

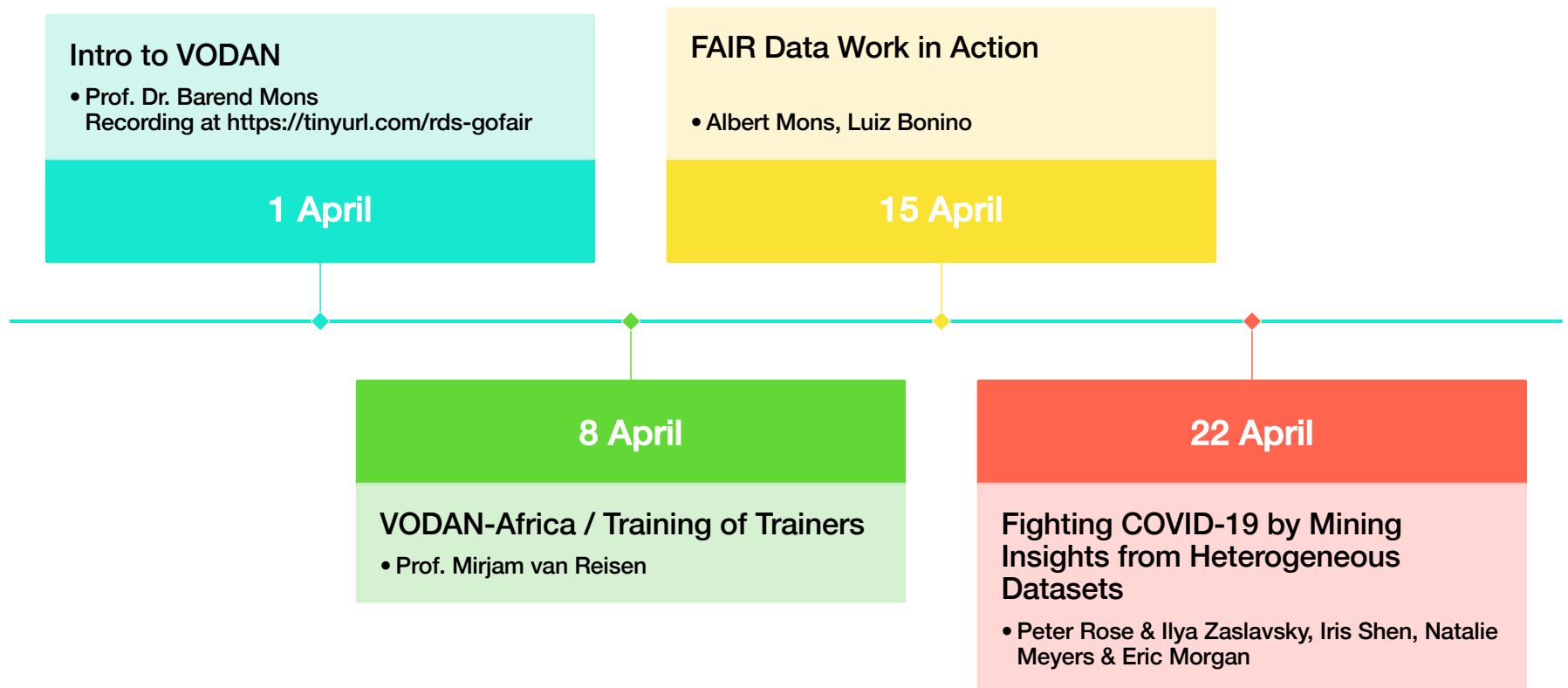
GO FAIR US Webinar Series:  
FAIR Data & COVID-19

Prof. Dr. Mirjam van Reisen  
VODAN-Africa / Training of  
Trainers (ToT)

8 April 2020

#VODAN

# Webinar Series Overview



# Virus Outbreak Data Network: Training of Trainers

## Key Messages



- By Prof. Dr Mirjam van Reisen, Professor 'Computing for Society', Leiden Institute for Advanced Computer Science, Leiden University
- In collaboration with Kampala International University, Mekelle University, Great Zimbabwe University, Olabisi Olanjo University, Leiden University, GO FAIR Foundation and Philips Foundation
- April 8, 2020 18:00 CET, Webinar @ Centre for Super Computers, [San Diego Super Computer Centre](#)

# Outline



Objectives



Training of  
Trainers



Next steps



# I. Objectives



CORONAVIRUS

# Coronavirus exposes lack of common data approach



NEWS IN BRIEF

TODAY, 17:41

Polish deputy PM resigns over May election date

TODAY, 16:21

Spain plans to introduce universal basic income

TODAY, 12:13

EPP calls for EU data centre for pandemic response

TODAY, 07:27

Three arrested in deadly French 'terror' attack

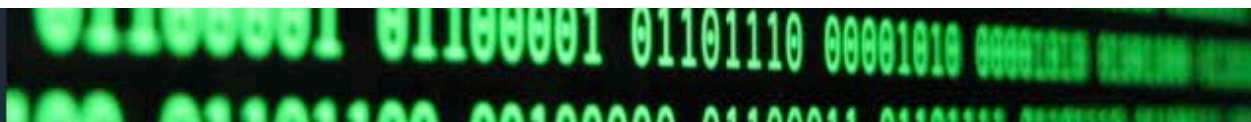
TODAY, 07:26

"One of the main problems is the duplication of data and lack of coordination between countries", said Mirjam van Reisen, who is part of the [Virus Outbreak Data Network \(VODAN\)](#) at the University of Leiden.

PRIVACY POLICY

CORRECTIONS

ADVERTISING



Some experts believe that the EU's current approach to data collection may be missing an opportunity to fight the coronavirus outbreak (Photo: [Christiaan Colen](#))

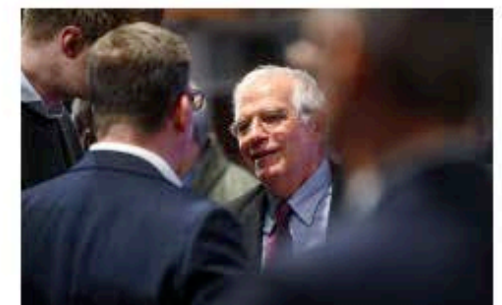
By **ELENA SÁNCHEZ NICOLÁS**  

BRUSSELS, TODAY, 07:21

The enormous differences between coronavirus cases reported worldwide raises questions on how countries are tracking their coronavirus outbreaks - or deliberately underreporting them.

3. APR, 17:04

EU waives customs duties, VAT on vital medical imports





## Lack of common data approach: defining the problem

---

- Increasing (unconnected) digital health data:
  - Patient data
  - Research data
  - Published articles
- CORD-19 is a resource of over 47,000 scholarly articles, including over 36,000 with full text, about COVID-19, SARS-CoV-2, and related coronaviruses
- Too much for manual handling, but current internet not equipped for *machine*-readability of data
- Data need to be prepared for human- and machine-readable Findability, Accessibility, Interoperability, Re-usability (FAIR)
- This constitutes the basis for the Internet of FAIR Data And Services



Internet of  
FAIR Data and  
Services -  
Connecting  
Principles

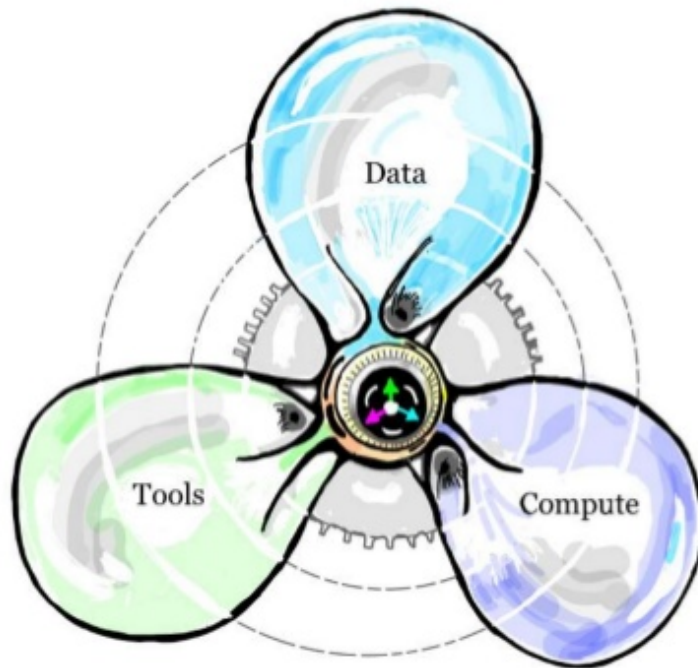
Australian Research Data Commons



Data Science meets AI:

Quality data for Federated AI Ready (FAIR) solutions

### The Internet of FAIR data and Services



- Distributed data
- Data remain in their place of origin
- Data are managed and stewarded in their place of origin
- Allows contextualized data analytics
- Fake data are detectable and traceable
- Allows Data sovereignty within legal and policy framework of the location where data is produced
- Connects fragmented data depositories
- Algorithms (tools) visit the data
- Computes and analytics on the basis of data visiting

# Specific benefits of VODAN for Africa



Avoid digital data removal to warehouses elsewhere



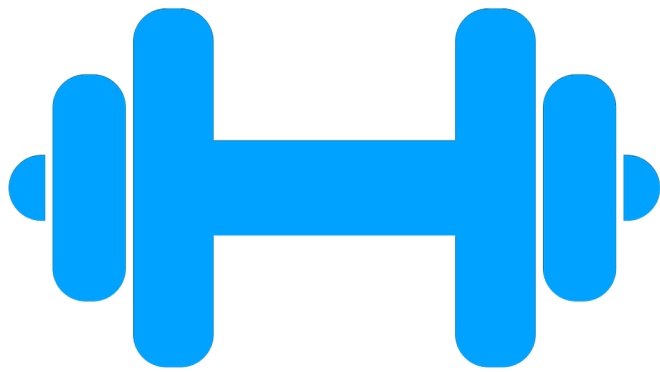
Strengthen data-informed Health Systems



Ensure data ownership and handling



Strengthen digital data stewardship and tooling



## II. Training for Trainers:

Building a Community fit  
for purpose

# VODAN:

## Establishing a solid network of trust and purpose



# Set up information channel with MoH / Bureau of Health



Approach

the MoH / Bureau of Health



Provide

information on VODAN and who  
are involved



Arrange

how you will communicate  
progress



Ask

their feedback



Request

endorsement



# Determining the governance & regulatory framework

Identify	the relevant policy documents and regulations on medical data handling;
Categorise	categorise the do's and don'ts of the medical data regulatory framework;
Involve	involve relevant people who understand laws and regulations on medical data handling;
Explain	the purpose of the task to the experts in MoH and hospital/clinic;
Ask	ask colleagues in MoH & hospital & clinics to review the document and improve it;
Check	if you have all the right people on board to determine the regulatory framework;
Involve	any senior people if necessary to do final check;
Present	a regulatory framework

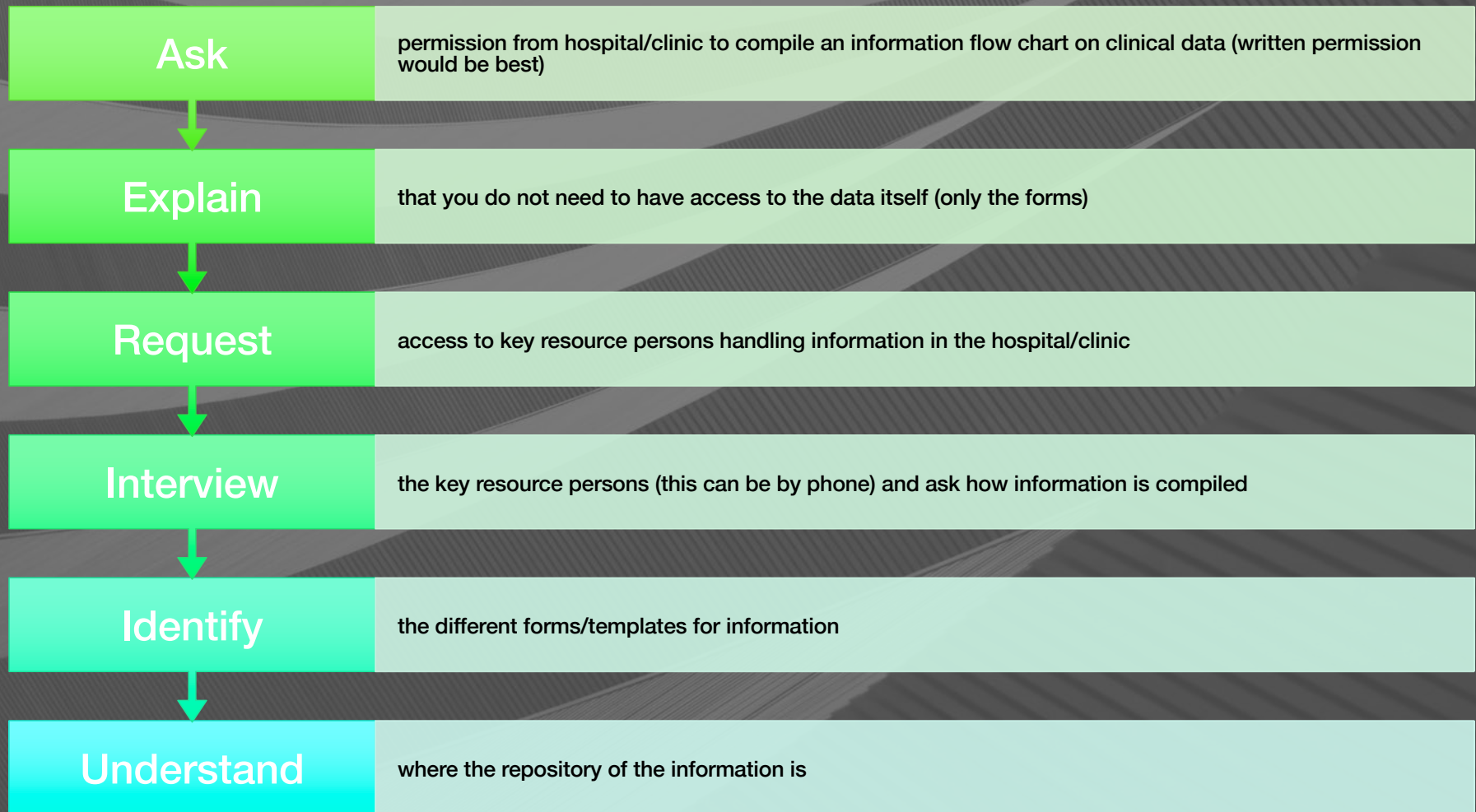


## Training of Trainers:

Questions to consider for identification of the hospital or clinic (health data producer)

- What is the hospital/clinic you will work with?
  - Preferably there is a working relationship and trust with the hospital management
  - Ensure clear communication on the reason for the involvement of the hospital
  - The data-manager of the hospital/clinic is invited to the training so that knowledge is transferred
  - Alternatively follow up training is arranged with the data-manager of the hospital/clinic
  - Multiple hospitals/clinics can participate, but for the first step it should be a manageable number of participating outfits – expansion will happen after the first successful steps are implemented

# Compilation of information flow charts





**MINISTRY OF HEALTH**

**OUTPATIENT REGISTER**  
(HMIS FORM 031)

**HEALTH FACILITY:** LIRA REGIONAL REFERRAL HOSPITAL

**DISTRICT:** LIRA

**HEALTH SUB - DISTRICT:** LMC

**OPENING DATE:** 16/08/2019

**CLOSING DATE:** 03/09/2019

**FEMALE**

Category	0-28days		29days-4 yrs		5-14yrs		15yrs & above		Total	
	M	F	M	F	M	F	M	F	M	F
New attendance	13	10	56	51	145	124	238	219	451	343
Re-attendance	00	02	44	10	26	35	45	75	149	110
Total Attendance	13	12	100	61	171	159	283	294	600	453

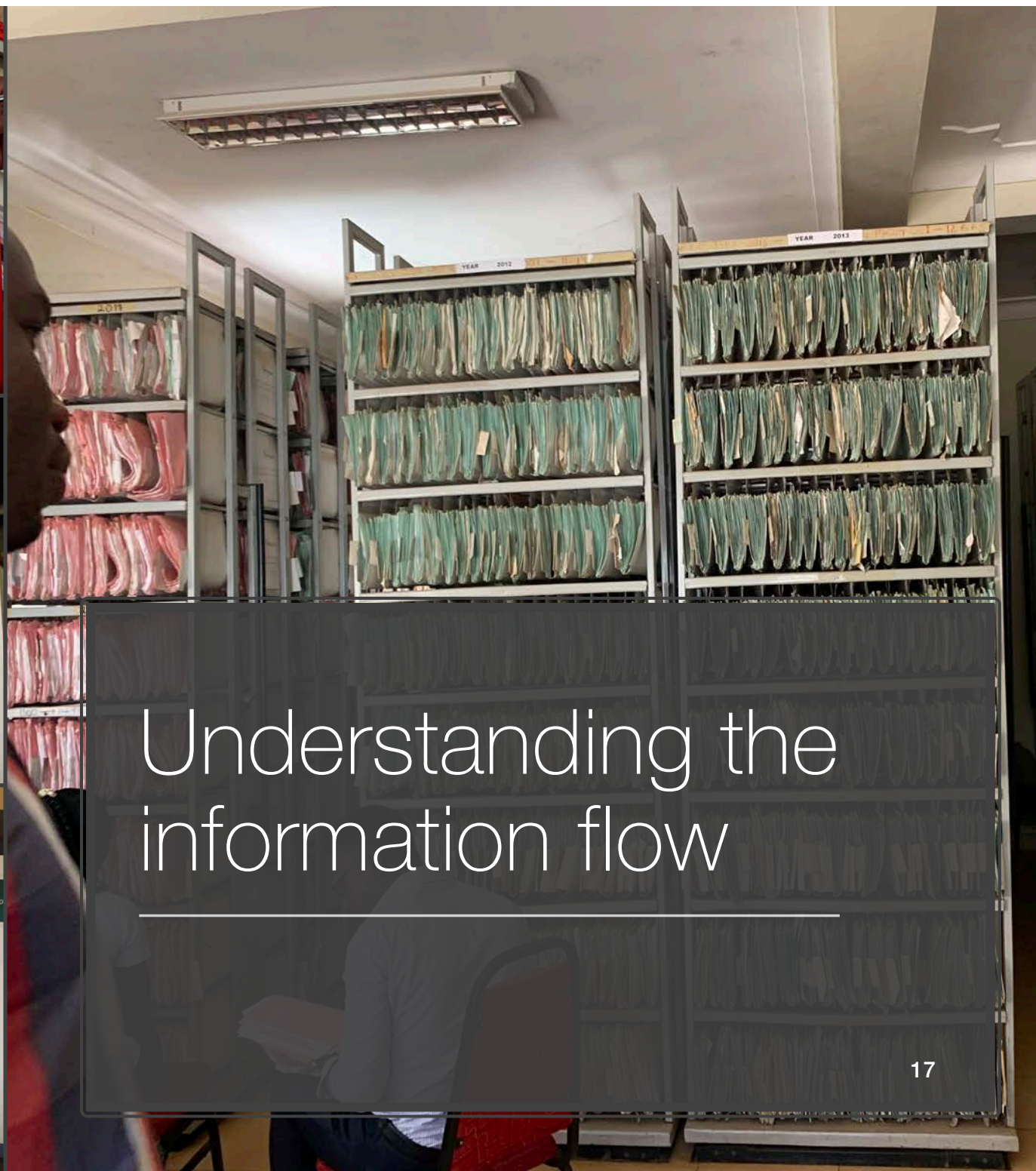
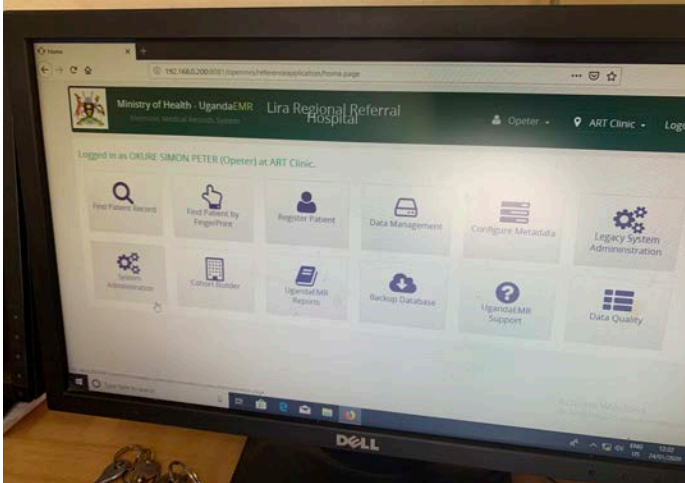
  

Category	0-28days		29days-4 yrs		5-14yrs		15yrs & above		Total	
	M	F	M	F	M	F	M	F	M	F
Referrals to unit	01	04	24	12	24	36	04	04	53	80
Referrals from unit	01	07	07	08	07	07	07	07	53	80

**1.3. OUTPATIENT DIAGNOSES FOR THE MONTH**

Diagnosis	0-28days		29days-4 yrs		5-14yrs		15yrs & above		Total	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1.3.1 Epidemic-Prone Diseases										
1. Acute Flaccid Paralysis	00	00	00	00	00	00	00	00	00	00
2. Acute Encephalitis (suspected rabies)	00	00	00	00	00	00	00	00	00	00
3. Cholera	00	00	00	00	00	00	00	00	00	00
4. Dysentery	00	00	00	00	00	00	00	00	00	00
5. Guinea Worm	00	00	00	00	00	00	00	00	00	00
6. Malaria	00	00	125	71	86	173	10	14	221	262
7. Measles	00	00	00	00	00	00	00	00	00	00
8. Bacterial Meningitis	00	00	00	00	00	00	00	00	00	00
9. Neonatal Tetanus	00	00	00	00	00	00	00	00	00	00
10. Plague	00	00	00	00	00	00	00	00	00	00
11. Yellow Fever	00	00	00	00	00	00	00	00	00	00
12. Other Viral Hemorrhagic Fevers	00	00	00	00	00	00	00	00	00	00
13. Severe Acute Respiratory Infection (SARI)	00	00	00	00	00	00	00	00	00	00
14. Adverse Events Following Immunization (AEFI)	00	00	00	00	00	00	00	00	00	00
15. Typhoid Fever	00	00	00	00	00	00	00	00	00	00
16. Presumptive MDR TB cases	00	00	00	00	00	00	00	00	00	00
Other Emerging Infectious Diseases specify e.g. Influenza like illness (ILI), SARS										
1.3.2 Other Infectious/Communicable Diseases										
17. Diarrhoea - Acute	00	00	00	00	00	00	00	00	00	00
18. Diarrhoea - Persistent	00	00	00	00	00	00	00	00	00	00
19. Urinary tract infections	00	00	00	00	00	00	00	00	00	00
20. Genital ulcers	00	00	00	00	00	00	00	00	00	00
21. Sexually Transmitted Infection due to (Sexual Genital Tract Infection)	00	00	00	00	00	00	00	00	00	00
22. Other Sexually Transmitted Infections	00	00	00	00	00	00	00	00	00	00
23. Urinary Tract Infections (UTI)	00	00	00	00	00	00	00	00	00	00
24. Intestinal Worms	00	00	00	00	00	00	00	00	00	00



Understanding the information flow



# Training of a COVID-19 Data Community

## 1. To establish the basics of the FAIR Data Point:

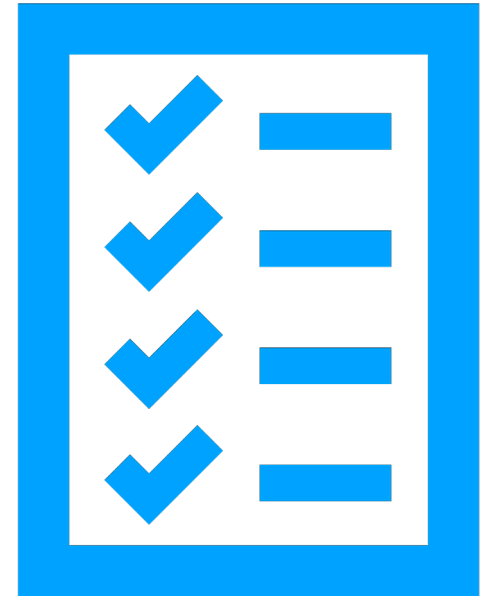
Officials in health administration (governance, regulatory framework)  
Health professionals (to work with the information flow and data handling in hospitals and clinics)  
Data scientists, semantic data modelers, interdisciplinary data stewards (to define and model the data in human and machine-readable ontologies and determine informatical models)

## 2. In subsequent phase:

Experts in Machine Learning and Artificial Intelligence, text mining tooling  
Experts in Algorithm auditing  
Experts with Legal knowledge for regulatory embedding of the use of the FAIR Data Point  
Clinicians and health analysts for health data analytics  
Researchers, experts for the formulation of hypotheses and judging cardinal assertions derived from data analytics



### III. Next Steps: implementation



# Decision-making in FAIRification of the COVID-19 Data Community

- Application of FAIR Principles based on:
  - Technical specifics (infrastructure)
    - code red
  - local authorities (governance and regulations) and domain specialists (doctors, virologists, researchers)
    - code blue



Technical infrastructure (generic operations)  
Data (domain-specific content)

## Box 2 | The FAIR Guiding Principles

### 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 standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization 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

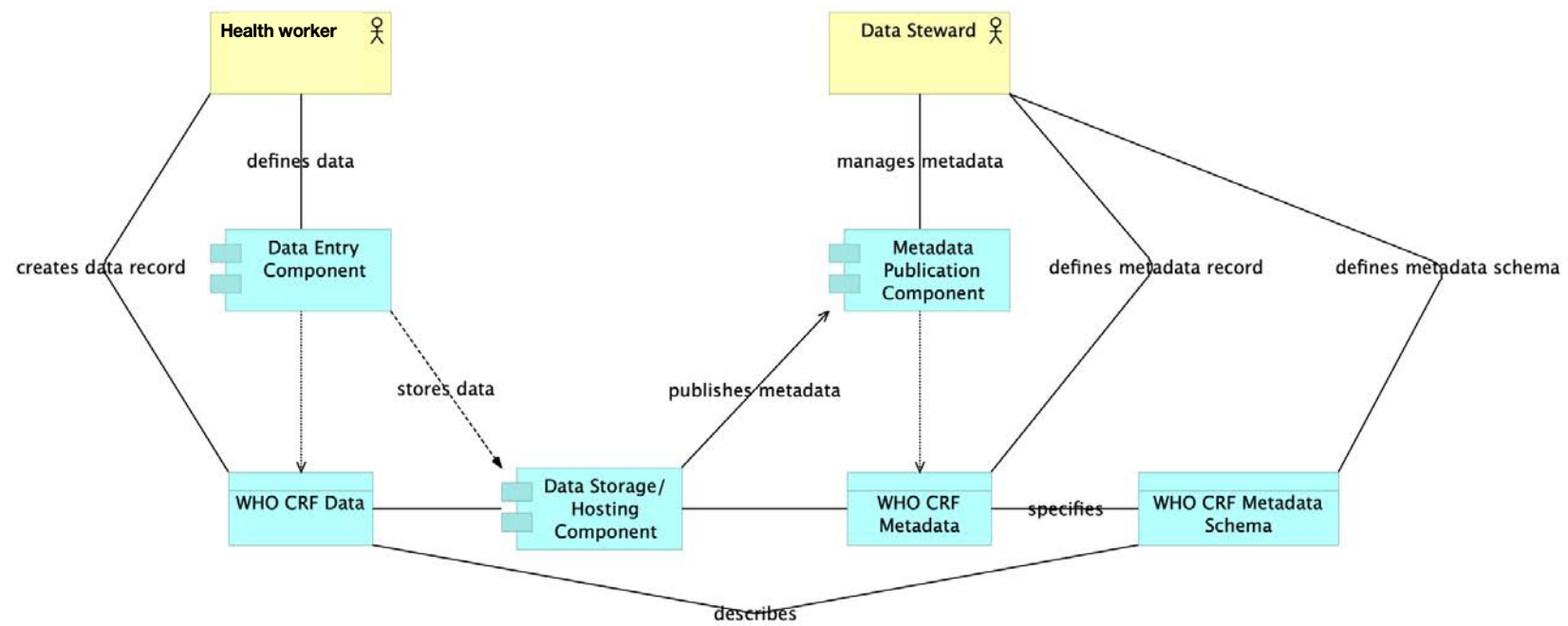
### To be Reusable:

- 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 license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards

The COVID-19 Data Community determines:  
Data as Open as Possible and as Closed as Necessary  
As Distributed as Possible, as Centralised as Necessary

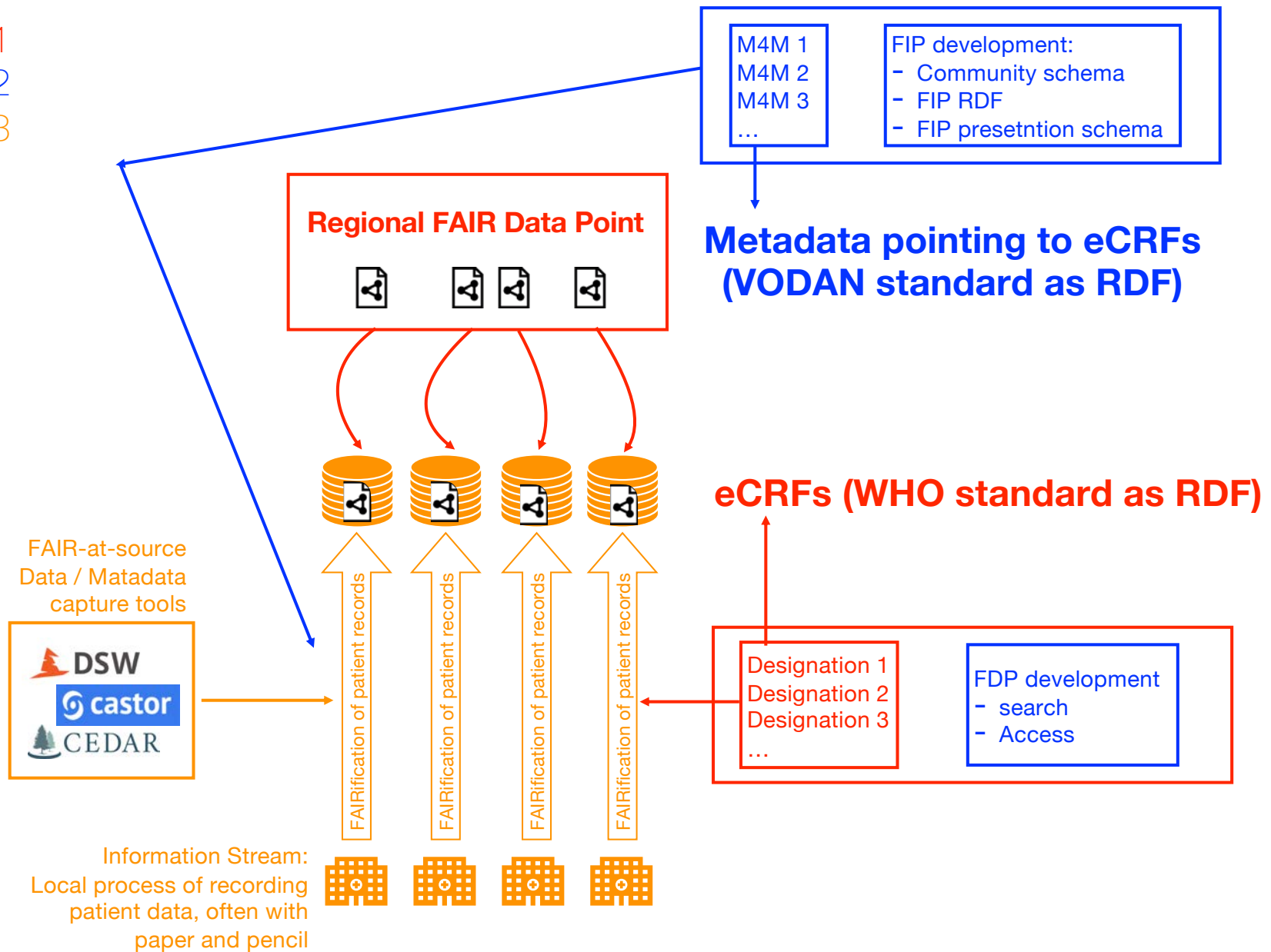
## Information structure of COVID-19 semantic model of patient data:

1. Digital Data Entry at hospital/clinic
2. Machine Readable interoperability of existing digital entries (ensuring interoperability and reusability)
3. Defining metadata (ensuring findability and establishes conditions for accessibility)



# VODAN Project: Social-Technical Specification

Phase 1  
Phase 2  
Phase 3





There is no FAIR Data  
without Machine-  
actionable Metadata





## Metadata for Machines Workshops

Domain Experts + Metadata Experts = Machine-actionable Metadata

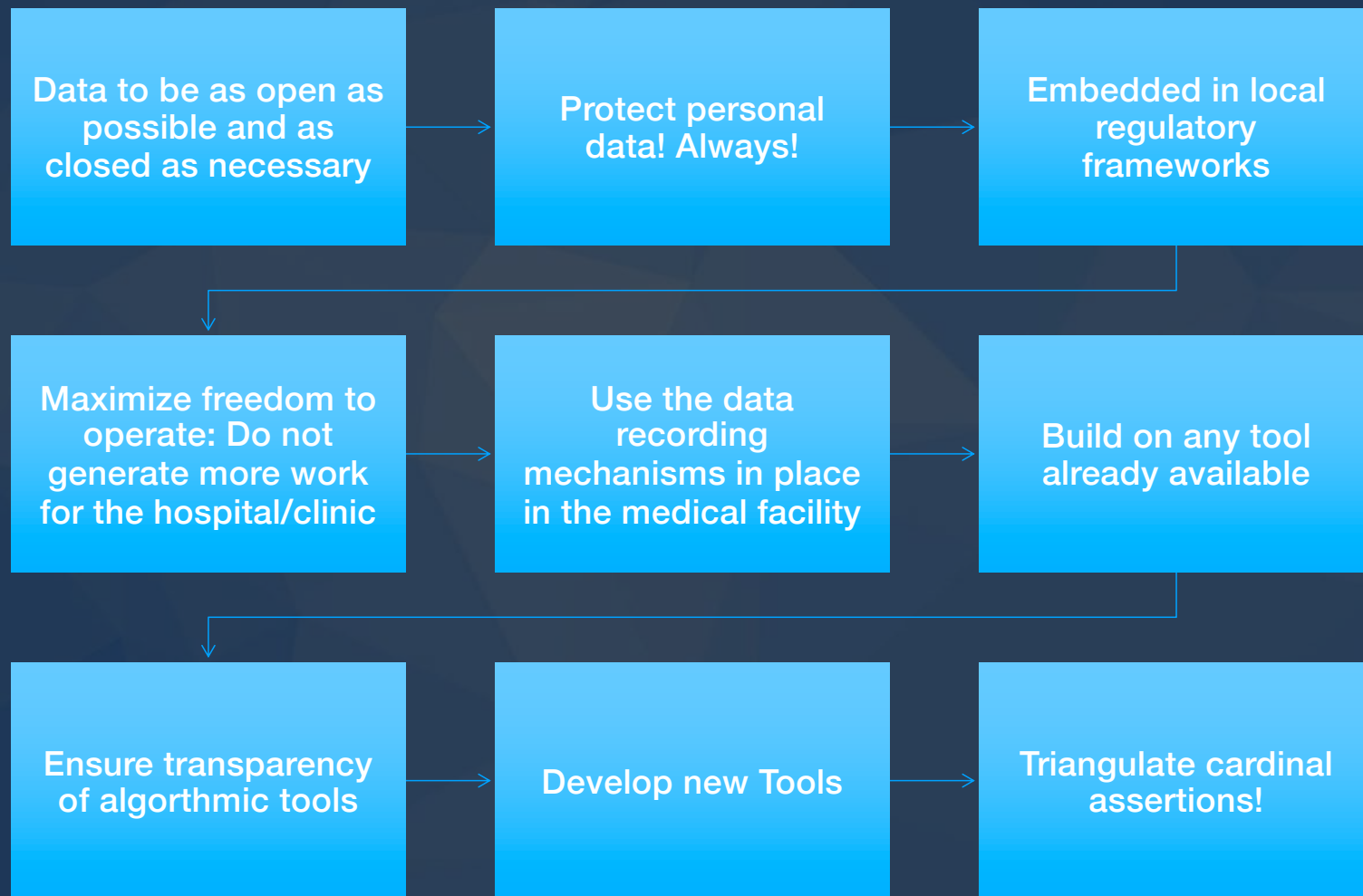
The creation of community specified, reusable metadata templates that prevent the reinvention of the wheel and drive convergence and interoperability.



# Metadata for Machines Workshops

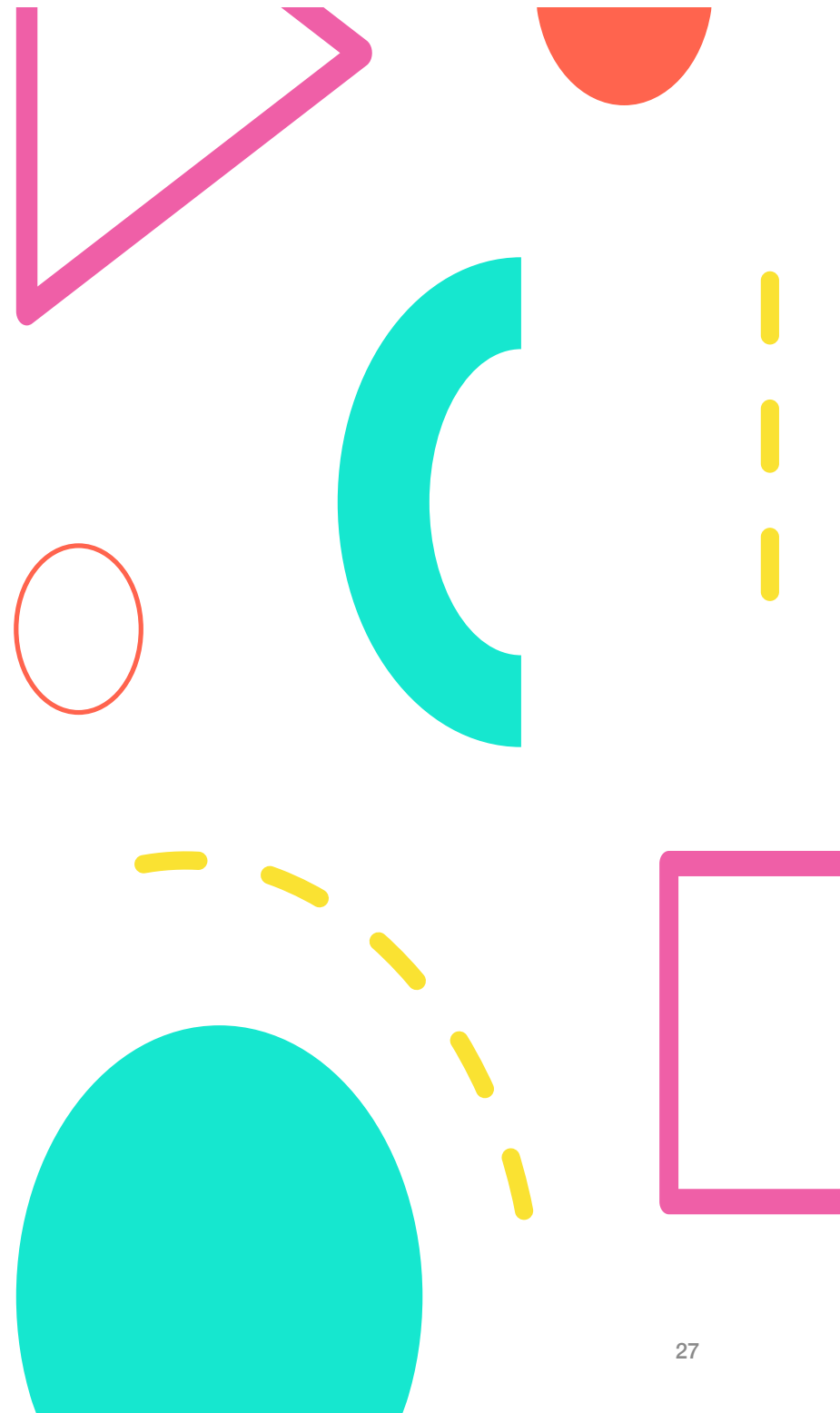
Question	FAIR Principle	Metadata decisions	National	Regional	Hospital
1	F1	What is the persistence policy for identifier systems used for data elements?			
2	F2	What are the minimal data elements needed to ensure Findability?			
3	A2	What is the persistence policy for the metadata?			
4	R1.1	What usage license(s) will be used?			
5	R1.2	What are the minimal provenance metadata needed to ensure reuse?			
6	R1.3	Give a machine-actionable FIP.			

# Key Aims of the FAIR Principles



# Virus Outbreak Data Network – Africa and Ambassadors

- PI: Mirjam van Reisen, Leiden University  
[mirjamvanreisen@gmail.com](mailto:mirjamvanreisen@gmail.com)
- Coordinator: Francisca Oladipo, Kampala International University  
[francisca.oladipo@kiu.ac.ug](mailto:francisca.oladipo@kiu.ac.ug)
- Technical Support Group lead: Mariam Basajja, Leiden University/Kampala International University  
[mariam.basajja@gmail.com](mailto:mariam.basajja@gmail.com)
- Support ambassadors Implementation Network: Aliya Aktau, Leiden University Aliya Aktau  
[aleka.aktau@gmail.com](mailto:aleka.aktau@gmail.com)
- GO FAIR International Support and Coordination Office: Erik Schultes



# Join and stay tuned:



Tune in next week for: VODAN, FAIR data in action

Join VODAN: [tinyurl.com/join-vodan](https://tinyurl.com/join-vodan)

Welcome at VODAN Training of Trainers: <https://www.vodan-totafrica.info/>

Join RDA-COVID19 WG:

[www.rd-alliance.org/groups/rda-covid19](https://www.rd-alliance.org/groups/rda-covid19)

Data (Science) COVID-19 Resources:

<https://www.academicdatascience.org/covid>

Join the conversation: [gofair.slack.com](https://gofair.slack.com)



[www.westbigdatahub.org/](https://www.westbigdatahub.org/)





**Thank you!**

---