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

Basajja, M.

Appendix B

Summary

This thesis is about making health data in digital form accessible for Health Systems in Africa, and in particular in Uganda. In order for data to be computer readable and reusable, the data should be Findable, Accessible, Interoperable and Reusable (FAIR). This is commonly referred to as the FAIR-principles. This thesis presents the design and implementation of the FAIR principles to digital health data in Africa, with a specific focus on Uganda. Health data, when in digital form, remain reusable and beneficial to others. In Uganda and other African countries, health data are collected and stored in paper format using Health Management registers. For the Uganda case, the aggregated data are routinely submitted to the DHIS2 system as required by the Uganda Ministry of Health. Data are not easily accessible and reusable by researchers and health professionals at the point of care.

The Virus Outbreak Data Network (VODAN) has been a crucial part of the research presented in this thesis and as such our research became part of the VODAN-Africa and Asia implementation network in March 2020. This network which aims to address current and immediate challenges in the use and connectivity of digital health data. The VODAN-Africa and Asia initiative is central to an expanding network of equal partnerships between African and European universities. These partnerships produce or utilize high-quality, high-volume, machine-actionable data, bringing African health insights into the global medical knowledge base to improve health and healthcare in Africa. The VODAN Africa and Asia architecture is a set of integrated tools with triple-store databases where datasets are stored and made available for discovery, access, and reuse under clearly defined terms.

Our study is driven by the need to bridge the gaps in digital health systems in Africa by exploring the transformative potential of the principles of FAIR data. The purpose of FAIR data support design is to overcome prevalent concerns about the interoperability and reusability of digital health data by researchers and different stakeholders. This is achieved by implementing remote querying facilities in Uganda and other member countries of VODAN-Africa, ensuring that data remain within health facilities while enhancing the quality of care. Understanding the need for FAIR data requires recognizing its potential to transform the current data landscape. Adherence to FAIR principles ensures that data become more findable, accessible, interoperable, and reusable. This thesis is structured into nine chapters, beginning with an introduction in Chapter 1, which outlines the research problem, the objectives, and the significance of the principles of FAIR data. The general problem we are addressing is 'How can data integration, supported by the FAIR data principles, address the challenges of data interoperability and reuse

within the African digital health sector?' In Chapter 2, we review the VODAN Africa and Asia initiative and the global implementation of FAIR principles. This chapter provides a global perspective on the transformative role of the FAIR principles in scientific data management and stewardship with a specific focus on Africa. This highlights the persistent challenges within African healthcare systems, as illustrated, underscoring the urgent need for inclusive and equitable FAIR practices in diverse scientific domains. We examine how the VODAN initiative addresses key challenges, including barriers with respect to collaboration, limitations in data reuse, sustainable development, and the co-creation of knowledge. According to the multiple stream framework of Kingdon, the corresponding theory of agenda setting is applied in the African context to explore the adoption of FAIR Data, highlighting the crucial role of policy actors in facilitating this transition.

In Chapter 3, we focus on investigating the data management practices and information flow in Ugandan health facilities through focus group discussions and interviews. We discuss the current challenges of health data management in Uganda, the limitations of paper-based systems, and the potential role of the FAIR architecture in improving data accessibility and reuse.

In Chapter 4, we assess Uganda's regulatory framework for digital health and examine the feasibility of the adoption of FAIR principles. The critical challenges paper-based data systems pose in Ugandan health facilities are highlighted, especially amid health crises such as the COVID-19 pandemic. We argue that aligning Uganda's digital health strategy with the FAIR principles could significantly improve data governance, interoperability, and reusability. To assure a good alignment, between the FAIR standards and Uganda's health data regulations, we specifically address current gaps.

In Chapter 5, we explore how the FAIR principles can enhance interoperability in Uganda's digital health initiatives. We assess the interoperability of two crucial digital health dashboards: the Digital Health Atlas Uganda (DHA-U) and the Uganda Digital Health Dashboard (UDHD), along with their associated vertical digital health tools within the Ugandan context. The initial phase involved assessing the level of FAIRness of DHA-U, the UDHD, and the vertical digital health solutions within these hubs using the FAIR Evaluation Services tool; an automated web-based service. Using a FAIR Data point, the level of FAIRness of the digital health hubs was enhanced. This facilitated integration and alignment with FAIR principles. Subsequently, the level of FAIRness was reassessed after the FAIRification process, using FAIR Evaluation Services.

In Chapter 6 we introduce the VODAN-in-a-Box (ViaB) system and demonstrate a proof-of-concept using the ViaB system for COVID-19 data transformation and access. To make this possible FAIR Data Points are used. These are distributed data repositories that host machine-actionable data and metadata that adhere to the FAIR Guidelines. In participating institutions, the FAIR Data Point is deployed through the ViaB that comprises a dockerised bundle of tools. ViaB is a set of multiple open-source and FAIR-enabling services with the goal to support the gathering of World Health Organization (WHO) electronic case report forms (eCRFs) as FAIR data in a machine-actionable way, but without exposing or transferring the data outside the facility. Following the execution of a proof-of-concept, ViaB was deployed in Uganda and at Leiden University.

In Uganda and other African counties, the DHIS2 platform is commonly used. In Chapter 7, we compare FAIR data with DHIS2 for privacy-preserving health analytics. The lack of flexibility in semantic data production was observed in the first proof-of-concept. This could be solved

by localizing existing platforms, such as the Center for Expanded Data Annotation and Retrieval (CEDAR), for locally deployable versions. We examined the comparative strengths and limitations of DHIS2, and FAIR data in ensuring data privacy while allowing accessibility and reusability for effective health analytics. We elaborated on the integration of the VODAN-Africa architecture to enhance the Findability, Accessibility, Interoperability, and reuse of clinical and research data across various healthcare facilities in Uganda. 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.

In Chapter 8 we examine the integration of machine learning models with FAIR health data for advanced analytics and decision making. Using anonymized datasets from VODAN-Africa, specifically from Uganda, i.e., Uganda's Outpatient Department (OPD) records and Uganda's integrated antenatal care (ANC) information. These data sets are structured and standardized according to the FAIR principles and are used to train various machine learning models for tasks such as predicting pregnancy risk factors, assessing fetal conditions, detecting anomalies and performing clustering analyzes. We have highlighted the effective integration of Machine Learning with FAIR data in a federated architecture in healthcare analytics and underscored the significance of high-quality, accessible, and reusable data to support collaborative research and enhance public health outcomes.

Finally, in Chapter 9, we enumerate the conclusions and future outlook of the research. The use of FAIR data principles provides a viable solution to improving data interoperability and reuse in Uganda's digital health sector. The FAIR principles can enhance the efficiency and effectiveness of health data use by addressing gaps in the policy landscape, strengthening data integration and analysis, and enabling data reusability through privacy-preserving access. Consequently, the findings of this study provide a foundation for future initiatives aimed at strengthening data management approaches in Uganda and comparable contexts.