

Incomplete COVID-19 Data: The Curation of Medical Health Data by the Virus Outbreak Data Network-Africa

Mirjam Van Reisen^{1,2,3,4†}, Francisca Onaolapo Oladipo^{4,5,6}, Mouhamed Mpezamihigo^{4,5}, Ruduan Plug^{2,4}, Mariam Basajja^{2,4}, Aliya Aktau^{2,4}, Putu Hadi Purnama Jati^{4,7}, Reginald Nalugala^{4,8}, Sakinat Folorunso^{4,9}, Samson Yohannes Amare^{4,10}, Ibrahim Abdulahi^{4,11}, Oluwole Olumuyiwa Afolabi^{4,5}, Ezra Mwesigwa^{4,12}, Getu Tadele Taye^{4,10}, Abdulahi Kawu^{4,11}, Mariem Ghardallou^{4,13}, Yan Liang^{2,4}, Obinna Osigwe^{4,5}, Araya Abrha Medhanyie^{4,10}, Munyaradzi Mawere¹⁴

¹Leiden University Medical Centre, 1310, the Netherlands

²Leiden University, 1310, the Netherlands

³Tilburg University School of Humanities and Digital Sciences, 1310, the Netherlands

⁴VODAN-Africa, 260101

⁵Kampala International University, P.O. Box 20000, Uganda

⁶Federal University Lokoja, Nigeria

⁷Central Bureau of Statistics, 81119, Indonesia

⁸Tangaza University College, P.O.Box 15055, Karen Nairobi Code 00509, Kenya

⁹Olabisi Onabanjo University, P.M.B 2002, Ago-Iwoye, Ogun State, Nigeria 120005, Nigeria

¹⁰Mekelle University, P.O. Box 231, Ethiopia

¹¹Ibrahim Badamasi Babangida University, 911101, Nigeria

¹²Makerere University, P.O. Box 7062 Kampala, Uganda

¹³Université de Sousse, 4002 Sousse, Tunisia

¹⁴Great Zimbabwe University, P.O. Box 1235 Masvingo, Zimbabwe

Keywords: FAIR Guidelines; FAIR Equivalency; Federated data; Digital health; GDPR; VODAN-Africa

Citation: Van Reisen, M., Oladipo, F., Mpezamihigo, M., Plug, R., Basajja, M., Aktau, A., Purnama Jati, P.H., Nalugala, R., Folorunso, S., Amare, Y.S., Abdulahi, I., Afolabi, O.O., Mwesigwa, E., Taye, G.T., Kawu, A., Ghardallou, M., Liang, Y., Osigwe, O., Medhanyie, A.A., Mawere, M.: Incomplete COVID-19 data: The curation of medical health data by the Virus Outbreak Data Network-Africa. *Data Intelligence* 4(4), 673–697 (2022). doi: 10.1162/dint_e_00166

Submitted: March 10, 2021; Revised: June 10, 2022; Accepted: July 15, 2022

[†] Corresponding author: Mirjam van Reisen, Leiden University Medical Centre (LUMC) (Email: m.e.h.van_reisen@lumc.nl, mirjamvanreisen@gmail.com; ORCID: 0000-0003-0627-8014).

ABSTRACT

The incompleteness of patient health data is a threat to the management of COVID-19 in Africa and globally. This has become particularly clear with the recent emergence of new variants of concern. The Virus Outbreak Data Network (VODAN)-Africa has studied the curation of patient health data in selected African countries and identified that health information flows often do not involve the use of health data at the point of care, which renders data production largely meaningless to those producing it. This modus operandi leads to disfranchisement over the control of health data, which is extracted to be processed elsewhere. In response to this problem, VODAN-Africa studied whether or not a design that makes local ownership and repositing of data central to the data curation process, would have a greater chance of being adopted. The design team based their work on the legal requirements of the European Union's General Data Protection Regulation (GDPR); the FAIR Guidelines on curating data as Findable, Accessible (under well-defined conditions), Interoperable and Reusable (FAIR); and national regulations applying in the context where the data is produced. The study concluded that the visiting of data curated as machine actionable and repositing in the locale where the data is produced and renders services has great potential for access to a wider variety of data. A condition of such innovation is that the innovation team is intradisciplinary, involving stakeholders and experts from all of the places where the innovation is designed, and employs a methodology of co-creation and capacity-building.

ACRONYMS

| | |
|----------|---|
| CEDAR | Center for Expanded Data Annotation and Retrieval |
| COVID-19 | coronavirus disease of 2019 |
| eCRF | electronic COVID report Form |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| FE-Score | FAIR Equivalency Score |
| GDPR | General Data Protection Regulation |
| VODAN | Virus Outbreak Data Network |
| WHO | World Health Organization |

1. INTRODUCTION

Following the announcement of the discovery of a new coronavirus disease of 2019 (COVID-19) 'variant of concern', Omicron (B.1.1.529), the Co-Chair of the African Vaccine Delivery Alliance, Ayoade Alakija, expressed outrage over the inadequate and incomplete action taken to include Africa in global efforts to curb the COVID-19 pandemic. This shone a spotlight on the dramatic inequality of vaccination rates: 63% of people in high-income countries were fully vaccinated by October 2021, compared to only 1.4% of people in low-income countries [1]. The B.1.1.529 variant of COVID-19 was first reported on 24 November 2021 by the National Institute for Communicable Diseases (NICD) [2] in South Africa, and was later classified as the Omicron variant by the World Health Organization (WHO) [3]. The inequality of vaccination has been condemned by WHO, with organisations such as Amnesty International calling for global vaccine equity.

Lack of and incomplete data on COVID-19 constitutes a critical shortcoming, affecting global efforts to curb the SARS-CoV-2 pandemic. It has resulted in a lack of clarity about the situation and obscured the problems posed by COVID-19 on the African continent. Lack of understanding of the challenges in Africa, missed opportunities to set up production facilities for vaccines in Africa, and limited allocation of vaccines to Africa, have undermined efforts to get the pandemic under control on the continent—and globally. The failure to address vaccination needs in Africa is partly due to the of lack of reliable data on COVID-19 infection and hospitalisation rates, which has resulted in the urgency of vaccination in Africa remaining low. While vaccination is needed to protect populations from COVID-19, irrespective of the scale of infection, it is also needed to curb the emergence of new variants of concern. Africa has become a blind spot in the quantification of the global pandemic, which makes the formulation of measures and distribution of resources for health facilities in Africa particularly challenging. Without adequate data on COVID-19 from Africa, global efforts to curb the pandemic are incomplete and new variants of concern may continue to emerge and spread globally before proper local measures are able to be taken.

With the absence of solid data, insights into the extent to which the 1.3 billion people in Africa have been affected are speculative [4–6]. It also remains unclear whether COVID-19 has affected the African continent less than other continents, or if the extent of infections has simply remained undetected [7]. If access to reliable measurement is limited, COVID-19 rates of infection in Africa will remain unquantifiable. As well as incomplete or missing data, data related to the incidence, treatment and severity of COVID-19 in Africa is often biased towards the more affluent regions, while impoverished regions see little testing, as people do not have the resources to seek medical help. In these circumstances, the number of cases and fatalities may go unreported, which explains the high variance in the tests per case ratio across the continent. For all of these reasons, it can be assumed that it is highly unlikely that Africa has been less affected by COVID-19 than other places.

‘Incompleteness’ is explored as a concept by Francis Nyamnjoh, who explains the world’s predicament as ‘incomplete’, while exploring the idea that COVID-19 invites the recognition of digital technology as a conduit to complement human capacity [8]. Digital transitions require new paradigms “to bridge the gaps between technological and human elements in digital service innovation” [9], and a recognition of epistemic ownership, which can also be referred to as agentic capability [10], the ability to understand the world and to act upon that understanding. The observation that data is incomplete, does not imply that something like ‘completeness’ can be attained. Incompleteness, as conceptualised by Nyamnjoh, refers to a state of understanding, bounded by ontology, that is characterised by “fluidity, compositeness of being and a capacity to be omnipresent in whole or in fragments” [8].

Incompleteness is translated as any understanding that requires a certain ordering. Key ordering principles are reflected as (logical) units of analysis. Nation states remain a key ordering principle. How the world is doing on COVID-19 is typically represented at the level of continents and countries. A comparison of leading COVID-19 trackers (WHO coronavirus dashboard [11]; John Hopkins COVID-19 dashboard [12]; The Conversation data map of Africa [13]; and Humdata COVID-19 cases data [14]) reveals that not one

of these trackers show differences within countries, however large these are. The data is based on averages of partial and biased representation of some places in nations that widely vary in size, population, demographics, and geography, undermining the intelligibility of differences. Such representations can be misleading [15]. The response measures taken to fight the pandemic reflect the presentation of data as naturally ordered within the unit of a nation state and, hence, have consisted of the closing of borders, national colour codes and mobility restrictions.

In the transcendence from digital data points to knowledge, these ordering principles represent images of the human thought process, formalised as semantic data. Digital data is a set of observations that have provenance and emerge from context, and can transcend into information, knowledge and understanding based on selective representations, modelled as ontologies, that mirror socially understood concepts. Data can only ever be relevant when accompanied by semantics that give data meaning, and the accuracy and validity of semantics is fundamental to the quality of data.

This raises the question: can we consider different ordering principles? Maeda and Nkengasong (2021) [4] found that data is based on testing and surveillance with a highly variable capacity across and within countries in Africa. The expectation is that data badly reflects the situation in each locality and that accurate estimates based on means would require representative sampling in population studies, which are generally lacking [4]. Studies that analyse data within countries show the variations obtained with a higher degree of data granularity. John Hopkins publishes a map of data within the United States [12], which shows that COVID-19 has affected domestic populations differently, with populations in vulnerable social economic situations being disproportionately affected [16]. In addition, Sze et al. (2020) [17] found that vulnerability to COVID-19 infection is not the same in different geographic locations and for different populations in the United States. The finding that populations in poverty are disproportionately affected begs the question of how low-income communities within low-income countries are affected by COVID-19 and how we can know this [18].

To complement the data on COVID-19 with a greater heterogeneity in terms of locations and communities could be a way of assisting diverse communities to deal with the pandemic. A higher degree of data diversity can also be regarded as in the global interest, as a way of getting a grip on any new SARS-CoV-2 variants of concern that may emerge, particularly in areas that are now not represented in the data or from which provenance is not detectable [19]. But, in order to do so, we need a different unit of analysis, for example, as a starter, the health facility where health data is produced and care is provided.

In summary, the incompleteness or absence of reliable and representative data from Africa is relevant for two main reasons: (i) it dampens the urgency to address healthcare gaps and needs in Africa related to the pandemic, and (ii) it exposes Africa and the world to SARS-CoV-2 variants of concerns. This may create a vicious cycle that will maintain the conditions under which the pandemic was able to spread in the first place and delay the transition to a post-COVID-19 era.

2. VIRUS OUTBREAK DATA NETWORK-AFRICA

This Special Issue describes the collaborative work of the Virus Outbreak Data Network (VODAN) Africa, which was established soon after the COVID-19 pandemic was announced. The research reported covers the period from March to December 2020, which is the first phase of the VODAN-Africa research programme. The objective of this phase was to explore the cause of structurally missing data from Africa, and identify and inventory alternative data-handling strategies that could overcome the structural problems. The hypothesis is that data is missing due to a global sampling bias, not due to stochastic processes.

The second research phase ran from January 2021 to March 2022. During this phase (which is ongoing the time of writing), the network deployed a minimum viable product (MVP) of patient health records produced in health facilities. The objective of the second phase is to bring the results into actionable strategies, software and deployment. This phase will be reported in a future publication.

VODAN was established under the GO FAIR Foundation, as a collaboration on the curation of data relevant to the COVID-19 epidemic as Findable, Accessible (under well-defined conditions), Interoperable and Reusable—or FAIR [20]. VODAN-Africa was developed as part of the GO FAIR Implementation Network Africa, which primarily investigates culturally sensitive approaches to digital data handling in Africa. VODAN-Africa is a growing network of researchers from over 10 African countries, as well as from other continents. These researchers have expertise in different disciplines related to health and medicine, data science and computer science, computer engineering and social sciences. The network is hosted by Kampala International University in Uganda.

As the COVID-19 lockdowns forced most researchers to work from home, the network focused on the use of ICT for collaboration. From the start of the network, the multidisciplinary group met online every week on Zoom, which had the best reach in many of the low-bandwidth areas in which the members of the network are located, allowing people in different geographies to collaborate. To increase quality in low-bandwidth areas, the one-hour meetings were held in audio without video. Every week the team of some 30 researchers took stock of the progress made and discussed the way forward. The advent of new digital communication platforms and the expansion of collaborative technologies proved to be an important asset for international collaboration with researchers in remote locales.

The initial group of universities represented by VODAN-Africa included 11 universities and 2 centres of expertise in 9 countries: Kampala International University (Uganda); Great Zimbabwe University and Solidarmed (Zimbabwe); University de Sousse (Tunisia); Mekelle University and Addis Ababa University (Ethiopia); Olabisi Onabanjo University, Ibrahim Badamasi Babangida University and Data Science Nigeria (Nigeria); University of Liberia (Liberia); Tangaza University (Kenya); Kilimanjaro Christian Medical College (Tanzania); and East Africa University (Somalia). The team also collaborated with Leiden University and Tilburg University in the Netherlands, the GO FAIR Foundation (a centre of research expertise), Stanford University, the Centre for Super Computing in San Diego, and the Chinese Academy of Sciences. The primary question that the research groups focused on was the possibility of increasing the capacity to generate novel data within Africa, in order to strengthen the understanding of the global spread of infections of COVID-19, while simultaneously strengthening local health systems in a sustainable manner.

3. METHODOLOGY

The research employed an ethnographic approach, following an action research design. The investigators sought to integrate stakeholders—such as the administrators of health facilities, data processors in facilities, doctors and physicians, policymakers in the various ministries of health and political representatives—in their regular consultations and exchanges. An ethnographic design is a way of exploring a problem by engaging relevant stakeholders in a particular problem and is especially relevant when dealing with what is known as a ‘wicked design problem’—a problem that requires structuring and definition [21]. Such a problem may have different solutions, depending on the pathway chosen to explore the problem in the first place. An important determinant, therefore, is who is consulted in the process of exploring the problem. The choice made in this research was to consult a wide array of practitioners. The consultations included conversations and meetings, training sessions, participation in testing and participation in the meetings of the research group. Stakeholders related closely to the collaborative effort of defining the problem, structuring it and exploring solutions. Action research allows investigators to carefully document a design process-in-the-making.

As noted earlier, the research presented in this Special Issue documents the work of the first phase of the research performed by VODAN-Africa. The objective was to explore the ‘fuzzy front’ end of this ‘wicked problem’ [22]. The exploration included the following objectives: (i) to understand the problem of the incompleteness of global health data on COVID-19 and its causes, especially in relation to Africa; (ii) to understand the related context of this problem; (iii) to explore the potential of designing solutions for more complete data by increasing data interoperability using the FAIR Guidelines; and (iv) to explore avenues for further research. The third question, on the potential of a solution using the FAIR Guidelines, splits into two very different sub-questions, namely: (i) the potential of using the FAIR Guidelines given the regulatory context and political policy priorities in Africa, and (ii) the technical feasibility of the application of the FAIR Guidelines.

4. FAIR GUIDELINES IN DATA PROTECTION THROUGH PUBLIC POLICY

4.1 Public Agenda

In this research, the framework of Kingdon is used to study the potential for the adoption of the FAIR Guidelines to improve the handling of sensitive health data in Africa. Kingdon (1984) describes the public agenda as emerging from what he calls a ‘primeval soup’ of ideas, generated by a “community of specialists” [23], which can be researchers and academics, officials, or interest group analysts, from which alternative ideas and proposals for the public agenda may emerge. He refers to ‘policy communities’ as “specialists in a given policy area” [23]. Kingdon distinguishes three streams that need to be aligned for any alternative to reach the policy agenda: the problem, policy and political streams. In the problem stream, a problem is identified, framed and understood in a certain way; the policy stream identifies the alternative solutions available to handle the problem, and the political stream is associated with the priority given to the matter at hand. In addition, Kingdon identifies what he refers to as the ‘policy entrepreneur’—a person or a group who has a special interest in a particular policy domain or issue and inputs time, energy and other resources in exchange for a potential impact on the direction of the public agenda [23].

4.2 The Primeval Soup of Digital Data

The concept of the FAIR Guidelines is concerned with the metaphorical primeval soup in relation to digital data objects. The concern of FAIR data is that the lack of findability, machine-readability and semantic meaning of digital data renders data meaningless—mere ‘floating’ digital objects that can only regain meaning with concentrated efforts to shed light on them by connecting them with other digital objects through which semantic meaning can be derived. The FAIR Guidelines were a response to the exponential growth of digital data and the observation that the World Wide Web (www) misses a machine-actionable equivalent necessary to create access, as well as give meaning, to digital data objects. The FAIR Guidelines are a way of bringing life to the ‘primeval soup’ of unconnected digital data objects, which are like free atoms that have not yet found their place in living, well-connected objects that are part of a meaningful ecosystem. According to the theory of FAIR data, the problem can best be solved by the creation of linked machine-actionable metadata, and the fruits of this strategy will be maximised if it can connect digital data repositied in multiple repositories, available in various locations. Hence, the FAIR Guidelines are based on the expectation that meaningful knowledge can be produced by linking digital data available across different locations, which are bridged through the Internet by elastic and federated virtual computational techniques.

4.3 Data Provenance, Residence and Localisation

The concept of structured data originating from an inherently chaotic and contextual process links well with the idea that digital data is always produced in a localised surrounding. The definition of this surrounding is the product of social categorisation. The ownership and control of data relates to the idea that digital data is produced in a specific place, relates to specific data subjects and that a degree of ownership and control over this data follows from its provenance. Following Mawere and Van Stam [24], ownership and control of data in Africa is defined at national and sub-national levels through the regulatory frameworks in each context. However, ownership and control of data in African countries is limited by the extractive practices of monopolistic platforms, which move data away from the place where it is produced [24], creating value elsewhere and resulting in the loss of the value of this data to the place where the data was produced [19, 25].

The ownership and control over data produced within the European Union (EU), or in relation to it, is a critical notion that informs the provisions of Europe’s General Data Protection Regulation (GDPR) [26]. The GDPR informs a new ethic of ownership and control of data based on the concept of the ‘citizenship of data’, which is rooted in an understanding of where data is coming from (provenance) and where and for what it is being used (meaning). The collection of data must be proportional to its intended use. The legal frameworks consider that data can be both ‘in rest’ or ‘in motion’ [27]. The GDPR and other privacy laws, such as, for instance, the laws of California, consider rights and obligations in relation to the disclosure and use of personal data. Data residency refers to the geographical location where data can be stored, based on the regulatory framework that applies. Data localisation laws refer to the regulations that specify how data can be collected, processed, stored or transferred within a country [28]. The recognition in recent legislation of the relevance of ownership and control of data has consequences for data handling in Africa,

specifically in relation to the ownership and control of the data, where the data should or could reside, and how we can optimise practices that recognise the provenance of the data. According to the African Union Digital Transformation Strategy for Africa (2020–2030), this recognition is likely to generate respect for the ownership and control of data “even though Africa is at the moment less restrictive, soon it will be necessary to ensure localization of all personal data of Africa’s citizens” [29]. The Alliance for Accelerating Excellence in Africa has also underscored the importance of the protection of data provenance [25].

4.4 Personal Health Data

In the EU, the ownership and control over data is greater when it concerns personal data belonging to a data subject who is a citizen of the EU and who enjoys the protection of EU laws. A higher level of protection is awarded to health data, which is classified as sensitive personal data (Article 53, preamble of the GDPR) [26]. The GDPR sets clear restrictions on the processing of digital data for health-related purposes, which is to be processed “only where necessary to achieve those purposes for the benefit of natural persons and society as a whole” [26]. The purpose of such processing is identified as in:

...the context of the management of health or social care services and systems, including processing by the management and central national health authorities of such data for the purpose of quality control, management information and the general national and local supervision of the health or social care system, and ensuring continuity of health or social care and cross-border healthcare or health security, monitoring and alert purposes, or for archiving purposes in the public interest. [26]

The legislation builds on the notion that health systems are a central component of national social protection systems, with the responsibility of providing “safe, high quality, efficient and quantitatively adequate healthcare to citizens on their territory” [26]. The legislation explicitly identifies health data as under the jurisdiction of national legislation and a national responsibility. This has direct implications for the physical location of data. The relevance of safeguarding data and African interests, while supporting the custodians of the data, is increasing, and African legislators are starting to recognise this through the promulgation of legislation regarding the protection of personal information, including health-related sensitive personal data [25].

5. OVERVIEW OF THIS SPECIAL ISSUE

The objective of the research reported in this Special Issue was to explore the reasons why data is structurally missing from Africa and, based on the findings, explore the theory that the FAIR Guidelines might be a way forward in creating new and alternative designs with the capacity to solve some of the problems around incomplete data in Africa.

5.1 Terminology and Regulatory Frameworks

In Article 3 of this Special Issue, Plug et al. [30] present an overview of the terminology relevant to the FAIR Guidelines, which is used as the conceptual framework for this research. This article presents a

systematic set of terminology defined to offer greater conceptual clarity on the issues discussed in this Special Issue. This is followed by a presentation by Stocker et al. [31] in Article 4 on the background to the fast rise of the FAIR Guidelines on the EU policy agenda. These authors observe that the FAIR Guidelines responded to a concern recognised by EU policymakers, namely, the loss of valuable data due to lack of curation strategies. As this concern was rising, the need for personal data protection was simultaneously gaining urgency, and the FAIR Guidelines present a potential solution, driving the issue onto the policy agenda. However, Stocker et al. [31] warn that the proof is in the pudding and that, unless the FAIR Guidelines are transformed into a set of tools, they may be removed from the agenda, as an impractical proposition.

In Article 5, Lin et al. [32] investigate the adoption of the FAIR Guidelines in non-Western geographies. The authors find that FAIR is recognised in the languages investigated, although to a lesser degree than in the English literature. The findings suggest that there might be potential for the application of the FAIR Guidelines in non-Western geographies.

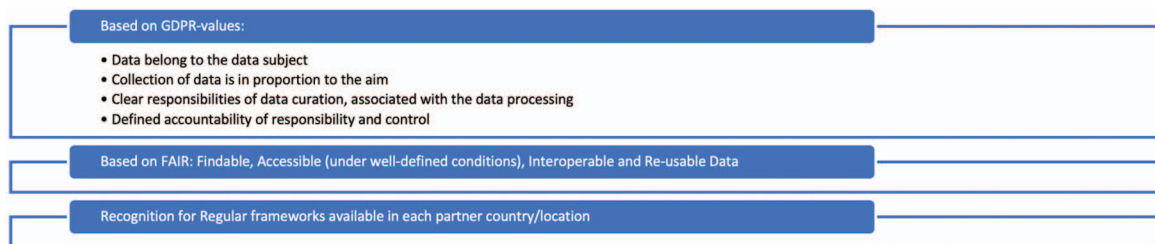


Figure 1. Legal framework for VODAN-Africa.

5.2 FAIR Equivalency

This is followed by six articles (Articles 6–11) that discuss the regulatory frameworks in six countries—Uganda [33], Indonesia [34], Ethiopia [35], Zimbabwe [36], Nigeria [37], and Kenya [38]—in relation to the application of FAIR Guidelines to the digitalisation of health data. Following this, Purnama Jati et al. [39] compare the FAIR Guidelines with open data, which is the original basis for the Satu Data policy in Indonesia (Article 11).

The researchers for this group of articles identified the EU GDPR and FAIR Guidelines as foundational frameworks for baseline data protection, in addition to the national regulatory frameworks in each participating country in Africa. A first task was to explore the extent to which the FAIR Guidelines were in line with national regulatory frameworks. The FAIR Guidelines can be understood as a spectrum—the degree of ‘FAIRness’ can range from very high to low, and the appropriate degree of FAIRness can differ from situation to situation. The facets of the FAIR Guidelines are [40]:

To be 'Findable':

- F1: (meta)data are assigned a globally unique and persistent identifier
- F2: data are described with rich metadata (defined by R1 below)
- F3: metadata clearly and explicitly include the identifier of the data it describes
- F4: (meta)data are registered or indexed in a searchable resource

To be 'Accessible':

- A1: (meta)data are retrievable by their identifier using a standardised communications protocol
 - A1.1: the protocol is open, free, and universally implementable
 - A1.2: the protocol allows for an authentication and authorisation procedure, where necessary
- A2: metadata are accessible, even when the data are no longer available

To be 'Interoperable':

- I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2: (meta)data use vocabularies that follow FAIR Principles
- I3: (meta)data include qualified references to other (meta)data

To be 'Reusable':

- R1: (meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1: (meta)data are released with a clear and accessible data usage licence
 - R1.2: (meta)data are associated with detailed provenance
 - R1.3: (meta)data meet domain-relevant community standards

In Article 6, Basajja et al. [33] offers a methodology to measure the degree of alignment with the FAIR Guidelines, called FAIR Equivalency, which is replicated in articles 7 to 11 [34–38]. FAIR Equivalency was determined by investigating the documents constituting the regulatory framework of the country to see how closely they are aligned with the FAIR Guidelines, as defined in the 15 facets of FAIR. The mention of the 15 FAIR facets (or equivalent terminology) in each of the policy documents was analysed by assigning codes (to the text) and labels to the appropriate FAIR facet (i.e., F1, F2, F3, F4; A1, A1.1, A1.2, A2; I1, I2, I3; R1, R1.1, R1.2, R1.3). The mention of an equivalent notion of the FAIR facet in a policy document was assigned the measure '1', and the absence of equivalency was assigned '0'. A FAIR Equivalency Score (FE-Score) was calculated as the sum of scores on all 15 facets for all policy documents. This methodology was carried out by researchers in Uganda [33], Ethiopia [35], Zimbabwe [36], Nigeria [37] and Kenya [38], as well as Indonesia [34], and is also included in an overview of comparative results presented in Table 1.

The analysis shows a high or very high degree of FAIR Equivalency in the countries studied. The percentage ranged from 70% (in Nigeria) to 100% (in three countries: Ethiopia, Kenya, Zimbabwe). Indonesia had a score of 75% and Uganda of 83%. A score of 100% means that the researchers found FAIR-like principles in all of the documents checked. A score of 100% does not mean that all of the FAIR facets are recognised in all of the policy documents, but it means that all of documents mention at least one FAIR Equivalent.

This means that in the policy documents, the FAIR Guidelines are recognised as relevant, at least to a certain degree. Given the positive score of FAIR Equivalency, it was concluded that the FAIR Guidelines did not contradict the existing national policy frameworks and that it was worthwhile to explore FAIR as a potential framework for increasing COVID-related health data from Africa.

Table 1. FAIR Equivalency comparison.

| Variable | Uganda | Indonesia | Ethiopia | Zimbabwe | Nigeria | Kenya | Average of percentage (row) |
|---------------------------------------|------------------------|---------------------|------------------------|------------------------|-----------------------|------------------------|---------------------------------|
| Total documents | 12 | 8 | 11 | 8 | 10 | 15 | - |
| Fair Equivalent documents | 10 (83%) | 6 (75%) | 11 (100%) | 8 (100%) | 7 (70%) | 15 (100%) | 88% |
| F1 | 5 (41%) | 1 (12%) | 4 (36%) | 1 (12%) | 7 (70%) | 9 (60%) | 38% |
| F2 | 3 (25%) | 1 (12%) | 1 (9%) | 4 (50%) | 4 (40%) | 5 (33%) | 28.17% |
| F3 | 0 | 0 | 1 (9%) | 1 (12%) | 6 (60%) | 0 | 13.5% |
| F4 | 4 (30%) | 2 (25%) | 1 (9%) | 5 (62%) | 0 | 8 (53%) | 29.83% |
| A1 | 9 (75%) | 3 (37%) | 9 (81%) | 6 (75%) | 7 (70%) | 5 (33%) | 61.83% |
| A1.1 | 6 (50%) | 2 (25%) | 2 (18%) | 3 (37%) | 6 (60%) | 9 (60%) | 41.67% |
| A1.2 | 7 (58%) | 4 (50%) | 3 (27%) | 6 (75%) | 7 (70%) | 8 (53%) | 55.5% |
| A2 | 0 | 0 | 0 | 0 | 0 | 8 (53%) | 8.83% |
| I1 | 8 (67%) | 5 (62%) | 9 (81%) | 4 (50%) | 5 (50%) | 2 (13%) | 48.67% |
| I2 | 1 (8%) | 0 | 0 | 3 (37%) | 1 (10%) | 10 (67%) | 26.5% |
| I3 | 4 (30%) | 0 | 2 (18%) | 3 (37%) | 1 (10%) | 5 (33%) | 24.5% |
| R1 | 4 (30%) | 1 (12%) | 8 (72%) | 2 (25%) | 7 (70%) | 5 (33%) | 32.5% |
| R1.1 | 4 (30%) | 1 (12%) | 2 (18%) | 7 (87%) | 3 (30%) | 6 (40%) | 47.67% |
| R1.2 | 4 (30%) | 0 | 5 (45%) | 3 (37%) | 5 (50%) | 4 (40%) | 32.33% |
| R1.3 | 6 (50%) | 3 (37%) | 10 (90%) | 3 (37%) | 6 (60%) | 3 (20%) | 40.17% |
| 'Findable' principle in document | Average: 24% | Average: 12.25% | Average: 15.75% | Average: 34% | Average: 42.5% | Average: 36.5% | Average for 6 countries: 31.96% |
| 'Accessible' principle in document | Average: 45.75% | Average: 28% | Average: 31.5% | Average: 46.75% | Average: 50% | Average: 49.75% | Average for 6 countries: 45.04% |
| 'Interoperable' principle in document | Average: 35% | Average: 20% | Average: 33% | Average: 41.3% | Average: 23.3% | Average: 37.67% | Average for 6 countries: 30.10% |
| 'Reusable' principle in document | Average: 35% | Average: 15.25% | Average: 56.25% | Average: 46.5% | Average: 52.5% | Average: 33.25% | Average for 6 countries: 39.17% |

Source: Created by Purnama Jati, 2021

The total number of policy documents included ranged from 8 in Indonesia and Zimbabwe to 15 in Nigeria. Of the four components of the FAIR Guidelines, the requirement that data be 'Accessible' had the highest mention (in equivalent terms), while the requirement that it be 'Interoperable' had the lowest mention. Facet A1 of accessibility ([meta-]data are retrievable by their identifier using a standardised

communications protocol) had the highest mention. A very low score was obtained for the facets A2 ([meta-]data are accessible, even when the data are no longer available) and F3 ([meta]data clearly and explicitly include the identifier of the data it describes).

Kenya, Uganda, Zimbabwe and Indonesia scored the highest for data being 'Accessible'. Ethiopia and Nigeria scored highest for it being 'Reusable'. 'Findable' had the lowest equivalency score in Ethiopia, Uganda, Zimbabwe and Indonesia. In Indonesia, 'Accessible' received the greatest attention, however, of the facets, it was facet I1 ([meta-]data use a formal, accessible, shared, and broadly applicable language for knowledge representation) that was most frequently referred to. Zimbabwe also highlighted accessibility the most often, but of all the facets R1.1 ([meta]data are released with a clear and accessible data usage licence) was mentioned the most frequently. Kenya also had the highest score for the requirement that data be 'Accessible', but the facet with the highest mention was the equivalent of I2 ([meta]data use vocabularies that follow FAIR Principles). The highest score for each of the guidelines among the six countries was Nigeria for data being 'Findable', Kenya for 'Accessible', Zimbabwe for 'Interoperable' and Ethiopia for 'Reusable'. It was concluded that the FAIR Guidelines are aligned with the national policy documents of the respective governments and regulatory bodies, which constitute the regulatory frameworks of the six countries studied [33–38].

5.3 Exploring the Information Flows of Digital Health Data: The Problem of Data Extraction

In the subsequent section of the Special Issue, Basajja and Nambobi [41] offer an analysis of information streams in selected health facilities in Uganda. This article demonstrates that data are produced in health facilities without the data being used for care practices within the facility. The data that is produced is sent away to the concerned ministry of health for policy purposes and further afield to platforms outside the African continent. Data does not return to improve practices at the original point of care. This article concludes that, unless data is analysed and visualised within the health facilities to support the quality of care, the production of data is meaningless. The practice of tying in digital data for reasons that do not benefit the facility where it is produced is diagnosed as a critical problem in health data architecture today.

Basajja et al. [42] measure the interoperability of digital health solutions in Uganda by examining existing digital health solutions in Uganda, namely, the Digital Health Atlas Uganda (DHA-U) and Uganda Digital Health Dashboard (UDHD), using the FAIR Evaluation Services tool. This study concludes that FAIR maturity is low in digital health solutions in Uganda. The movement of data through vertical upward streams of competing solutions to platforms generally outside the country and the continent is not conducive to horizontal data integration at the policy level, or at the point of care.

The articles by Basajja and Nambobi [41] and Basajja et al. [42] both point to the problem of health data extraction. They note that data that is produced do not always serve purposes at the point of care. The extraction of data can be regarded as the removal of 'value' from Africa to solely benefit places outside the continent. It is concluded that innovation aimed at expanding medical health data from Africa should restore the local ownership and use of data.

5.4 Proof of Concept of a FAIR-Based Health Data Architecture for COVID-19

The next section of this Special Issue explores the design of an alternative data architecture for COVID-19 based on FAIR Guidelines. Basajja et al. [43] presents the proof of concept of the first test carried out in September 2020 to place FAIR Data Points in selected African locations and curate and reposit data locally. The data produced in machine actionable format were based on the WHO electronic COVID Report Form (eCRF) and reposted in FAIR Data Points discoverable over the Internet. The proof of concept constituted a test to send queries over the Internet in the form of an algorithm and compute findings by visiting the data reposted in local containers within the places where the data was produced. The proof of concept was carried out by Basajja in September 2020, with a data visiting experiment of FAIR Data Points placed in two continents, one at Leiden University Medical Centre in Leiden and another one at Kampala International University in Uganda.

The proof of concept of data visiting, which was carried out with the software produced by DSWizard and installed in selected sites, was technically successful and showed the possibility of retaining data ownership with data repositing of machine actionable data in locale, while adding the possibility of querying the data across different sites, countries and continents with approved algorithms. The experiment also revealed a range of issues that would need to be addressed in any further iteration of a solution based on the FAIR Guidelines, as was intended. These included, among other issues the following:

- The need for data production that was flexible and could be adapted to the input carried out at point of care in health facilities.
- Interest on the part of health facilities to understand and analyse the data they produce for improved care, and the focus on data most relevant to point of care operation (such as outpatient data records).
- The importance of including local engineers to incorporate sensitivity to the local engineering realities, challenges and possibilities.
- The realisation that the FAIR-based data infrastructure requires new skills in terms of data science and data curation and, hence, the need to invest in capacity building.
- The observation that co-creation with interdisciplinary experts and stakeholders connected across Africa, and in collaboration with partners in Asia, Europe and the United States, proved to be a fast-track to innovation.

Purnama Jati et al. [44] describes the considerations based on the proof of concept for further development of the data infrastructure regarding the critical question of access and control to the data. Ghardallou et al. [45] discuss a study to create FAIR data objects for data obtained through scientific data collection and to make this data machine actionable and interoperable with the patient data obtained in VODAN-Africa health facilities. This article focuses on data on COVID infections obtained from refugees and migrants, who usually do not have access to medical facilities and testing. These are vulnerable populations, in which infection may remain undetected and which have no structural access to treatment or vaccines. This article describes how the FAIR curation of data can make these interoperable with patient data. This will increase the representativity of vulnerable social groups in the data.

In the next article, Folorunso et al. [46] discuss a workflow in which analytical algorithmic tools can be used on FAIR data. The study shows an in-country geographic analysis of COVID-19 infections in Nigeria, a country with a population of over 200 million, and attempts to clarify these variations. The study shows that an analysis of the data can reveal large differences in infection rates within countries, enabling better targeted policy measures.

5.5 Capacity Building

In the following two articles, Oladipo et al. [47] and Akindele et al.) [48] discuss efforts to enhance the capacity of data curation, following FAIR data. Oladipo et al. [47] look at a new curriculum specifically set up to grasp FAIR data curation as a new area of instruction in computer science. Akindele et al. [48] examine the potential for reaching students in Africa through distance learning, using digital tools, precipitated by the COVID-19 lockdowns.

6. DISCUSSION: SPECIFICATIONS AND REQUIREMENTS FOR FURTHER IMPLEMENTATION

This Special Issue documents the first phase of the implementation of VODAN-Africa. The following important conclusions were reached during this phase. The study of the regulatory framework for (digital) health in five countries in Africa (Ethiopia, Kenya, Nigeria, Uganda, Zimbabwe), as well as Indonesia, showed an interest in the values associated with the FAIR Guidelines. The greatest interest shown was in relation to the accessibility and interoperability of data. It was further found that data extraction from Africa was considered a key problem to be addressed, specifically the non-use of data at the point of care where the data extracted was generated (health facilities). The studies revealed an urgent need to redirect efforts for data use to improve the quality of care. In addition, it was found that the parallel efforts of digital health solutions were focused on data analytics within narrowly defined objectives, which lacked attention to horizontal data integration within and across facilities.

Based on these preliminary findings, it was hypothesised that the FAIR Guidelines could be used as a conceptual framework for a new approach to health data curation. The approach follows the workflow of FAIR data proposed by Jacobsen et al. [49], who proposed the 'FAIRification' of both data and semantic data, divided into three phases: pre-FAIRification, FAIRification, and post-FAIRification (see Figure 2).

The efforts reported in this Special Issue are concentrated on phase 1 (pre-FAIRification) and phase 2 (FAIRification), with a proof of concept implemented as a data visiting effort in the post-FAIRification phase.

During the FAIRification phase, the team made use of the DSWizard. The selection of this tool was based on the requirement that data should be curated in the locale in which it was produced, and hosted with an exposure to the Internet so as to be findable and reachable for algorithmic queries. The data production was based on the WHO eCRF, a standard document to record patient data related to COVID-19. While the proof of concept was successful, the experiment showed that a review and adaptation was needed of the specifications and requirements of any software to be use or produced. The adaptations to specifications and requirements were identified as following:

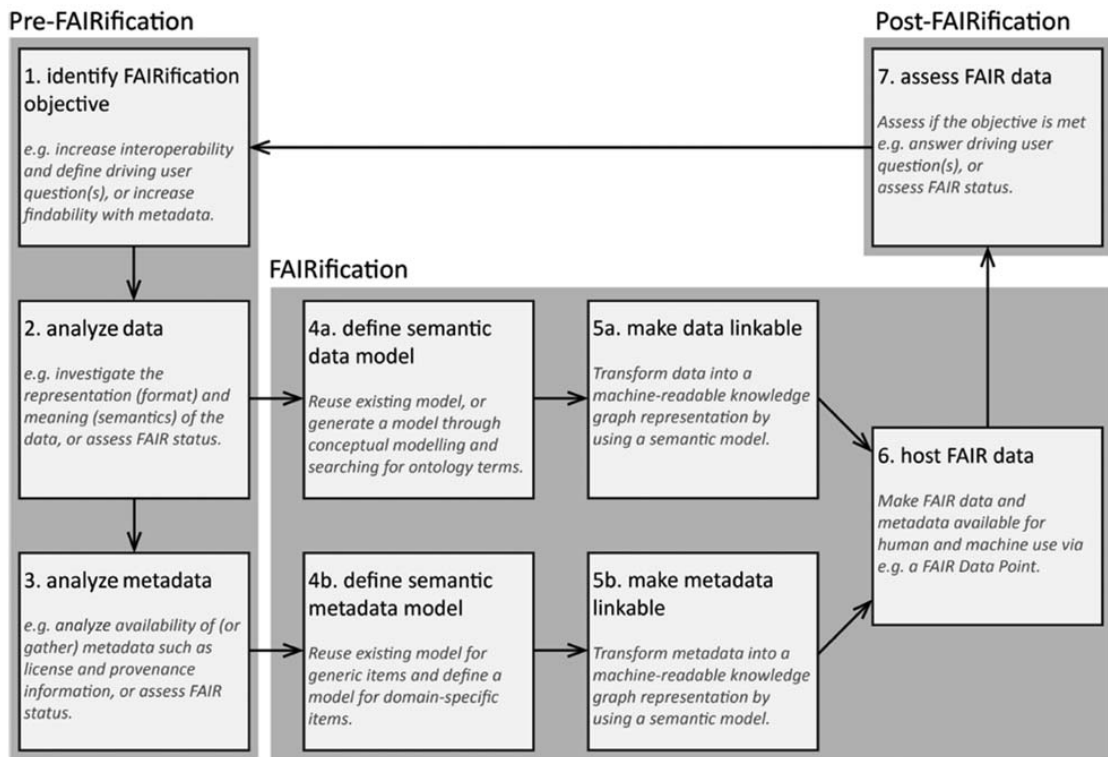


Figure 2. FAIRification workflow model [49].

- A more direct link with actual workflows within the health facilities is needed.
- The workflows have to be adaptable to variations in the workflows of different health facilities.
- The VODAN-Africa procedure needs to be fully compatible with existing workflows in health facilities to avoid the duplication of data input.
- A flexible production of machine-actionable templates of patient data and scientific health data needs to be available to adapt to actual work practices and to replace the inflexible standard of using just one WHO eCRF template, which most health facilities did not use.
- The localisation of data repositing and ownership by the facility over the data handling process is a fundamental requirement for the success of the design.
- Data visualisation within the health facilities was identified as a critical need for the data to be able to inform health practices at the point of care (where the data is generated).
- Software tools need to be adapted to African realities and engineers in the locale should be integrated to strengthen co-creation innovation practices.
- The capacity building of FAIR data curation and FAIRification processes, including the use and adaptation of available software, is a critical need to ensure future sustainability and expansion.

| | |
|-------------------|---|
| Ensure | Ensure health data produced in a facility add value to the improved care of patients in that facility. |
| Promote | Ensure data ownership in the place where data is produced. |
| Contribute | Contribute to better healthcare in vulnerable communities. |
| Create | Create sustainable solutions that can be managed by the existing ecosystems (including capacity building systems). |
| Nurture | Nurture strong female and diverse leadership in IT, innovation and health. |
| Strengthen | Strengthen co-creation with technical communities in the places of collaboration. |
| Enhance | Strengthen stakeholder involvement and build communities of purpose. |

Figure 3. Core values of VODAN-Africa.

7. LOOKING AHEAD TO PHASE 2: FURTHER DEVELOPMENT OF A VODAN-AFRICA ARCHITECTURE

The assessment of these requirements in Phase 1 led the VODAN-Africa team to develop a new architecture for the second phase of the VODAN-Africa implementation [19, 49]. This architecture includes a one-stop data entry, facilitated by an editor through templates that generate machine-actionable language (RDF or JSON), with data repositied in containers within the health facility, with an automated output to the health information system in use by the health facility, and an output to the facility dashboard where data is visualised for use by the medical team in the facility. The dashboard also includes the visualisation of generic data obtained through data visiting, computed over all VODAN-Africa facilities, to provide additional context and value to these data. The architecture enables interoperability with other data, including data obtained through scientific efforts and curated to be interoperable with the VODAN-Africa facilities.

The research team identified that this architecture could be realised if it could develop a localised version of the software of the Center for Expanded Data Annotation and Retrieval (CEDAR). Based at Stanford University, CEDAR was established in 2014 to create a computational ecosystem for the development, evaluation, use, and refinement of biomedical metadata. In the second phase, the VODAN research team established a collaboration for co-creation with the CEDAR team to design a localised version that could be installed within health facilities in Africa.

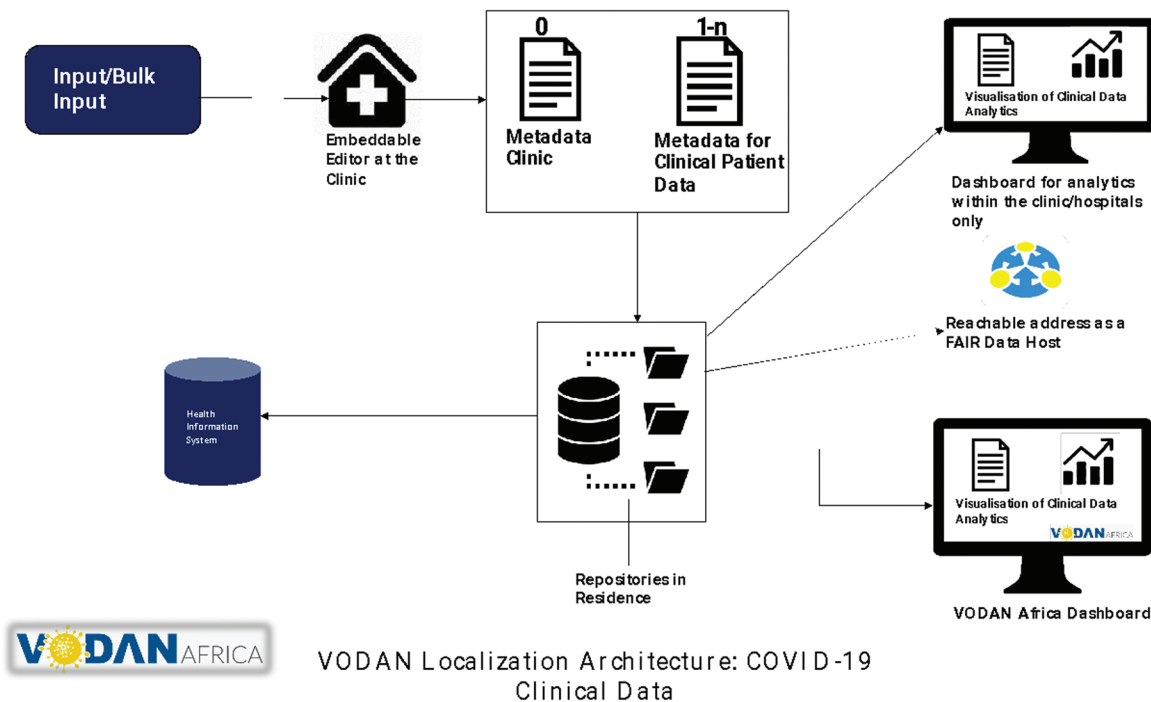


Figure 4. VODAN-Africa localisation architecture for patient health records [19].

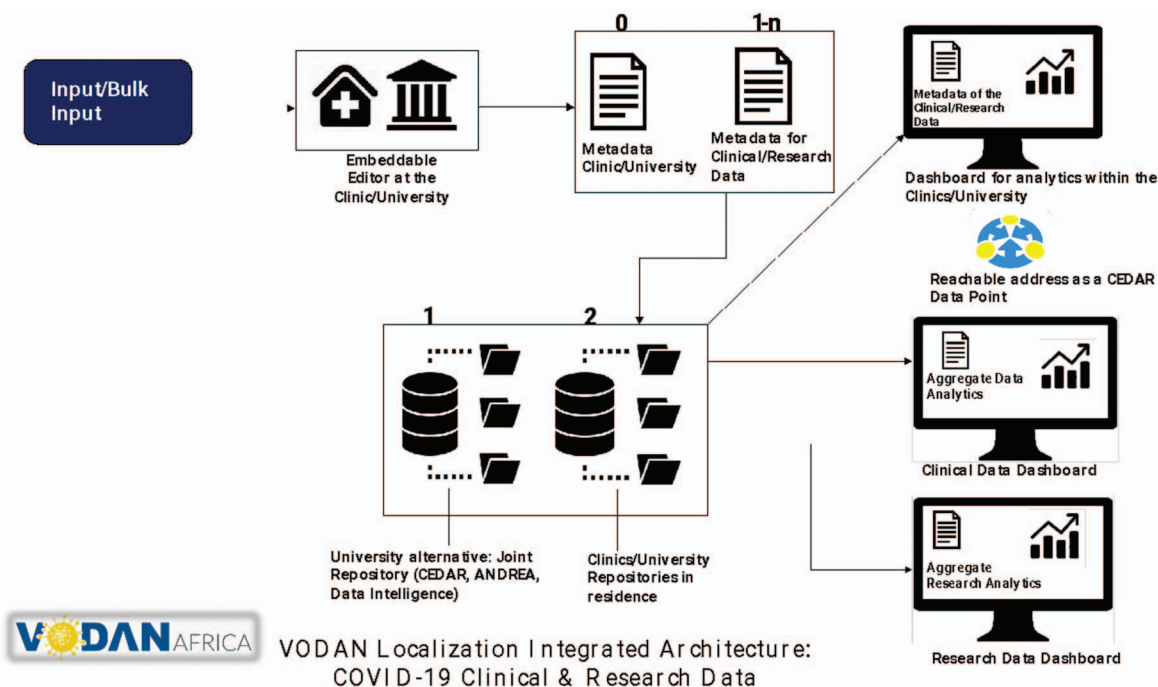


Figure 5. VODAN-Africa localisation architecture for patient health records combined with scientific data [19].

The basic infrastructure of the CEDAR-based FAIR Data Point was the creation of a visualisation on the Dashboard based on requests sent through the interface to the back-end service of metadata repositored within the facility, creating a front end visualisation. The FAIR Data Point does not need to be visible publicly on the Internet, but can be installed as a point reachable through CEDAR, through which data in location can be reached (with a prerequisite of access being granted).

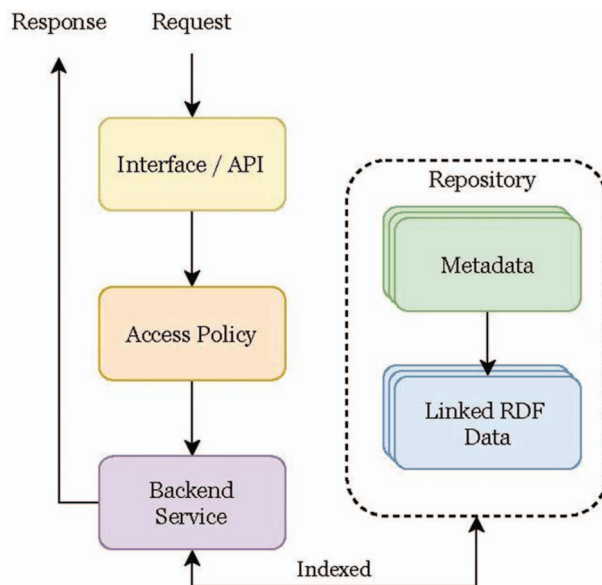


Figure 6. CEDAR as a FAIR Data Point [50].

The overarching architecture is based on the various components of the CEDAR software, but redesigned to follow localisation requirements. Figure 6 shows the architecture in which CEDAR is engineered as a FAIR Data Point, and federated with the software for localised machine actionable data production, reachable over the Internet in a closed CEDAR community. This is the basic architecture for realisation of localised deployment in phase 2. This is further expandable with a dynamic search function, which is architecturally conceptualised in Figure 7.

The system generates data at different levels of sensitivity for different purposes. The personal data are protected, encrypted and repositored within the health facility. The processed data and aggregate data can be visited for computational purposes. This leads to the following architecture to support the visualisation of data at the point of care (in the health facility) computed from data produced and repositored in the health facility (internal dashboard) as well as data aggregated through data visiting from health facilities in VODAN-Africa Network (External Dashboard).

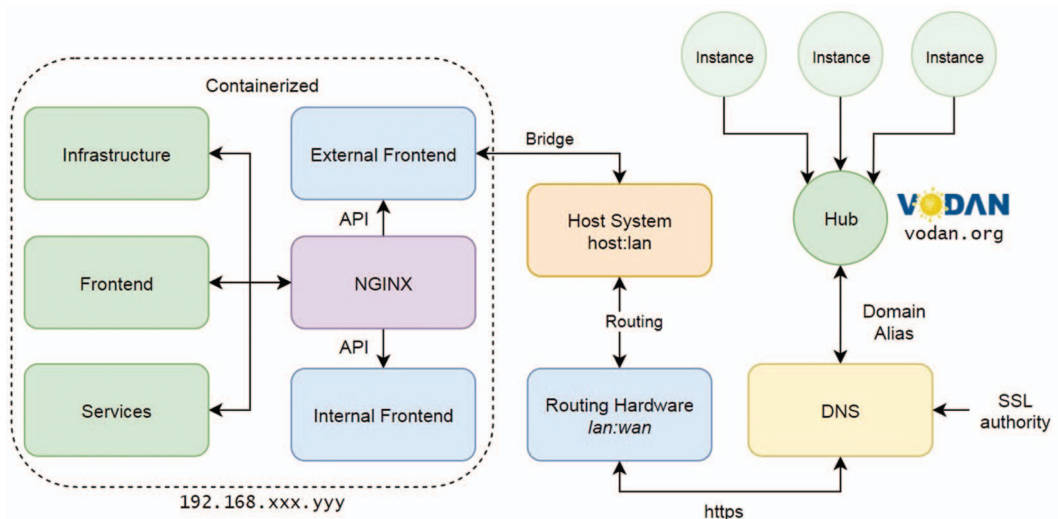


Figure 7. Conceptual architecture for data analysis and visualisation of locally produced and reposit machine-actionable data [51].



VODAN-AFRICA INTERNAL & AGGREGATE DASHBOARD & HIS INTEGRATION

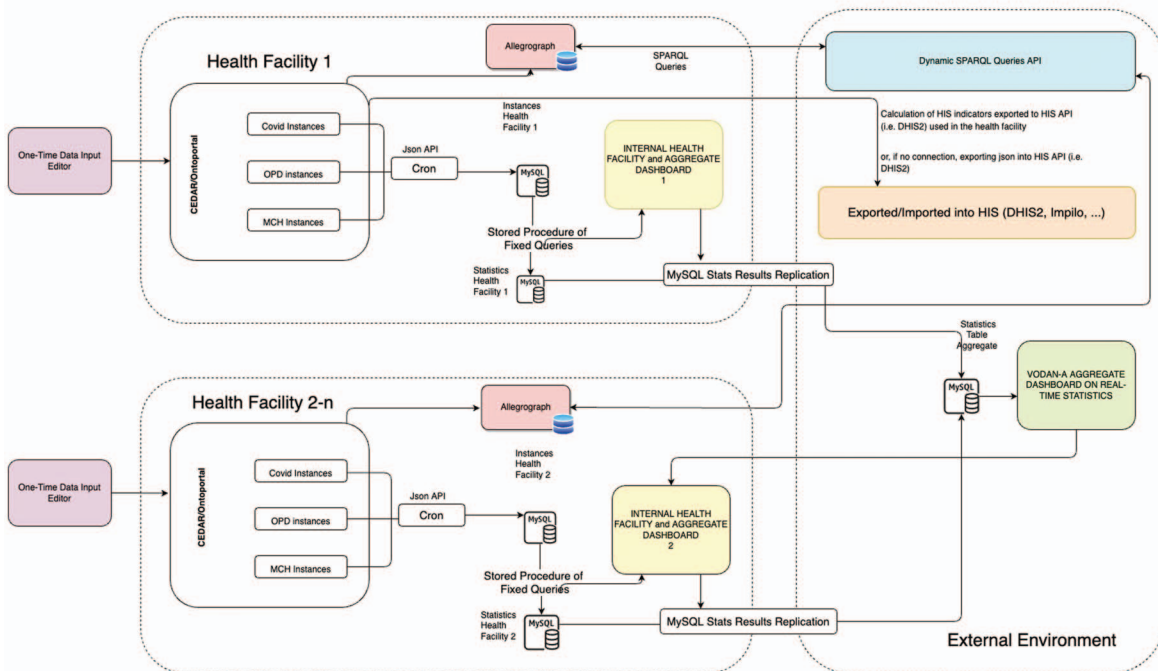


Figure 8. Conceptual architecture for data analysis and visualisation of locally produced and reposit machine-actionable data.

The design of this data infrastructure, as envisaged, will aim to simultaneously serve the purpose of: (i) informing decisions at the point of care based on the data produced in the facility, and (ii) increasing the data available for analysis, including at the global level, without compromising the ownership of the data. While all personal data remain within the facility, the aggregate data is obtained by federated computations through data visiting and can be shared in aggregate form as this data is depersonalised and de-identified.

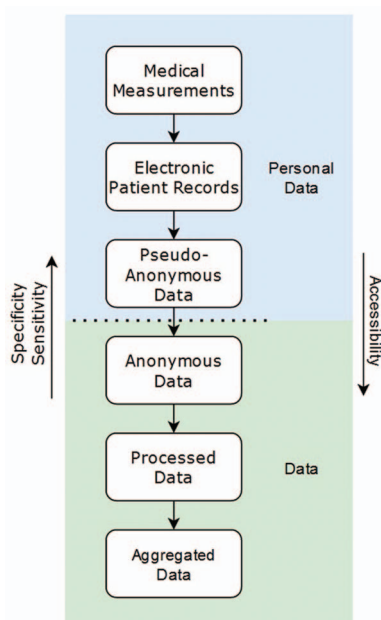


Figure 9. Accessibility spectrum of processed personal data [50].

The second phase is being rolled out with the participation of Leiden University and over 80 health facilities in 8 countries in Africa.

For future design the architecture will make better use of the machine-actionable semantic enrichment of the data, by storing of the data in a triple store and to install through docker to maximise flexibility of operation and allow dynamic queries, all within the agreed access and permission control associated with regulatory frameworks in each place.

8. CONCLUSIONS

The analysis of the evolution of the COVID-19 pandemic in Africa suffers from incomplete data. This negatively impacts on the capacity to adequately respond to COVID-19 at the point of care. Data partitioned as averages per country, based on limited samples, hide the variations across countries and communities. Few population studies within Africa have been carried out with a representative research design or sampling methodology that covers the entire continent.

In this Special Issue, the research network VODAN-Africa investigated the reasons why the availability, variety and veracity of COVID-19 data from Africa are inadequate. The lack of ownership of the data produced is reflected in the low degree of relevance of such data to the users. This raises the question: Would it be possible to enhance data collection in Africa by using the FAIR Guidelines for data production in health facilities? The studies in this Special Issue found that the FAIR Guidelines are aligned with policy directions in all of the countries that were studied, which is confirmed by the FAIR-Equivalency of their regulatory frameworks.

In current research paradigms, digital health data is extracted from health facilities without rendering any sustainable use of the data by the facilities for improvement of quality of care at the point of care. In addition, digital health applications suffer from a vertical architecture that does not prioritise the horizontal integration of data inside health facilities. To make data relevant, information flows in health facilities need to be redirected to make use of the facility central to the purposing of such data. To test this, the team at VODAN-Africa conducted a data visiting experiment of data produced as machine actionable, curated and repositied in location and visited with computational queries over the Internet. With this, the proof of concept of data visiting was successfully established.

The studies in this Special Issue looked at whether or not a data architecture based on Findable, Accessible, Interoperable and Reusable (FAIR) data would provide opportunities for horizontal research data integration, linking together data from multiple studies into a sustainable data ecosystem. A FAIR Equivalency analysis of regulatory frameworks showed a positive inclination to embrace the FAIR Guidelines in the studied African countries, revealing the potential for new data governance paradigms that enable more equitable international research cooperation within the health domain. Hence, it can be concluded that the FAIR concept could help envision a design that would protect data ownership and enable data use within the health facilities, while simultaneously allowing data analysis across facilities through data visiting. A successful proof of concept of data visiting across two continents, Africa and Europe, was carried out. The assessment of the first research phase suggests specifications and requirements for further development. These include the co-creation of the inclusive innovation process with designers based in Africa, more flexibility to adapt ontology-based metadata specifications according to the research design to various workflows in different health facilities in Africa and the capacity building to curate data as machine actionable semantic objects.

In a further assessment, it was found that adaptations were necessary to construct an architecture that would be useable in health facilities with adaptability to the operational variations in the clinics. A second phase has started based on the conclusions drawn from the first phase. This second phase is based on a collaboration between VODAN-Africa, Leiden University and Stanford University to develop a localised version of the software developed by CEDAR.

Compliance with regulatory frameworks and GDPR-based personal data protection provide a strong foundation to defend data privacy. The proposed VODAN architecture employs the FAIR Guidelines as the conceptual framework to inform the design of a localised data curation and repositing of (patient) health

data, useable for data analysis within the facilities. This enables collaboration while retaining data in a strong framework of ownership, control and retention of data where data is produced. The architecture is enhanced with data visiting capability to create a more comprehensive set of data from varied places in Africa, available for multiple site data analytics.

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.

AUTHORS' CONTRIBUTIONS

Mirjam van Reisen conceptualised this article, wrote the first and second drafts, and approved the final draft. **Francisca Onalapo Oladipo** contributed to the research reported in this article. **Mouhamed Mpezamihigo** provided support for the coordination of the research reported in this article. **Rudian Plug** conceptualised key ideas for this article and provided the architecture for elements of the research reported. **Mariam Basajja** carried out research concerning the problem explored in this article and conceptualised the methodology of FAIR Equivalency reported in this article. **Aliya Aktau** coordinated the capacity building of researchers in the research group that carried out the work for this article. **Putu Hadi Purnama Jati** coordinated the work with the health facilities and stakeholders in each of the participating countries, conducted the comparison of FAIR Equivalency in six countries, and wrote the text and the table of results. **Reginald Nalugala** provided ideas on the adoption strategy within Kenya. **Sakinat Folorunso** provided expertise on the use of the FAIR data model. **Samson Yohannes Amare** conducted research on the requirements for the innovation reported on in the article. **Ibrahim Abdulahi** provided insights into the adoption potential in Nigeria. **Oluwole Olumuyiwa Afolabi** provided expertise in the realisation of the software adaptation to localised use in African countries. **Ezra Mwesigwa** provided expertise on the visualisation of data and the architecture to link locally repositied data to dashboards. **Getu Tadele Taye** researched the potential to work with the DSW and CEDAR platform in health facilities. **Abdulahi Kawu** (a data steward) carried out research. **Meriem Ghardallou** was in charge of investigating the inclusion of scientific data on COVID-19 prevalence in the architecture. **Yan Liang** worked on an early architecture for the solution discussed in this article. **Obinna Osigwe** assisted in regular reporting on research outcomes. **Araya Abrha Medhanyie** coordinated research between digital health and computer science experts. **Munyarazdi Mawere** made comments on the article.

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] Amnesty International: COVID-19: Omicron variant must spur world leaders to share vaccines. [Online]. Amnesty International (2021). Available at: <https://www.amnesty.org/en/latest/news/2021/12/covid-19-omicron-variant-must-spur-world-leaders-to-share-vaccines/>. Accessed 14 December 2021
- [2] NICD: New COVID-19 variant detected in South Africa [Online]. National Institute for Communicable Diseases (25 November 2021). Available at: <https://www.nicd.ac.za/new-covid-19-variant-detected-in-south-africa/>. Accessed 12 December 2021
- [3] WHO: Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern [Online]. World Health Organization(26 November 2021). Available at: [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern). Accessed 12 December 2021
- [4] Maeda, Justin M., Nkengasong, John N.: The puzzle of the COVID-19 pandemic in Africa. *Science* 371 (6524), 27–28 (2021). DOI: 10.1126/science.abf8832
- [5] Wadvalla, Bibi-Aisha: How Africa has tackled COVID-19. *BMJ* 370, m2830 (2020). <http://dx.doi.org/10.1136/bmj.m2830>
- [6] Mbow, M., Lell, B., Jochems, S.P., et al.: COVID-19 in Africa: Dampening the storm? *Science* 369(6504), 624–626 (2020). DOI: 10.1126/science.abd3902
- [7] AIV: The Netherlands and the global approach to COVID-19. Letter, Advisory Council on International Affairs (AIV), The Hague (2020)
- [8] Nyamnjoh, Francis B.: A post-COVID-19 fantasy on incompleteness and conviviality. In: C. Besteman, H. Cabot, B. Kalir (eds), *Post-COVID fantasies*, *American Ethnologist* (27 July 2020). Available at: <https://americanethnologist.org/features/pandemic-diaries/post-covid-fantasies/a-post-covid-19-fantasy-on-incompleteness-and-conviviality>. Accessed 9 December 2021
- [9] Häikiö, J., Koskela-Huotari, K.: Bridging the gap between technical and human elements in digital service innovation. In: *BLED 2014 Proceedings* 14 (2014). Available at: <https://aisel.aisnet.org/bled2014/14>. Accessed 9 December 2021
- [10] Buskens, I., Van Reisen, M.: Theorising agency in ICT4D: Epistemic sovereignty and transformation-in-connection. In: M. Mawere (ed.), *Underdevelopment, Development and the Future of Africa*, Langaa RPCIG, Bamenda, pp. 394–432 (2016).
- [11] WHO: WHO coronavirus (COVID-19) dashboard [Online]. World Health Organization (n.d.). Available at: <https://covid19.who.int/>. Accessed 9 December 2021
- [12] John Hopkins University: COVID-19 dashboard [Online]. Coronavirus Resource Center (2021). Available at: <https://coronavirus.jhu.edu/map.html>. Accessed 9 December 2021
- [13] The Conversation: Data map: Tracking COVID-19 waves in Africa [Online]. The Conversation Media Group (2 November 2021). Available at: <https://theconversation.com/data-map-tracking-covid-19-waves-in-africa-170884>. Accessed 9 December 2021
- [14] Humdata: Novel coronavirus (COVID-19) cases data [Online]. Humanitarian Data Exchange (HDX) (n.d.). Available at: <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases> Accessed 9 December 2021

- [15] Taylor, L., Broeders, D.: In the name of development: Power, profit and the datafication of the global South. *Geoforum* 64, 229–237 (2015). <http://dx.doi.org/10.1016/j.geoforum.2015.07.002>
- [16] Tai, D.B.G., Shah, A., Doubeni, C.A., Sia, I.G., Wieland, M.L.: The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. *Clinical Infectious Diseases* 72(4), 703–706 (2021)
- [17] Sze, S., Pan, D., Nevill, C.R., et al.: Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis. *EClinical Medicine* 29, 100630 (2020). DOI: 10.1016/j.eclinm.2020.100630
- [18] Spivak, G.C.: “Can the subaltern Speak?” (Revised edition). In: R.C. Morris (ed.), *Can the Subaltern Speak?: Reflections on the History of an Idea*, Columbia University Press, pp. 21–78 (2010). Available at: <http://www.jstor.org/stable/10.7312/morr14384.5>. Accessed 24 December 2021
- [19] Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihigo, 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
- [20] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al.: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(1), 1–9 (2016). <http://www.nature.com/articles/sdata201618>
- [21] Björklund, T.A.: Initial mental representations of design problems: Differences between experts and novices. *Design Studies* 34(2), 135–160 (2013). <https://doi.org/10.1016/j.destud.2012.08.005>.
- [22] Sanders, L.: On modeling: An evolving map of design practice and design research. *Interactions* 15(6), 13–17 (2008). <https://interactions.acm.org/archive/view/november-december-2008/on-modelingan-evolving-map-of-design-practice-and-design-research1>
- [23] Kingdon, J.W.: *Agendas, alternatives, and public policies*. HarperCollins Publishers, New York (1984)
- [24] ACM Web Science Conference, University of Southampton, Association for Computing Machinery, & Association for Computing Machinery: *WebSci '20: 12th ACM Conference on Web Science Companion*, July 6–10, 2020, Southampton, United Kingdom, Association for Computing Machinery (ACM), New York, NY (2020). <https://doi.org/10.1145/3394332.3402823>
- [25] Alliance for Accelerating Excellence in Science in Africa (AESA): *Recommendations for data and biospecimen governance in Africa*. ASP Policy Paper 3, African Academy of Sciences (2021). Available at: <https://www.aasciences.africa/sites/default/files/Publications/Recommendations%20for%20Data%20and%20Biospecimen%20Governance%20in%20Africa.pdf>. Accessed 13 May 2021.
- [26] 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
- [27] Bauer, J., Petroni, M.J.: Chapter 7. Data Ethics Guidebook [Online]. *Causeit Guide to Digital Fluency, Digital Fluency* (2018). Available at: <https://www.digitalfluency.guide/data-ethics/data-sovereignty-and-data-politics>. Accessed 13 December 2021
- [28] Techopedia: Data Sovereignty [Online]. *Technopedia* (7 September 2021). Available at: <https://www.techopedia.com/definition/33304/data-sovereignty>. Accessed 13 December 2021
- [29] African Union: *The digital transformation strategy for Africa (2020–2030)*. African Union, Addis Ababa (n.d.). Available at: <https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf>. Accessed 22 December 2021
- [30] Plug, R., Liang Y., Aktau, A., Basajja, M., Oladipo, F.O., Van Reisen, M.: Terminology for a FAIR Framework for the Virus Outbreak Data Network-Africa. *Data Intelligence* 4(4), 698–723 (2022)
- [31] Stocker, M., Stokmans, M., Van Reisen, M.: Agenda setting on FAIR Guidelines in the European Union and the role of expert committees. *Data Intelligence* 4(4), 724–746 (2022)

- [32] Lin, Y., Purnama Jati, P.H., Aktau, A., Ghardallou, M., A., Nodehi, S., Van Reisen, M.: Implementation of FAIR Guidelines in selected non-Western geographies. *Data Intelligence* 4(4), 747–770 (2022)
- [33] Basajja, M., Van Reisen, M., Oladipo, F.O.: FAIR Equivalency with regulatory framework for digital health in Uganda. *Data Intelligence* 4(4), 771–797 (2022)
- [34] Purnama Jati, P.H.: FAIR Equivalency in Indonesia’s digital health framework. *Data Intelligence* 4(4), 798–812 (2022)
- [35] 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)
- [36] Chindoza, K.: Regulatory framework for eHealth data policies in Zimbabwe: Measuring FAIR Equivalency. *Data Intelligence* 4(4), 827–838 (2022)
- [37] Kawu, A.A., Joseph, E., Abdullahi, I., Maipanuku J., Folorunso, S., Basajja, M., Oladipo, F.O., Ibrahim, H.L.: FAIR Guidelines and data regulatory framework for digital health in Nigeria. *Data Intelligence* 4(4), 839–851 (2022)
- [38] 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)
- [39] Purnama Jati, P.H., Lin, Y., Nodehi, S., Cahyono, D.B., Van Reisen, M.: FAIR versus open data: A comparison of objectives and principles. *Data Intelligence* 4(4), 867–881 (2022)
- [40] GO FAIR: FAIR Principles [Online]. Available at: <https://www.go-fair.org/fair-principles/>. Accessed 30 April 2020
- [41] Basajja, M., Nambobi, M.: Information streams in health facilities: The case of Uganda. *Data Intelligence* 4(4), 882–898 (2022)
- [42] Basajja, M., Nambobi, M., Wolstencroft, K.: Possibility of enhancing digital health interoperability in Uganda through FAIR data. *Data Intelligence* 4(4), 899–916 (2022)
- [43] Basajja, M., Suchanek, M., Taye, G.T., Amare, S.Y., Nambobi, M., Folorunso, S., Plug, R., Oladipo, F.O., Van Reisen, M.: Proof of concept and horizons on deployment of FAIR Data Points in the COVID-19 pandemic. *Data Intelligence* 4(4), 917–937 (2022)
- [44] Purnama Jati, P.H., Van Reisen, M., Flikkenschild, E., Oladipo, F.O., Meerman, B., Plug, R., Nodehi, S.: Data access, control, and privacy protection in the VODAN-Africa architecture. *Data Intelligence* 4(4), 938–954 (2022)
- [45] Ghardallou, M., Wirtz, M., Folorunso, S., Touati, Z., Ogundepo, E., Smits, K., Mtiraoui, A., Van Reisen, M.: Expanding non-patient COVID-19 data: Towards the FAIRification of migrants’ data in Tunisia, Libya and Niger. *Data Intelligence* 4(4), 955–970 (2022)
- [46] Folorunso, S., Ogundepo, E.A., Basajja, M., Awotunde, J.B., Kawu, A.A., Oladipo, F.O., Abdullahi, I.: FAIR machine learning model pipeline implementation of COVID-19 data. *Data Intelligence* 4(4), 971–990 (2022)
- [47] Oladipo, F.O., Folorunso, S., Ogundepo, E.A., Osigwe, E., Akindele, A.T.: Curriculum development for FAIR data stewardship. *Data Intelligence* 4(4), 991–1012 (2022)
- [48] Akindele, A.T., Tayo, A.O., Taye, G.T., Amare, S.Y., Van Reisen, M., Berhe, K.F., Gusite, B., Edozie, E.: The impact of COVID-19 and FAIR data innovation on distance education in Africa. *Data Intelligence* 4(4), 1013–1032 (2022)
- [49] Jacobsen, A., Kaliyaperumal, R., da Silva Santos, L.O.B., Mons, B., Schultes, E., Roos, M., Thompson, M.: A generic workflow for the data FAIRification process. *Data Intelligence* 2(1–2), 56–65 (2020)
- [50] Plug, R., Liang, Y., Basajja, M., Aktau, A., ... Van Reisen, et al.: FAIR and GDPR compliant population health data generation, processing and analytics. *Semantic Web Applications and Tools for Healthcare and Life Sciences* (2022, forthcoming)
- [51] Plug, R.: Presentation. Unpublished (2022)