

VODAN in a Box: Proof of Concept

Mariam Basajja

Leiden Institute of Advanced Computer Science
Leiden University, Netherlands

Mariam Basajja is currently pursuing a PhD in Computer Science at Leiden University in the Netherlands. She obtained a masters degree in Advanced Computing, machine Learning and Data mining from the University of Bristol, UK with a distinction and also acquired a masters of science in computational Intelligence from University of Nairobi, Kenya. She also holds a bachelors in Applied Computer Technology with a concentration in Software Engineering from United States International University Africa with a distinction. Her main areas of interest are around FAIR data, big data, databases, data analytics, cloud computing, ICT Infrastructure and others.

Marek Suchánek

Faculty of Information Technology,
Czech Technical University in Prague, Czechia

Marek Suchánek is a PhD student at Faculty of Information Technology, CTU in Prague and also at Faculty of Business and Economics, University of Antwerp. He gained both master and bachelor degrees from FIT CTU in Prague with a distinction. His main field of study is connecting conceptual modelling with sustainable software development using Normalized Systems. He is part of the Data Stewardship Wizard development team from the very beginning but participates also in other projects related to software design and development. Marek also has experience with computer networks (winner of Cisco NetRiders 2015) and software testing & quality (IEEE QRS 2017 prize).

VODAN-in-a-Box (ViB) is a composition of multiple **FAIR-enabling and open-source services** with a single goal: to support **gathering electronic Case Report Forms as FAIR Data** in a machine-actionable way but without a need to expose or transfer data outside the facility (so-called on-premise solution). As the name suggests, it is related to the activities of GO FAIR's implementation network called **Virus Outbreak Data Network**. It aims to **promote FAIR Data** during the current and also future virus outbreaks. The most significant partner **VODAN-Africa**, a network of universities and hospitals in Uganda, Ethiopia, Nigeria, Kenya, Tunisia, and Zimbabwe, participated actively in **M4M workshops** and on proving the concept of ViB solution. **Valuable feedback and beneficial experience** emerged during the deployment across African countries for both local experts and developers of the solution. Due to the **Proof of Concept** that we present in this poster, future development and use of VODAN-in-a-Box can be outlined.

Deployment On-Premise

The ViB deployment is designed to be relatively simple, even though it consists of **multiple interconnected services**. It is distributed as an **open-source GitHub repository** with necessary configuration files and docker-compose.yml. Therefore, **Docker and docker-compose** are required for running the ViB both locally as well as in production mode for higher security and availability. ViB connects **CRF Wizard (specialized Data Stewardship Wizard, DSW), FAIR Data Point (FDP), and triple stores**. Both FDP and DSW are highly-configurable and **actively developed open-source projects**. To deploy ViB in production mode, the teams from VODAN-Africa network had to prepare **DNS records** for those services including generating **SSL certificates** for secure access. Some configuration steps are needed to be done manually due to security reasons (e.g. setting secrets, passwords, and permissions) or local configuration (e.g. own metadata layers in FDP). Thanks to **successful collaboration within the network** and provided documentation, there were only minor issues during the deployment-related to some misconfiguration or problems acquiring certificates. **Kampala International University deployed the first ViB in production on 22nd July 2020**, just a one day after its full demonstration during a regular M4M workshop.

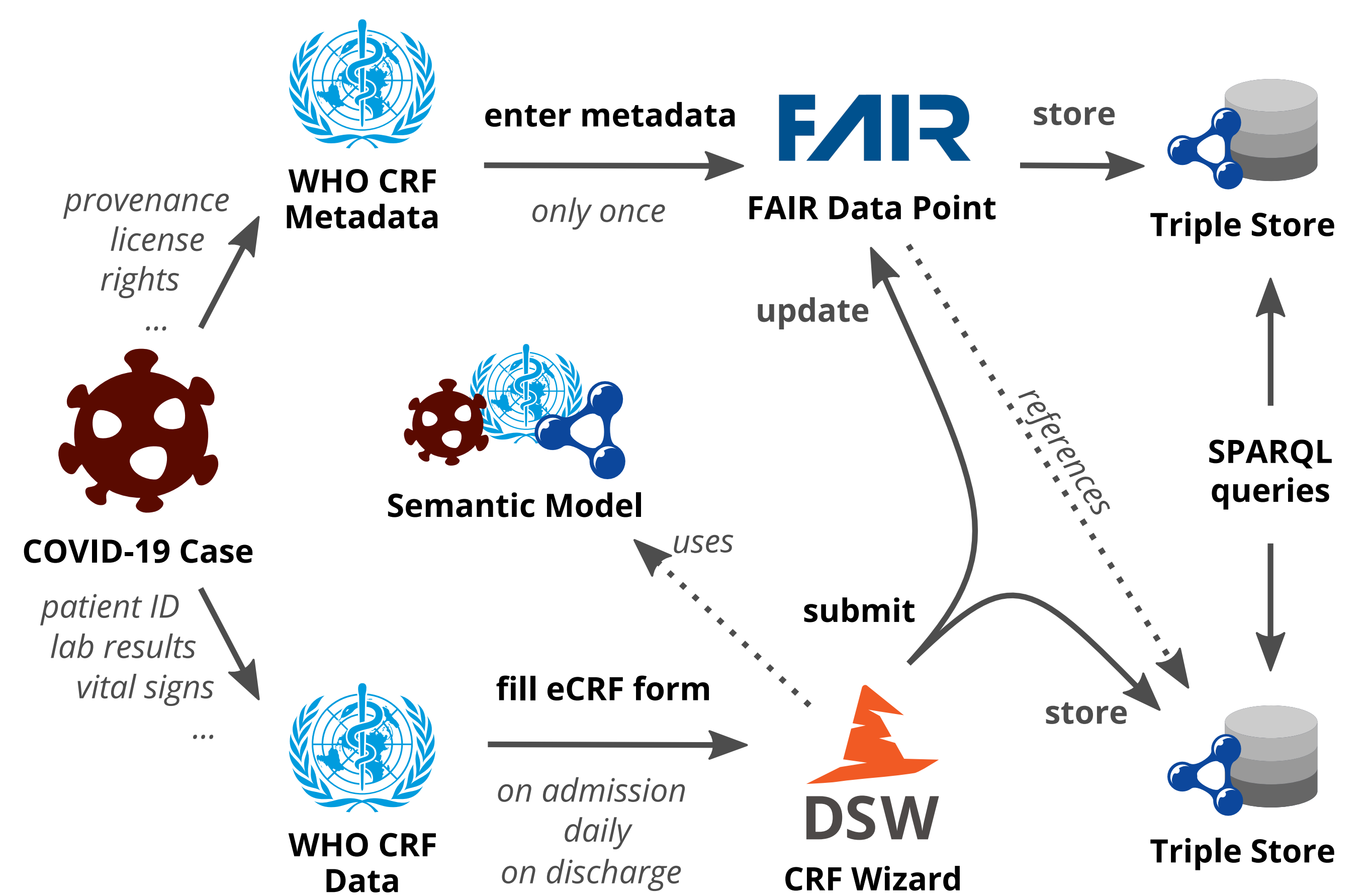
Proof of Concept

On 29th June 2020, a query test between the Netherlands and Africa with data held in residence **was carried out successfully**. The already installed FAIR datapoints support automation of cross country data exchange of observational **COVID-19 patient data**. Allows hospitals, facilities and researchers in different countries to **query FAIR compliant data** at the same time while fully complying with **local data protection and governance laws**. Mock data was fed into the VODAN-in-a-Box using the CRF Wizard tool whose semantic model was originally based on the **WHO COVID-19 form standard**. The test was demonstrated by Mariam Basajja and Erik Flikkenschild.

This was done using **SPARQL** which is an RDF query language. A three-database **distributed SPARQL query** was entered into the AllegroGraph query interface, which is also part of the ViB. Within this testing query, three service call requests were made using the SPARQL endpoints of the three countries' (Uganda, Kenya, and Netherlands) to get and display the facility name against the number of COVID-19 patients.

Future Steps and Call to Action

In the ongoing Covid-19 pandemic, very few data are publicly available in Africa. It is not easy to tell whether these data do not exist in reality, or they do exist, but access to them is very limited due to several reasons. The establishment of the Covid-19 FAIR Data Points shall help address the challenges of **data accessibility and interoperability** that gravely affect public health responses to disease outbreaks. It will also serve the following purposes: give Africa an **important role in the fight against the Coronavirus pandemic**; **ensure data ownership** (African data resides in Africa and avoid digital data removal to warehouses elsewhere); enable Africa to be a **huge resource of verified data and strengthen data-informed health systems** for Africa and the world. In addition to this, **further ViB development is being planned** to allow local customizations and other data types. This everyone can **try it on their own** and **even improve due to its open-source nature**.



```
##start##
#defaultView:Map
PREFIX vodian: <http://purl.org/vodian/whocovid19crfsemdatamodel/>
PREFIX obo: <http://purl.org/obo/owl/obo/>
PREFIX vodian_inst: <http://purl.org/vodian/whocovid19crfsemdatamodel/instances/>
PREFIX wd: <http://www.wikidata.org/prop/direct/>

SELECT DISTINCT ?facility_name (COUNT(DISTINCT ?crf) AS ?num_of_patients) {
  ?crf a vodian:who-covid-19-rapid-crf ;
  obo:BFO_0000051 ?module_1 ;
  obo:BFO_0000051 ?module_3 .
  ?module_1 a vodian:Module_1 ;
  obo:BFO_0000051 ?module_2 a vodian:Facility_name ; vodian:has_literal_value ?facility_name .
}
```

facility_name	num_of_patients
"The Zambezi Hospital"	"1"
"KIUTH"	"13"
"The Zambezi Hospital "	"1"
"KMITC"	"1"
"LUMC"	"4"
"Kenya Medical Training College (KMITC) Isolation Center"	"1"
"Zambezi Hospital"	"1"
"kenya"	"1"
"Kenya Medical Training Center (KMITC)"	"1"
"ZAMBEZI HOSPI"	"1"

VODAN FAIR Data Point

Demonstration WHO COVID-19 CRF catalog

A catalog containing a number of datasets created using the VODAN in a Box toolset. These datasets have mocked-up metadata for demonstration purposes only.

Datasets

ACME Hospital COVID-19 CRF synthetic dataset

Dataset containing the case reports of COVID-19 for the ACME Hospital

[DROID_0080600](#) [MONDO_0100096](#) [RADKjIGPL_2mE90jB3YQX6WGGdCC8ZwPkxElGHsXOJE](#)

[who-covid-19-rapid-crf](#) [SIO_001410](#) [Q84263196](#)

Issued 08-07-2020 Modified 25-07-2020

Metadata Issued 08-07-2020 Metadata Modified 25-07-2020

Version 0.1

Language en

License cc-by-nc-nd4.0

Issued 08-07-2020

m.basajja@liacs.leidenuniv.nl
suchama4@fit.cvut.cz

docs.vodan.fairdatapoint.org

github.com/VODAN-Tech

