RESEARCH PAPER

Data Access, Control, and Privacy Protection in the VODAN-Africa Architecture

Putu Hadi Purnama Jati[1†], Mirjam van Reisen[1,2,3,4], Erik Flikkenschild[1,3], Fransisca Oladipo[4,5,7], Bert Meerman[6], Ruduan Plug[1], Sara Nodehi[1]

[1]Leiden University, 2331 GL Leiden, the Netherlands

[2]Tilburg University, P.O. Box 90153 5000, the Netherlands

[3]Leiden University Medical Center (LUMC), Leiden University, 1310 Leiden, the Netherlands

[4]Virus Outbreak Data Network-Africa

[5]Federal University, 260101 Lokoja, Nigeria

[6]GO FAIR Foundation, 2333 AA Leiden, the Netherlands

[7]Kampala International University, Uganda

Keywords: VODAN; VODAN-Africa; FAIR Data and Services; FAIR privacy framework; FAIR certification; FAIR Guidelines

Citation: Purnama Jati, P.H., Van Reisen, M., Flikkenschild, E., Oladipo, F., Meerman, B., Plug, R., Nodehi, S.: Data access, control, and privacy protection in the VODAN-Africa architecture. Data Intelligence 4(4), 938–954 (2022). doi: 10.1162/dint_a_00180
Submitted: March 10, 2021; Revised: June 10, 2022; Accepted: July 15, 2022

ABSTRACT

The Virus Outbreak Data Network (VODAN)-Africa aims to contribute to the publication of Findable Accessible, Interoperable, and Reusable (FAIR) health data under well-defined access conditions. The next step in the VODAN-Africa architecture is to locally deploy the Center for Expanded Data Annotation and Retrieval (CEDAR) and arrange accessibility based on the 'data visiting' concept. Locally curated and reposited machine-actionable data can be visited by queries or algorithms, provided that the conditions of access are met. The goal is to enable the multiple (re)use of data with secure access functionality by clinicians (patient care), an idea aligned with the FAIR-based Personal Health Train (PHT) concept. The privacy and security requirements in relation to the FAIR Data Host and the FAIRification workspace (to produce metadata) or dashboard (for the patient) must be clear to design the IT architecture. This article describes a (first) practice, a reference implementation in development, within the VODAN-Africa and Leiden University Medical Center community.

† Corresponding author: Putu Hadi Purnama Jati, Leiden University (Email: putu.hadi.purnama.jati@umail.leidenuniv.nl; ORCID: 0000-0002-6533-3709).

ACRONYMS

AI	artificial intelligence
CEDAR	Center for Expanded Data Annotation and Retrieval
DHIS	District Health Information System
DPA	data processing agreement
DPO	data protection officer
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile
GDPR	General Data Protection Regulation
IT	information technology
MVP	minimum viable product
VODAN	Virus Outbreak Data Network
VODAN-IN	Virus Outbreak Data Network Implementation Network
WHO	World Health Organization

1. INTRODUCTION

The potential of technology in the health sector is great. During the World Health Organization's (WHO's) 58th World Health Assembly (WHA58.28) in 2005, relevant stakeholders noted the potential impact that advances in information and communication technologies (ICTs) could have on healthcare delivery, public health, research, and health-related activities, for the benefit of both low- and high-income countries. Of particular interest are plans to implement national electronic public-health information systems and improve the capacity for surveillance of, and rapid response to, disease and public health emergencies.

To explore the growth and impact of eHealth across countries WHO has launched the Global Observatory for eHealth [1]. With the rapid growth in eHealth, there is an urgent need to balance data sharing benefits with privacy rights and ethical and regulatory criteria [2]. The FAIR Guidelines (that data should be 'Findable', 'Accessible', 'Interoperable', 'Reusable') identify clear criteria that promote the manual and automated depositing, discovery, sharing, and reuse of data in contemporary data publishing environments [3]. FAIR is not synonymous with open data—the 'A' in FAIR denotes accessibility, but only under well-defined conditions, which means it can be free and open, but must be validly protected under certain circumstances to protect privacy, national security, or competitiveness [4]. FAIR is a promising concept that has the potential to ensure responsible access to health data by recognising the protection of subjects and the need to share data on which decisions can be made, as data are transformed into information and used by improved advanced computing technologies [2, 5].

Africa has been largely absent from global data science in health in the past, resulting in health workers, doctors, nurses, and laboratory scientists interested in health knowledge not having access to leading health data [6]. The COVID-19 pandemic has presented an opportunity for African countries to accelerate their

commitment to better data management. This commitment is critical to enable Africa to effectively fight the pandemic—and future pandemics—and deliver better healthcare.

To take advantage of this opportunity, the Virus Outbreak Data Network (VODAN)-Africa has adopted the FAIR Guidelines to manage COVID-19 data [6]. The VODAN-Africa initiative, supported by the GO FAIR Foundation, involves universities and hospitals in Uganda, Ethiopia, Nigeria, Kenya, Tunisia, and Zimbabwe, and the Leiden University Medical Centre. The main objective of VODAN-Africa is to create COVID-19 FAIR Data Hosts in African countries to make Africa a source of data on the COVID-19 pandemic, protected by laws and regulations in each country. The local installation of the Center for Expanded Data Annotation and Retrieval (CEDAR) has been developed to undertake customisations based on the local verification of patient data.

However, particular attention needs to be paid to the processing of the health data of patients, as it is sensitive by nature [2]. The architecture built by VODAN-Africa needs to pay more attention to the principle of 'accessibility' under well-defined conditions and patient data should be carefully shared by filtering the stakeholders who can access or query the data in the repository. To do this, several factors—such as local regulations and the European Union's (EU's) General Data Protection Regulation (GDPR)[①]—need to be considered in the architecture to provide a compliant system for the participants involved in VODAN-Africa.

Accordingly, this article proposes an access and control framework with six access control layers. The convergence between the framework and the VODAN-Africa architecture discussed in this article is intended as a general model for developing FAIR Data Hosts among VODAN-Africa participants.

This architecture aims to control the data in four distinct ways: (i) the clinical data is protected with secure algorithms; (ii) the data remains in the locale where it is produced and does not leave the health facility; (iii) the data is governed by the regulations in place in each specific locale, including also the GDPR and FAIR Guidelines, which VODAN-Africa has adopted, and (iv) the health facility is entirely in charge of data access, control and security under the governing regulations of the specific country.

2. STUDY DESIGN

VODAN-Africa generally follows an ethnographic research design, which is oriented towards the stakeholders in the programme's development [6]. One of the project's tasks is to recognise that design is a social process that happens in a particular environment and that understanding the environment and the people in it is critical to solving design problems [7]. In this particular project, the design problem is twofold. The first part is that, as yet, there has been no practical application of FAIR data in the health sector. The second part of the problem is that theoretical writing on the application of FAIR in the health sector is rooted in a Eurocentric perspective on health and digital data. Van Reisen et al. [6, 7] have identified the need to apply the ideas to other geographies, which will allow researchers to better understand the validity of FAIR in different settings.

① This regulation covers on personal data held by private companies and public authorities to standardise the laws throughout the EU.

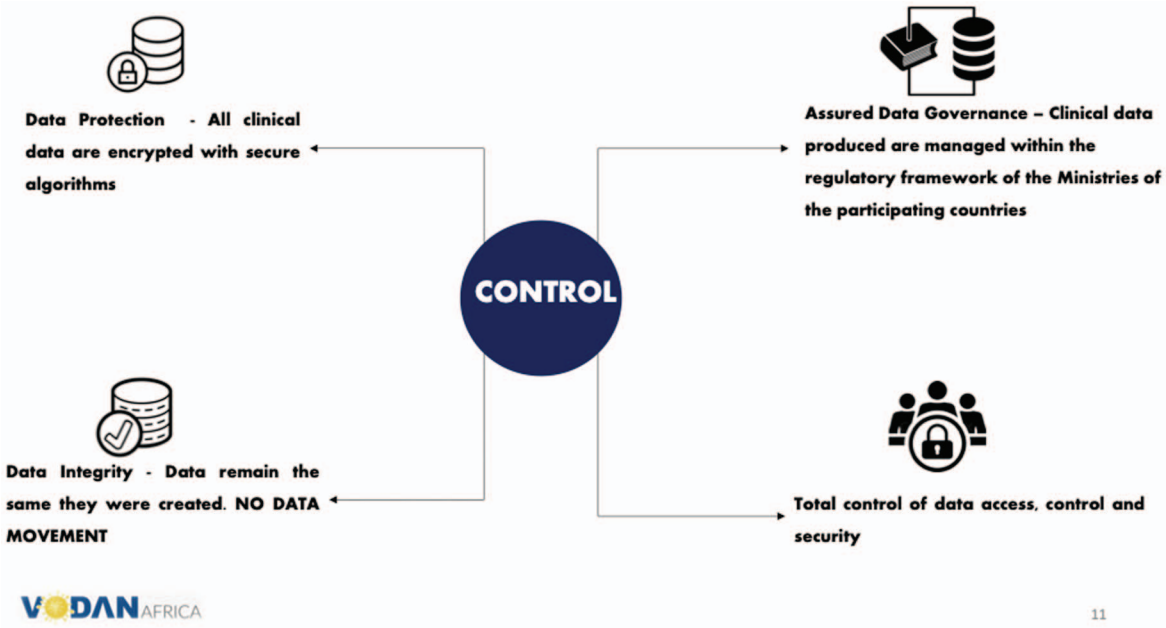

Figure 1. VODAN-Africa control architecture [6].

This research was designed as a case study, which is well suited to identifying the particularities of a certain situation. It focused on the introduction of machine-readable FAIR data to enhance the interoperability of health data in selected health facilities in nine countries: Zimbabwe, Uganda, Nigeria, Kenya, Tanzania, Somalia, Liberia, Tunisia, and Ethiopia. The study was conducted in cooperation with the Leiden University Medical Center, which served as a reference point for the project to discuss alternative approaches, compare obstacles and difficulties, and exchange ideas. The objective of the study was to explore the situational specifics in the introduction of FAIR as a basis for advancing the use of digital health data to improve decision-making and enhance the quality of healthcare in health facilities.

3. CONTEXT

The COVID-19 pandemic poses a significant challenge for all countries in determining the most effective way of combating the pandemic. Scientific models are critical resources for forecasting, predicting, and responding to biological, social, and environmental crises such as this pandemic [8]. As of the time of writing (early 2021), the COVID-19 pandemic is far from over. GO FAIR has highlighted that data storage and data reuse during this pandemic (and in previous instances) has been sub-optimal, and access to knowledge on past and present epidemics is not always equally available to all populations and countries affected by the pandemic [9]. Therefore, it is urgent to identify essential data trends by using machine learning and artificial intelligence (AI). In addition, to achieve the desired goals, every part of the data process must be FAIR compliant.

The VODAN-Implementation Network (IN) (one of the implementation networks of GO FAIR) aims to create a 'community of communities' that can quickly design and construct a genuinely international and interoperable data network infrastructure that promotes evidence-based responses to virus outbreaks, such as the current COVID-19 pandemic [10]. There are three phases of the VODAN project that include social-technical specification. The first phase is building a regional FAIR Data Host that enables data owners to view FAIR metadata and data. The FAIR Data Host stores machine-discoverable and interoperable data sets and their detailed metadata. The datasets can reside inside or outside the FAIR Data Host or its equivalent, such as CEDAR [11]. The second phase is developing the FAIR Implementation Profile (FIP) for different environments and conditions and various communities, addressing the FAIR Data Host as infrastructure (the desirable, as opposed to essential, FAIR Guidelines). This FIP reflects the implementation strategies of different communities to maximize the reuse and interoperability of data between and within the existing FAIR enabling tools [12].

The last phase (phase three) ensures the FAIRification of patient data so that every source data meets the FAIR Guidelines by using metadata capture tools to convert patient data in local clinics/hospitals, which are often in manual form, into a machine-interoperable format. This phase involves storing and accessing personal data, which requires an access control policy for FAIR data stewardship.

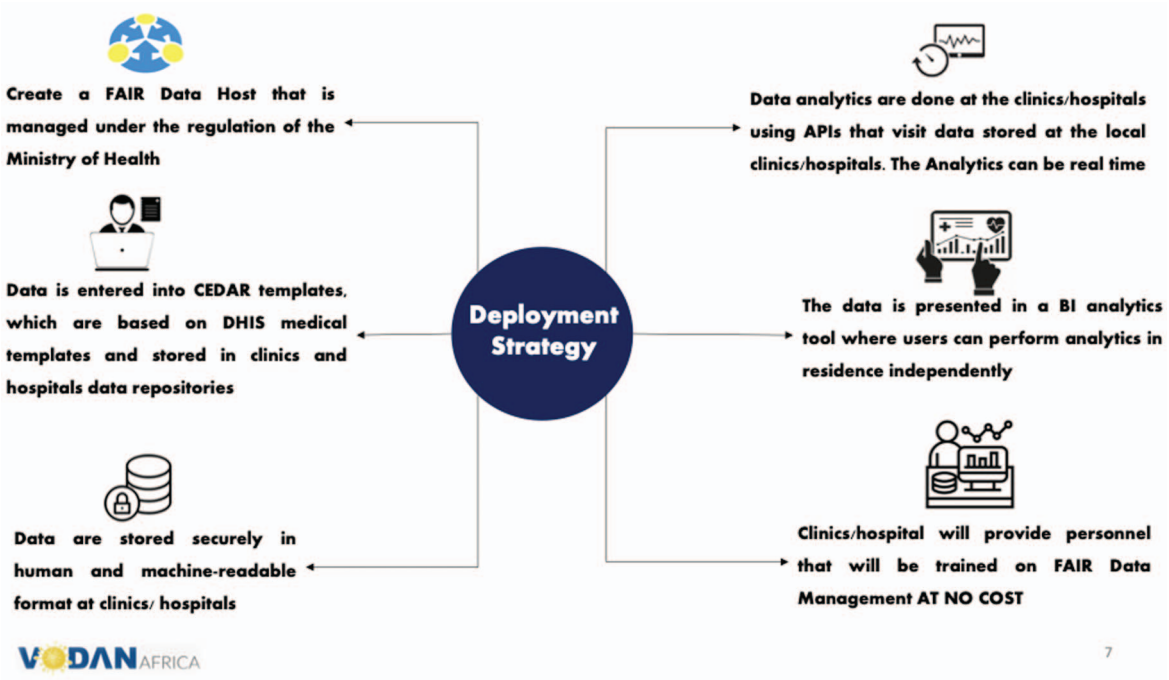

Figure 2. VODAN-Africa deployment strategy [6].

The access and control arrangements constitute a critical aspect of developing a trustworthy architecture for collecting, processing and transferring data securely in the FAIR Data Host or its equivalent, which is CEDAR-based. The successful use of access and control procedures is regarded as a technical challenge and a challenge of capacity building [6].

4. DATA VISITING PRIVACY FRAMEWORK

Before an institution can safely provide data for data-visiting or re-use, data privacy protection should be considered using a privacy assessment method. Among other things, the implementation of the FAIR Guidelines should comply with local laws and regulations, as stipulated by the local regulatory institutions. This relates to the regulatory context in which VODAN-Africa is implemented.

In VODAN-Africa, this analysis has been carried out using what is referred to as 'FAIR Equivalency' analysis; this measures the alignment of the relevant regulatory framework with the FAIR Guidelines. An analysis has been carried out for Uganda [12], Kenya [13], Nigeria [14], Ethiopia [15], and Zimbabwe [16], as well as outside Africa in Indonesia [17]. In addition to these analytical tools, the authorities relevant to health data management in the geographies incorporated in VODAN-Africa are required to support and approve the data handling activity.

At the next layer, within the data handling community, VODAN-Africa's 'community agreements' are generically based on the standard of personal data protection, respecting the GDPR. The GDPR aims to guarantee the free flow of data with the most stringent level of protection awarded to personal data, both within Europe and outside. The country coordinators in VODAN-Africa, together with the VODAN-Africa Board and communities, have signed a data processing agreement (DPA) with VODAN-Africa. In this agreement they recognise their duty to exercise responsibility over data protection within the execution of VODAN-Africa data management. Data handling within the participating health facilities is regulated through data use agreements between the facility and VODAN-Africa. These agreements govern the responsibilities, obligations, and rights of all institutions and persons involved in data handling operations across the different sovereignties and geographies at all different levels. These can be referred to as the community agreements that relate to the IT reference architecture in VODAN-Africa. These agreements are predecessors to the Personal Health Train (PHT) concept and the clinical IT goal architecture, with potentially a dynamic mode and global reach in the future. Closest to the data is the data producer (usually the health facility), which exercises its institutional regulatory framework regarding its data processing.

The data visiting privacy framework (see Figure 3) addresses the generic access control requirements for the secondary use of data for different uses, such as AI (data scientists), personal medicine (doctors), and open science (researchers). In other words, due to the differences that exist in the interpretation of the GDPR within institutions and countries, it is hard to form an interoperable system. Therefore, the data visiting privacy framework is defined in order to create a win-win situation for all stakeholders.

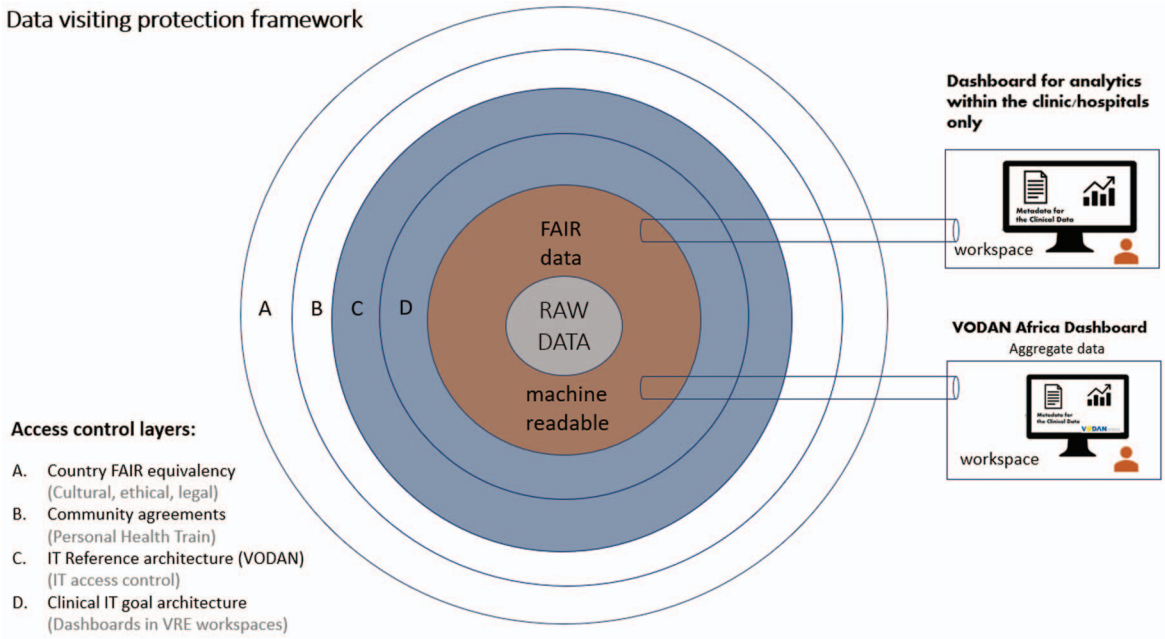

Figure 3. Data visiting privacy framework: The four access layers of the privacy model used by VODAN-Africa.

In order to gain trust in privacy, a combination of measures—such as the GDPR, ethical policies, technology, and institutional (data transfer) agreements—are required. From a legal perspective, three major issues are delaying innovation for data-driven re-usability:

- The absence of a policy in which ethical/legal discussion points are translated into an institutional position. Concerning the secondary use of data, its use within AI, personalised medicine, and in the open science context should be handled in separate ethical discussions.
- Trust in the anonymity of data is a challenge. As soon as you start harvesting data from different data sources, the indirect traceability of the identity of data becomes a problem. To address this, reliable anonymization technologies are required.
- The absence of a community-wide harmonised privacy-by-policy view on IT systems, resulting in expensive and time-consuming point-to-point solutions between institutions.

To overcome these problems, the FAIR privacy framework is defined in four layers, and each layer has a specific protection goal. Figure 3 shows the four layers of the data protection framework, which are as follows:

- **Layer A (ethics and compliance):** Risk assessment should cover national legal requirements and ethics. A generic questionnaire covering all the necessary components is required, which will be developed within the VODAN community in the future.
- **Layer B (VODAN-Africa):** VODAN-Africa defines the access rules that must be developed, showing what a user is allowed (and not allowed) to do. This timetable will result in fully automated access

to (anonymized or pseudonymized) personal data (one credential for all logins) or a defined step between processes for manual use (manual login).

- **Layer C (IT reference architecture):** A community reference architecture is needed to position innovative IT services such as ontology-based access control. A semantic ontology can be an informal conceptual framework with conceptual forms and their named relationships, if any, in natural language [18]. In the semantic web, policy management could be deployed to define rules for accessing a resource and provide users to interpret and comply with these rules [19]. Ontology-based access control (OBAC) is developed to create, modify, and query semantically rich policies.
- **Layer D (clinical IT goal architecture):** Each participating member in a community has to implement the institutional access control IT reference architecture following the agreed outcomes and decisions from the other layers. The challenge here is to adhere to the GDPR anonymization definition. The GDPR only concerns personal data that are related to natural persons. The identification of an individual is made via that information directly or indirectly [20]. Because in anonymized data, it is not possible to identify data objects (directly or indirectly), the GDPR does not consider anonymized data as personal data. None of the patient data can be identified anonymously, making the system guarantee that data is safe for analysis. However, the pseudonymization of data may be necessary if the patient needs to be identified for a medical purpose, such as a medical treatment.

5. PRIVACY ASSESSMENT PROCEDURE (INSTITUTIONAL LEVEL)

The privacy framework described in the previous section is primarily developed for (higher) management, senior advisors, and decision-makers to address *privacy by design concerns*. The primary purpose is to discuss institutional concerns regarding security and privacy (often in an ethical domain) in non-technical terms. To start the process, a questionnaire was developed to assist in decisions about the protection of institutional data, in other words, to decide on data stewardship responsibilities, in line with the GDPR. The data stewardship responsibilities relate to the objectives of VODAN-Africa, which are to (i) store patient data in residence, (ii) create computational feedback from data within clinics, (iii) create interoperability through data visiting across, expressed in computed (frequency) analytics for simple data visualisations, and (iv) strengthen capacity and capability within the health facilities to support steps (i)–(iii), while maintaining data sovereignty. In VODAN-Africa, the 'access and control' issue goes back to each facility's 'control and processing' functions, which are complemented by national oversight as an umbrella in a politically sovereign region. The controller, which may be the health facility director, or someone assigned by the director to exercise this responsibility, determines the purposes and means of processing patient data, and the data processors (within and/or outside the clinic) assist in the realisation of the computing of the queries on the dashboard, under the instruction of the data controller in the clinic. The data controller in the clinic works under the guidance of the supervisory authority, presumably a ministry of health or an entity authorised by the ministry of health, possibly assisted by a data protection officer (DPO). In VODAN-Africa, the country coordinators also exercise responsibility as DPOs within the countries that they represent.

As VODAN-Africa is a dynamic real-time data computing system, it needs 'humans in the loop' to create trust in the system—especially regarding the data control—and these humans are always accountable to

the hierarchy. Data processing can be stopped or reviewed at any time, as per the instruction of the data controller, and/or the supervisory authority. To achieve this, a clear understanding of data control tasks, data processing tasks, and the role of the DPO and supervisory authority is needed for each location. In addition, the capacity to adapt the dashboard—and which data is available for viewing by whom, in each situation and each time—is critical.

For access and control to be viable on the work floor, the adaptation of the dashboard in clinics needs to be a relatively trivial operation. This could be achieved by the instruction from the data controller in the clinic about tasks that the data processors fulfil. The operation to adapt the dashboard should be relatively intuitive and transparent. After all, the data controller will be responding to various decisions made in the hierarchy, including political decisions. In our common understanding, this would be a human process and could be regarded as part of a capacity-building process for people to understand their responsibilities in terms of discharging their obligations, providing instruction, and responsiveness in relation to their position in the hierarchy.

It is conceivable that the development of the minimum viable product (MVP) will require these roles to be automated. The access and control issues could also be based on written agreements for data control and processing, customised for each facility, which forms the basis of the dashboards. However, a critical element of access and control is that the dashboard is adaptable to respond to changing requirements in the clinic so that administrators can effectively exercise access and control within a dynamic environment in compliance with applicable data regulations.

6. ACCESS CONTROL APPROACHES IN VODAN ARCHITECTURE

Figure 4 shows the architecture integrated template based on CEDAR and the tools that focus on the use of metadata templates, which define the necessary data attributes to describe specific biomedical experiments [21].

The architecture follows the District Health Information Software (DHIS) template for vocabulary control or any alternative health information system in use by the facility. Following the (D)HIS template is important, as it allows an aligned reporting system with what the ministry of health in the country requires. In order to enable this alignment, VODAN-Africa has analysed how health facilities in each country may be obliged to comply with such systems and has created a component through which this de-identified data can be automatically produced to comply. A local repository ensures that the clinic or hospital retains its data so that it can conduct in-residence data analysis. There are three outputs of the data repository: (i) a dashboard for analysis of the data within the clinic or hospital, (ii) a CEDAR instance for publishing the availability of the health facility (with FAIR curated data) to the community (findability), and (iii) a dashboard showing generic data based on VODAN-Africa's stakeholder agreement. Data security must be understood at each transition point and to protect storage in the chain. Identified challenges in each architectural domain could be an essential approach to formulate policies for access and control.

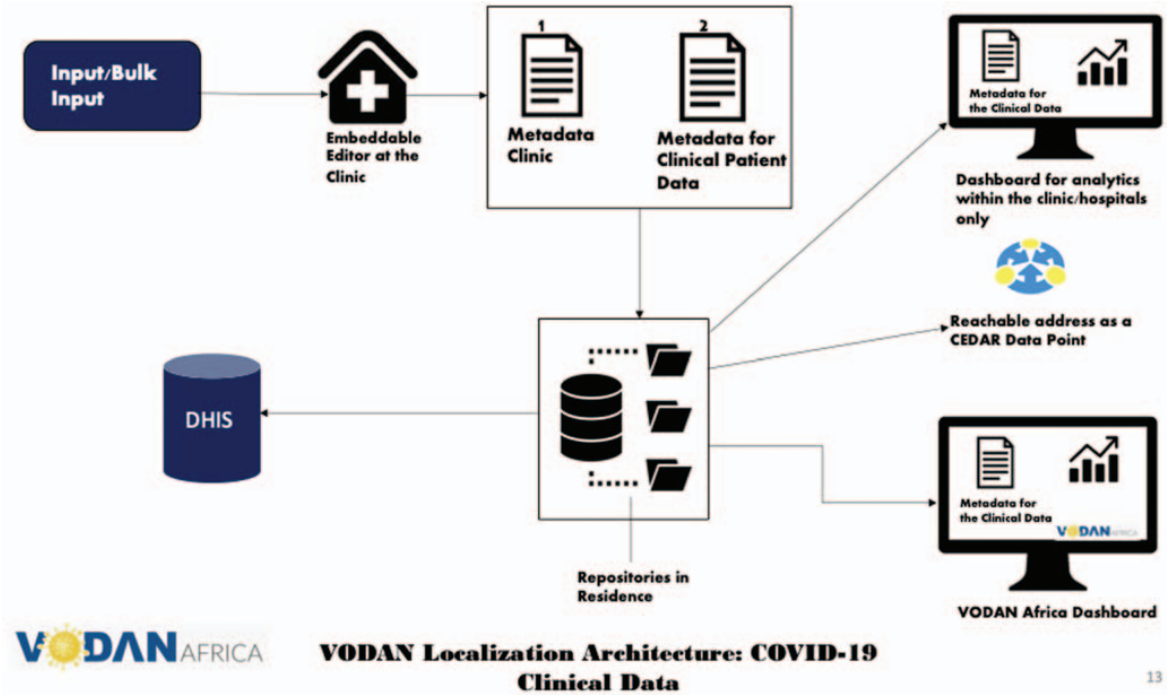

Figure 4. VODAN-Africa CEDAR localization architecture [6].

Access and control are meant to restrict the behaviour and operations that a legitimate user can perform. Access control also limits what a user can do directly and what programs running on behalf of the user can do [22]. An architecture that implements FAIR should identify what users can perform in the access and control procedure, because a safe architecture must be transparent and secure to gain trust from every stakeholder. We realise that no best practice fits all systems to protect user privacy in building access and control. Therefore, we need to define policies and mechanisms suitable for the architecture of FAIR and CEDAR integration. Policies are high-level guidelines that specify how access controls are determined and access decisions are taken, while mechanisms are low-level software and hardware functions configured to implement a policy. Policies establish high-level rules for access control and decisions, while low-level software and hardware frameworks enforce such policies [22]. The architecture policies are the first layer of the data privacy framework, which covers legal requirements and ethics. Six layers of control have been developed, based on the VODAN-Africa legal document and the current architecture (Figure 5).

6.1 Layer 1: VODAN Community Agreement Document

In this layer, we define who the parties involved are and with what purpose limitations. The outcomes are fixed in a DPA. The DPA forms the top level of data access and control, on which all further control and access patterns are based. The DPA recognises sovereign local regulations, which are specified by the regulatory body of the nation in which the data resides, including by the respective ministries and local

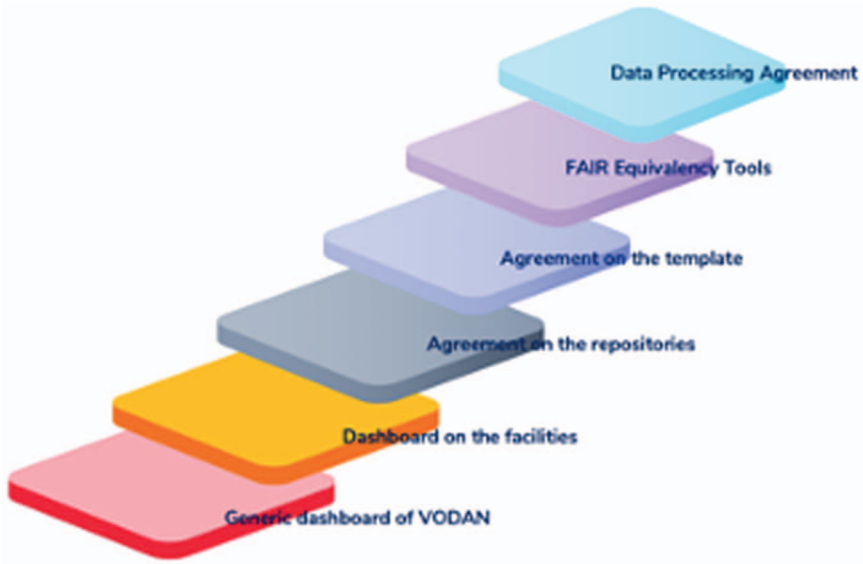

Figure 5. VODAN six layers of control [6].

regulatory officials. Therefore, the community agreement is a formal agreement that considers the various regulatory frameworks, which may differ from country to country, and should accommodate local laws and regulations. The agreement is based on the joint controller regulatory framework described in Regulation (EU) 2016/679 of the European Parliament and of the Council, Chapter IV, Section 1 (GDPR) [23]. In addition to regulations on oversight by a DPO, this framework stipulates that two or more stakeholders make decisions on 'why' and 'how' data is used and stored, which are under the authority of the data processing and controlling parties that implement data protection by design (Chapter 4, Article 25) [23], by issuing guarantees and guidelines in the form of a certified mechanism specified in the agreement. The format in which these decisions are made, with the relevant conditions, must be well documented with transparent conditions compliant with the overarching regulations. This framework provides further rules for the data processors (defined in GDPR as the parties acting on the data) and the data controllers (the parties that house the data in residence). It must be noted here that controlling data does not automatically imply data ownership, as relevant ministries or the health facility may delegate open access or external data to be controlled internally with full legal rights, dependent on the licence under which the data has been procured.

6.2 Layer 2: FAIR Equivalency Tool

FAIR implementation in digital health can address the lack of sustainability of digital medical initiatives, fragmentation, and the lack of integration solutions, as long as the FAIR Guidelines are adopted/implemented with contextual awareness [24]. FAIR Equivalency aims to assess the governance framework regarding digital health/eHealth policies in a specific context, for example, in a national or local context, to see if they (in)directly adhere to the FAIR Guidelines. Determining FAIR Equivalency allows us to understand the

limitations in place due to different national policies and regulations. In short, FAIR Equivalency checks what is allowed and not allowed in a specific context when developing a FAIR architecture.

Not only in addition to, but also as part of, the joint controller agreements, FAIR provides a strong conceptual basis for data protection by design. Data processing can only occur under the data controller's supervision and certification by strictly defining an access pattern and utilising data visiting. The data controller (typically the supervisory board of the database, in addition to the DPO) and the data processor are jointly responsible for data security, for which each agreement to process data is handled as per the architectural documentation of VODAN-Africa and according to the regulatory oversight by the government supervising body.

EU2016/679 Chapter 4, Section 2, Articles 32–34 (GDPR) [23] provides top-level regulations, which FAIR implementation must abide by for secure data communication, storage, and handling, including stipulations on the mandatory reporting of data breaches. As a consequence, for an implementation to be FAIR Equivalent in terms of access and control, it must be ensured that, within the second layer, the architecture of implementation allows for access and control agreements from the first layer, while ensuring that the access and control of data is both secure and transparent for both stakeholders and supervisory authorities.

6.3 Layer 3: Agreement on Templates

In this layer, we specify which templates (metadata) a particular health facility will use. The community agrees to use these machine-readable templates in the clinic/facility. Therefore, possible privacy concerns about, for example, the use of personal data items, can now be developed based on the agreement.

Once the regulatory requirements have been met, access and control agreements have been made, and the architecture is FAIR compliant, securely storing and handling private information should be part of data protection by design. As templates, which may stipulate which personally identifying information is gathered or stored, are defined by health facilities in residence, the ultimate responsibility for regulating data collection and storage falls to the DPO of the facility.

All templates, or changes therein, should be certified by the fata controller and the DPO, as indicated in EU2016/679 Chapter 4, Section 1, Article 30 (GDPR) [23]. This article stipulates that the representative of the joint controller, in this instance the data controller of the health facility and the DPO, shall maintain a record of all data processing activities concerning personally identifiable data that belongs to a data subject and uphold relevant safeguards for data protection regarding these data.

6.4 Layer 4: Agreement on Repositories

This layer relates to the agreement of the clinic with a sustainable repository. Based on this, the responsibilities are defined in relation to access and the security of these data. The following question is addressed: Where on the globe is the machine-readable data health stored (or is the data stored in residence

or somewhere else in the country)? Furthermore, in this layer, based on the approved policy, decisions are made as to who has access to that repository for what purpose. In VODAN-Africa, the data is stored in residence, by design, in other words in the location where the data is produced.

All potentially identifying information may only be stored following the provisions of the GDPR on the protection of the data subject. EU2016/679 Chapter 3, Sections 1–2, Articles 12–15 contain provisions about the rights of data subjects from the nation states on which these articles are legally binding [23]. These articles stipulate that data subjects may retrieve their own data if requested. This regulation must be incorporated into the access design, in addition to measures that allow for confirmation that the individual is the data subject. This may pose challenges as the data controller, and data subject may reside in different nations and fall under different regulatory frameworks. In VODAN-Africa, this problem is resolved in that the data is always stored in residence as per the GDPR framework, which adheres to the highest requirements for data protection. In addition, because of these regulations, for full compliance, there must be a logging system for each repository that indicates which parties have accessed personally identifiable information, which is also in accordance with EU2016/679 Chapter 4, Section 1, Article 30 on the collection of personally identifiable data [23].

6.5 Layer 5: Facility Dashboard

There are two types of dashboards connected to data repositories. The first dashboard is for each health facility for data analysis purposes. A method needs to be developed for how to protect the dashboard in relation to the local directory services, such as the MS-Active Directory. Moreover, the construction of the dashboard requires agreement on accessibility on a per-use case basis with a different selection of required data sources (based on third layer templates).

The facility dashboard is the second-to-lowest level component and is where local data processors and the data controller interact according to the access and control framework. For example, an in-residence researcher may ask permission from the representative of the data controller, in this case the supervisor of the facility database, to access sensitive data for research purposes. As these exchanges of data happen in-residence or within known research groups that are local legal entities, such data requests are typically only subject to local regulations, for which the data controller is the executive regulator responsible for overseeing data access and control and the relevant ministries establish the supervising body that creates the guidelines for the data controller to comply with.

6.6 Layer 6: Generic Dashboard of VODAN

The second dashboard in the architecture aims to provide a generic dashboard to make data available to the community (such as for researchers, data scientists, and the government). As a result, a community agreement representing the informed consent of citizens is required in relation to what will be included in the dashboard, with its purpose limitation. Another requirement could be an opt-out system, a kind of light system in which the facility can implicitly agree to provide data for general information on this dashboard.

The generic dashboard forms the lowest level access layer, where all interactions between the joint controllers of the various stakeholders take place. This level is most critical for access and control, as any failure to safeguard or comply may overrule earlier safeguards. Examples of this are requirements for the verification of access using certificates or special access tokens.

A full regulatory framework on external access, as required for data access between nations, is given in EU2016/679 Chapter 5, Articles 44–50 [23]. However, as with data visiting no legal transfer of actual data is performed, a new access pattern framework needs to be defined to provide safeguards on the authorisation and use of such data. This would depend on the legal interpretation of these regulations on data visiting and become possible subject to newly developed stipulations. For full compliance, adhering to Articles 44–50 would provide a strong initial framework for data protection and regulation purposes.

7. IT DEPLOYMENT

This roadmap provides milestones for a demonstrable MVP, testing the rails and tracks with actual data. VODAN-Africa will address this in collaboration with the relevant parallel programs. Translating the VODAN reference architecture into an institutional target IT architecture is the objective. Approved architectures for FAIR Data Hosts are ready to be interconnected and adjusted according to the privacy assessment requirements; in other words, all lights in the FAIR Data Host access control authentication matrix (truth table) must be green to access the data.

8. FAIR CERTIFICATION

To increase the 'trust level', steps can be taken towards FAIR endorsement and certification. Within the certification process, there must be full transparency and community support for the standards that must be met. In addition, it is important to recognise that certification brings together governance, policy, legal boundaries, and technical applicability. An initial study to investigate the automated assessment/certification of VODAN FAIR Data Hosting (FAIR Data Host) has begun.

Multiparty data projects centred around data-sharing and/or data-visiting rely heavily on trust among the different stakeholders. When multiple parties work towards a new architecture for a complex data exchange environment, all participants must adhere to specific agreed rules, regulations, and codes of conduct. To increase the 'trust factor' in these data projects, certification could be desirable. A well-defined certification schema, built up and agreed upon by the community involved, could be of great value to ensure proper adherence to the defined and agreed rules.

Building up a certification programme requires multiple parties to collaborate and gradually develop the specified formal requirements. A good approach is to start with requirements and guidelines that are relatively easy to adhere to in the first phases, gradually building up the schema to a more sophisticated level. It is important to include the vast majority of participants in the initial stages.

9. CONCLUSION AND FUTURE WORK

This article lays out a roadmap for developing security, access, and control IT architecture for FAIR Data Hosts, including creating a FAIRification environment with a secure connection to health data sources. The first step is to use a privacy framework specifically designed for addressing stakeholder concerns about the global multi-purpose reuse of FAIR data. The focus is on determining the required access control technology (addressing the 'A' in FAIR) in relation to the sensitivity of data in transit and storage. The roadmap starts with a risk assessment of ethical, cultural, and legal issues to inform the design of VODAN's infrastructure architecture for secure FAIR data curation. Data stewards within the health sector must have access to health data (privacy-sensitive), for example, in order to metadata the dataset (input) or analyse data from a particular perspective (output).

A data privacy framework (consisting of four layers) is proposed to ensure access, control, and security in the VODAN architecture. The implementation of the privacy framework should be done by identifying access and control guidelines using three types of documents: the DPA between the coordinators of VODAN in each of the African countries (and beyond), FAIR Equivalency with regulations in place in each location, and an agreement on the architecture (template, repositories, and dashboard). VODAN-Africa has implemented a FAIRification process by using CEDAR's metadata capture tools. The architecture created should make the CEDAR tools converge with the security framework to guarantee secure data.

In the VODAN-Africa architecture, six layers are identified through which access and control over the data handling are arranged. The layers correspond well with the data protection regulation of the EU, which distinguishes clear roles and obligations, which taken together, amount to an acceptable level of checks and balances to guarantee the highest possible level of data protection. VODAN-Africa is in full compliance with the GDPR, stringently arranging data protection based on the principle of repositing data in location and allowing data access through data visiting.

This research provides a direction to perform privacy security in the VODAN architecture and contributes to a trusted FAIR workspace for data stewards. Future research should answer several questions, such as how the framework can be implemented in different cultures and environments and detailed mechanisms to fulfil needs in different environments. In addition, an evaluation is needed to check whether or not the architecture has been successfully implemented to ensure the access, control, and security of the patient data. Furthermore, it is worth considering a comprehensive certification schema to increase trust in the VODAN infrastructure.

ACKNOWLEDGEMENTS

We would like to thank Misha Stocker for managing and coordinating this Special Issue (Volume 4) and Susan Sellars for copyediting and proofreading. We would also like to acknowledge VODAN-Africa, the Philips Foundation, the Dutch Development Bank FMO, CORDAID, and the GO FAIR Foundation for supporting this research. Finally, we acknowledge the district local government authorities, particularly the district health officers and the data management personnel who availed information that was crucial to the success of this study.

AUTHORS' CONTRIBUTIONS

Putu Hadi Purnama Jati (putuhadi2808@gmail.com, 0000-0002-6533-3709): Writing—original draft preparation, investigation, conceptualization. Mirjam van Reisen (mirjamvanreisen@gmail.com, 0000-0003-0627-8014): Supervision, Writing—review and editing, validation, conceptualization, project administration. Erik Flikkenschild (e.flikkenschild@lumc.nl, 0000-0002-7285-1651): Writing—original draft preparation, validation, resources, investigation, conceptualization. Fransisca Oladipo (francisca.oladipo@kiu.ac.ug, 0000-0003-0584-9145): Supervision, validation, conceptualization, project administration. Bert Meerman (bert_meerman@hotmail.com, 0000-0002-0071-2660): Writing—original draft preparation, validation, resources, investigation, conceptualization. Ruduan Plug (ruudplug@gmail.com, 0000-0001-5146-6116): Writing—review and editing, resources, investigation. Sara Nodehi (sara.nodehi95@gmail.com, 0000-0002-2919-1336): Writing—review and editing, resources, investigation.

CONFLICT OF INTEREST

All of the authors declare that they have no competing interests.

ETHICS STATEMENT

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

REFERENCES

[1] WHO; International Telecommunication Union: National eHealth strategy toolkit overview. International Telecommunication Union (ITU), Geneva (2012). Available at: https://apps.who.int/iris/bitstream/handle/10665/75211/9789241548465_eng.pdf?sequence=1&isAllowed=y. Accessed 29 April 2021

[2] Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., et al.: The 'A' of FAIR—As open as possible, as closed as necessary. Data Intelligence 2(1–2), 47–55 (2020). doi: 10.1162/dint_a_00027

[3] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al.: Comment: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1), 1–9 (2016). doi: 10.1038/sdata.2016.18

[4] Purnama Jati, P.H.: Improving Satu Data Indonesia with FAIR elements: A model to extend Satu Data Indonesia principles in COVID-19 data management. Unpublished Master's Thesis, Leiden University (2020)

[5] European Commission, Directorate-General for Research and Innovation: Turning FAIR into reality: Final report and action plan from the European Commission Expert Group on FAIR Data. Publications Office of the European Union, Luxembourg (2018). doi: 10.2777/54599

[6] Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihgo, M., Folorunso, S., Schultes, E., et al.: Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. Advanced Genetics 2(2) (2021). doi: 10.1002/ggn2.10050

[7] Van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A.O., Kirkpatrick, C., Mons, B.: Towards the tipping point of FAIR implementation. Data Intelligence 2, 264–275 (2020). doi: 10.1162/dint_a_00049
[8] Barton, C.M., Alberti, M., Ames, D., Atkinson, J.-A., Bales, J., et al.: Call for transparency of COVID-19 models. Science 368(6490), 482–483 (1 May 2020). doi: 10.1126/science.abb8637
[9] GO FAIR: Virus Outbreak Data Network (VODAN) [Online]. GO FAIR (n.d.). Available at: https://www.go-fair.org/implementation-networks/overview/vodan/. Accessed 29 April 2021
[10] Mons, B.: The VODAN-IN: Support of a FAIR-based infrastructure for COVID-19. European Journal of Human Genetics 28(6), 724–727 (2020). doi: 10.1038/s41431-020-0635-7
[11] Dutch TechCentre for Life Science (DTL): Find FAIR Data tools—Data Stewardship [Online]. (2021). Available at: https://www.dtls.nl/fair-data/find-fair-data-tools/. Accessed 29 April 2021
[12] Basajja, M., Van Reisen, M., Oladipo, F.: FAIR Equivalency with regulatory framework for digital health in Uganda. Data Intelligence 4(4), 771–797 (2022)
[13] Thea, E.I., Nalugala, R., Nandwa, W., Obwanda, F., Wachira, A., Cartaxo, A.M.: FAIR Equivalency, regulatory framework and adoption potential of FAIR Guidelines in health in Kenya. Data Intelligence 4(4), 852–866 (2022)
[14] Kawu, A.A., Joseph E., Abdullahi, I., Maipanuku J., Folorunso, S., Basajja, M., Oladipo F., Ibrahim, H.L.: FAIR Guidelines and data regulatory framework for digital health in Nigeria. Data Intelligence 4(4), 839–851 (2022)
[15] Taye, G.T., Amare, S.Y., Tesfit G.G., Medhanyie, A.A., Ayele, W., Habtamu, T., Van Reisen, M.: FAIR Equivalency with regulatory framework for digital health in Ethiopia. Data Intelligence 4(4), 813–826 (2022)
[16] Chindoza, K.: Regulatory framework for eHealth data policies in Zimbabwe: Measuring FAIR Equivalency. Data Intelligence 4(4), 827–838 (2022)
[17] Purnama Jati, P.H.: FAIR Equivalency in Indonesia's digital health framework. Data Intelligence 4(4), 798–812 (2022)
[18] GO FAIR: Build the FAIR Implementation Profile—How to go FAIR [Online]. GO FAIR (n.d.). Available at: https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/. Accessed 10 January 2021
[19] Health-ri: The PHT concept [Online]. (n.d.). Available at: https://pht.health-ri.nl/pht-concept. Accessed 13 January 2021
[20] NetSec.news: What is the definition of a HIPAA covered entity? NetSec News (9 October 2017). Available at: https://www.netsec.news/definition-hipaa-covered-entity/ Accessed 29 April 2021
[21] CEDAR: CEDAR, Better metadata means better science [Online]. (2021). Available at: https://metadatacenter.org/. Accessed 29 April 2021
[22] Can, O., Bursa, O., Unalir, M.O.: Personalizable ontology-based access control. Gazi University Journal of Science 23, 465–474 (2010)
[23] European Parliament and Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union L119/1, pp. 1–88 (2016). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&rid=2. Accessed 7 September 2021
[24] Basajja, M.: FAIR Data Guiding Principles: Adoption in the ehealth environment in Uganda. Data Intelligence (2021)