Leids Universitair Medisch Centrum

17 November 2023

TOWARDS AN AFRICA HEALTH DATA SPACE

VODAN PHASE III

Mirjam van Reisen, Joëlle Stocker, Samson Yohannes Amare, Ruduan Plug, Aliya Aktau, Putu Hati Purnama Jati, Tesfit Gebremeskil, Kai Smits, Mariam Basajja, Abdullahi Kawu, Getu Tadele, Nambobi Mutwalibi, Natascha Buchs.

VODAN AFRICA

TABLE OF CONTENT

ACKNOWLEDGMENTS

It is a great pleasure to deliver the 2023 Annual Report of VODAN-Africa, which presents the outcomes of the deployment of VODAN-Africa V2.

Vodan-Africa has enjoyed the support of Philips and the Philips Foundation from 2020, when we began the first deployment to create digital patient data within a FAIR-based architecture for pandemic surveillance and knowledge integration.

The deployment of VODAN-A V01 was completed with a Proof of Concept in 2020. This showed the potential of a federated architecture in which data could be visited across countries and continents and inclusive of low resource areas, often under-represented in the data.

The Requirements and Specifications of VODAN-A V02 stipulated that a dynamic architecture should be set up, adapted to the different contextual situations in Africa and relevant to the specific needs in each health facility. At the end of 2022, Vodan-Africa had been deployed in 90 health facilities across eight countries. VODAN strived to ensure the curation of data. This report sets out the next phase following the deployment which was realised in 2022.

VODAN-Africa is proud to present this report as we embark on the review of Requirements and Specification for VODAN-Africa V03, which will start in 2024. We are grateful to everyone who has enabled the progress we present in this report. VODAN-A has grown into a lively, solid network, that delivers granular quality data for the generation of knowledge for better care, for support of decisions by AI generated models for the surveillance of patients, the integration of data with different provenance and the re-use of data for further health research in the Africa Health Data Space.

Mirjam van Reisen,
PI VODAN-Africa,
Chair FAIR Data Science, LUMC

WORLD UNBOUND: THE CURATION OF DATA AS MEANINGFUL ONTOLOGIES

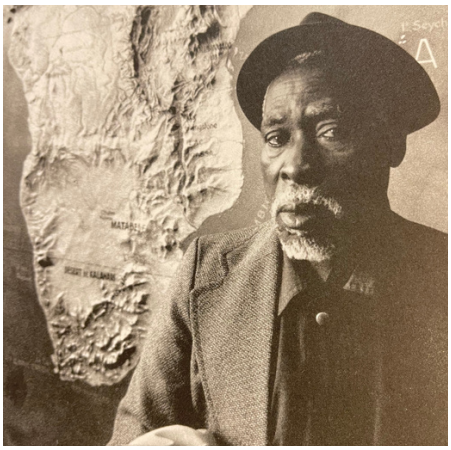
Central in the work of VODAN is data curation, in collecting and managing data so that it is accessible in different applications, while held and possessed by the community within which it was created. To understand what the act of data curation fully entails, we need to have a closer look at the essence of curation. In the work by Jean-Paul Martinon, The Curatorial: A Philosophy of Curating (2013), two significant analogies are presented. One is a Latin fable of the entity "Cura", who moulds humans from clay found on a riverbank. The fable describes a quarrel on the ownership of these humans, does it belong to the one who moulded it (in this case: Cura), to the one who created it (here the God Jupiter), or to the one from whose flesh it was created from (the earth)? In the end a compromise is made: at the end of their life, the spirit of humans returns to Jupiter who gave them life, the body returns to the earth from which it was moulded, whilst Cura gets to possess and care for them while they live.

The second analogy is an unfinished work of art by Stéphen Mallarmé entitled "This is", which according to Martinon, is the ultimate act of curation. Some important aspects of this artpiece are that it displays the work of others, that it brings the past and present together, that it enables a constellation of meaning which is not achievable through any other artform, that it combines human agency and the absolute, that it is centred from the viewer, that it is multi-sited (there is no centre of significance), and that it does not have predetermined rules, grammar or syntax, but rather needs to find its own language.

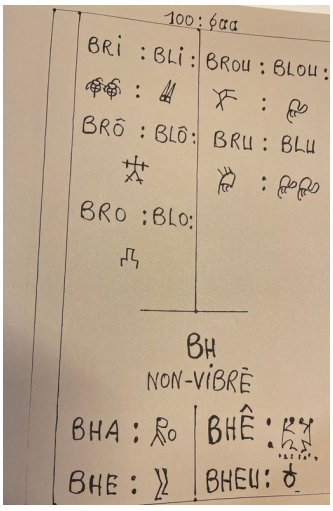
Both these examples give an interesting insight in what curation, and data curation are. The fable puts two important aspects forward: the caring of what you are curating, and the ownership thereof. "This is" focuses on what the meaning is of the space where data is curated to, perhaps a data space.

In the broader literature, curation is often centred around usability. This is especially clear if it is compared to two similar terms: archiving and preservation. Whereas all three are linked to the collection of things, the act of curation emphasises the need for discovery and reuse (Yakel, 2007). Thus, crucial in data curation is ensuring that the data has value as an ontology, that it is intelligible, so that that the data is usable and reusable as meaningful assets.

The African artist, scholar, linguist and teacher, Frédéric Bruly Bouabré is a curator of ontologies in a multi-disciplinary and highly creative approach. Using ontologies to describe the language of the Bété community in Côte d'Ivoire, he has become a renowned artist. His exhibition "World Unbound" shows a beautiful collection of artefacts to make sense of the Béte language in a new and systematic way. As Bouabré develops his work, he goes on to create ontologies of cultural artefacts, social practices, and anything around him, to which he wants to give meaning.



The work of Bouabré is so compelling as he works in a cycle of linguistic transformation "from speech to image and then to text" (Nzewi, 2022: 8) . Bouabré demonstrates that this art of curation "enlightens, rules, explains, and orders the eternal laws of well-meaning reason" (Bouabré, 2013: 132). This curating work of art fraternizes all the members of 'great humanity"' in order to alleviate the 'too many sufferings' of the individual during his passage on 'earth'" says Bouabré in the book "A man in the world, a giant touches the sun" (2013). Bouabré reveals the work of ordering simple things that are as a decolonising and liberating act:



> "Writing, he understood, confers epistemic power – the power to classify, to name, and perhaps even to claim ownership."
>
> (Nzewi, 2022: 12)

Showing the existence and relevance of the pictographic African writing systems he shows that a taxonomy is a crucial step to create, to think, to write, and to confer meaning. Importantly Bouabré alludes to an African root of classification to give meaning and to signify (Nzewi, 2022). VODAN follows in the enlightened African tradition of care, of curation of every instance, which exists only ever in relation, based on the insight that nothing exists on its own, but exists only in connection to something else.

BASIC FEATURES OF THE FAIR DATA ARCHITECTURE

The basis of the architecture is FAIR data and data that is FAIRified at the point of creation of the data instance. This is called FAIR by design or de Novo FAIRification. The essence of the architecture is a one-time data entry which generates multiple parallel use trajectories of the data. The data is held in residence, in a triple store which in this case is an allegrograph. The triple store is a secure reposit where the data is in a safe container, without backdoor. If permission is granted, the data is visitable by algorithms, thereby allowing for federated data computations. For dynamic queries through data visiting in the triple store, additional permission for each exercise is required.

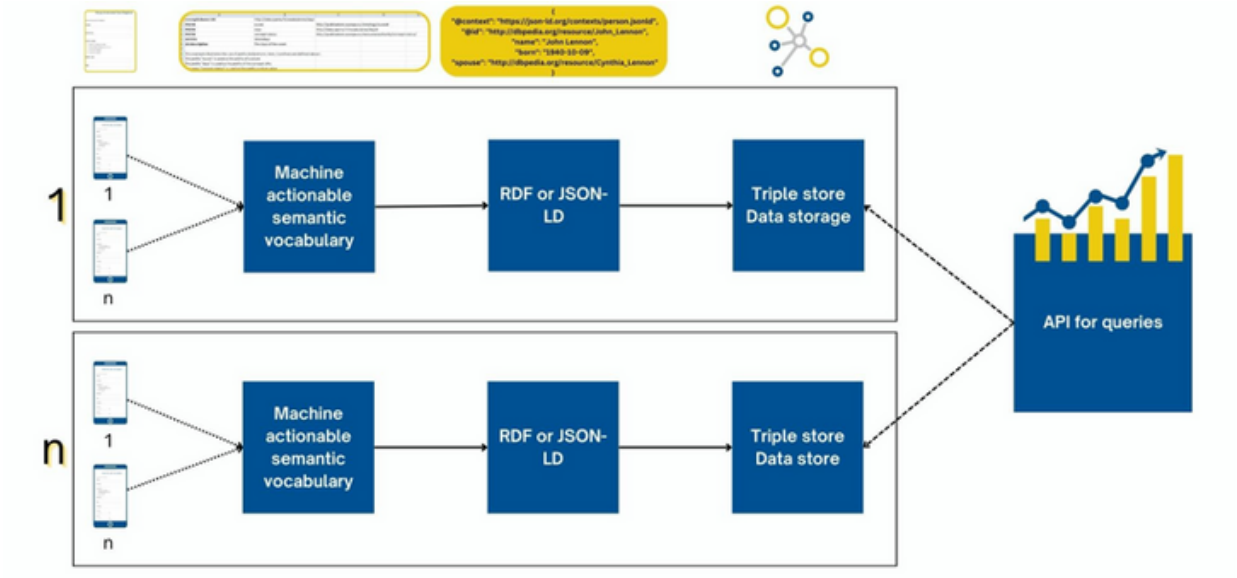


Figure 1. FAIR Data production and processing.

The legal requirements are based on local and national regulatory adherence in each place. The gold standard for personal protection of GDPR and data reuse for visualisation in the dashboard in the health facility and the community is arranged in a data use agreement which is signed by the administrator of the health facility, the ministry of health, the country coordinator of VODAN overseeing the data handling in that health facility, and the VODAN representative.

This architecture has been tested in a natural environment. It has been deployed in 90 health facilities in 8 countries, with deployment realised at different levels. This architecture in a real-life setting, has been used to study the deployability and adoption potential of the VODAN architecture.
The architecture consists of federated mini-services that are created on the basis of CEDAR software. These are among others:

- Template creation of machine actionable semantically annotated ontologies for federated use.
- The creation of new machine actionable semantic vocabularies in our portal for federated use.
- The creation of a FAIR Implementation Profiles (FIPs) for the sub-communities to document the arrangements on the FAIR-facets, and to specify how the repositories are accessible through the FAIR data points of the VODAN community.
- The services were adapted for specific circumstances such as low connectivity use, to support data production in all health facilities.
- Additional services were created such as bulk upload, data entry by phone, upload of data in the HIS2, and upload of data in local triple store.
- The visualisation of the data which appears on a dashboard in the health facility.
- The upload of data for surveillance in the VODAN community dashboard.
- The tracking of data production in each facility and in the VODAN community.
- Two templates were created, one for Ante Natal Care (ANC) and one for Out Patients Department (OPD).

All the steps in the architecture concern federated tools for federated use in the facility with a harmonisation across the community.

Figure 2. The 14 FAIR Facets to develop semantic machine-actionable digital assets

The following requirements were defined for vODAN-Africa V02:

R1: Flexible data production (based on VODAN controlled vocabulary)
R2: Localization of the Metadata System
R3: Bulk input of data in data production platform
R4: Useability and demonstration of value

The following specifications were defined for VODAN-Africa V03 tools:

S1: Open Source
S2: Programmability and adaptability
S3: Own maintenance
S4: Availability for Training
S5: Convergence with other FAIR developers to increase efficiency

Based on these tools specifications it was decided to use CEDAR as the platform from which to develop V02, intended as a Minimum Viable Product (MVP). The original goal was for the MVP to be tested in 20 health facilities in Ethiopia, Uganda and Zimbabwe.

The following figure shows the timeline of the deployment and creation of the Proof of Concept (2020) and the MVP (2021-2022) and the testing of the MVP in 2023.

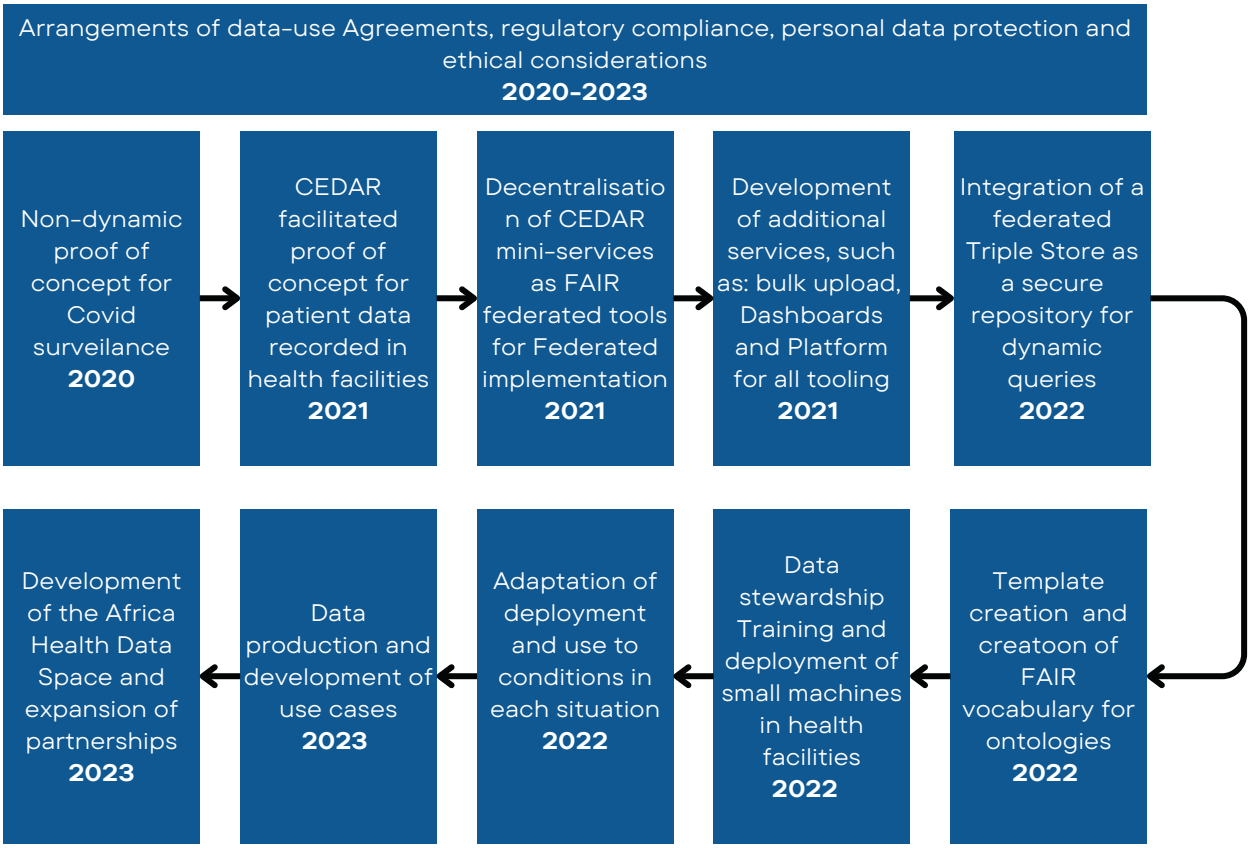


Figure 3. VODAN Development and Creation Process 2020-2023.

FHIR versus FAIR

The Fast Healthcare Interoperability Resources (FHIR) offers relevant tooling for the exchange of information between hospital and medical facilities for the exchange of information. FHIR in combination with other service providers can provide a FHIR & FAIR architecture.

The main characteristics of a robust FAIR architecture are:

1. The architecture is shaped according to the hourglass with a connecting layer linking data to knowledge generation
2. The enrichment of the data instances with robust controlled semantic attributes
3. The convergence in machine actionable coding language, such as RDF or Json LD
4. The description of the FAIR facets in a FAIR Implementation Profile
5. The creation of FAIR Data Points for data visiting
6. The ability to use a triple store to create graphs from the data and perform SparQL queries, and perform AI and Machine Learning operations
7. The ability to visit the data in a federated structure on the internet, allowing explicit control over data production, storage and data-visiting
8. A maximum freedom of operation for data curation
9. A maximum freedom of operation for knowledge generation, application and service
10. A maximum ability to control access, control and arrange permission-processes for data visiting

FHIR is equivalent to the FAIR facets of F1, F3, A1, A1.1. and A1.2.. Interoperability and Reuse of data-instances is not directly facilitated by FHIR.

Figure 4. FHIR equivalence with FAIR

The FAIR Data creation process

Using CEDAR, the FAIR data creation is implemented by creating a template for FAIR data creation. The template creation involves the following steps:

- Adding fields to the template based on existing controlled vocabularies
  - Designing the template through selection of fields and field properties;
  - defining the ontologies contained in the field
  - linking these to existing semantic controlled vocabularies.
- Adding fields to the template based on newly created controlled vocabularies
  - Designing the template through selection of fields and field properties;
  - creation of new vocabularies for ontologies contained in the field;
  - linking these to bioportal as controlled vocabularies;

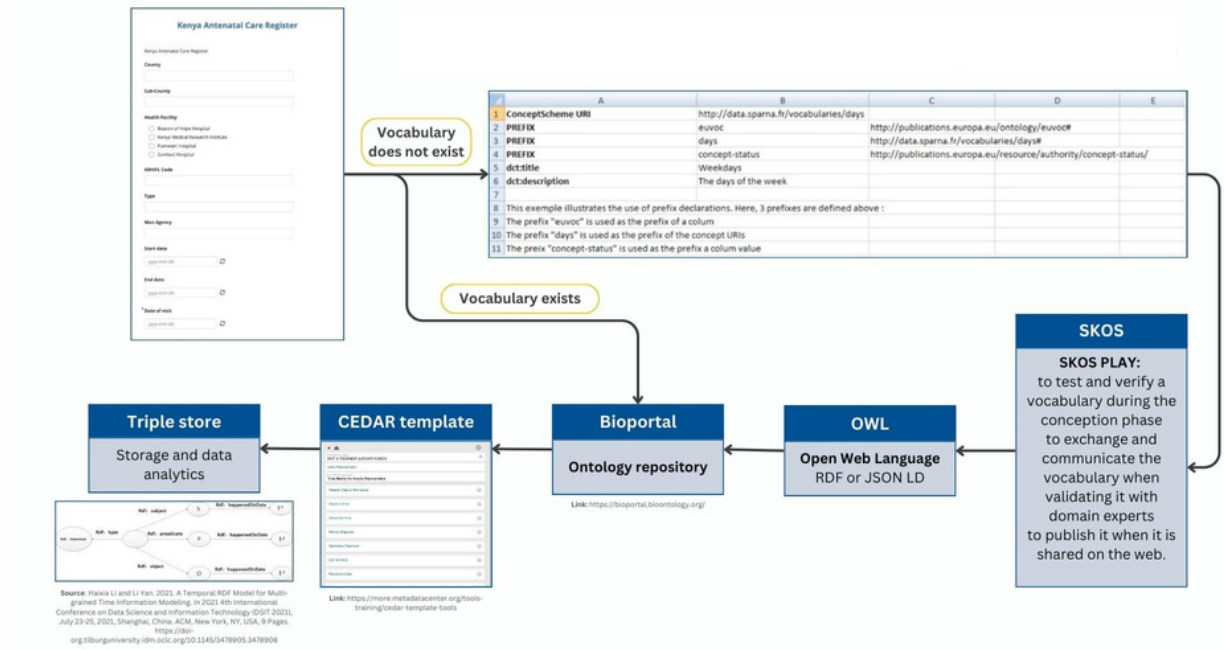


Figure 5. FAIR data creation process

There are several options for the creation of new controlled vocabulary:

1. through an excel-sheet prepared for controlled data creation, uploaded through SKOS for the testing and verification of the vocabulary, and checking the conformity with the OWL Web Ontology Language - a standard of the W3W standard of the World Wide Web (Semantic Web) to define semantic ontologies. The vocabulary is then uploaded in bioportal.
2. through the programme **protégé** for the creation of vocabularies, according to OWL standard, a free open-source ontology editor.

Both procedures support RDF schemas and triple representations.

Once the template is completed, the data instances can also be uploaded by bulkupload (https://github.com/VODANA/CEDAR_bulk_upload).

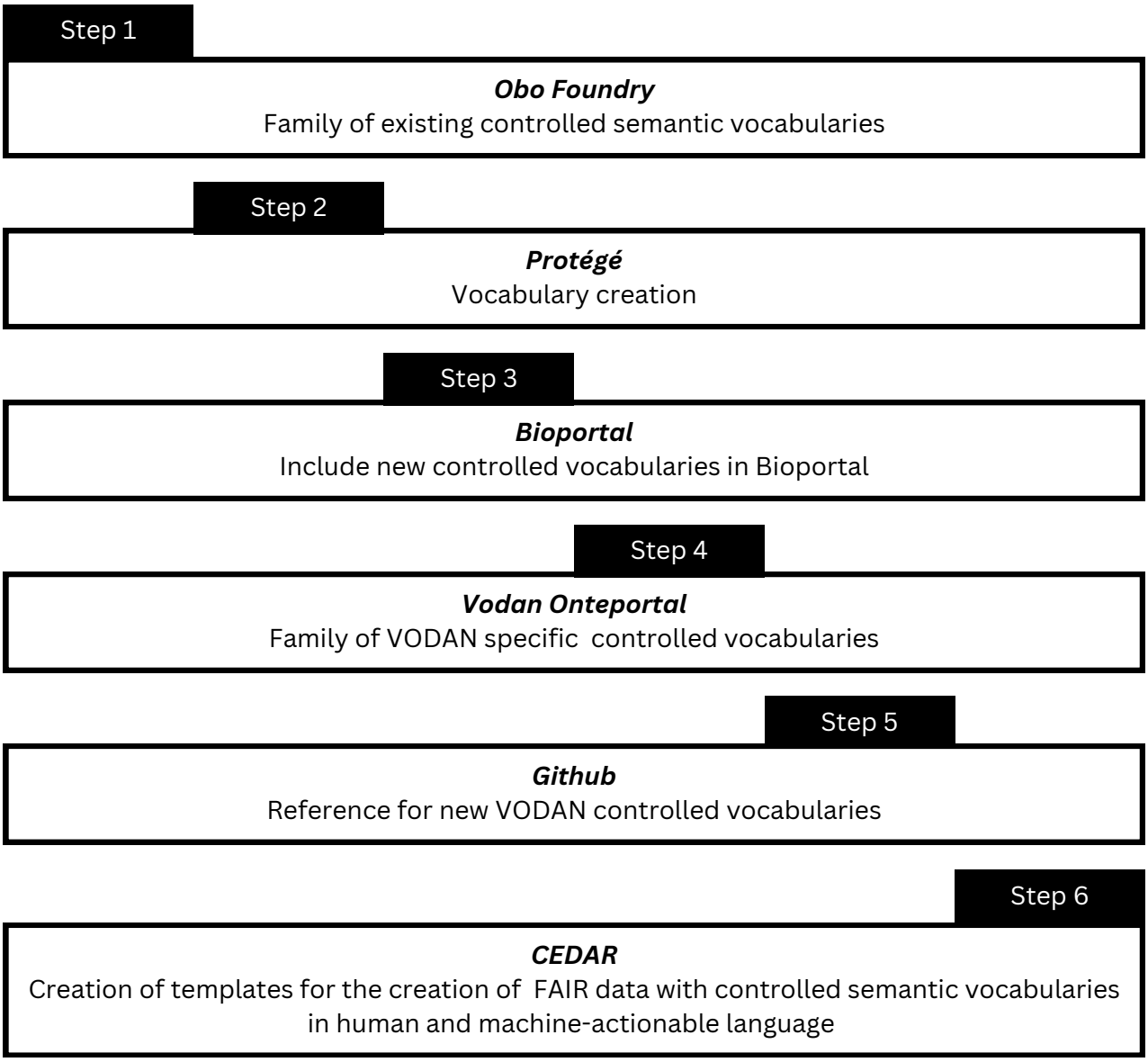


Figure 6. Controlled Semantic vocabularies for ontologies.

The Obo Foundry (https://obofoundry.org/) provides a set of well defined persistent controlled ontologies. VODAN uses the Obo Foundry for the selection and creation of new ontologies, which are then uploaded in protégé (https://protege.stanford.edu/) and bioportal in a vodan segment of the onteportal store of ontologies. While original set up for the field of biomedicin, it now also expands into ontologies of social and other domains.

The procedure for vocabulary creation draws on the specification of OWL which provides constructs for classification. Named things, which are classes, properties and individuals have unique identifiers (URIs) on the web.

The VODAN ontologies refer back to VODAN Github (https://github.com/VODANA). The process supports the template creation in CEDAR, which links fields to machine-actionable semantic vocabularies. The CEDAR deployment in health facilities is arranged through a federated tool:
https://github.com/VODANA/CEDAR_bulk_upload

DEVELOPMENT OF A QUALITY DATA PIPELINE

A total of 90 health facilities were included in different stages of the deployment of VODAN to support data production. In 2023 VODAN-Africa finished the V2 deployment with the production of a FAIR quality and ethical data pipeline.

Following the deployment phase, all facilities were approached for a systematic review of the deployability of the VODAN architecture (VODAN architecture v02). All the country coordinators and data clerks were asked to fill in a questionnaire and interviews were conducted with stakeholders on the technical limitations as well as any problems or concerns over the deployment, the use of the software for data production and the deployment or adoption concerns and how they were tackled.

Identified issues included access to electricity and connectivity, erroneous error messages generated by the software system, experience and training of the data stewards, and the need for a support system. Based on this, it was decided to rank the best facilities for a quality data pipeline as a basis for data-production that can serve for the testing of use cases. The quality data pipeline is created by federated data-production four countries: Ethiopia, Kenya, Nigeria and Uganda.

Table 1. List of Health Facilities which produce federated patient data for the VODAN-Africa Quality Data Pipeline (2023)

Health Facility	Country
Railway Clinic	Nigeria
GH Lapai	Nigeria
OOU	Nigeria
FUL	Nigeria
PUMWANI	Kenya
Beacon of Hope	Kenya
Lira Hospital	Uganda
Ark specialist Hospital	Uganda
Kampala International University	Uganda
Hoima referral Hospital	Uganda
Ayder Referral Hospital	Ethiopia - Tigray
Mekelle Hospital	Ethiopia - Tigray

The regular data production of electronic patients data were based on Ante Natal Care and Out Patient Department records.

In addition, other data have been integrated in the pipeline. This data includes the following:

- Data-sets on interviews of COVID-19 prevalence and challenges experienced by migrants who had little if any access to health facilities in Tunisia;
- Data-sets on interviews of trauma experienced by victims of human trafficking in Libya, Niger, Tunisia, Sudan, and Ethiopia;
- Data-sets of DHIS2 data on prevalence of Female Genital Schistosomiasis (FGS), HPV vaccination and prevalence of HIV in Uganda (10-year data);
- Data-sets on a study of vaccination of babies against diarrhea;
- Data-sets on neonates risks of mother-child transmission.

The incorporation of these data sets in the VODAN-Africa data pipeline has demonstrated:

- the relevance of FAIRification for data-reuse and analysis across different data-sets for increased insights;
- the possibility to enhance electronic patient data with other types of data to capture situations of persons in low-resource situations, who are not accessing health facilities;
- the relevance of FAIRification of scientific data to enhance long-term value of this data;
- the ethical relevance of re-using this data, given the costs of obtaining such data.

The integration of this data has provided good learnings on the feasibility of the FAIRification process, the costs and the gains to be made of it.

ARCHITECTURE OF THE AFRICA HEALTH DATA SPACE

The landscape of the Africa Health Data Space consists of three layers. At the point of creation, the data is made FAIR, which means that it is machine actionable and semantically annotated. It is reposited with those attributes within the place where the data is produced or in any case that it is controlled in the place where the data is produced.

In the service layer, the software is offered to make the data findable, accessible under well-defined conditions, interoperable and reusable. The service layer also includes the hosting of the data and the reachability of the data through a FAIR data point. These services are provided as federated tools and as tools that are also FAIR.

In the top layer, the new insights are derived through algorithmic queries for the data producers as third-party customers. The user experience as API as well as data visualisation are also arranged in this layer.

In the health data space, value is added in the bottom layer through data curation in the form of FAIR data stewardship. In the middle later, value is added through FAIR and federated tooling and hosting. In the top layer, new value is created by the computational services offered to generate intelligence over the data. Whereas in the bottom later, value is added **to** the data, in the top layer, new value is generated **from** the data.

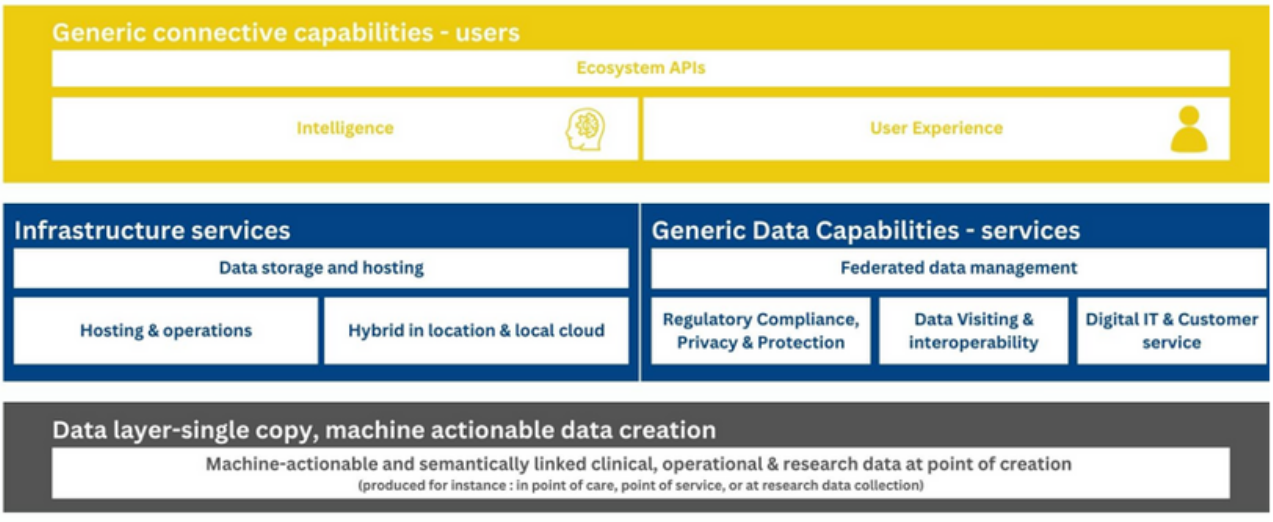


Figure 7. A proposed landscape for the Federated Health Data Space.

The quality Data Pipeline of VODAN-Africa is a federated data space, in which health facilities reposit the data in a triple store, which does not have a back end. The access to the repository is controlled by the health facility.

The data can be visited upon permission by the health facility. In the VODAN-Africa V02 deploument the access is arranged by data use agreement, which is signed by the Health Facility administrator, the Ministry of Health, the VODAN-Africa Country Coordinator and VODAN-Africa. For specific queries and operations, for which access to the data is requested, permission must be given on an explicit case-by-case basis.

The totality of the federated repositories make up the Africa Health Data Space, which is an intelligent, secure and Africa-based space. The objective is to prioritise the use of data for insights obtained at point of care where the data is produced . In addition, surveillance and machine-learning operations over the data performed in different health facilities, can benefit decision-making specific to and tailored for the health facility and can also be the basis for personalised care (see use cases).

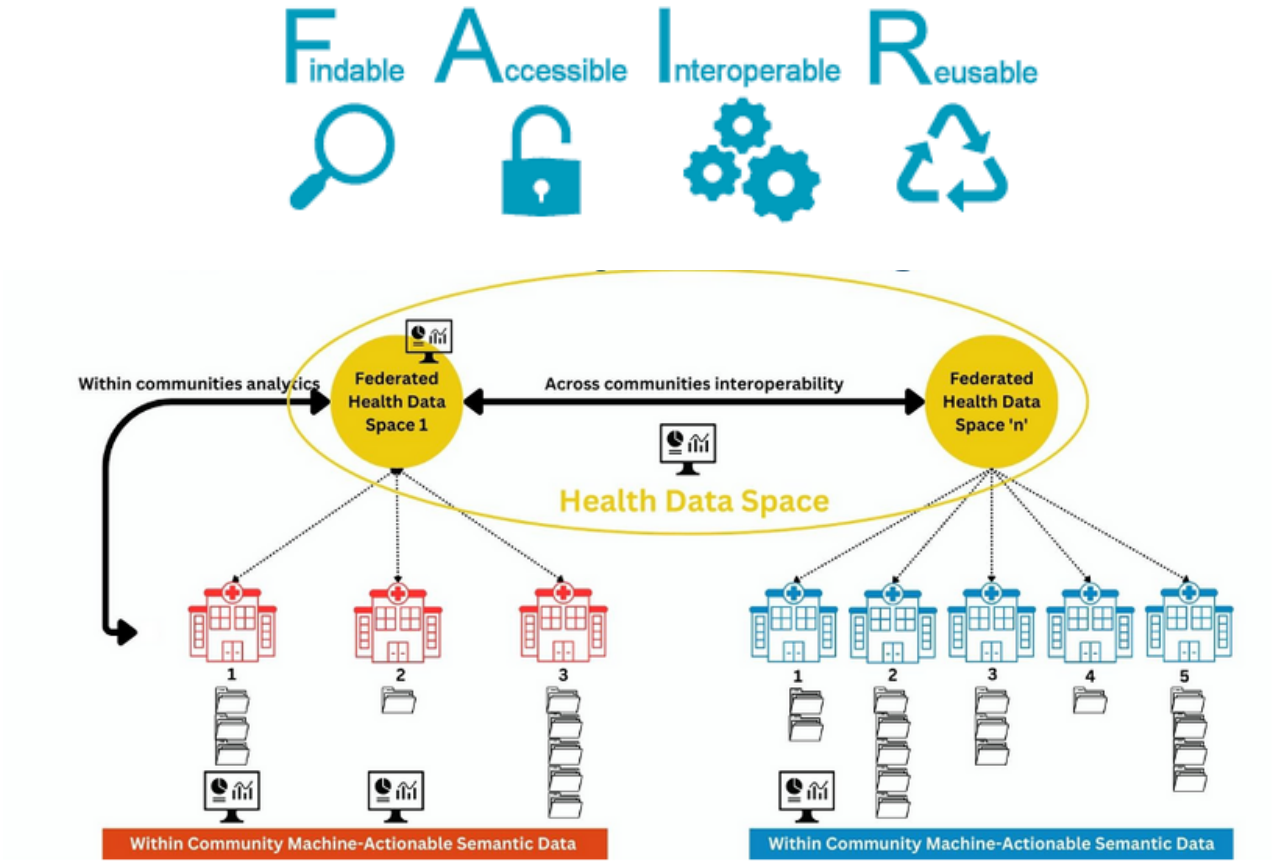


Figure 8. FAIR-OLR Federated analysis & learning.

DEVELOPMENT OF USE CASES IN THE AFRICA HEALTH DATA SPACE

The efficacy of the federated data infrastructure was assessed through the execution of several use-cases.

Pandemic Early Warning

The key aspect of this use-case involves adapting federated data infrastructures, allowing stakeholders to locally contribute and analyse data while maintaining privacy and security. This decentralised approach enables collaborative analysis of large-scale datasets without centralising data sharing, aligning with FAIR-data principles. The use of federated data in global pandemic early warning systems has numerous benefits, including aggregating diverse datasets for a comprehensive understanding of pandemic factors. Federated data also ensures strong privacy safeguards, minimising unauthorised access risks and fostering trust among stakeholders. Advancements in distributed analytical techniques for federated data are crucial for realising the full potential of early warning systems, promoting collaborative analysis without sharing raw data. These techniques enhance data quality, standardise analysis methods, and facilitate interoperability, resulting in more replicable and repeatable scientific outcomes. Overall, these advancements contribute to the development of robust predictive models and global-scale early warning systems for detecting and monitoring infectious diseases.

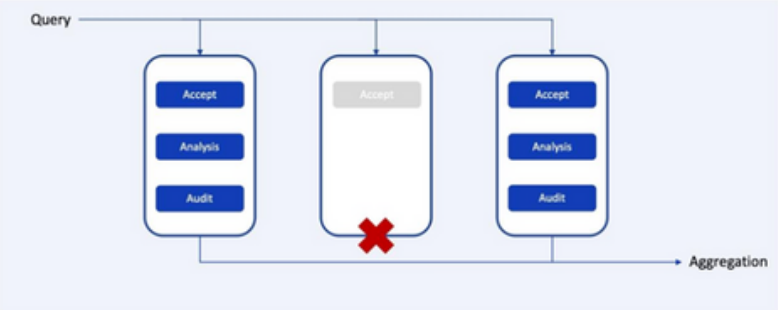


Figure 9. FAIR Federated Analysis of Aggregation Based on Permissions per Locale where Data is Produced for Queries

Integration of Data on Incidence of COVID-19

This use-case focuses on FAIRifying data related to the impact of COVID-19 on migrants, refugees, and asylum seekers. The three-phase process involves transforming a database, compiled from interviews, articles, and reports, into FAIR data to address the lack of representation of vulnerable communities, particularly those in Africa. Recognizing the mobility of these groups and the need for better access to services and inclusion in COVID-19 prevalence data was considered crucial. By making the data FAIR-compliant, it ensured cross-analysis accessibility under well-defined conditions throughout the data lifecycle. The study, conducted in Tunisia, Libya, and Niger, utilised 118 interviews and 565 reports on COVID-19 prevalence among migrants. FAIR curation, following VODAN parameters, involved creating new templates and vocabularies. Analysis revealed a low COVID-19 incidence (5%) and limited healthcare access for migrants. Despite border closures, the study uncovered the development of new, riskier routes by smugglers. This use-case demonstrates how federated curation with VODAN parameters facilitates analysis across diverse datasets.

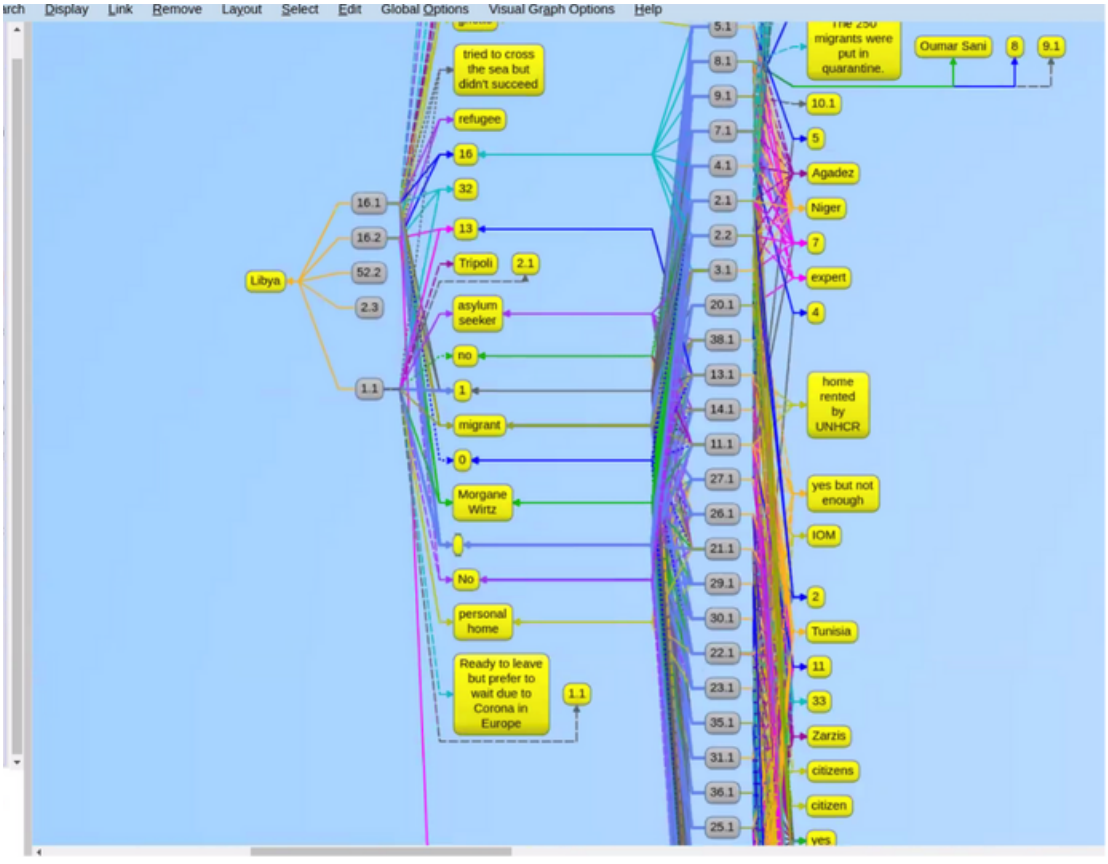


Figure 10. Sample of a Knowledge Graph Created in Allegrograph from FAIRified Data on COVID-19 among Refugees, Migrants and Asylum Seekers in Tunisia, Libya and Niger

Regional Dashboard for Monitoring at Bureau-level in Tigray

The Tigray Regional Health Bureau oversees health facilities in the region. This study explored the potential reuse of data to monitor key health performance

indicators, such as diagnostics and treatment. Health facilities employed a dashboard at the facility level, enabling the Tigray Regional Health Bureau to monitor aggregated health data in near real-time (Van Reisen et al., 2022). The VODAN parameters organise data curation at the source as machine-readable data and metadata, retained within the producing health facility. Data and metadata are then transferred from local facilities to regional and national systems, following defined ownership and regulations. The use of dashboards provides benefits to both facilities and the regional administration, leading to enhanced decision-making, improved efficiency, and ultimately better patient care.

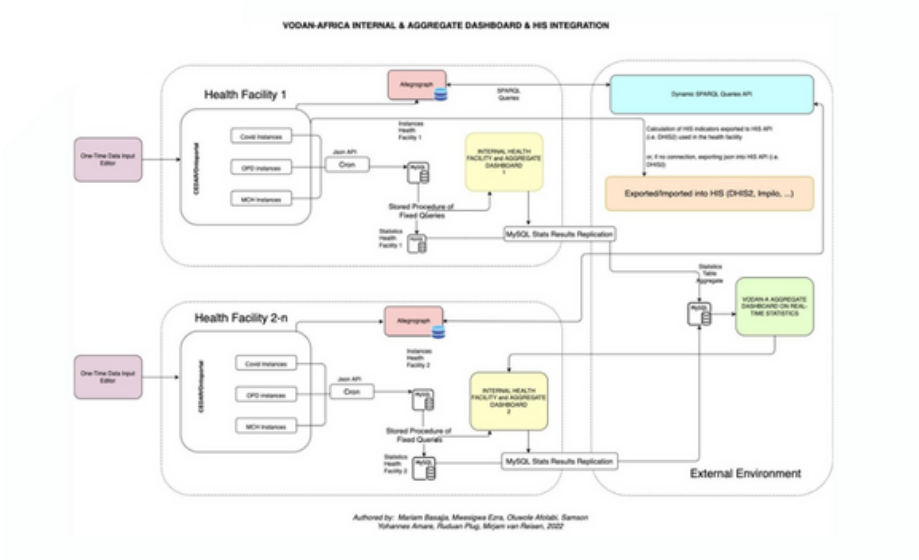


Figure 11. VODAN-Architecture for a One-Data Entry Machine-Actionable Semantic Curation for a Multiple Functionalities Architecture Based on Queries through Data-Visiting of Federated Local Depositories in Allegrograph

Syphilis Case in Ayder Referral Hospital, a University Hospital

The application of VODAN data was tested in the routine screening and treatment of syphilis during antenatal care (ANC) visits at Ayder Comprehensive and Specialized Hospital in Ethiopia, a health facility utilising VODAN data curation. VODAN offered a software platform facilitating data production, visualisation, and remote queries using the RDF query language SPARQL. The platform enabled a straightforward visual representation of syphilis cases (reactive and non-reactive) and adolescent pregnancies. By refining the visual query, warning signs for congenital syphilis (CS) were identified. CS occurs when a mother with syphilis passes the infection to her baby during pregnancy. This patient data analysis provided early warning signs, triggering the treatment of CS.

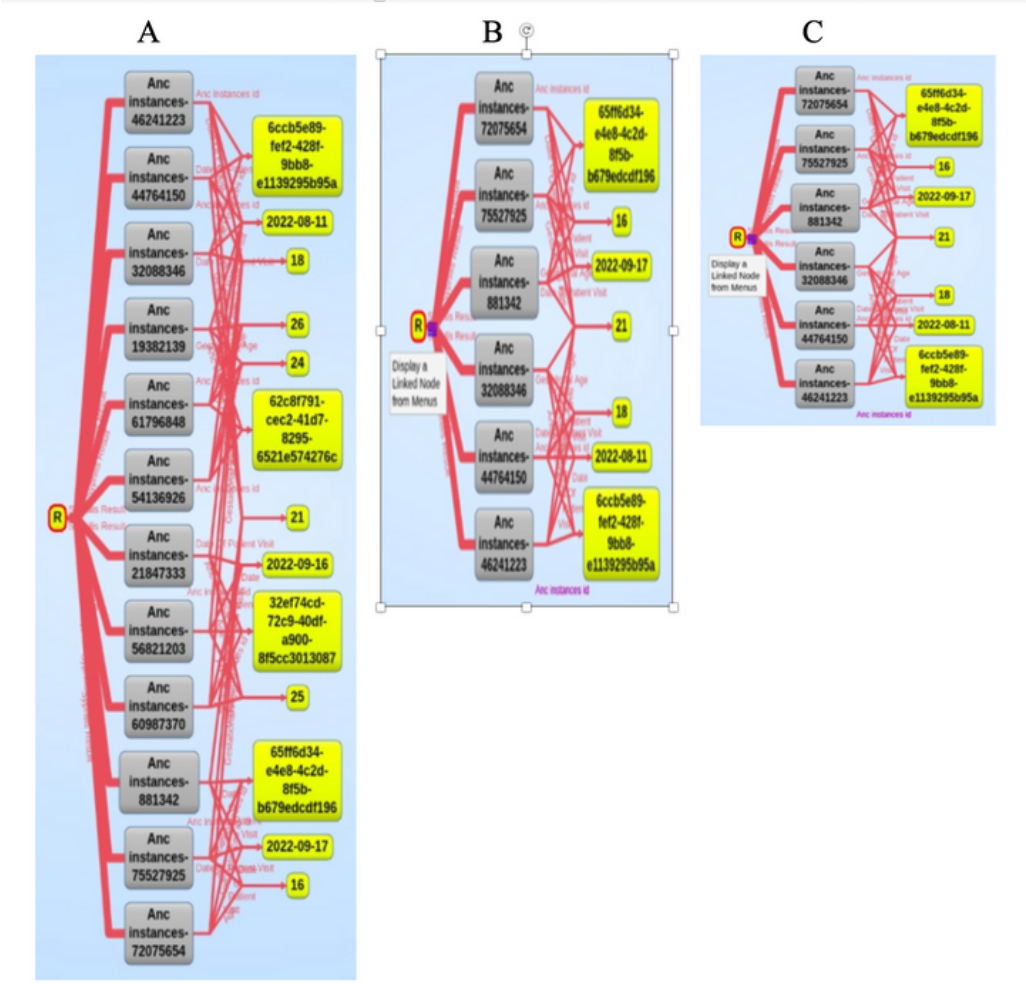


Figure 12. (A) All Syphilis Reactive Cases; (B) Syphilis Reactive Adolescent Cases; (C) Congenital Syphilis (CS) at Risk Cases

FAIR Based Data Architecture for Scaling-up Health System Implementation Research: The Saving Little Lives Project

The Maternal, Adolescent, Reproductive, Child Health (MARCH) research center at Mekelle University, Ethiopia, implemented VODAN principles in the Saving Little Lives project, a scaled-up research initiative across the four major regions of Ethiopia. The research aimed to (i) incorporate data from a greater number of health facilities, (ii) utilize various VODAN data points for research data inclusion, (iii) integrate federated locally stored data within health facilities, and (iv) standardize machine-actionable semantic data production across different sites in Ethiopia.

The project employs standardized indicators accessible across all sites, facilitating alignment with VODAN standards for tools and formats. This approach enables the creation of a dashboard for effective monitoring of data and information by the team coordinator. Furthermore, the adoption of VODAN principles for data ensures reusability, allowing the utilization of project data for future research and projects long after the project's completion.

Figure 13. Sample of a Knowledge Graph Created in Allegrograph from FAIRified Data on Labour and Delivery from Saving Little Lives Project in Tigray

Retrospective FAIRification to Identify the Drivers of Perinatal Care Outcomes in Kenya and Tanzania (MomCare)

This case study illustrates how the content and structure of data contribute to actionable insights (value creation) for the MomCare program, a perinatal care intervention spanning pregnant women in Kenya and Tanzania. MomCare connects vulnerable populations with health services through a mobile health application for accessibility, patient registration, and billing data (Aksünger, 2022). Although currently, aggregated data guides resource allocation, advanced analytics focusing on specific use-cases, such as identifying individual risk factors or analysing interdisciplinary factors at a country level, are not derived from the data. Presently, the data is not curated as machine-actionable semantic data at the source.

Through an explorative design, this case study highlights the limitations of retrospectively applying the FAIRification process. The establishment of a semantic model and the creation of a triple-store graph database resulted in the loss of provenance and quality of semantic data that could have been provided at the source. Consequently, this case study offers clear recommendations on the importance of data curation as machine-actionable semantic instances at point of production.

FAIR Data Curation of Clinical Research Data on Vaccines

This use-case focuses on establishing FAIR metadata for the Female-only Controlled Human Schistosoma Mansoni Infection Model (CoHSI2). The study's metadata pertains to clinical research involving controlled human infection with female schistosoma mansoni cercariae. The study aimed to assess the efficacy and safety of infection in healthy volunteers, utilizing primary and secondary outcome measures and monitoring associated adverse events. This curated metadata contains crucial information for monitoring intervention impact and improving adverse event management related to schistosomiasis infection.
The meticulously curated metadata has been made publicly accessible through a FAIR Data Point. This FAIR Data Point, hosted on an Azure instance in the Netherlands, ensures the availability of (meta)data in a machine-readable semantic format. Widely used ontologies in the field enhance the quality of the (meta)data. Key metadata fields include patient registration visit, event description, ICD-10 code, onset and end date and time, severity, relationship to the CoHSI, and classification of the event as a serious adverse event. Additionally, information about the treatment given in response to the adverse event is provided.

Integration of Cardio-related Patient Generated Health Data

Monitoring vital signs at home is crucial for patients with cardiac-related conditions like hypertension. Key indicators such as pulse rate, blood pressure, body temperature, and respiration rate, typically checked during episodic health facility visits, are increasingly monitored at home for chronic diseases. Interoperable home-generated data, integrated with health facility records, enables effective individual patient monitoring.

Utilizing the VODAN data-pipeline, a low-end technology was identified for measuring vital signs in hypertensive patients at home, following VODAN parameters for parsing and curation. An interactive voice response (IVR) system, tailored to local languages, serves indigenous patients in Africa for healthcare (Demena et al., 2020). The produced data is parsed to the VODAN-based digital patient health data system in the health facility, ensuring secure storage and interoperability with the patient's unique identifier. This allows for the identification of abnormal readings, aiding health workers in prompt response to patients

exhibiting concerning signs. A real-life implementation of this architecture will be tested as per the presented Figure 1.6 (Kawu et al., 2023).

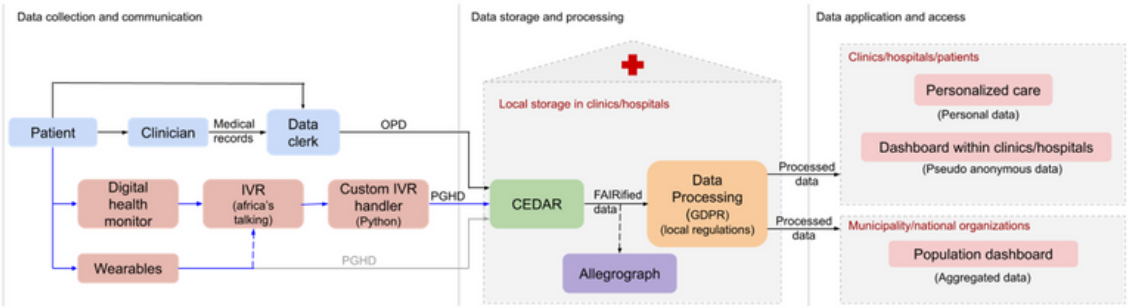


Figure 14. Proposed Architecture of Federated Patient Generated Health Data (PGHD) Interoperability with Patient Data Curated at Health Facility

The Value of Digital Patient Data in low-connectivity areas

One of the experiences of VODAN is that particularly in health facilities that are remote or that are in conflict-related situations, the model is especially relevant for the following reasons:

- Lack of diagnostic tests where the data can help identify patients at risk, for instance, to determine the risk of mother-child transmission of syphilis, HIV, or other infectious diseases.
- It is very difficult to enter the data at national or global level, so the access to the data within the health facility is a real asset to the doctors or health workers.
- Those health facilities are often poorly resourced but they have capabilities for data entry and therefore can gain additional financial resources by offering it for data visiting in the health data space.
- There tends to be a bias in AI product on health because these populations are often not represented in the data on which the AI is computed. This helps prevent such biases.

SUSTAINABLE MODEL FOR UPSCALING

Value addition is created over the data through data curation and additional services for hosting the data in the places where the digital data instances are created. In the VODAN context, this is usually the health facilities, but it can also be extended to university hospitals, universities and other relevant data producers. This data can be engineered into applications that facilitate services at point of care as well decision making by administrations, physicians or health workers, to support insights that help diagnostic and treatment considerations. Additionally, this data can be used for pandemic surveillance and for reviewing trends that offer relevant health insights. Another application is daily updates and visualisations that support understanding of development at point of care.

The data produced in this manner contributes to enhancing the quality of healthcare services at the point of care. These insights can be generated when permission is granted at the regional level, allowing regional health administrations across facilities with shared objectives to extract valuable insights from the data. This process facilitates the development of collaborative applications and the monitoring and surveillance of data trends within the community, extending even across national boundaries in federated formats. Hence, the production of this data by which value is added through particular services will assist these service provisions at point of care.

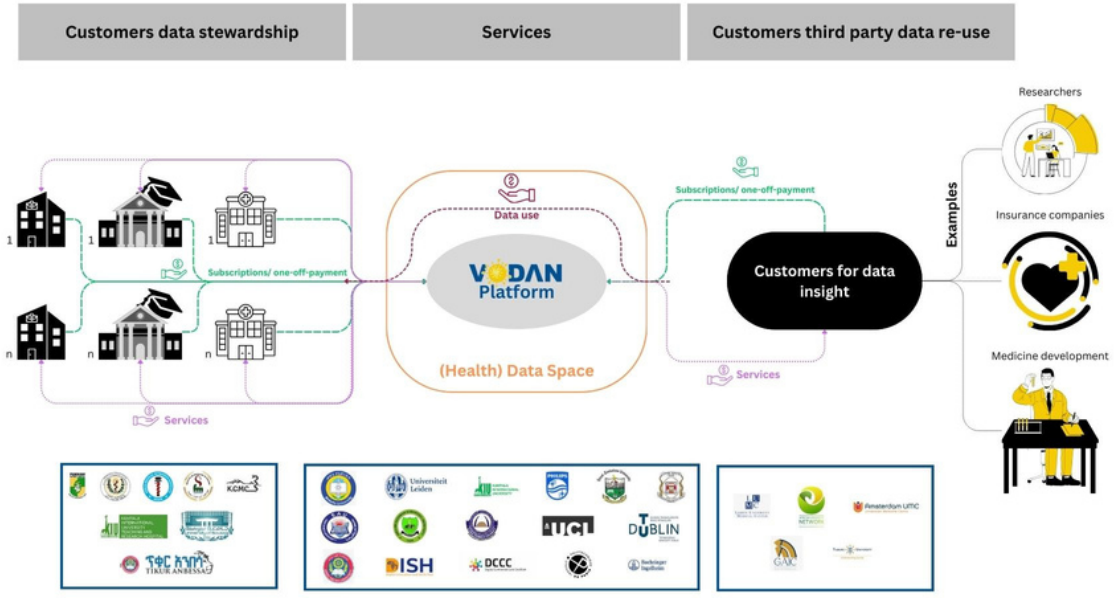


Figure 15. FAIR Sustainability model

The FAIR sustainability model is a platform business model of FAIR services. In this model the data curation and hosting services are covered by resources which are generated from the third party insights obtained from the FAIR data.

VODAN is in a position to offer services to assist health facilities in generating and hosting FAIR patient data.

These services include:

- Data stewardship
- Training and support of data clerks
- Data quality management
- Software to support the hosting of these services
- The implementation of queries over the data to generate new insights
- Visualisation of the data in APIs.

Since the data can be accessed under clearly defined conditions, this architecture offers room for third party customers to benefit from this data and to channel new resources that can cover the costs of the data production and the services necessary to host the data. This is critical for the sustainability of the FAIR data space. The third party customers would need to have an agreement with VODAN on the interoperability and the reuse of the data available in the health data space. The parties that offer the data must also provide consent for such operations to be carried out and will financially benefit from the data visiting.

ORGANISATIONAL DEVELOPMENT

VODAN-Africa is founded under the senate of Kampala International University (KIU) and established as a foundation reporting to the vice-chancellor of KIU as the chair of the VODAN-Africa foundation. The organisation functions under the legal framework of Uganda, whereas the foundation is established under Belgian law and seated in Brussels. The programme is governed by the VODAN-Africa Board, which consists of a Chair, Executive Director and all Country Coordinators, responsible for implementation in their respective geographies. The foundation reports directly to the VODAN-Africa Board under KIU and has the power to handle finances under the supervision of the VODAN-Africa board.

The Board is supported by a Steering Committee which includes overseas partners to VODAN-Africa. The Steering Committee has established a Consortia of partners who provide funding or other contributions to VODAN-Africa. The figure below visualises this organisational set-up.

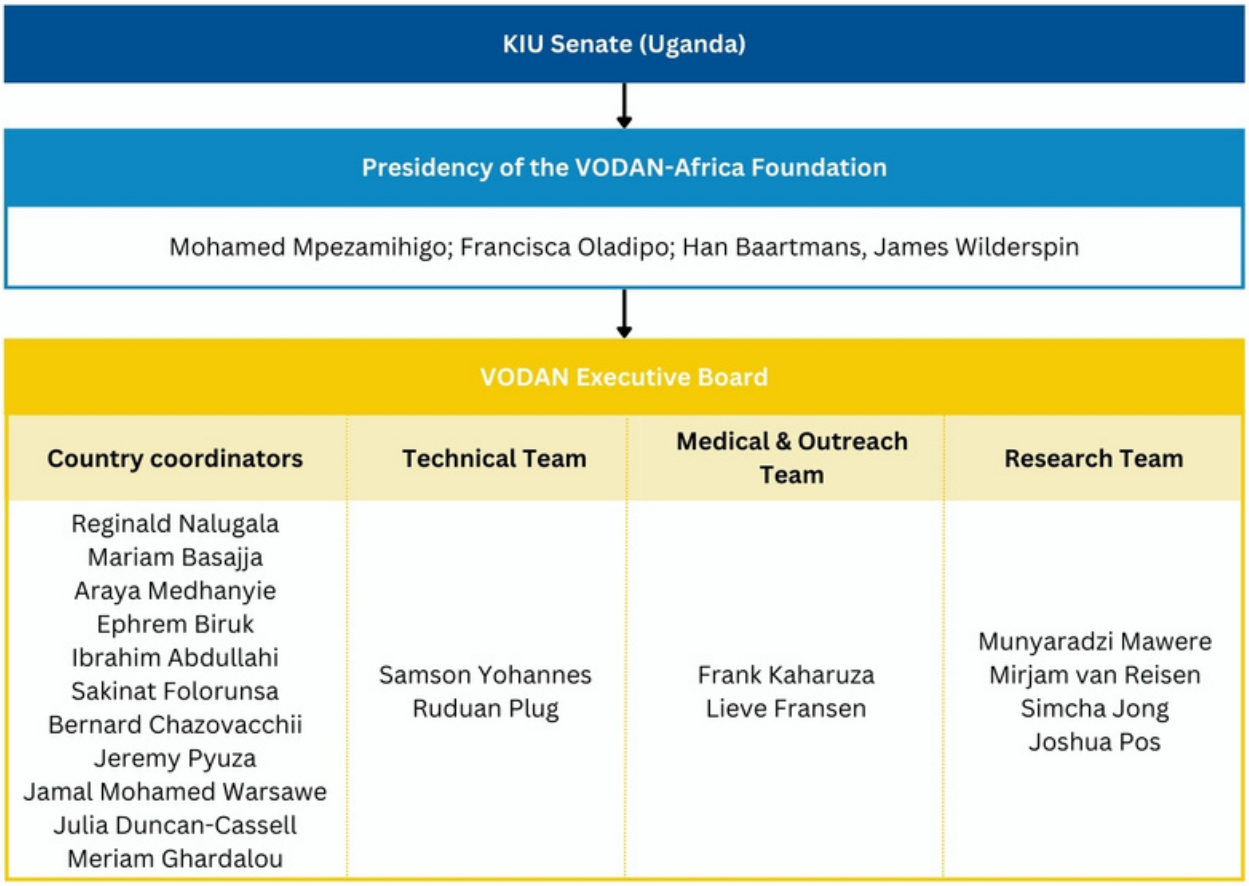


Figure 16. VODAN governance structure.

SUPPORT

- A start-up grant by Philips Foundation;
- A Google grant to develop the workflow;
- A grant to demonstrate Proof of Concept by Invest International, Philips Foundation, CORDAID and Go-FAIR Foundation;
- A grant by IMS to strengthen VODAN-Africa data stewardship in Africa;
- A grant by IMS to strengthen data production for refugee patient pods in Africa;
- A contribution by Philips to the development of workflows;
- A grant of Nuffic on the development of data stewardship training curricula;
- A contribution by LUMC research for the development of a workflow on vaccine effectiveness in neonates;
- An in-kind donation by Philips for the development of a business plan;
- A contribution from Accenture for one PhD on disruptive business models to investigate the Africa Health Data Space proposition implemented with the Polytechnical Institute Paris;
- An in-kind donation by Microsoft to assist in Machine Learning analytics over federated data through data visiting;
- An in-kind donation of Achmea to provide an expert on the second phase of the business development;
- An in-kind donation of Software Innovation Group to advise on software development;
- An in-kind donation of Achmea to advise on Requirements and Specifications of the software adaptation;
- A collaboration with the Dutch government on a PhD studying automated permission control in the VODAN-Africa architecture;
- A collaboration with Amsterdam University Medical Service on one PhD on federated analytics for surveillance through data visiting in VODAN-Africa;
- A collaboration with Wageningen University on Data Stewardship Training;
- A collaboration with Leiden Institute of Advanced Computer Science (LIACS) offering one PHD;
- A collaboration with Dublin Technical University hosting one PhD and supervision;
- A collaboration with Ayder hospital in Ethiopia offering facilities to the technical research and engineering team;
- A collaboration with Tangaza University, hosting the coordination of VODAN-Africa;
- A collaboration with the IBBUL University in Nigeria, supporting the data production and deployment team in Nigeria;
- An in-kind collaboration with Kampala International University offering the communication and development infrastructure for VODAN-Africa.

WHAT PEOPLE SAY ABOUT US

UNESCO (2021)

"For instance, the principles underlying research at the Kampala International University in Uganda, centre around 'collaboration within context', bolstering its lead in the multi-country Virus Outbreak Data Network (VODAN) to manage African COVID-19 pandemic data within the precepts of data sovereignty. However, valiant efforts are hampered by an overreliance on expertise and funding from outside the continent."

DCCC (2023)

"VODAN Africa -which aims to unleash the full potential of health data – currently deployed in 90 healthcare facilities across 8 African countries."

"According to Africa CDC's Jean Philbert NSENGIMANA (Phil) , government's biggest responsibility is to now create an enabling environment for public-private partnerships in digital health, to foster meaningful collaboration without hindering innovation."


DCCC
Digital Connected Care Coalition

Accelerating the digital health transformation towards Universal Health Coverage

Fall | Issue 9 | November 2023

Recaps, up-coming events and interesting readings

Digital Health is a collaborative affair - that was the resounding message during the Africa Health Tech Summit. We need to team up for interoperability, connected care chains, health data access, and more - all are needed to reach better and universally accessible patient care.

In the words of Dr Amit N. Thakker: "Collaboration is the new Innovation!" and as DCCC, we invite all digital health stakeholders, from healthcare payers to providers and from governments to pharma and tech companies to join forces with us as we work to implement inclusive, impact and scalable digital health solutions, like the ones presented in this newsletter.

As always - more impactful work is in the making. We will share the latest through our LinkedIn page and in our next newsletter. On to better accessible, equitable healthcare!

For the Core Team - Jeroen Maas

10-YEAR TIMEFRAME

VODAN has established a ten-year time frame with milestones, against which to measure progress towards the development from Proof of Concept in 2020 to the full fledged integration of the Africa Health Data Space in the Global Data Space, to serve public and private sector customers.

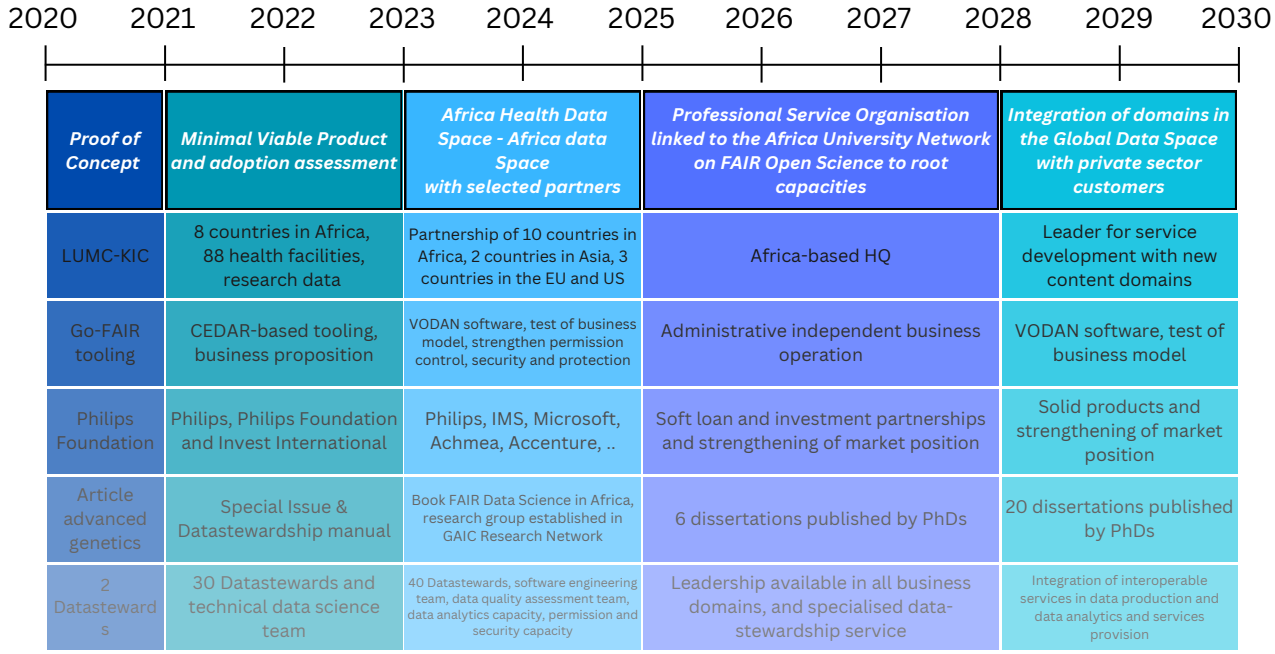


Figure 17. VODAN's 10-year timeframe.

In order to achieve this VODAN needs to invest in a range of areas:

- development of the federated software of mini services to be more user friendly and adapted to context;
- to facilitate the input of data with the intelligence layer obtained through the CEDAR services;
- expansion of the research in key areas, such as automated permission; federated data analytics and integration of legacy systems;
- development of use cases to demonstrate the value of the data curation;
- building up of the sustainable business case;
- establishment of a service oriented organisation;
- the creation of a high-level consortium to support the governance of the Africa health data space;
- ensuring that tools across geographies are FAIR and standards are established and agreed to support this.

INTERDISCIPLINARY INSIGHTS IN THE DATA SPACE

Data Spaces of other disciplines would be structured in the same way as the Africa Health Data Space, presented above, consisting of three layer: the data creation layer, the service layer and the intellgence layer.

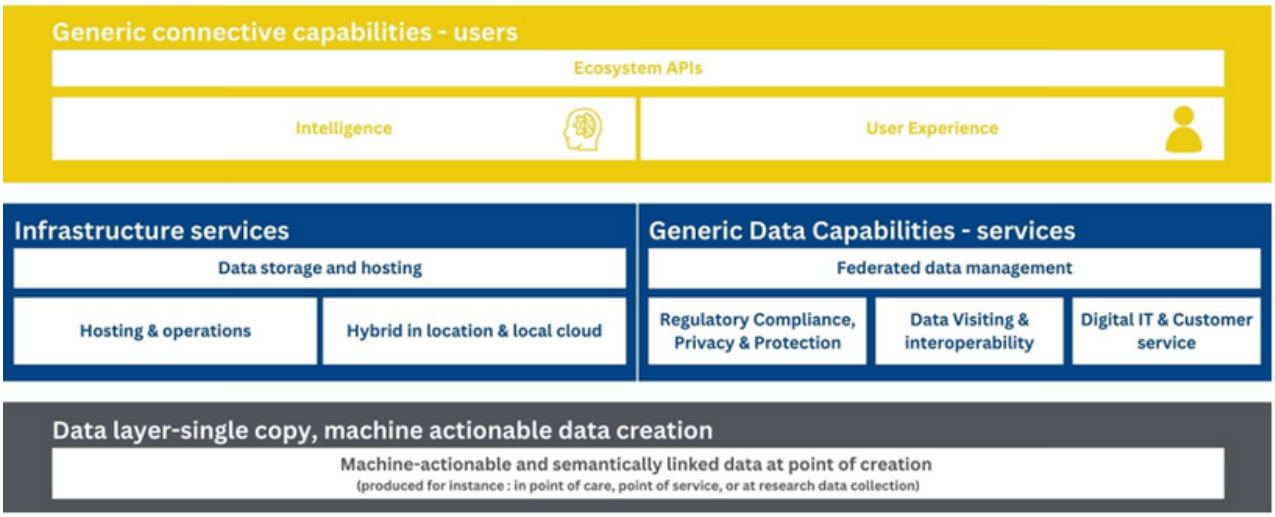


Figure 18. The proposed landscape fo the federated data space.

Within the Federated Data Space, the data is protected and it is ensured that it is only used with consent of those from which the data is produced. The data is held within the communities where it originates from, but interoperable across different communities.

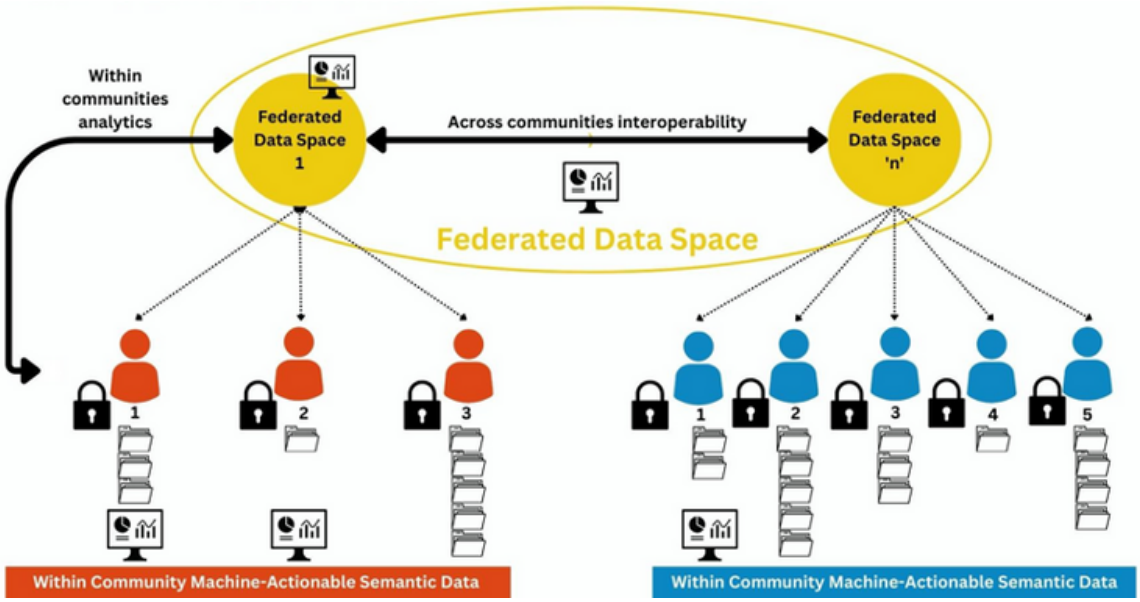


Figure 19. FAIR-OLR Federated Personal Pods.

Interoperability Across Sectors Analysis: Data on Human Trafficking

The human trafficking research data use-case focuses on developing FAIR tools to share data while safeguarding the identity of victims and survivors. This approach allows controlled access and reuse of data under defined conditions. Human trafficking data, particularly regarding health outcomes, lacks coordinated FAIRification efforts. Existing tools for finding trafficking victims on the web have not translated into the necessary tools for FAIRifying research data, especially concerning refugees (Hultgren et al., 2016). VODAN standards were applied for health outcomes data curation, involving the construction of ontologies and tools for converting data into a FAIR VODAN-compliant format. The researchers explored creating knowledge graphs from the data and testing their interoperability with other datasets.

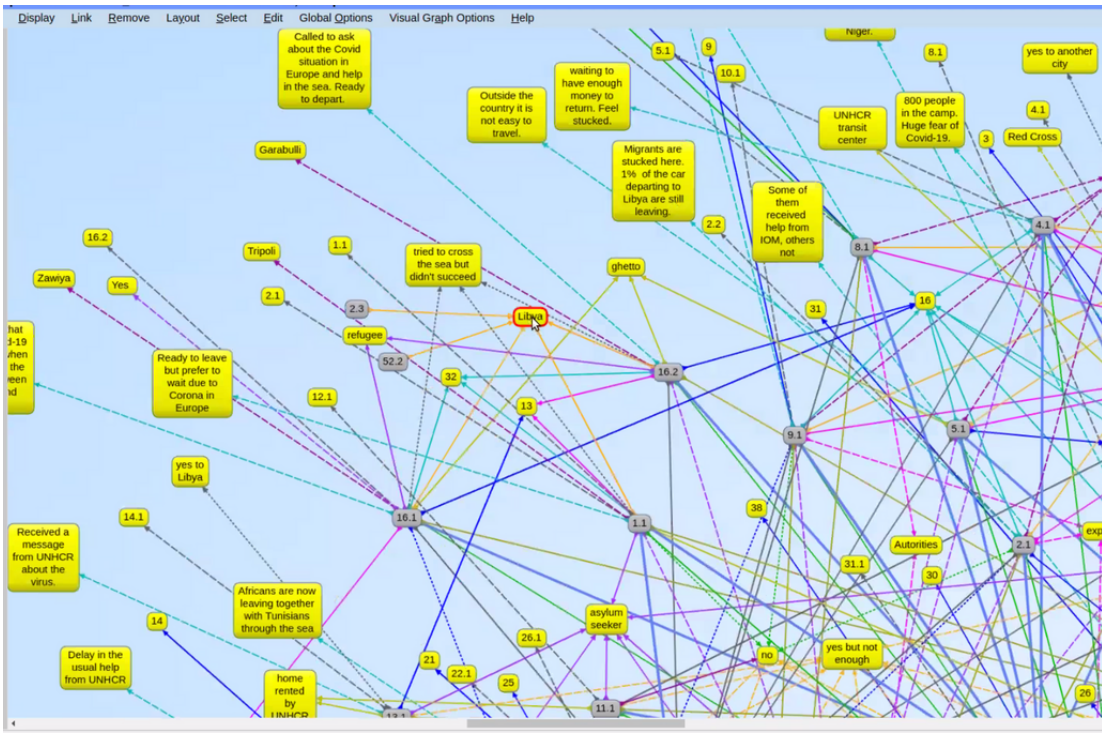


Figure 20. FAIR Data obtained from interviews with migrants regarding access to health facilities during the Corona pandemic in 2020.

This use-case demonstrates the interoperability and the re-usability of the FAIRified data. The data can be integrated with patient data to complement the knowledge about the challenges that different population groups were facing during the COVID-19 pandemic. The use-case is easily replicable with new data.

This research was supported by a grant of the Dutch research organisation NWO. The research was carried out in 2020. The results demonstrate the relevance of de novo FAIRification at the point of creation of a data instance. This allows the data to maintain qualities of provenance, and original attributes of the data-instance, when it was created.

More use cases across different domains include

Levels of toxicity and the prevalence of certain cancers.

Water quality and cholera

Use of media data on epidemic trends and patient data.

Examples of third parties customers include

New data from underrepresented communities for development of products by insurance providers.

The side effects of medicine by pharmaceutical companies.

The implementation of medical studies in the data.

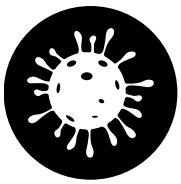
RESEARCH INNOVATION AND TRAINING

Fieldlabs

Master students FieldLabs have greatly contributed to the research and innovation and demonstrating the relevance of FAIR-by-Design practices through concrete use cases. The following FieldLabs were run in 2023.

Development of a regional administration patient data dashboard for pandemic monitoring based on federated algorithm

- **Getu Tadele and Samson Yohannes**



FAIR-data integration and exploration of FGS, HPV and HIV prevalence in Uganda

- **Mildred Akandinda, Bwaga Ibrahim and MariamBasajja**

Use-case of a risk analytics of patients with syphilis for maternal health care in multicountrydata-production

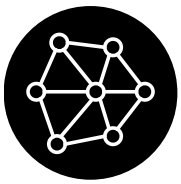
- **Samson Yohannes**

Vascular disease risk analytics through interoperability of FAIR patient records with wearables

- **Abdullahi Kawu and Rens Kievit**

Data interoperability of relevant information pertaining to elderly with dementia in The Netherlands

- **Lars Schrijver and Ria Landa**



Development of a FAIR-based ontology to identify nodes and graphs of perpetrators in human trafficking data

- **Kai Smits**

Development of a FAIR-based ontology to identify risk places in human trafficking data

- **Kai Smits**

FAIR Data stewardship training

A critical aspect of the upscaling of VODAN will be the potential to train datastewards.

VODAN-Africa has developed a range of resources to train datastewards, who can work with stakeholders and assist in the creation of FAIR-data instances.

As part of the DISH e-learning project, the Mekelle University team has developed a specific course dedicated to training of datastewards. The test of the course will be completed in 2023 and this test-run will help to improve the quality of the course. The course material includes explanatory ciurricula, powerpoint presentations, video material, demonstration material and quizes.

Wageningen University has developed a course on FAIR Data Stewardship, which follows the same approach as the DISH e-learning project. Thiscourse is also available.

DISH

https://drive.google.com/drive/folders/12zZN6WkFs0H3kQMiVdZeRtorDTmlml25
This course provides a structured skilling of data stewardship across domains to gain certification at three levels of data stewardship.

WAGENINGEN
UNIVERSITY & RESEARCH

https://www.wur.nl/en/activity/the-data-science-and-artificial-intelligence-dsai-carnival.htm
This is a course for advanced researchers on research data FAIRification.

GO FAIR
M4M

https://www.go-fair.org/how-to-go-fair/metadata-for-machines/
This is a course to introduce FAIR processes in a community that wants to discover the practice of FAIR-data within the community

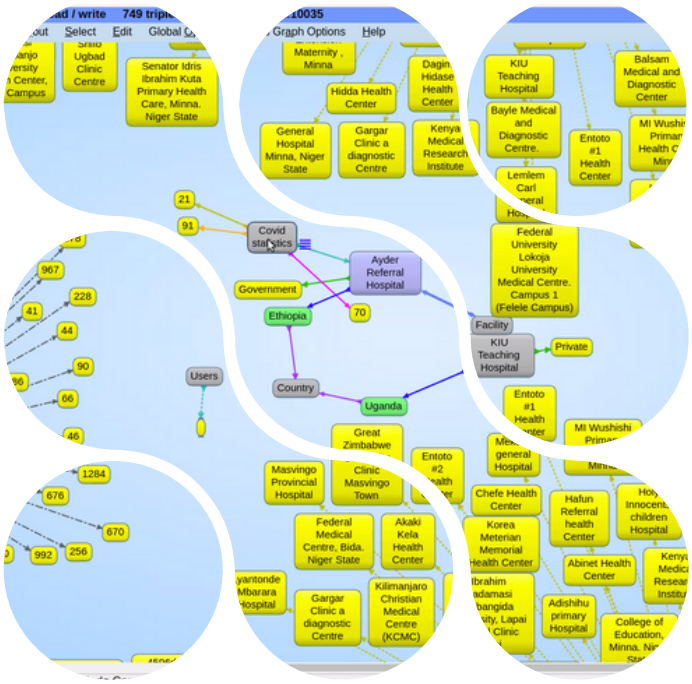
Researchers

VODAN relies on research to advance the knowledge required to innovate in intelligent architectures for knowledge integration.

The research carried out by VODAN allows the team to investigate specific problems, test solutions and document the research.

What makes VODAN distinct is the following:

- An artistic open environment, with collaboration from different settings and across domains to strengthen creativity;
- A commitment to be inclusive, especially oriented towards environments that experience difficulties of inclusion in the digital era;
- A non-comformist radical dedication to test solutions in any space, whatever the difficulties, in the belief that this will lead to the creation of understanding and new knowledge;
- An interest to incorporate highly diverse researchers;
- The desire to work in ethnographic approaches that connect various stakeholders to investigate the problem and develop pathways towards solutions;
- A rigorous evaluative peer-process towards joint exploration of problems, potential solutions and to test these in different settings;
- Sharing of insights across domains to help investigation from different angles and disciplines to help imaginative approaches;
- A joint responsibility of team building among researchers, to help and assist each other and to build collaboration based on the various strengths available;
- An orientation towards mentoring and training of students and youth to become leaders in this field.



Abdullahi Kawu

FAIR-data based Interoperable Digital Generated Data

Aliya Aktau

Vocabulary creation for interoperable FAIR-data

Charles Kahiro

Models for ethical value creation on federated FAIR Data production

Getu Tadele Taye

Modeling a regional surveillance health system of FAIR-data

Ibrahim Bwaga

Identifying communities of differentiated risk profiles

Joëlle Stocker

FAIRification of soundscape data for resilience modeling

Kai Smits

Human Trafficking analytics in Libya

Kudakwashe Kindoza

Deployment factors of federated FAIR-Data for interoperable solutions

Liya Manu

Integration of harmonised federated FAIR Data information for off-line use

Mariam Basajja

Creation of a FAIR-data based digital information system in Uganda

Mildred Akandinda

Identifying communities of differentiated profiles in Uganda

Morgane Wirtz
Migrants health analytics in Tunisia
Natascha Buchs
Business disruption through the FAIR-data based Federated Data Space
Putu Hadi Purnama Jati
GDPR-based access and control permission architecture
Rens Kievit
Automated permission controls for a GDPR compliant secure dynamic architecture
Ruduan Plug
Statistical models for federated FAIR Data models of privacy data produced
Samson Yohanes Amare
Federated software services for FAIR-data
Tesfit Gebremeskel
Modeling of vocabularies of federated FAIR-data

DOCUMENTATION AND PUBLICATIONS

[1] van Reisen, M., Amare, S.Y., Plug, R., Tadele, G., Gebremeskel, T., Kawu, A.A., Smits, K., Woldu, L.M., Stocker, J., Heddema, F., Folorunso, F., Kievit, R., Medhaniye, A.A. (2023 - forthcoming). Curation of Federated Patient Data: A Proposed Landscape for the Africa Health Data Space. *Elsevier*.

[2] Van Reisen, M., Amare, S.Y., Nalugala, R., Taye, G.T., Gebreselassie, T.G., Medhanyie, A.A., Schultes, E. & Mpezamihigo, M. (2023). Federated FAIR principles: Ownership, localisation and regulatory compliance (OLR). *In: FAIR Connect*, 1(1), 1-7. IOS Press. DOI: https://doi.org/10.3233/FC-230506

[3] Amare, S. Y., Taye, G. T., Gebreselassie, T. G., Plug, R., & van Reisen, M. (2023). Realizing health data interoperability in low connectivity settings: The case of VODAN-Africa. *In: FAIR Connect*, 1(1), 55-61. IOS Press. DOI: https://doi.org/10.3233/FC-221510

[4] Kawu, A. A., O'Sullivan, D., Hederman, L., & Van Reisen, M. (2023). FAIR4PGHD: A framework for FAIR implementation over PGHD. *In: FAIR Connect*, 1(1), 35-40. IOS Press. DOI: https://doi.org/10.3233/FC-230500

[5] Cardoso Silva Ferreira, A., & van Reisen, M. (2023). Hourglass-based interoperability through nanopublications in VODAN-A. *In: Fair Connect*, 1(1), 5-11. IOS Press. DOI: https://doi.org/10.3233/FC-221512

[6] Gebreslassie, T. G., van Reisen, M., Amare, S. Y., Taye, G. T., & Plug, R. (2023). FHIR4FAIR: Leveraging FHIR in health data FAIRfication process: In the case of VODAN-A. *In: FAIR Connect*, 1(1), 49-54. IOS Press. DOI: https://doi.org/10.3233/FC-230504

[7] Van Reisen, M., Stokmans, M., Mawere, M., Basajja, M., Ong'ayo, A., Nakazibwe, P., Kirkpatrick, K., Kudakwashe, C. (2020). FAIR practices in Africa. *In: Data Intelligence 2(2020)*, 264–275. doi: 10.1162/dint_a_00049 (eds. Mons, B., Jacobsen, A. & Schultes, E.). MIT Press, http://www.mitpressjournals.org/dint

[8] Van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A., Kirkpatrick, K., Mons, B., (2020). Towards the Tipping Point of FAIR Implementation. *In: Data Intelligence 2(2020)*, 264–275. doi: 10.1162/dint_a_00049 (eds. Mons, B., Jacobsen, A. & Schultes, E.). MIT Press, http://www.mitpressjournals.org/dint

[9] Jacobsen, A., Miranda Azevedo, de R., Juty, N., Van Reisen, M., ..., Schultes, E. et. al. (2020). FAIR principles: interpretations and implementation considerations. *In: Data Intelligence 2(2020)*, 264–275. doi: 10.1162/dint_a_00049 (eds. Mons, B., Jacobsen, A. & Schultes, E.). MIT Press, http://www.mitpressjournals.org/dint

Special Issue Articles

[10] Van Reisen, M., Mons, B.: Introduction to the Special Issue: Data Intelligence on patient health records. *Data Intelligence 4(4)* (2022). doi: 10.1162/dint_e_00165

[11] 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)* (2022). doi: 10.1162/dint_e_00166

[12] Plug, R., Liang, Y., Aktau, A., Basajja, M., Oladipo, F., Van Reisen, M.: Terminology for a FAIR framework for the Virus Outbreak Data Network-Africa. *Data Intelligence 4(4)* (2022). doi: 10.1162/dint_a_00167

[13] 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)* (2022). doi: 10.1162/dint_a_00168

[14] Lin, Y., Purnama Jati, P.H., Aktau, A., Ghardallou, M., Nodehi, S., Van Reisen, M.: Implementation of FAIR Guidelines in selected non-Western geographies. *Data Intelligence 4(4)* (2022). doi: 10.1162/dint_a_00169

[15] Basajja, M., Van Reisen, M., Oladipo, F.: FAIR Equivalency with regulatory framework for digital health in Uganda. *Data Intelligence 4(4)* (2022). doi: 10.1162/dint_a_00170

[16] Taye, G.T., Amare, Y.S., Gebremeskel, G.T., Medhanyie, A.A., Ayele, W., Habtamu, T., Van Reisen, M.: FAIR Equivalency with regulatory framework for digital health in Ethiopia. *Data Intelligence 4(4)* (2022). doi: 10.1162/dint_a_00172

[17] 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)* (2022). doi: 10.1162/dint_a_00176

[18] 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)* (2022). doi: 10.1162/dint_a_00179

[19] 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)* (2022). doi: 10.1162/dint_a_00180

[20] 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)* (2022). doi: 10.1162/dint_a_00181

[21] 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)*, (2022). doi: 10.1162/dint_a_00184

REFERENCES

DCCC. (2023). *Accelerating the digital health transformation towards Universal Health Coverage.* Issue 9. Available at: https://app.mailerlite.com/n3s9g4t0p0

Frédéric Bruly Bouabré, letter to André Magnin, November 1, 1988, in Frédéric Bruly Bouabré, vol. 4, l'homme dans le monde, un géant touche le soleil, ed. Magning, Yaya Savané, and Denis Escudier (Paris: éditions Xavier Barral, 2013), 132. Translation by Jeanine Herman. Cited in Nzewi, Ugochuwkwu-Smooth C. (2022). Frédéric Bruly Bouabré: World Unbound. *Moma*.

Martinon, Jean-Paul (ed.). (2013) The Curatorial: A Philosophy of Curating, *London: Bloomsbury.*

Nzewi, Ugochuwkwu-Smooth C. (2022). Frédéric Bruly Bouabré: World Unbound. *Moma*.

UNESCO, ICEE. (2021). *Engineering for sustainable development: delivering on the Sustainable Development Goals.* Available at: https://unesdoc.unesco.org/ark:/48223/pf0000375644
.

Yakel, E. (2007). Digital Curation. *OCLC Systems and Services: International Digital Library Perspectives*, 23(4), pp. 335–340. Available at: https://doi.org/10.1108/10650750710831466.

COLOFON

Authors: Mirjam van Reisen, Joëlle Stocker, Samson Yohannes Amare, Tesfit Gebremeskil, Kai Smits, Ruduan Plug, Putu Hati Purnama Jati, Aliya Aktau, Mariam Basajja, Abdullahi Kawu

Title: Towards an Africa Health Data Space. Vodan Phase III

Year of publication: 2023

Published by VODAN-Africa

Reference: Mirjam van Reisen, Joëlle Stocker, Samson Yohannes Amare, Tesfit Gebremeskil, Kai Smits, Ruduan Plug, Putu Hadi Purnama Jati, Aliya Aktau, Mariam Basajja, Abdullahi Kawu, Natascha Buchs, Getu Tadelle (2023). Towards an Africa Health Data Space. Vodan Phase III. *Report of Activities.* VODAN-Africa.

Overall coordination: Mirjam van Reisen, Prof FAIR Data Science LUMC
Images: Joëlle Stocker
Text and editing: Joëlle Stocker
Technical lead: Samson Yohannes Amare, Ruduan Plug
Ontologies: Aliya Aktau
Modeling of Electronic Health Data: Tesfit Gebremeskil
Review of the data production: Abdullahi Kawu, Putu Hadi Purnama Jati
Data production: Samson Yohannes Amare
Review of fields and quality of data sets: Mariam Basajja
Integration of interdisciplinary data: Kai Smits
Business model of the Health Data Space: Natascha Buchs
FAR Data Stewardship Training Course: Getu Tadelle, Aliya Aktau, Kai Smits
Data Stewardship support: Nambobi Mutwali

We thank all colleagues in VODAN for their contributions and support to the otucome achieved. We are grateful to Philips for the support and we hope this report contains valuable knowledge of the feasibility of a Health Data Space.

We thank all partners for their generous support.