patient that are associated with each diagnosis. The findings from analyzing the data on patient demographics can then be used in strategic planning for implementing interventions and allocating resources to effectively manage the spread and fatality of the diseases. By understanding which age groups and genders are most affected by particular diagnoses, healthcare providers and policymakers can make decisions that are better informed to improve the outcomes in public health in Uganda.

SPARQL Query 2

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

SELECT ?diagnosis ?ageGroup ?gender (COUNT(?diagnosis) AS ?numDiagnoses)
WHERE {
  ?record vodan:Age ?age.
  ?record vodan:Sex ?gender .
  ?record vodan:Diagnosis ?diagnosis .
  BIND (
    IF(xsd:integer(?age) < 18, "< 18",
      IF(xsd:integer(?age) >= 18 && xsd:integer(?age) <= 25, "18-25",
        IF(xsd:integer(?age) >= 26 && xsd:integer(?age) <= 35, "26-35",
          IF(xsd:integer(?age) >= 36 && xsd:integer(?age) <= 45, "36-45",
            IF(xsd:integer(?age) >= 46 && xsd:integer(?age) <= 55, "46-55",
              IF(xsd:integer(?age) > 55, "> 55", "Unknown"))))))
    AS ?ageGroup)
}
GROUP BY ?diagnosis ?ageGroup ?gender
ORDER BY DESC(?numDiagnoses)
```

diagnosis	ageGroup	gender	numDiagnoses
"UTI"	"18-25"	"Female"	"9"
"Gastritis"	"< 18"	"Female"	"8"
"Peripheral neuropathy"	"> 55"	"Female"	"7"
"PUD"	"18-25"	"Female"	"7"
"Spondylolysis lumbar region"	"> 55"	"Female"	"7"
"Late syphilis"	"> 55"	"Female"	"6"
"UTI"	"46-55"	"Female"	"6"
"Dental carries"	"18-25"	"Female"	"6"
"Hypertension"	"> 55"	"Female"	"6"
"UTI"	"> 55"	"Male"	"6"
"Common cold"	"< 18"	"Female"	"6"
"Malaria"	"< 18"	"Female"	"5"
"Acute nasopharyngitis"	"< 18"	"Female"	"5"
"STI"	"18-25"	"Female"	"5"

Figure 7.9: Output produced from running the SPARQL query to retrieve and count the number of specific diagnoses grouped by age groups and gender from the OPDUganda Repository

Patient Demographic Analysis in the Outpatient Registry Uganda Data

Various visualizations were created to analyze the patterns in the OPD Uganda patient data. The age distribution of the top 10 highly diagnosed diseases and infections was plotted using bar graphs. Age groups from the OPD Uganda dataset were in intervals of 4 years between

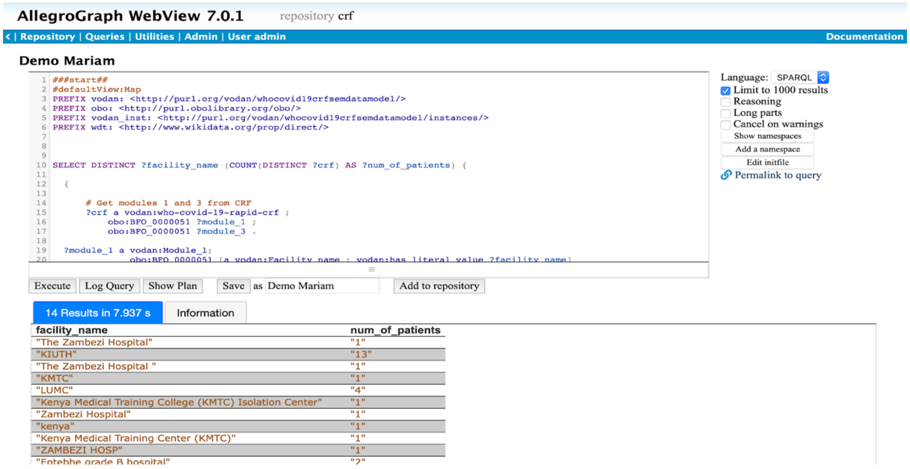

Figure 7.10: Output produced from running the SPARQL query to retrieve and count the number of patients from the OPDUganda Repository in each hospital

0 and 80+ years. These visualizations can be used to diagnose diseases associated with age, to determine the appropriateness of health services, and to distribute these services between certain ages [214]. In the same manner, bar graphs were employed to indicate the distribution of the number of the ten most frequently diagnosed diseases and infections by sex. It is useful for studying gender-specific health concerns, for developing specific health initiatives, and for assessing potential discrepancies in disease rates or in access to medical services between male and female populations [214]. A word cloud image was generated to visually represent the diagnoses and prescriptions given. This visualization technique allows for quick identification of the most frequent diagnoses and prescriptions by displaying them in larger font sizes. It provides an intuitive and engaging way to summarize the prevalence of various health conditions in the dataset, helping to identify key areas of concern at a glance [229]. A histogram was created to represent the age distribution of the entire group of data, the mean age, and the standard deviation of the sample. This provides an overall picture of the age distribution of the population and allows the determination of age groups of the population that are diagnosed most frequently to plan the availability of medical services and healthcare for the prevailing age groups [215].

Data Acquisition

Data was retrieved from the AllegroGraph repository named OPDUganda using SPARQL queries. The following query was run on the OPD Uganda repository to retrieve the age, sex, diagnosis, prescription, district of residence, blood pressure, and blood sugar levels from the OPD Uganda repository.

SPARQL Query 3

```
PREFIX opd: <http://dashboard.vodana.health/>
SELECT ?s ?age ?sex ?diagnosis ?prescription ?district
?blood_pressure ?blood_sugar
WHERE {
  ?s opd:Age ?age ;
     opd:Sex ?sex ;
     opd:Diagnosis ?diagnosis ;
     opd:Prescription ?prescription ;
     opd:District ?district ;
     opd:BloodPressureMmhg ?blood_pressure ;
     opd:BloodSugarMmL ?blood_sugar .
}
```

General Diagnoses and prescriptions from Uganda Outpatient data

Figure 7.11 illustrates the diagnoses and shows their frequency using the font size. Figure 7.12 shows the top prescriptions given using the font size.

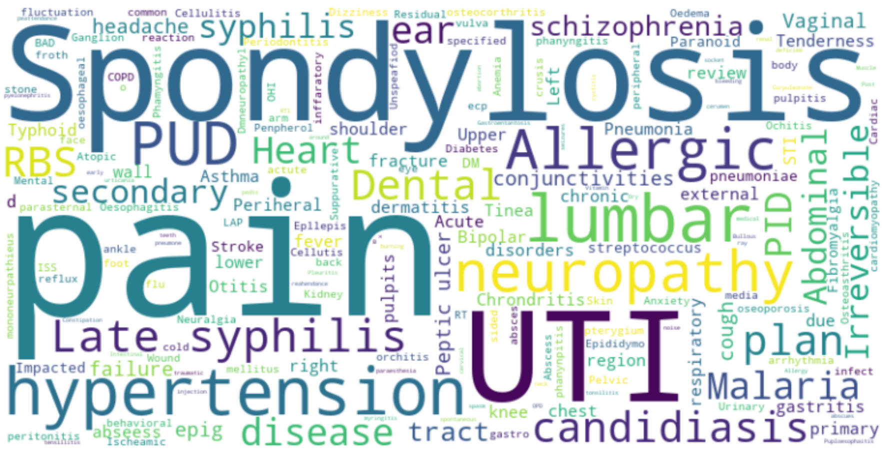

Figure 7.11: Word cloud visualization of the most frequently reported diagnoses in Uganda outpatient dataset, where word size corresponds to the relative frequency of each diagnosis

7.5 Findings: Comparison of DHIS2 and FAIR Data for Privacy-Preserving Health Analytics of Patient Data in Uganda

In this section, we address the DHIS2's role in transforming Uganda's health Information systems, FAIR's role in transforming Uganda's Health Information systems, the strengths and limitations of DHIS2 and FAIR data principles.

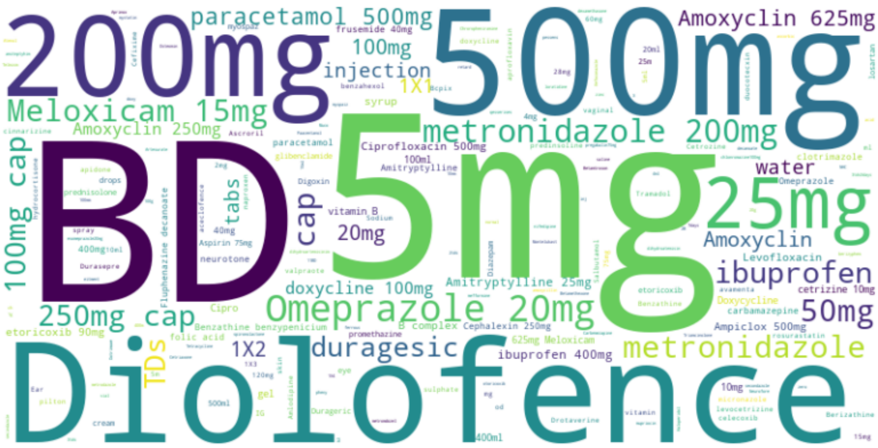

Figure 7.12: Word cloud visualization of the most frequently reported prescriptions in Uganda outpatient dataset, where word size corresponds to the relative frequency of each prescription

7.5.1 DHIS2's Role in Transforming Uganda's Health Information Systems

By implementing DHIS (District Health Information Software 2), Uganda has brought revolution in its health information management system via strong electronic support. DHIS2 has incorporated itself into healthcare practices, which is fundamental to disease immunization, surveillance, and maternal health programs within the country [283].

DHIS2 was introduced and has helped improve disease immunization, surveillance, and maternal health programs. DHIS2 has also made it possible to improve monitoring of vaccination rates, due to its real-time entry of data and reporting, which has resulted in more specific and focused immunization initiatives and used resources more efficiently. This has been exceptionally critical towards the containment of preventable diseases, particularly in the rural population.

In addition, DHIS2 has been instrumental in health surveillance programs in Uganda. Due to the union of data carried out by healthcare facilities nationwide, DHIS2 enables tracking disease outbreaks, enabling respondents in the field of public health to respond fast to emerging threats. It has also been useful for tracking infectious diseases such as malaria, cholera, and Ebola, which can be detected and contained in less time.

Through DHIS2, health care providers have enjoyed real-time access to vital information on pregnancy, childbirth, and post-delivery care in the maternal health sector. This has improved maternal services, as the required medical services are done at the expected time and pregnant women are well monitored during their pregnancy periods. The capabilities of the software to monitor the data related to antenatal visits, delivery results, and complications have helped to reduce the number of maternal and infant mortality rates.

In addition, the user-friendly nature of DHIS2 and its ability to integrate with any other health system have made its explosion to be employed by health workers of different levels. This has helped to create a strong healthcare data ecosystem in Uganda, which has ensured the sustainability of improving the delivery of data-driven health services.

Together, the functions of DHIS2 in Uganda's health information system serve as a good example of the benefits of digital technologies in healthcare by facilitating a change in workflow, efficient use of funds, and ultimately the health outcomes of low-resource countries.

Privacy Mechanism in DHIS2

DHIS2 incorporates several privacy mechanisms designed to protect patient data and ensure system security [264]. One of the main mechanisms is Role-based Access Control (RBAC), where users are granted access only according to their assigned roles in the healthcare organization. This control mechanism assists in discouraging unauthorized data access in the system. However, there is a problem with using RBAC, more specifically in the case of access control, where the controls are not sophisticated enough. This becomes problematic in settings with a diverse range of users, as the system may not fully account for the varying levels of access required by different roles within the healthcare environment [274].

Besides RBAC, DHIS2 uses data encryption to ensure the data is secure while transferring between the devices and the server to meet the organization's data confidentiality and integrity standards. This will reduce interceptions because data is encrypted while passing through a given network. However, there is still an important issue with the encryption scheme, more specifically, because the current system does not offer full end-to-end encryption. This constraint poses the data at risk both in the storage and processing phases in case of vulnerability in the system along the data journey [306].

In addition, the system has a feature of a user activity log, which records the activities of the users on the DHIS2 system. This monitoring assists in keeping a check on who is doing what, and who is up to which kind of something, to hold an individual or a party answerable, and to look out for signs of unauthorized activities. Since the user activities can be logged in DHIS2, the system can reveal the usage patterns and identify any abnormal behavior [50]. However, the system is limited in its technical capacity to analyze user logs adequately. This constraint hinders the possibility of identifying security breaches and responding to them promptly, with the result that the system is vulnerable to security threats. These challenges call for an even more robust strengthening of the DHIS2 privacy and security features in an environment that demands the highest degrees of health data confidentiality.

Case Study: DHIS2 and Immunization Campaigns in Uganda

In Uganda, the use of DHIS2 greatly helped a national polio immunization campaign because it provided real-time data tracking. Health workers in facilities across the country, including remote areas, used mobile devices to enter immunization data while the data were aggregated at the district and national levels. This was made possible because health officials were able to identify districts with low immunization rates and post additional hands, leading to a 92percent drop in polio cases [256].

However, problems were encountered, for example, weak Internet connection, especially in rural areas, resulting in the inability to complete real-time reporting and privacy issues. In identifying privacy issues, basic privacy measures such as encryption and access control were implemented in DHIS2, but end-to-end encryption and user authentication issues highlighted the need for stronger security protocols. Solving these connectivity and privacy problems is important to enhance the functional performance of DHIS2 and protect health data in future future immunization campaigns [256].

Challenges in DHIS2 Implementation in Uganda

Despite the widespread adoption of DHIS2 in Uganda's healthcare system, its implementation faces several challenges that hinder its full potential. These challenges span technical, human resource, and privacy/security aspects, all of which significantly impact the effectiveness of the system, particularly in rural and remote areas.

Technological Challenges
One of the limitations observed in the course of the implementation of DHIS2 in Uganda is the connectivity problem, characteristically marked by intermittent poor internet, especially in rural areas, which results in delayed data submission and synchronization. This impacts the efficiency and accuracy of the flow of health information. However, poor ICT infrastructure leads to slow processing and data errors. These challenges devalue the effectiveness of DHIS2 in generating real-time data for analysis and decision-making as well as a tool for healthcare enhancements [305].

Human Resource Constraints
Some of the limitations include Human resources and limited-trained staff to operationalize DHIS2, especially in rural areas of Uganda. The targeted groups of people are usually trained at the beginning of the project and do not receive consistent support later, which means they cannot make the most of the system's tools, incorporating the capabilities of data analysis and report generation of reports. Further, high turnover of staff amplifies these difficulties, alongside interrupting data acquisition and integrating recurrent training [309]. It results in poor use of DHIS2 and poor standardization in the reporting of such data, bringing a lack of harmony in policies formulated and criteria set, in decision-making, and the effectiveness of the system.

Privacy and Security Concerns
While DHIS2 includes privacy measures like encryption and role-based access control (RBAC), concerns remain about data security. The lack of robust end-to-end encryption leaves patient data vulnerable during storage and processing. Weak implementation of RBAC, with insufficient granularity, risks unauthorized access to sensitive information. Additionally, limited customization options hinder the compliance with the Uganda Data Protection and Privacy Act (2019), making it difficult to tailor the system to local healthcare and legal requirements, thus increasing the risks of non-compliance and potential legal liabilities [294][277].

7.5.2 FAIR's Role in Transforming Uganda's Health Information Systems

Global Relevance of FAIR Data Principles in Uganda's Context

The FAIR data (Findable, Accessible, Interoperable, and Reusable) are global principles guiding the management and sharing of data. Optimize data value through better search, availability, utilization, and recurrence, which is vital for resource-constrained health systems such as those in Uganda. The application of these principles improves access, improves the decision-making process, and possibly improves interdisciplinary collaboration nationally and internationally [145]. However, for their applications in Uganda, a lot of effort must be put into making it understandable within the context of Ugandan health, laws, and technology.

Findable: Facilitating Access to Health Data

The "Findable" part of the FAIR concept focuses on the set of rules according to which the data source must be discoverable just by using the proper metadata and structured indexes. Currently in Uganda, there are numerous scattered data sources and no consolidated database/website of health information. Its use would set up a single database in which all the report's different components could be traced [272]. For instance, DHIS2 has already assisted in exercising the centralization of data in health; the implementation of FAIR will likewise extend extensive metadata intake into the data-indexing process, supporting the decisions of involved health practitioners, policymakers, and international organizations [276].

Accessible: Breaking Down Barriers to Health Data Access

The "Accessible" principle allows data to be organized and to be accessed by identified and approved users only [290]. The flow of health data in Uganda is still faulty, meaning that different parties have unreliable means of getting the necessary information. Implementing FAIR principles would improve data availability for health officials, researchers, and policymakers. For instance, though Uganda's Ministry of Health utilizes DHIS2, data sharing is often limited. Adopting FAIR principles would standardize data sharing, making it accessible for researchers, NGOs, and international organizations. It would foster cooperation and raise productivity in responding to health crises such as the coronavirus pandemic [258].

Interoperable: Enhancing Collaboration and Integration of Health Data Systems

The principle "Interoperable" means that data must be formatted to fit and integrate with other systems. In Uganda, data integration problems arise from format, platform, and standards. These barriers would be solved in case Uganda embraced FAIR principles so that its data in the health sector would correlate with international standards, as well as be published to international organizations. This would give better integration with other systems such as the WHO Global Health Observatory, adding value to global disease surveillance and research. The interoperability would also enhance the movement of information between the health departments in Uganda, the various Governments, and other stakeholders in the health sector to boost healthcare coordination.

Reusable: Promoting Long-Term Value from Health Data

The "Reusable" principle makes it possible for data to be useful for multiple purposes, for example, for research, policy, and program purposes [290]. Most healthcare data in Uganda is still fragmented, and the information can only be used for reporting without extension. With FAIR principles in mind, data would be well described with comprehensible metadata and stored as usable formats for reuse. For instance, data collected in immunization campaigns could be used for vaccination analysis or monitoring of health status, to name but a few. This would improve data value; provide research-based decisions and popularity of Uganda in international collaboration and funding, of global health research [289].

Case Study Analysis: Implementation of FAIR Data Principles in Uganda through Open Science Framework (OSF)

The implementation of FAIR data principles has had a significant influence on the management of global health data and learning from such practices is valuable to Uganda's health data frameworks. An example of this is the Open Science Framework (OSF) [311], the core objective of which is related to the management and sharing of materials according to the FAIR principles. Such information is useful to understand how Uganda needs to apply these principles, specifically in connection with its current health data platforms comprising DHIS2.

The OSF is an all-encompassing platform that handles the life cycle of research work from data collection to sharing of data while at the same time keeping the data secure and making it readily available [286]. By incorporating the use of sophisticated encryption techniques and adopting RBAC methodology, OSF guarantees that research data of the sensitive kind can be shared safely and requires access control by the roles of users. This feature of OSF is especially valuable for Uganda because privacy considerations are important for handling often-sensitive data related to patients and public health in the country. Existing systems such as DHIS2, while indispensable in health data management, can present significant difficulties in guaranteeing strong privacy safeguards, including poor encryption mechanisms and user identification. Currently, OSF's strategy of encryption, alongside confined access as a method of protecting data, can be used by Uganda to enhance the security and privacy of health information [279].

Additionally, the scalability of the OSF is another key takeaway for Uganda. The OSF is capable of managing big databases in several research disciplines [286]. This scalability makes the platform suitable for managing the rising volumes of information that emerge in global health research, particularly in outbreaks or emergencies. The challenges Uganda has been experiencing when it comes to data aggregation, reporting, and analysis can learn from OSF architecture so that it can find a way of handling the big data that is produced by Uganda's health system. If the Ugandan government widens the use of DHS2 or adjusts it to other local health information systems through integration of other scalable data management solutions, then undoubtedly health analytics of Uganda could be enhanced for better and faster comprehensive illness surveillance, enabling more effective disease monitoring and faster decision-making [276].

The FAIR principles' application to the management of health data for Uganda could help increase openness, strengthen multipronged engagements between local and global health agencies, and generally boost health system outcomes [312]. So, following the example of the OSF and the lessons learned on privacy-preserving data sharing and scalability, Uganda could start the path toward improving data management to protect sensitive health data, but also make it available to conduct important biomedical and operational research as well as inform decision-making [293]. Although these FAIR principles can be applied in any country's health information systems, their integration into Uganda's systems can go a long way in improving health analytics, better data-informed policies, and a stronger health system in general.

Privacy Features of FAIR Data Principles

The principles of FAIR data are intended to facilitate the management and utilization of research findings and data in any field of science. Of all the principles, they have significant consequences in the aspect of privacy, especially within the healthcare setting, which involves information from patient records [259]. Among the principal concerns of the FAIR principles, privacy is equally important, and here it is accentuated by many methods that can provide data privacy and increase the general benefits derived from data at the same time. Such privacy measures include anonymization, consent management, and proper execution of data-sharing techniques and procedures [279].

Data Anonymization

Anonymization is among the most important privacy operations recommended by the FAIR principles. In health data management, particularly in developing countries like Uganda, handling and sharing of patient data requires anonymization to avoid revealing people's identities. Anonymization is the process of excluding or masking nominated personally identifiable information (PII) from data sets so that even if unauthorized parties access the data, the data cannot be used to identify specific individuals or business entities [280]. For patient anonymity, this follows both global best practices and the Ugandan Data Protection and Privacy Act enacted in 2019.
However, compared to Health Information Systems (HIS) such as DHIS2, which have comparatively few integrated anonymization options, the FAIR principles aim to generalize the anonymization of data [266]. The authors do not recommend particular software methods but stress that anonymization should be a component of any data management strategy. This position is rather beneficial, especially for developing countries such as Uganda, since different health systems that may require the implementation of anonymization may do it differently based on the nation's needs and its technological capabilities.

Consent Management

Consent management is another privacy aspect that the FAIR principles encourage. Consent involves actions of people (including patients) that authorize access to their data, use, and sharing. The principles of FAIR support the proper consent procedures in a way in which people's data are to be used, explained, and consent is sought before using those data. This is especially true in health facilities where access, storage, transmission, and use of data are accompanied by the highest levels of security [270].
Like most systems, DHIS2 is implemented within a certain architecture that can integrate or lack the broad option of consent management solutions, but the guidelines of FAIR data stress the need for consent to be included in the process of data sharing. This means that each time data is shared, especially for research purposes, consent must be verified and documented [268]. The framework provides versatility that facilitates the incorporation of multiple consent man-

agement solutions, which are legal in global standards, in addition to Uganda's data protection laws [307].

Secure Data Sharing Protocols

The FAIR principles focus on how data is shared, whether between data repositories, platforms, or between different organizations, to reduce the ability of unauthorized persons to gain access to the data. Data sharing poses a very sensitive issue in healthcare systems, as the violation of the Patient' Information Act has legal consequences, and ethical and financial implications are involved. To communicate data safely, the FAIR principles recommend the utilization of encryption and access protocols and data integrity verification processes [288].
Unlike DHIS2, which uses a set of security measures (such as role-based access controls and encryption), the FAIR principles do not impose the necessity of any particular software solution for the secure sharing of data. Researchers have emphasized the right systems to ensure the best practices in data security that can be developed in different ways according to the technology base within respective regions. In this regard, for Uganda, this flexibility provides an opportunity for local systems to incorporate data preservation methods that can be shared securely in conjunction with the FAIR principles that are compatible with Ugandan challenges such as weak internet connection, resource constraints, and infrastructure constraints [145].

Adaptability and Flexibility of the FAIR Framework

One of the most compelling aspects of the FAIR principles, especially in comparison to systems like DHIS2, is their adaptability. The FAIR principles do not bring recommendations on the software or tools that should be used for privacy [255]. They do this without prescribing the specific contingency, which enables healthcare providers and countries to implement privacy practices that exist locally and yet meet international benchmarks.
Such flexibility is desirable, more so in developing countries such as Uganda, where resources are relatively scarce. Although DHIS2 is a recognized platform in Uganda's health sector; though, it may lack the privacy required by the FAIR principles. On the other hand, with the help of the FAIR framework, Uganda can choose privacy solutions tailored to its conditions: from the deployment of other types of encryption to the more liberal concepts of consent and improved data masking methods. This flexibility enables Uganda to maintain patient confidentiality while keeping data open for use in public health [294].

Table 7.4: Comparison of DHIS2 and FAIR Data for Privacy-Preserving Health Analytics in Uganda.

Aspect	DHIS2	FAIR Data
Strengths	• User-friendly interface enabling widespread adoption in many health facilities. • Provides key performance metrics such as data completeness and reporting timeliness. • Facilitates evidence-based decision-making, even in rural areas with limited technical expertise.	• Superior interoperability for seamless data sharing across platforms and domains. • Promotes data reusability for research, policy development, and public health interventions. • Enhances global and regional collaboration through standardized data-sharing protocols.
Limitations	• Limited flexibility in accommodating diverse privacy protocols. • Weak role-based access control (RBAC) and insufficient encryption mechanisms expose data to breaches. • Rigid adaptation to dynamic local contexts and emerging privacy concerns.	• Requires substantial technical expertise, infrastructure, and resources for effective implementation. • Demands advanced tools for secure data sharing, consent management, and data governance. • Significant investments are needed for training, infrastructure upgrades, and alignment with international data standards.

Continued on next page

Table 7.4 continued from previous page

<table>
<tr><th>Aspect</th><th>DHIS2</th><th>FAIR Data</th></tr>
<tr><td>Implications for Uganda</td><td>Operational Impact
Effective in operational decision-making but constrained by privacy vulnerabilities and rigid structure.
Integration Potential
Integrating FAIR principles with DHIS2 could address privacy shortcomings, including weak encryption and access controls.
Challenges
Uganda faces resource constraints, outdated infrastructure, and a shortage of skilled personnel required for integration.
Recommendations
Pilot projects in selected regions could evaluate the feasibility of integrating FAIR principles into DHIS2 for applications such as immunization tracking.</td><td>Operational Impact
Highly beneficial for data sharing and reuse but challenging to implement in resource-constrained settings such as Uganda.
Integration Potential
FAIR-compliant systems could facilitate improved collaboration within Uganda and with international health organizations, strengthening health analytics and decision-making.
Challenges
Implementation requires substantial investment in capacity building, regulatory frameworks, and international partnerships for technical and financial support.
Recommendations
Collaboration with international partners and donor agencies could provide the expertise and funding needed to overcome implementation barriers.</td></tr>
</table>

7.6 Conclusion

This study demonstrated the critical need to convert Health Management Information System (HMIS) data from paper-based formats into machine-actionable, interoperable formats to enhance data reuse, accessibility, and real-time analytics. Through the implementation of FAIR (Findable, Accessible, Interoperable and Reusable) principles, data from various health facilities in Uganda—and across eight VODAN-Africa countries—was digitized using tools like the

DHIS2 and FAIRification platforms. In the FAIRification process, the data were transformed into standardized knowledge representations using semantic web technologies (e.g., RDF) and integrated into semantic models supported by existing and custom ontologies. These semantic models enhanced data interoperability and contextual understanding, allowing for more meaningful analytics.

The FAIR data terms were assigned persistent and resolvable web identifiers, enabling human and machine readability. The controlled vocabularies—originally legacy lists in printed HMIS registers—were published and mapped via Bio-portal using the LOOM algorithm, aligning them with international standards and enhancing their findability and reuse.

In contrast, while DHIS2 enables standardized data entry and has improved data comprehensiveness and decision-making within Uganda's healthcare system, it still suffers from limitations. Its reliance on aggregated data restricts granularity, limits the discovery of deeper insights (as evidenced in the less detailed heat maps), and poses challenges related to privacy, especially in rural and resource-constrained areas.

The deployment of FAIR-compliant tools within the VODAN-Africa infrastructure, including knowledge graphs and localized CEDAR platforms, enabled federated querying and dynamic visualization through dashboards. These tools illustrated the capability of health facilities to view real-time statistics (e.g., Outpatient 002, COVID-19, maternal care) while maintaining patient privacy and data security. This hybrid approach—integrating FAIR principles with DHIS2—presents a viable and promising pathway for Uganda to achieve more robust, privacy-preserving health data analytics that comply with the Data Protection and Privacy Act, 2019.

However, scaling this approach requires overcoming substantial barriers, including limited technical capacity, inadequate infrastructure, and weak regulatory enforcement. Further work is needed to explore advanced uses of FAIR data, such as machine learning for predictive modeling or disease co-occurrence analysis, which could further unlock the value of semantic, linked health data in Uganda and similar contexts. Cost and programmatic effectiveness plans to scale up the health information systems will be helpful for the successful development of the Ugandan Health Information System. It has also been observed that FAIR Data are granular compared to non-FAIR data (see Appendix 9.6). In conclusion, the integration of FAIR principles into DHIS2 represents a significant opportunity for Uganda to enhance its healthcare data ecosystem. By addressing existing challenges and leveraging international collaborations, Uganda can establish a robust, privacy-preserving, and interoperable health information system that sets a benchmark for other developing nations.