A FAIR Health Information System

Virus Outbreak Data Network (VODAN)-Africa

SWAT4HCLS

12 January 2022

Strengthening Data for COVID-19 Health Care and Research

Mirjam van Reisen, Prof FAIR Data Science, Leiden University Medical Centre

Improving Quality of Healthcare Through Data Ownership and Localization
Mission of VODAN-Africa

• VODAN Africa Initiative started in 2020 as a platform to enable access to critical data needed from Africa to fight COVID-19

• The Initiative is focused on improving health data analysis, under the regulatory provisions of the country, and strengthening national capacities for health data analytics as well as the use of health data at the point of care
Outline

1. The Problem
2. Establishing the VODAN-Africa Research group
3. Architecture and Results
4. Further research
1. The problem

- Africa does not appear adequately in the data on COVID-19
- Data from Africa is incomplete
- Data is not useable at Point of Care
Distribution of COVID-19 cases worldwide, as of week 52 2021
Distribution of COVID-19 deaths, worldwide, as of week 52 2021
14-day COVID-19 case notification rate per 100,000, weeks 51-52
Diversity of COVID infections in Africa within one country (example Nigeria)
Omicron variant detected around world

"The World Health Organization classified Omicron as a "variant of concern," due to the number of mutations that might help it spread or evade antibodies from prior infection or vaccination. The variant was first identified in South Africa and has also been detected across Europe and Asia.

Sources: GISAID; Reuters reporting

T. Hartman, 29/11/2021
Practices with Data from Low and Middle Income Countries: Extraction

• LMICs originally lacked strong regulatory frameworks on data management, and/or enforcement;

• Forms of data extraction from LMICs resulting in distrust from data producers;

• Data extraction for research has led to lack of return on data for research in LMICs;

• Data extraction of health data through health data management systems without return to point of care has undermined data use for improvement of quality of care;

• Due to lack of return, the data has little value in place of origin, which undermines quality and value of such data at the place where data is produced...
Hospital Data Flow

1. Start KIU Information flow
2. Reception
3. Search patient database by name/DOB
4. Client?
5. Patient receives his file with unique id from last visit
6. Patient is sent to the General Out-Patient DEPT for Bio data checkup
7. Patient sees clinician for diagnosis, written back of form 5, leave form 5
8. Manually enter data into MOH HMIS REGISTER BOOK
9. Update the hospital storage system (Digital)
10. IN/OUT patient files stored in shelves
11. Enter data into MOH DHIS2 system (Digital)
12. END
Virus Outbreak Data Network Problem Analysis: lack of interoperability, ownership and security of patient data

DHIS = District Health Information System
Introduction

Problem statement

Overview of the DHIS2 architecture, source: Mancj et al.

Current data management systems -> Problematic

Discover the world at Leiden University
Problem for Data Collection on COVID-19

• Limited data from LMICs;
• Unreliability and incompleteness of data;
• Lack of interest from LMIC partners in data-sharing through data warehousing;
• Data incompatibility and lack of interoperability of databases;

resulting in insufficient data diversity with bias towards western geographies.
The challenge

DATA UNAVAILABILITY IN VULNERABLE AND REMOTE SETTINGS

NEED FOR DATA LOCALISATION AND OWNERSHIP, DATA-CURATION IN RESIDENCE AND WITHIN REGULATORY FRAMEWORKS RELEVANT TO SETTINGS WHERE DATA IS PRODUCED

PRODUCTION OF DATA MUST BE RELEVANT TO LOCAL PRACTICES, BUT THIS IS OFTEN NOT THE CASE AT PRESENT

UNRESOLVED DIGITAL DATA CHALLENGES SUCH AS LACK OF INTEROPERABILITY

LACK OF CAPACITY AND TRAINING OF DATA STEWARDS AND CURRICULUM DEVELOPMENT FOR DATA SCIENCE
Medical Patient Records data curation as FAIR – interoperable with Research medical health data

- **GDPR**: Federation of Data Curation to ensure Ownership and Data Localisation as essential preconditions
- **Quality of data**: Centrality of Context and Provenance of Data
- **Ethical and Responsible Data**: Data used for improvement of health services at Point of Care
The warehousing of data removes data from place of origin, with loss of metadata and loss of access to understanding of contextual variables;

The data management practices in relation to LMICs do often no longer adhere to highest standards of procedures of data management such as under GDPR and other regulations;

GDPR declares need for **data ownership**, privacy procedures and purpose of collecting data to be specifically formulated in terms of purpose and adhered to.
Advantages of Federated Data Curation

- Data is generally becoming very large with complications for central repositing and preferences moving to distributed repositing;
- Data is generally becoming too large to investigate without metadata on provenance, clarification of data and provision of machine-readable access (under well defined conditions);
- Moving from paper-based to digital patient health records;
- Use, and strengthening of quality of data at point of care;
- Highly structured data from health facilities which are better distributed across Africa than COVID-testing and treatment Centers;
- Strengthening interoperability of Digital Health tools;
- Integration in regular health services and direct contribution to improvement of data analytics at point of care.
Advantages to federated data curation

- Compliance with regulations in every place/country
- Adhering to data sovereignty
- Linking data inspection and analytics to clearly elaborated and agreed procedures of ownership and responsibilities following GDPR
- Increased trust in data, use of data at point of care and increased data quality, as well increased availability of data from LMICs
- Enabling cross-border data visiting in different data sovereignties
DRAFT ARCHITECTURE OF THE EAST AFRICA OPEN SCIENCE CLOUD FOR HEALTH DATA ACCESS (EAOSCH)

- Health facilities in border areas in East Africa
- DHIS2 — District Health Information System 2
- Uganda
- Kenya
- Tanzania

Cloud Services

Clinical Record Interoperability

Research Tools

Country Portals and Dashboards

Multi Country Portal and Dashboards

Community Health Workes

DHIS2

LAB

CLINICS
2/ VODAN- Africa Research Group of Universities

PIs
Prof. Mirjam van Reisen
Prof. Francisca Oladipo
Wicked problem:

- Interdisciplinary research
- Integration of stakeholders
- Ethnographic Design Study
- Strong emphasis given to ecological validity
Phase 0: Development of a FAIR-based architecture

Data is curated, stored and used within the health facilities where it is produced

Data security is arranged by the health facility where data is produced

Efficiencies of data production, processing and reuse is increased by one editing entry into a machine readable editor

Data is Findable, Interoperable and Reusable

Data Access Permission is arranged within the context of each facility and recognizing national and other relevant regulatory frameworks
Requirements & Principles of VODAN - Africa

- Data need to include research data and clinical data
- Data is curated as machine-readable FAIR (Findable, Accessible under clear conditions, Interoperable, Reusable) objects;
- Data-collection includes research institutes and health facilities under strict guidance of Ministries of Health and Higher Education and regulatory frameworks in place in country;
- Data is held in residence – this is on premise or premise of choice of research institute and/or health facility;
- Data is visited over the internet through FAIR Data Hosts, machine-findable entries for data visiting if permission is granted, creating Federated-AI-Ready query opportunity;
- Open Source development avoiding Vendor Lock in;
- Data remain in residence and are available for support of the facility.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Uganda</th>
<th>Indonesia</th>
<th>Ethiopia</th>
<th>Zimbabwe</th>
<th>Nigeria</th>
<th>Kenya</th>
<th>Average of percentage (row)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total documents</td>
<td>12</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>‘Findable’ principle in document</td>
<td>Average: 24%</td>
<td>Average: 12.25%</td>
<td>Average: 15.75%</td>
<td>Average: 34%</td>
<td>Average: 42.5%</td>
<td>Average: 36.5%</td>
<td>Average for 6 countries: 31.96%</td>
</tr>
<tr>
<td>‘Accessible’ principle in document</td>
<td>Average: 45.75%</td>
<td>Average: 28%</td>
<td>Average: 31.5%</td>
<td>Average: 46.75%</td>
<td>Average: 50%</td>
<td>Average: 49.75%</td>
<td>Average for 6 countries: 45.04%</td>
</tr>
<tr>
<td>‘Interoperable’ principle in document</td>
<td>Average: 35%</td>
<td>Average: 20%</td>
<td>Average: 33%</td>
<td>Average: 41.3%</td>
<td>Average: 23.3%</td>
<td>Average: 37.67%</td>
<td>Average for 6 countries: 30.10%</td>
</tr>
<tr>
<td>‘Reusable’ principle in document</td>
<td>Average: 35%</td>
<td>Average: 15.25%</td>
<td>Average: 56.25%</td>
<td>Average: 46.5%</td>
<td>Average: 52.5%</td>
<td>Average: 33.25%</td>
<td>Average for 6 countries: 39.17%</td>
</tr>
</tbody>
</table>

Preparation: Measuring FAIR Equivalency with e-Health

- **Table 1. FAIR Equivalency comparison**
- **Source:** Created by Purnama Jati, 2021
Design Architecture Phase 1: Deployment On-Premise
- East Africa Open Science Cloud for Health in 2017 establishes the need for a within and across border health data sharing

- FAIR IN Africa adopts FAIR Principles for Federated deployment and increased data ownership in Africa in 2019

- VODAN-Africa Team demonstrates data visiting works across two countries and two continents, September 2020

- VODAN-Africa Team reaches consensus on the requirements and specifications for the clinical health data architecture in February 2021
Design Architecture Phase 2: Creating Flexibility with CEDAR semantic Metadata

The CEDAR Workbench is an essential component of open science, ensuring FAIR data and enhancing scientific reproducibility.
Data Protection - All clinical data are encrypted with secure algorithms.

Data Integrity - Data remain the same they were created. NO DATA MOVEMENT

Data Stewards capacity training limited to data input

Assured Data Governance – Clinical data produced are managed within the regulatory framework of the Ministries of the participating countries

Total control of data access, control and security

BENEFITS FOR HEALTH CLINICS

VODAN AFRICA
The data is presented in a BI analytics tool where users can perform analytics in residence independently.

Data analytics are done at the clinics/hospitals/universities using APIs that visit data stored at the local clinics/hospitals.

The Analytics can be real-time.

Create a FAIR Data Host that is managed under the regulation of the Ministry of Health and/or Higher Education.

Data is entered into CEDAR templates and stored in clinics and hospitals and/or university data repositories in the country.

Data is stored securely in machine-readable format at clinics/hospitals/universities.

Clinic/hospital Data Scientists are trained on FAIR Data Management.

Deployment Strategy

31
VODAN Localization Architecture: COVID-19 Clinical Data

Input/Bulk Input

Embeddable Editor at the Clinic

Metadata Clinic

Metadata for Clinical Patient Data

Health Information System

Repositories in Residence

Visualisation of Clinical Data Analytics

Dashboard for analytics within the clinic/hospitals only

Reachable address as a FAIR Data Host

VODAN Africa Dashboard
VODAN Localization Architecture: COVID-19
Research Data
VODAN Localization Integrated Architecture:
COVID-19 Clinical & Research Data

Input/Bulk Input → Embeddable Editor at the Clinic/University → Metadata of the Clinical/Research Data

1. University alternative: Joint Repository (CEDAR, ANDREA, Data Intelligence)
2. Clinics/University Repositories in residence

Dashboard for analytics within the Clinics/University
Reachable address as a CEDAR Data Point
Aggregate Data Analytics
Clinical Data Dashboard
Aggregate Research Analytics
Research Data Dashboard
Scope of VODAN-Africa:
- 85 Health Facilities
- 9 Countries
- 12 Universities
- Expanding to 15 countries
• Proof of Concept realised between Leiden University Medical Centre (LUMC) and clinics in Africa in September 2020;
• Successful machine-learning investigation of FAIR-data;
• Collaboration with Chinese Academy of Science on documentation and exchange of practices;
• Collaboration between four continents: Africa, Asia, Europe and US;
• 84 health facilities and universities in Africa have agreed to participate from 9 countries: Uganda, Tunisia, Liberia, Nigeria, Zimbabwe, Ethiopia, Somalia, Kenya and Tanzania;
• Convergence with 62 ZonMw funded COVID-19 research programmes, to be included in queries (ZonMw);
• FAIR-evaluator developed by CEDAR available (wellcome trust);
• Distance learning courses available (NUFFIC-funded);
• MVP of queries across the 84 facilities expected in February 2022.
Data Visiting Access and Control Layers
VODAN
Schema of a FAIR Data Point

Fig. 2. The basic framework of a FAIR Data Point.
VODAN-AFRICA FAIR HEALTH DATA ARCHITECTURE ALIGNING WITH WHO SMART GUIDELINES

WHO/GO FAIR/VODAN COMMUNITY

MINISTRY OF HEALTH

DATA STEWARDS AND TECHNOLOGY PARTNERS

HEALTH WORKFORCE

HEALTH FACILITY

CLINICAL HEALTH DATA INTEROPERABILITY

RESEARCH / UNIVERSITY

Research Community

Research Group

Data Stewards and Technology Partners

Researchers

Data

MAY 2021

PROF. MIRJAM VAN REISEN
PROF. FRANCISCA OLADIPWO
OBINNA OSIGWE

VODAN AFRICA
Data Production Editor – allowing flexible templates

Localised instance of CEDAR-Platform in 84 Health Facilities

One point data production editor

Templates with controlled vocabulary, currently: OPD, MCC, COVID-19 prevalence

Data reposited in the Health Facility (only)
Metadata produced in rdf and jason ld
(here on the measurement, provenance of location, method and units)
Federated querying process
Internal and External Dashboard
This is VODAN and this is what you can find in the Dashboard

OVERVIEW

FACILITY INFORMATION

VODANA DASHBOARD
Dashboard Overview

- **Countries**: 8
- **Facilities**: 85
- **Metadata**: 0+

**Participating Countries**
- Ethiopia: 30 Facilities
- Kenya: 4 Facilities
- Liberia: 1 Facility
- Nigeria: 21 Facilities
- Somalia: 19 Facilities
- Tanzania: 1 Facility

**Facility Overview**
Statistics of facilities by countries

- Somalia: 19 Facilities
Overview of VODAN facilities
Facility Dashboard (only access to own facility)
Aggregate data visited through the 84 facilities
(accessible by all health facilities and VODAN-Africa community)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Visits</td>
<td>530</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>15.42</td>
</tr>
<tr>
<td>Deaths</td>
<td>7</td>
</tr>
<tr>
<td>Admissions</td>
<td>363</td>
</tr>
<tr>
<td>Discharge</td>
<td>207</td>
</tr>
<tr>
<td>Emergency</td>
<td>524</td>
</tr>
<tr>
<td>Ambulance Call</td>
<td>351</td>
</tr>
</tbody>
</table>

**Generic KPIs of VODAN Africa**

**Emergency calls and Admissions # by month**

**Patient visits # month**
Capacity Building
Design Phase 3
Proposals for acceleration of COVID-19 Data from LMICs:

• Agreement of funders to include FAIR-based data curation on all COVID-19 relevant research funding;
• Collaboration on convergence on FAIR-based curation across all COVID-19 research through accredited providers of training and tools;
• FAIR & VODAN based acceleration of clinical data curation in LMICs;
• Strengthening of data regulations according to GDPR standards across LMICs;
• Strengthening of services by creation of Data Stewardship Competence Centers;
• Strengthening of capacities by distance learning on FAIR-based data curation for science;
• Seeking collaboration with Health Management Information platforms (such as DHIS) to converge with FAIR principles.
References

https://www.vodan-totafrica.info/


