

a Data Visiting architecture

with the perspective on trust and security to the Danish Data

Danish DM Advisory Forum, April 28th 2022



1. Introduction
2. Role of GFF-NL -----
3. Awareness and trust
4. Policies and GDPR
5. FAIR orchestration

What we do

Solutions making data universally Findable, Accessible, Interoperable, Reusable [FAIR] are beginning to emerge. The GO FAIR Foundation guides people and organisations on implementation choices fit to their purpose.



FAIR Principles

Working closely with the community, the GO FAIR Foundation offers explicit, expert interpretation of each of the FAIR Guiding Principles.



Implementation Criteria

GO FAIR interpretations of FAIR lead to specifications that guide FAIR implementation choices.



Qualification & Certification

GO FAIR criteria provide the foundation for schema development and certification.



Fellows in Residence Program

A network of leading experts engaging with the community to push the boundaries of FAIR.



GO FAIR Pioneers

A network of FAIR service providers, implementing solutions adhering to GO FAIR criteria.

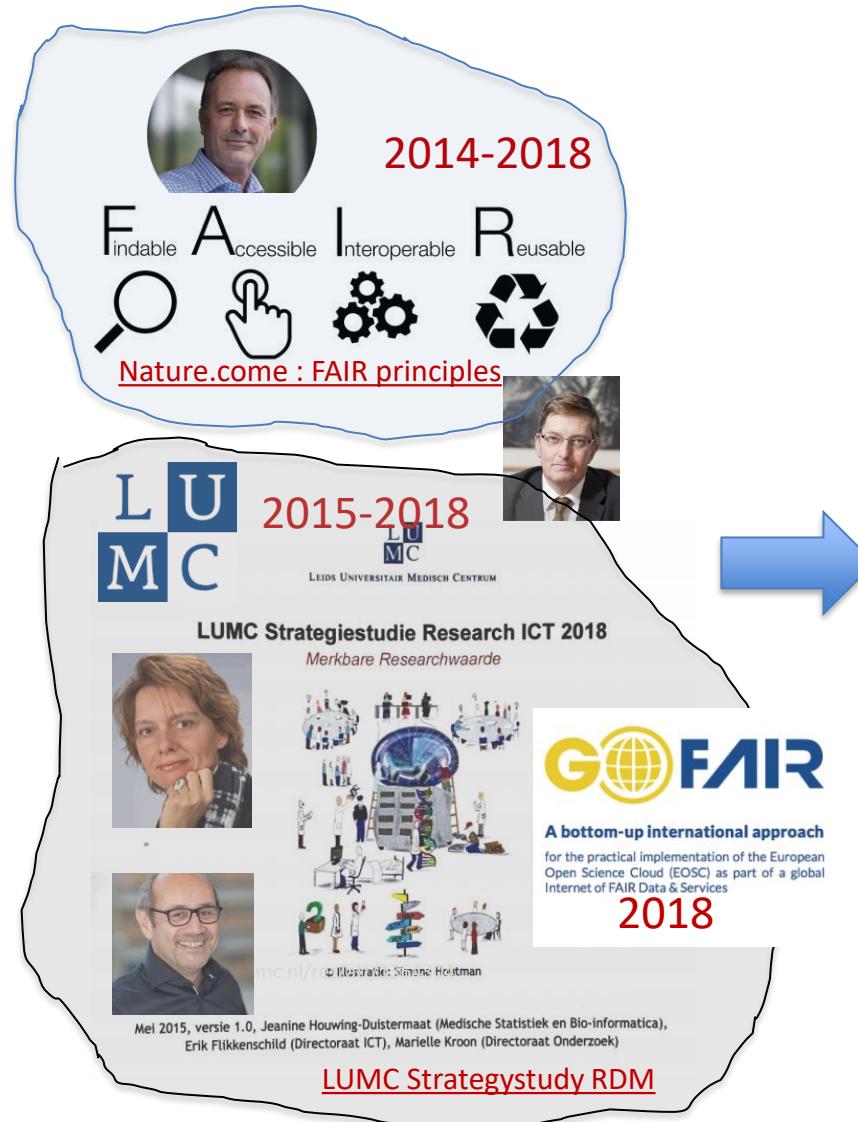


Services offered by the GO FAIR Foundation

The GO FAIR Foundation advises on project development in all its stages, from writing tenders to product deployments.



About me



Founded 2018

The GO FAIR Foundation: we're here to serve you!

[Home - GO FAIR Foundation](#)



[Team - GO FAIR Foundation](#)

The GO FAIR Foundation (GFF) was established in February 2018 as a separate legal entity under Dutch law in order to support the FAIR principles and metrics, and to support the [GO FAIR international support & coordination office \(GFISCO\)](#).

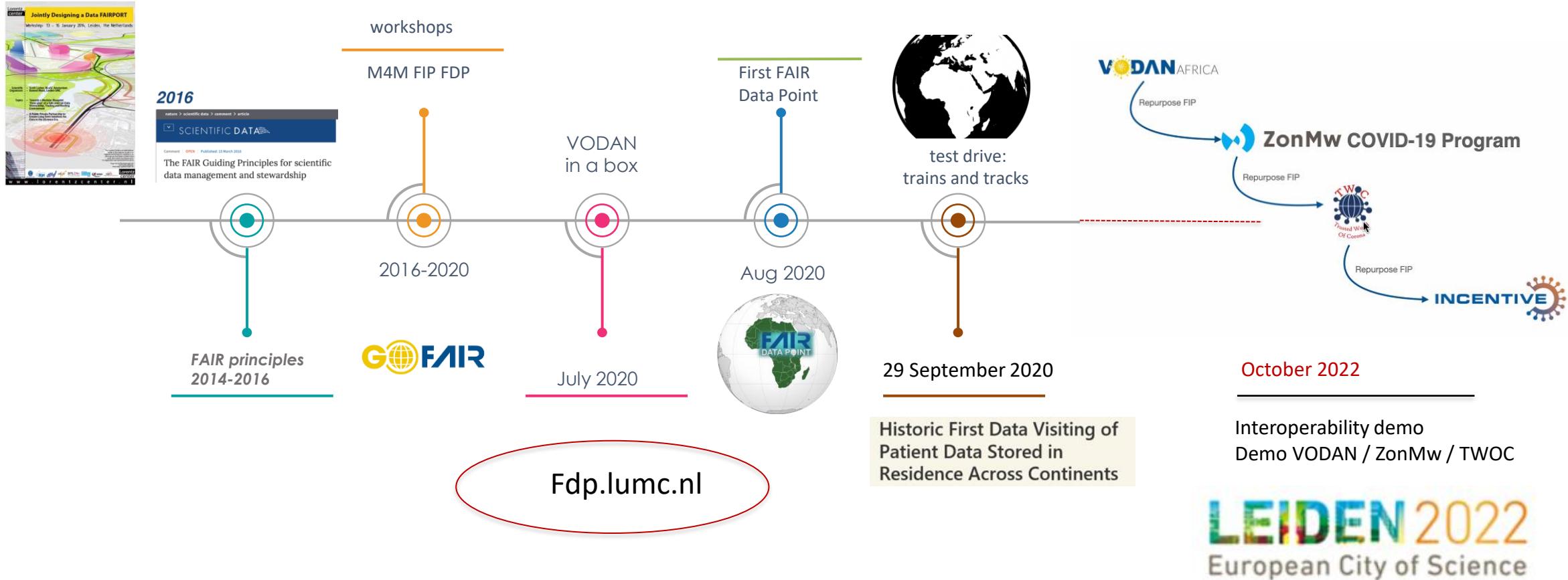
It is the Foundation's mission to play an important role in the development of an Internet of FAIR Data & Services.

The Foundation's main focal point is to develop certifications for a variety of GO FAIR tools and services.

[Mission - GO FAIR Foundation](#)



FAIR principles deployment in practice:





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

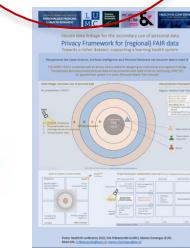


Box 2 | The FAIR Guiding Principles

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

FAIR
Enabled
Resources

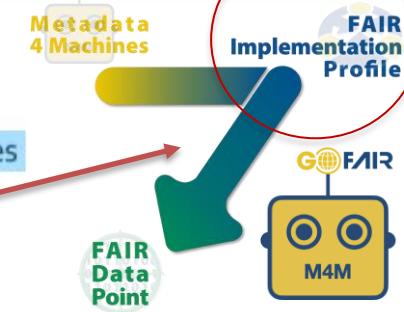


To be Accessible:

- 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 data visiting concept



Why a Personal Health Train?

Modern health care and research generates and requires vast amounts of data captured and stored at different locations. Managing these data in a centralized database is neither desirable nor feasible. Central data processing gives rise to issues on agreements for data-accessibility, data standardization and privacy. Researchers found ways to use health data from various sources in decentralized databases. They coined this approach, the Personal Health Train (PHT).

[More information](#)

[The PHT concept | The Personal Health Train \(health-ri.nl\)](#)

The data visiting concept roadmap



Awareness

The trust factor and its stakeholders

Method: FAIR eco system

Policies and GDPR

Governance requirements translated in access control

Method: The privacy network for regional data

FAIR orchestration

Architecture scenario's

Method: anDREa



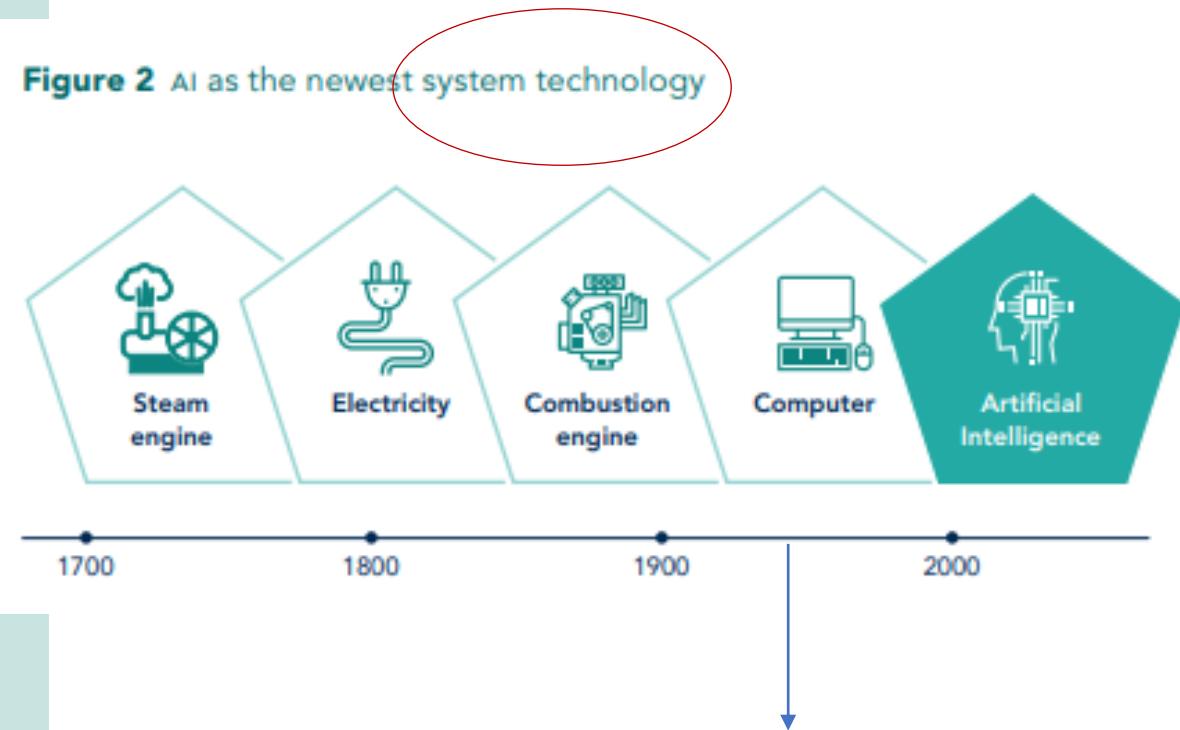
Mission AI

The New System Technology

WRR



Figure 2 AI as the newest system technology



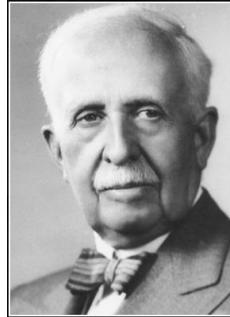
OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.*

Figure 3 Five overarching tasks for embedding AI in society

1st step (Institutional board rooms) Trust in the more value of the FAIR principles

Responsibilities are given to him on whom trust rests. Responsibility is always a sign of trust.

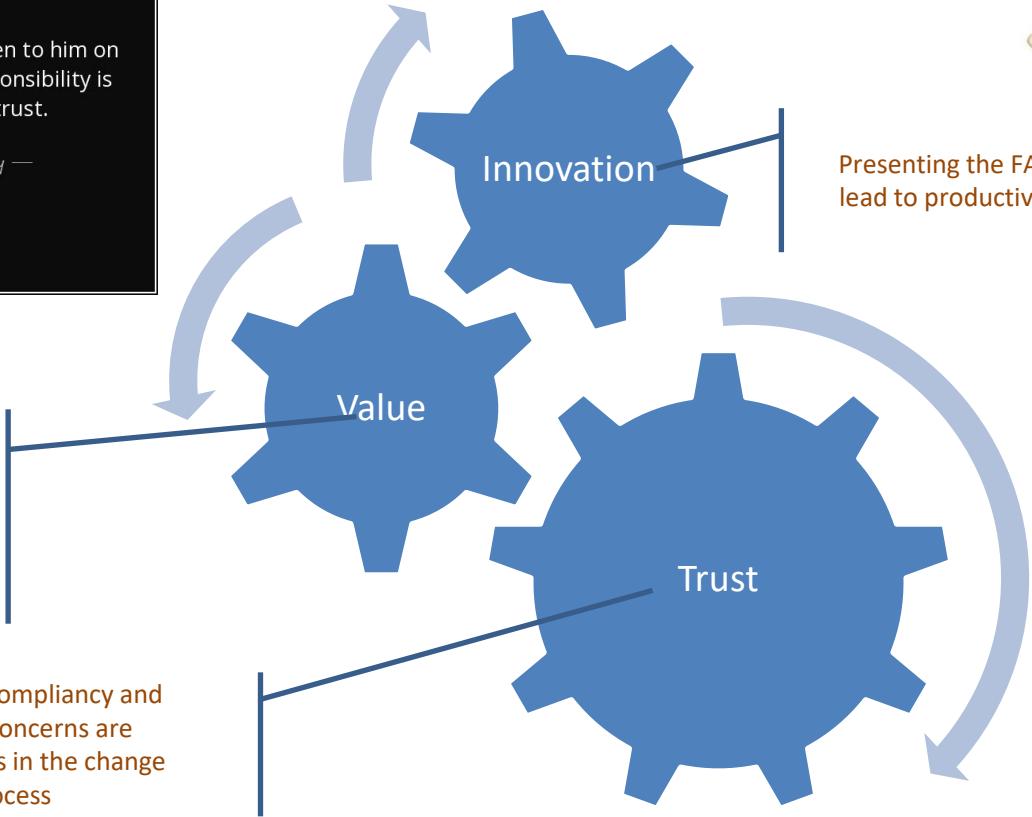
— James Cash Penney —

AZ QUOTES

(Example Academic Hospital)
Value driven care in combination
with
Open Science with access to
FAIR data are the key success
factors

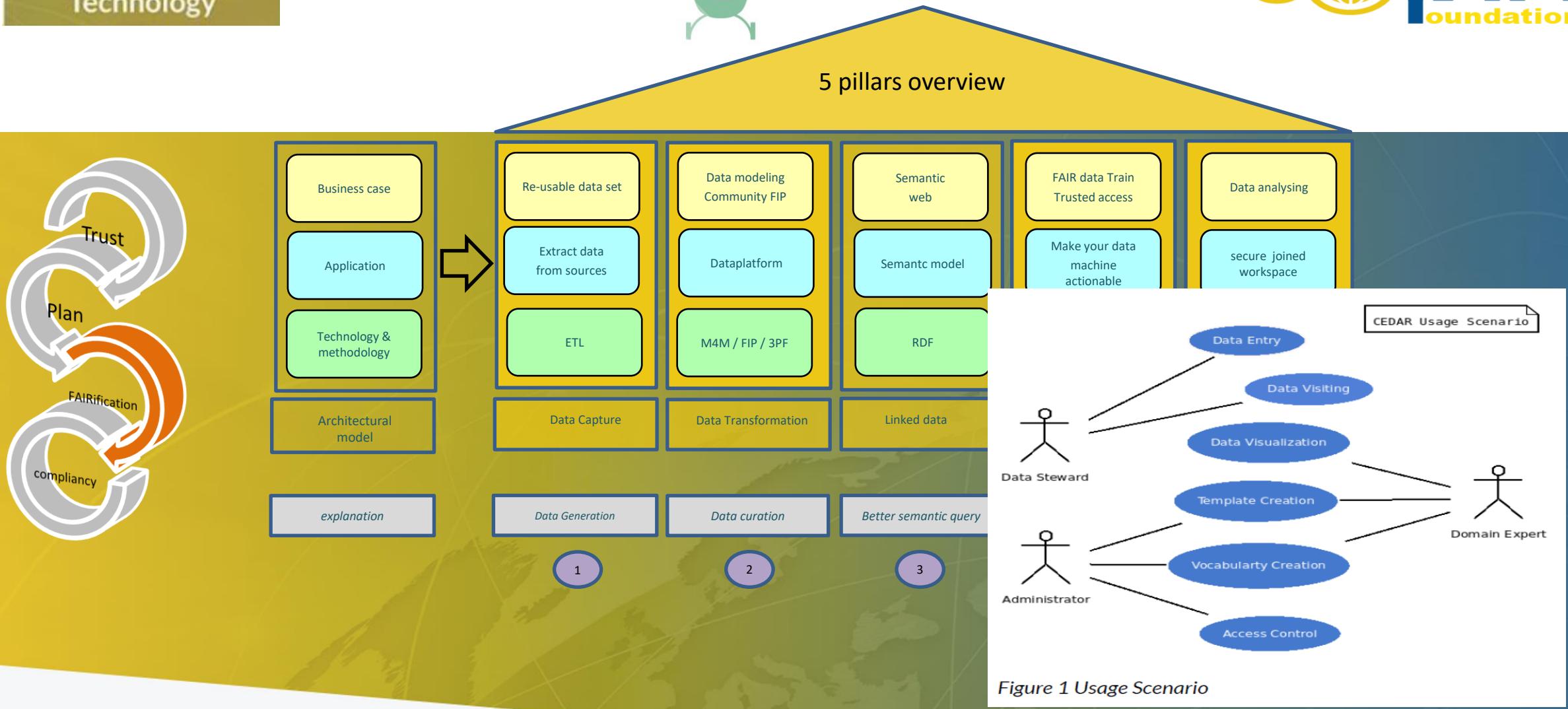


Addressing compliancy and
business concerns are
preconditions in the change
process



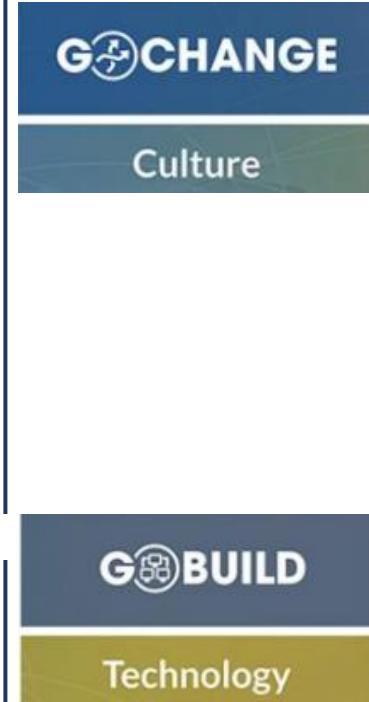
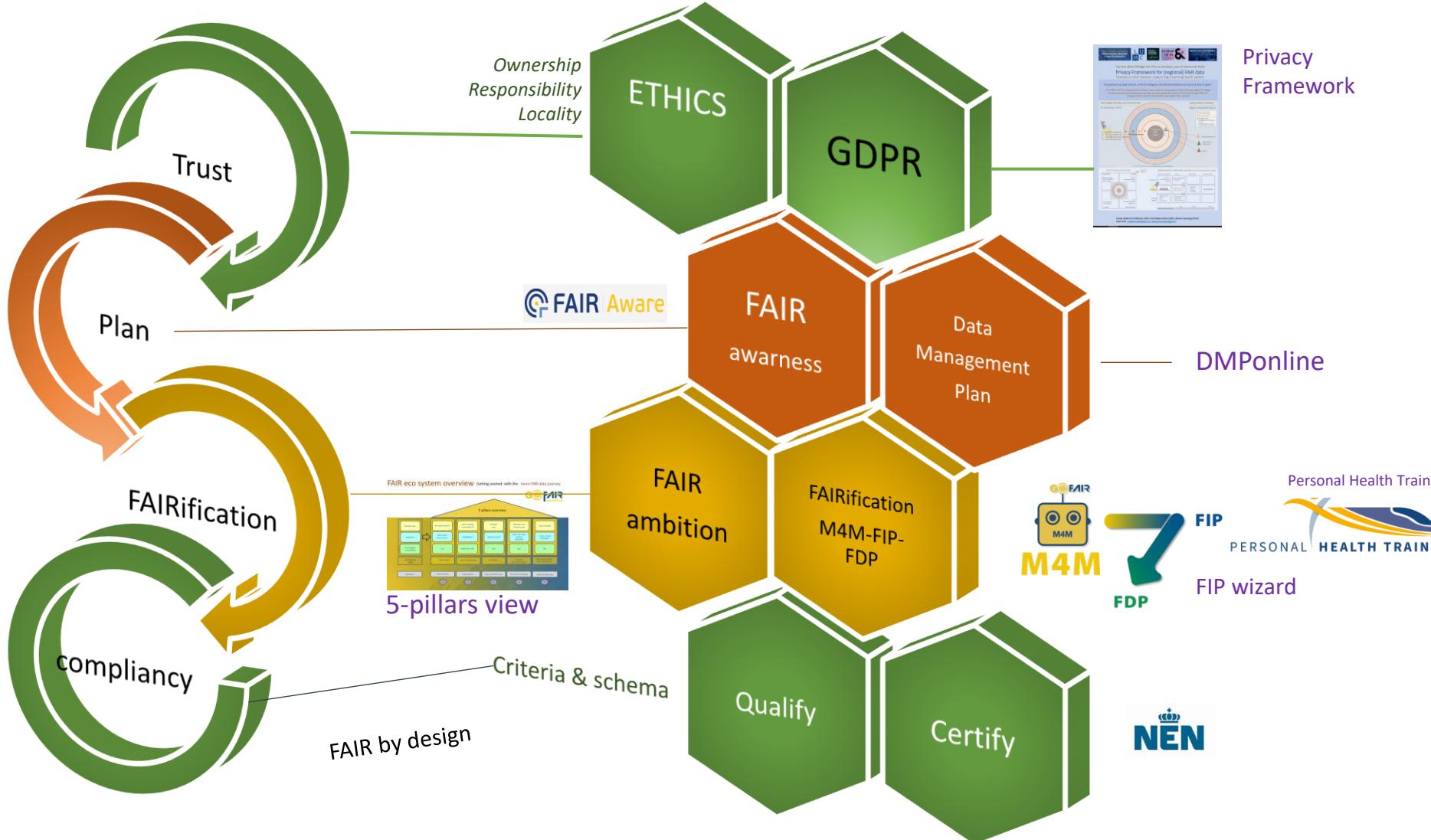
FAIRification roadmap

Getting started: with the more FAIR data journey



FAIR eco system building blocks

Methods (examples)

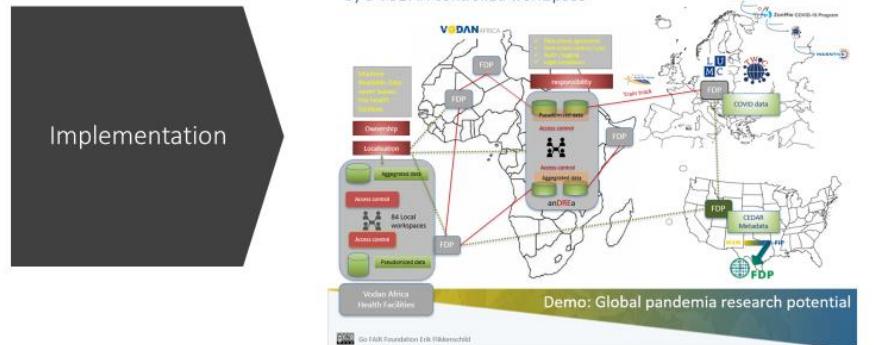




Conclusion

This is the beginning of a new internet. Africa and Europe can work together to provide an internet where ownership of the data in the place where data is produced, strong while collaboration can provide data analytics through data visiting

History of Cooperation: Data visiting Across Africa and Europe



Implementation



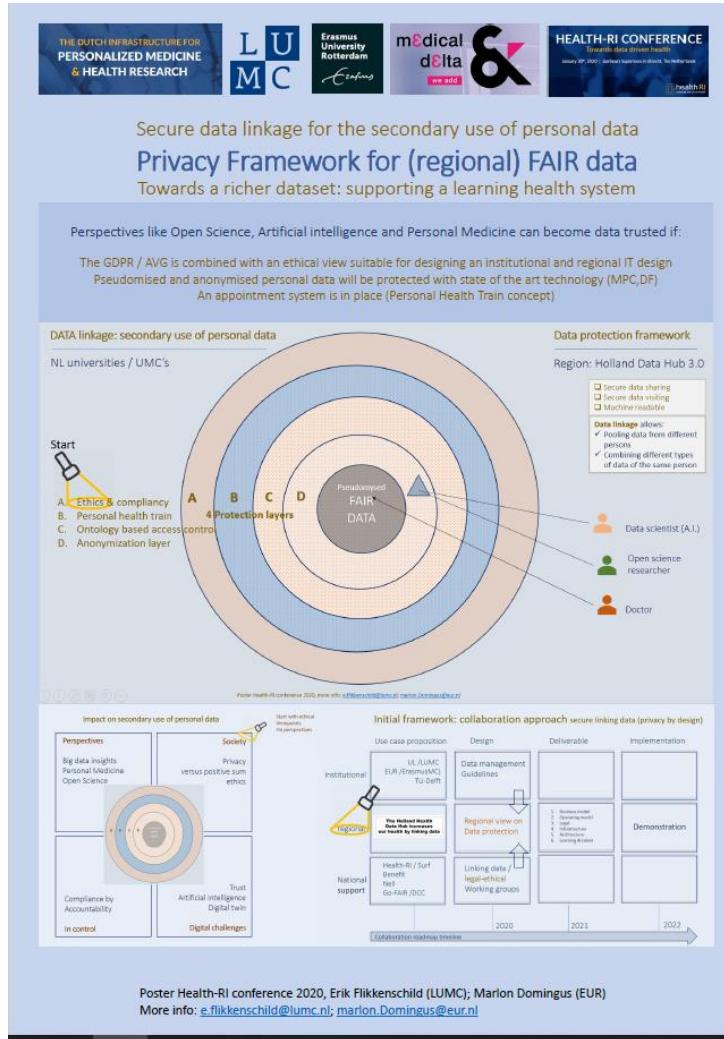
Critical Issues for AU-EU Collaboration on Health and Science

Critical Issues

- Regulatory Framework for Data Production and Management
 - FAIR and GDPR for ethical computer and data science together with an increasing awareness in Africa of the need to retain data ownership
 - Privacy and Data Protection Laws of African Countries
 - To end data extraction and colonisation of knowledge
 - Africa had been at forefront of designing the first distributed machine actionable interoperable data system in which the data is held in the repository in the health facility
 - This enables digital data analytics to improve health services at point of care
 - Makes data visiting possible through federated data in ten sovereignties across Africa and Europe

	Stakeholder	Concern (s)	Topic (s)	Method (s)
1	Governement	funding	FAIR data	
2	Board of Directors	Business case priority	FAIR principles	Strategic plan development
3	Management	Project priority	Project proposal	Tactical plan development
4	Privacy Officer	GDPR	DPIA kick off	Privacy framework workshop
5	Security Officer	ISO 27001	Security assessment	certification
6	IT architect	Institutional compliancy	IFDS Reference architecture	
7	Data owners	Data license violation	Data Acces Agreement	MVP demonstration
8	Data stewardship	No training	Data Competence Centre	NL-Health-RI coordination
9	Data requestor	No access to data	Open data, certified IFDS , community building	Go FAIR Implementation networks





Roadmap

- Ethical review per perspective**
Define the institutional view on the positive sum of a social perspective, gaining the proper balance between common interest (e.g., the patient) against the individual privacy aspects;
- Investigate privacy and ownership concerns**
Use the Privacy Framework for communication purposes with (research) management and institutional policy makers;
- Design the I.T access control architecture**
Anonymity is protected by including 4 access control protection layers in the design (addressing the A in FAIR); e.g., an institution can demand the presence of a technology anonymization layer (Multi party Computation);

Privacy Framework for (regional) FAIR data | Zenodo



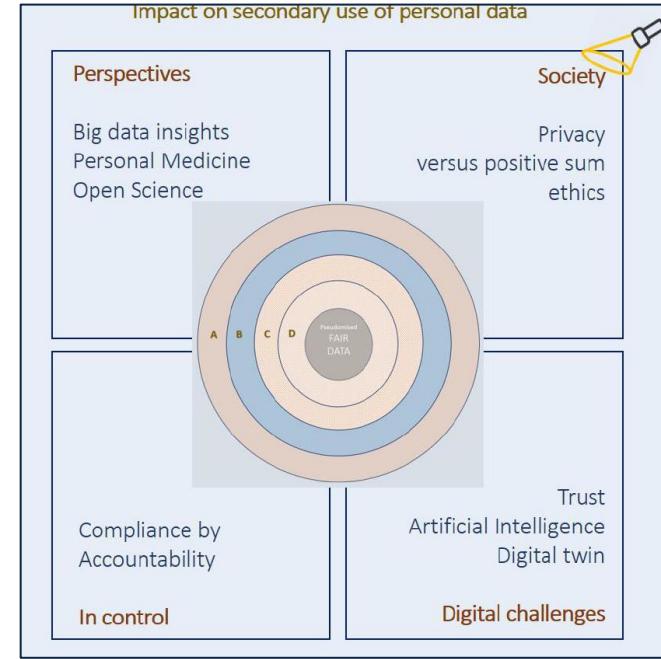


First step: impact analysis (re-use of data)

Update February 2022

Legal restrictions
and use of
anonymisation
filters

Ownership
Responsibility
Data Localisation



Pandemia challenges
Climate challenges

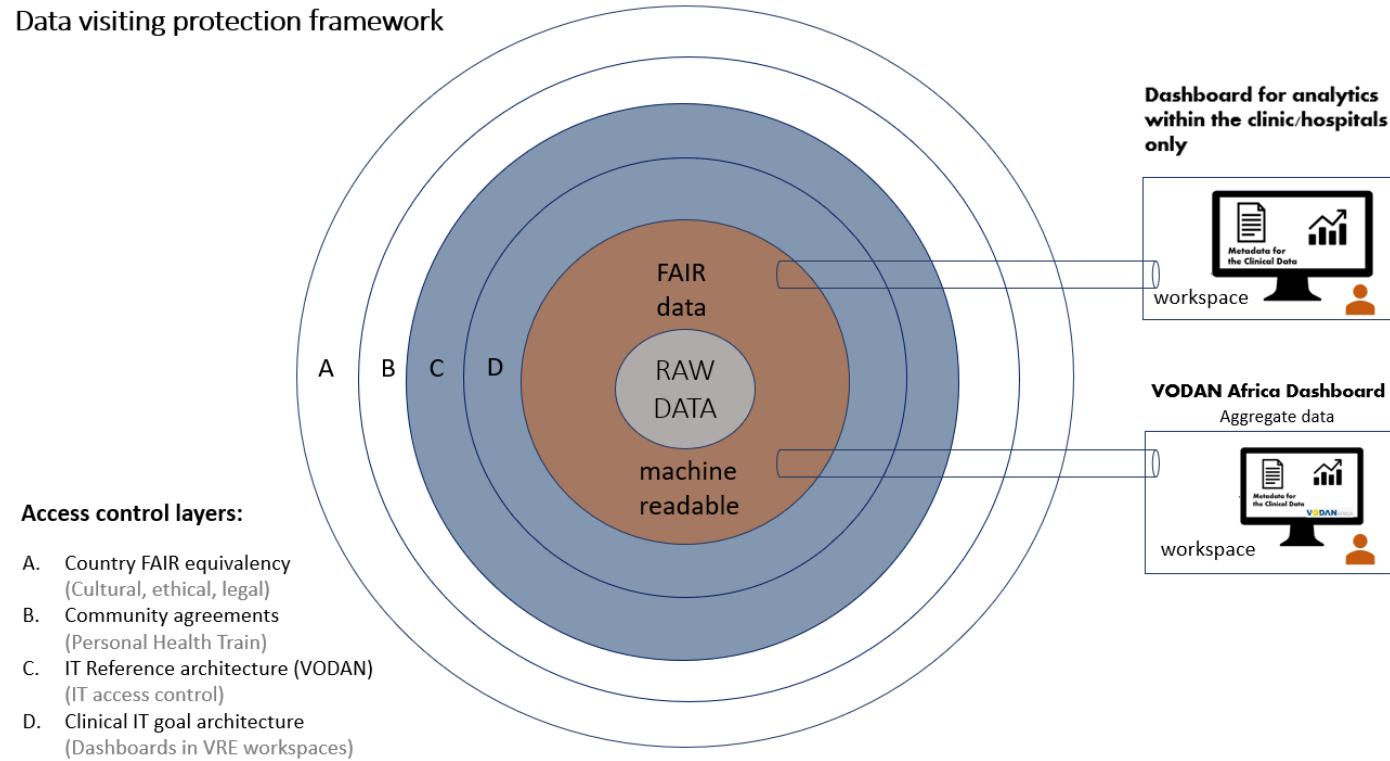
Citizen data

Stakeholders



Access Control measures : a design per perspective

Data visiting protection framework



VODAN-Africa since GDPR really fits with standards. we are going to understand roles which are identified by GDPR. These roles are named Data controller/data processor/data protection officer/supervisory authority. These four functions need to be distinguished in each country for each level.

Data Intelligence (data-intelligence-journal.org)



2022 work in progress



FAIR DIGITAL OBJECTS FORUM

starting with the basics

Part 1: A basic view on FDO's as MAUI's

Machine actionability at the core

FAIR Digital Object Framework Documentation Working Draft

This version: 30 March 2021 Date of last edit: 06 April 2021 Editor/Author: Luiz Bonino Bonino de Silveira Sartori (University of Twente, Leiden University Medical Center, GO FAIR)

Luiz Bonino, Barend Mons, Erik Schultes (GO FAIR, LUMC, LACDR, UT, TNO, DTL)

International context: From FAIR principles to FAIR Digital Objects

nature

Explore content ▾ Journal information ▾ Publish with us ▾ Submit

WORLD VIEW | 25 February 2022

Invest 5% of research funds in ensuring data are reusable

It is responsible to support research and data stewardship.

Barend Mons

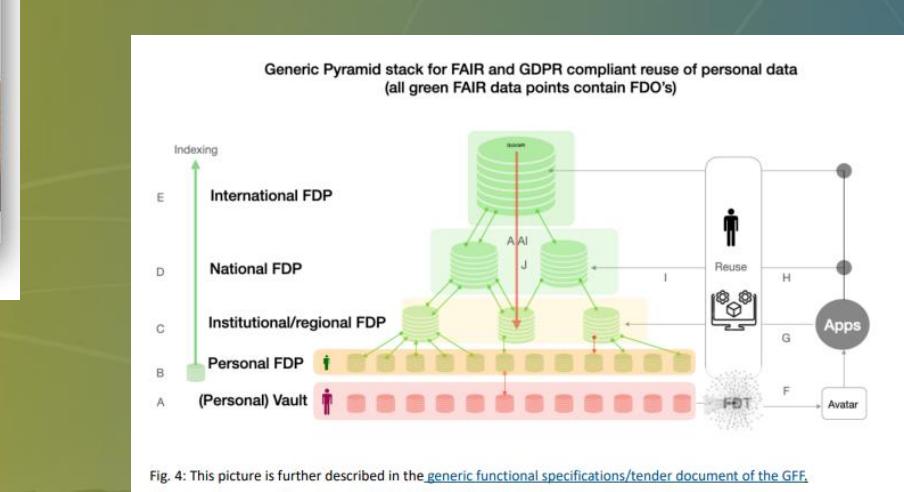
Part 2: FAIR compliant data stewardship

Machine actionability at the core

DATA STEWARDSHIP OPEN SCIENCE M4M FIP FAIR GO FAIR

Barend Mons, Erik Schultes, Erik Flikkenschild, Bert Meerman (GO FAIR, LUMC, LACDR, TNO, DTL)

FAIR (infrastructure) will not fly without proper and professional Data Stewardship and support for experimental researchers



Health~Holland
SHARED CHALLENGES, SMART SOLUTIONS

Barend Mons: Vision on FAIR based infra for JRC

THEME ARTICLE: SPECIAL SECTION ON THE 75TH ANNIVERSARY OF THE IEEE COMPUTER SOCIETY

75 Years of Astonishing Evolution of IT: 1946–2021

George Strawn, US National Academy of Sciences, Engineering, and Medicine

as rapidly as it did. How soon will FAIR data conquer the world? Will the combination of AI and robotics eventually do all our work? What will be the long-term impact of gene editing tools such as clustered regularly interspaced short palindromic repeats (CRISPR), and biotechnology, or of nanotechnology and quantum technology? What other revolutionary changes are right under our noses that we will only see after the fact? In conclusion, I will predict that we will see as many IT and IT-related changes in the next 30 years as we have seen in the last 75!

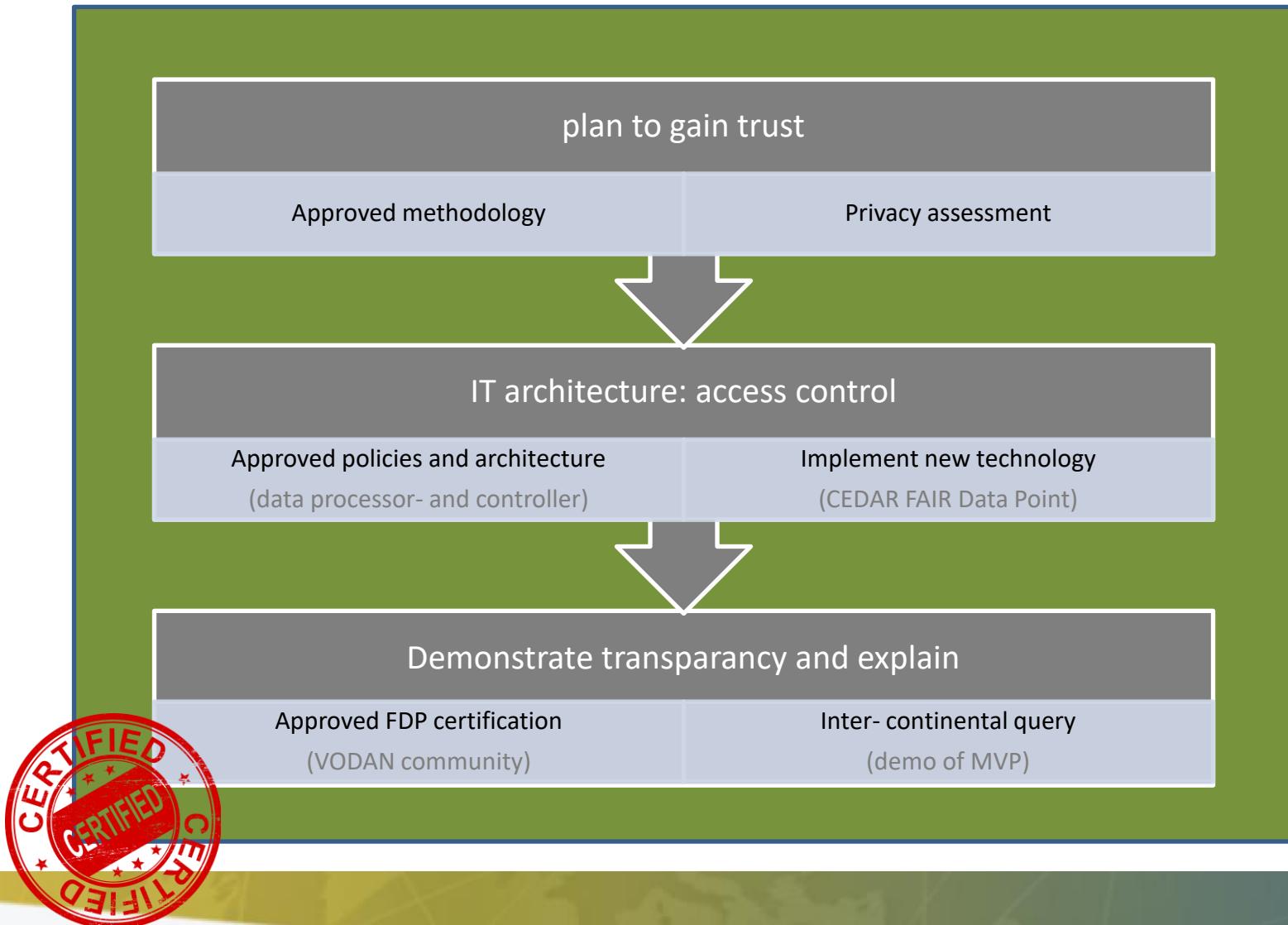
Part 3: FAIR compliant infrastructure

Machine actionability at the core

Barend Mons, Erik Schultes, Erik Flikkenschild, Bert Meerman (GO FAIR, LUMC, LACDR, TNO, DTL)



VODAN trust roadmap

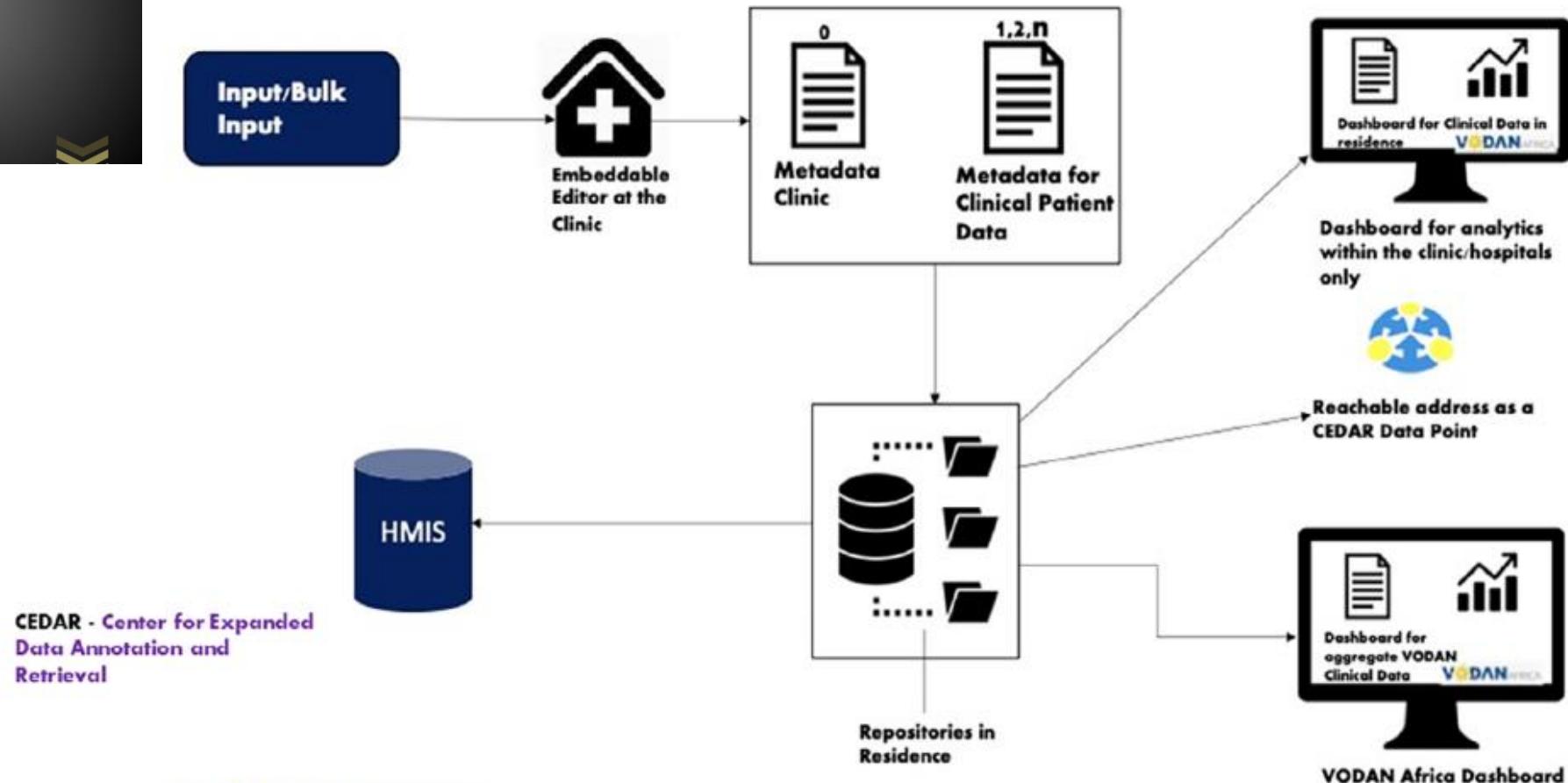


Integrating FAIR IT
services starts with
security & privacy
(by design)

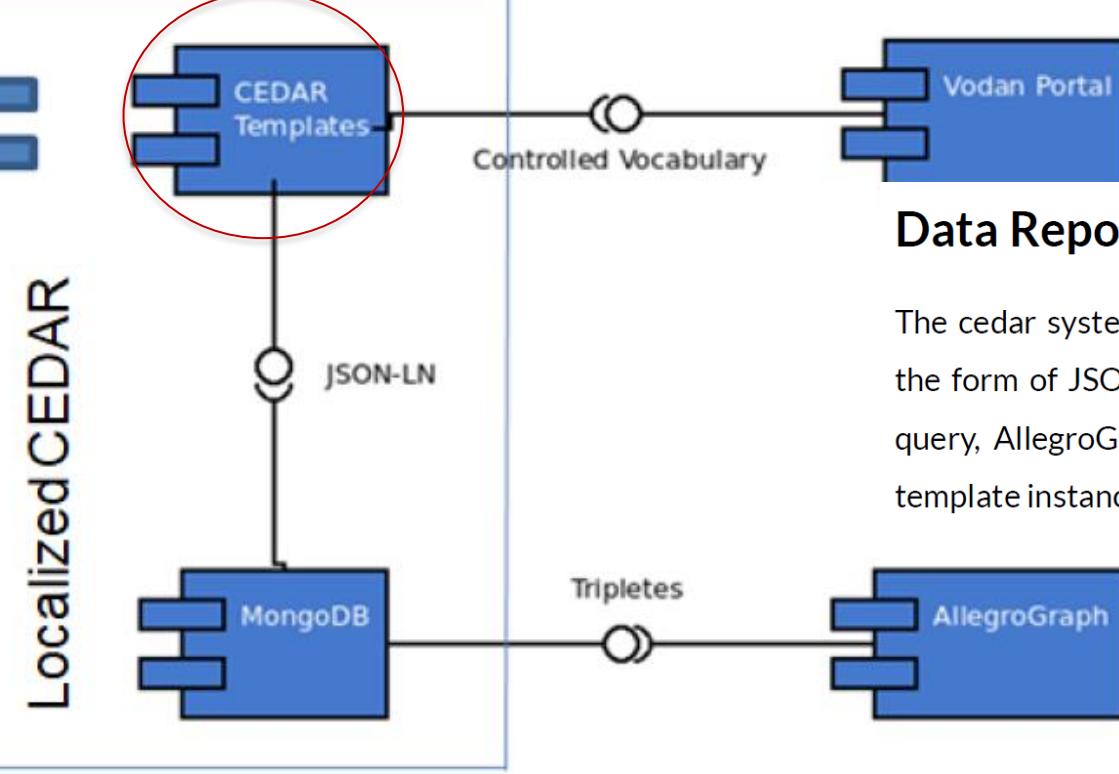




VODAN Africa
Architecture
Description &
Documentation



VODAN-Africa Architecture: Clinical Data (Van Reisen, Oladipo & Osigwe, 2021)



Data Repositories in Residence

The cedar system currently saves data (template instances) on a local MongoDB server in the form of JSON linked data (JSON-LD). To allow for data visiting and conduct distributed query, AllegroGraph is regarded as a preferable triple store. There are two ways to query template instances (JSON-LD) produced in a facility using AllegroGraph.

Figure 2 Data Repository Components

Justification for using Allegrograph as a Repository for data visiting

AllegroGraph efficiently performs the three most important tasks for a triple store: to load, store, and query data.

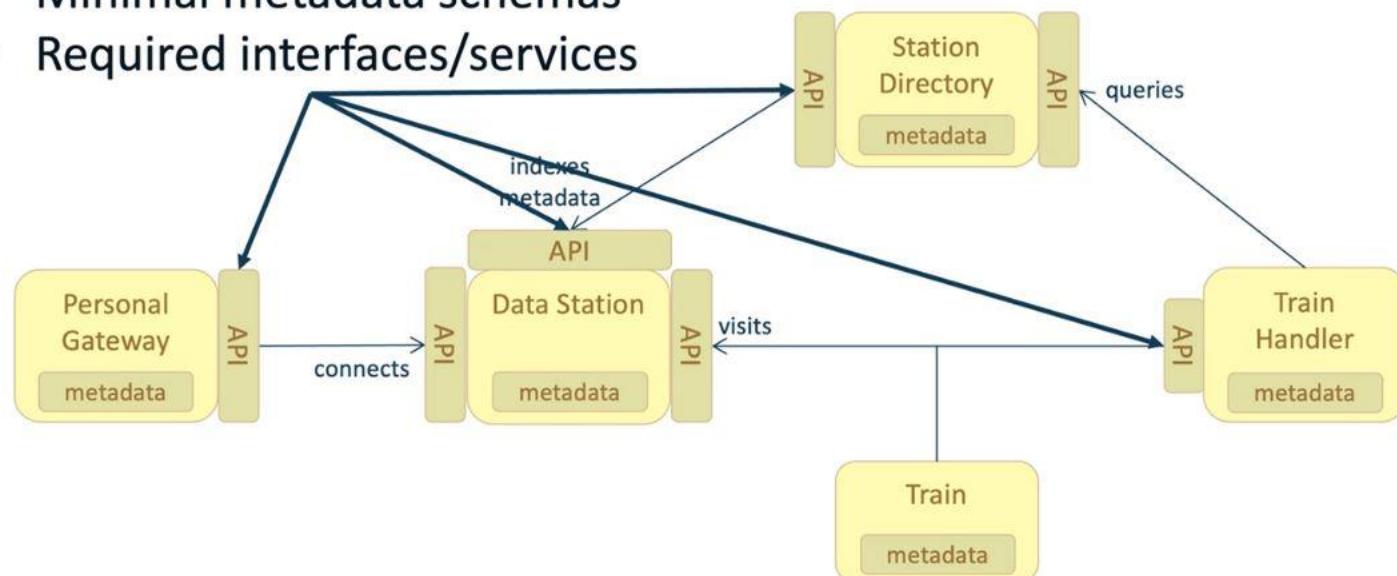
- Loading of triples, through its **highly optimized RDF/XML and N-Triples parsers**, is best-of-breed, particularly on **large files**. With just standard x86 64-bit hardware, it can load gigabytes of RDF data in minutes.
- Storage is **persistent**, including between application launches in on-disk binary trees. There is no additional serialization or deserialization overhead.
- Querying is both **flexible** and **performant**. Multiple indices support fast access through a simple triple-level API, Allegro Prolog, or SPARQL (W3C standard RDF query language).

FAIR Compliance of the Architecture

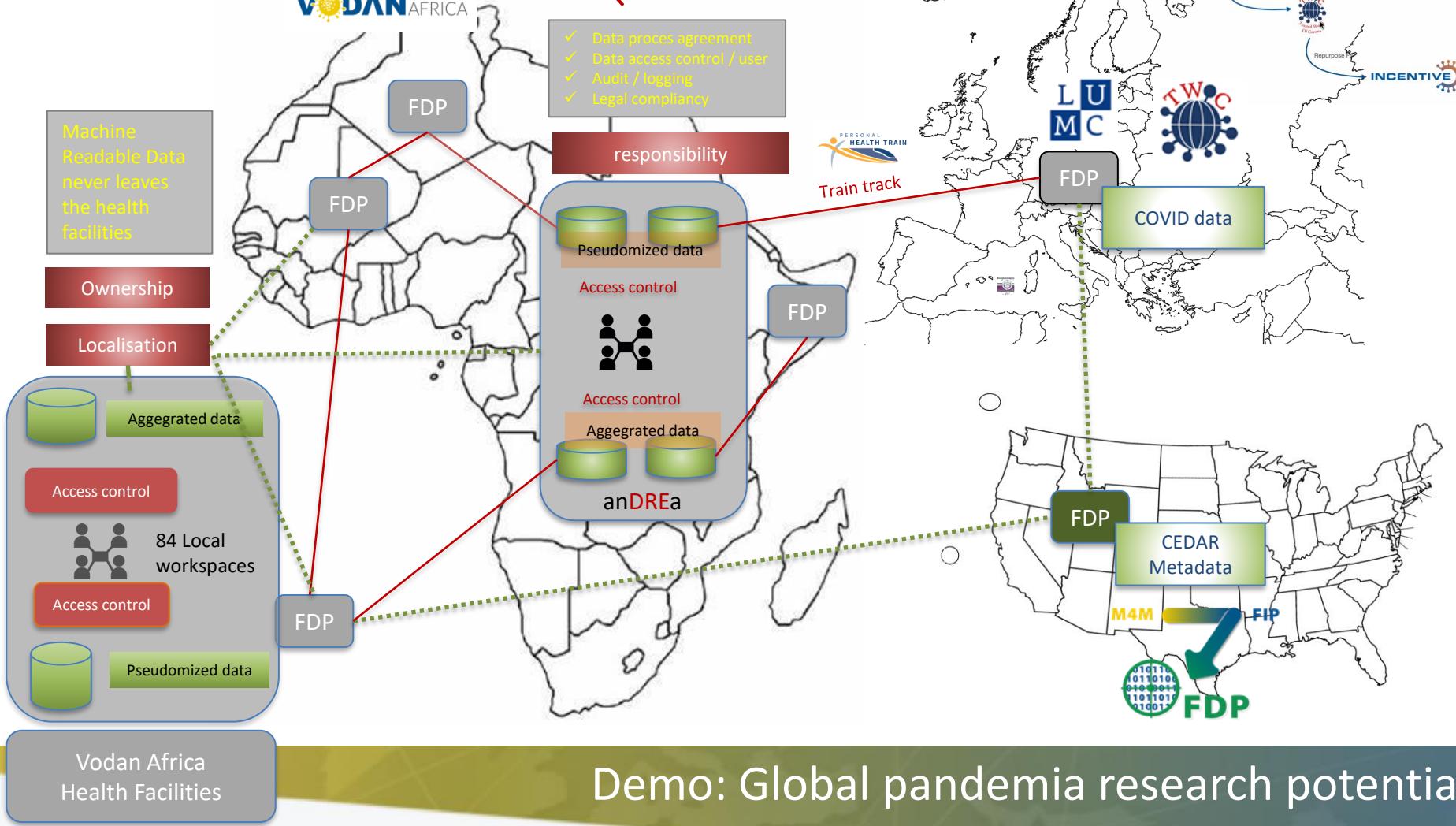
FDT – certification possibilities

- Minimal metadata schemas
- Required interfaces/services

#	Principle/Facet	Compliance
F1	(Meta)data are assigned a globally unique and persistent identifier	Given the (meta) data is produced using CEDAR templates, unique and persistent identifiers are generated automatically.



FAIR Data Points (FDP) connected: by a VODAN controlled workspace



Governance and access control domains

Go FAIR Foundation implementation criteria:
Ownership, data Localisation, Responsibility (OLR)

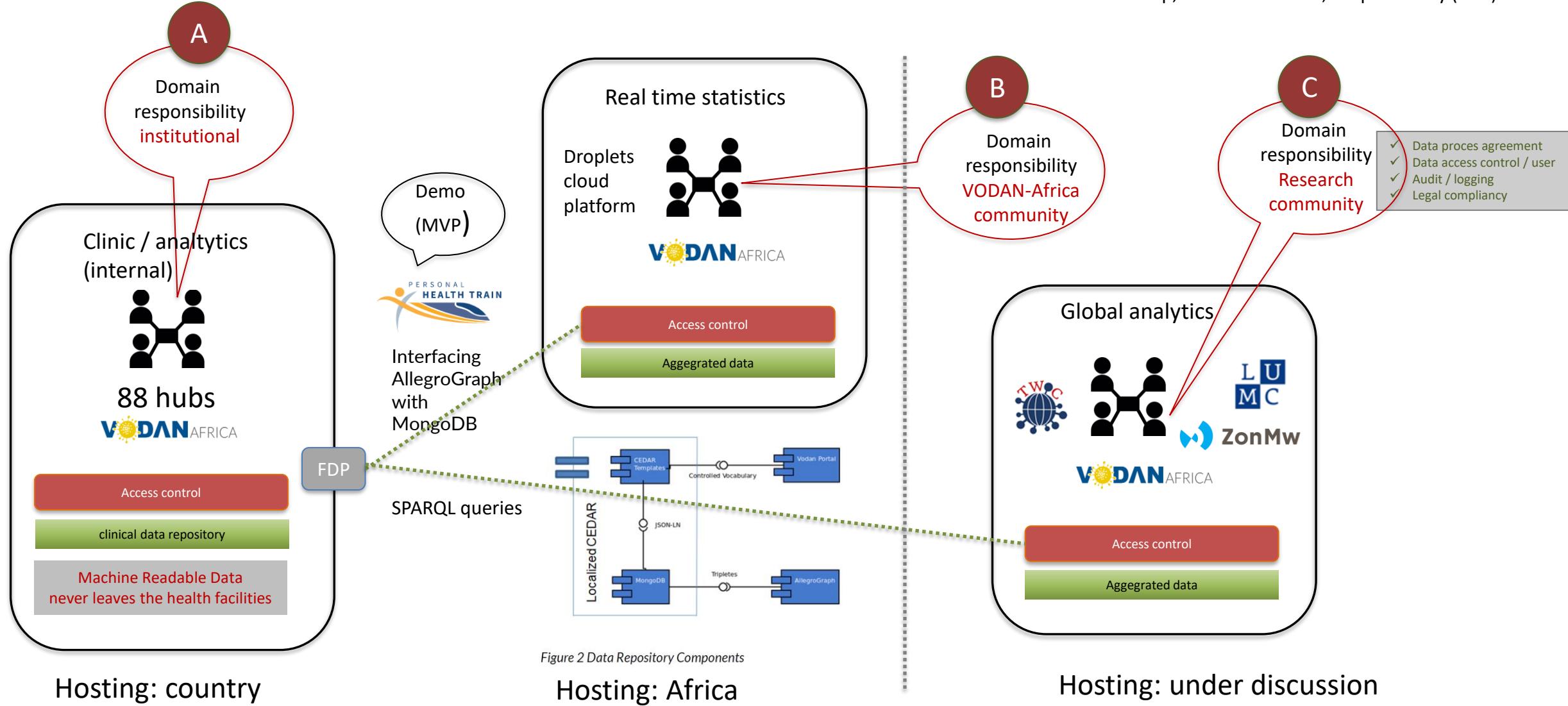


Figure 2 Data Repository Components

Discussion view VODAN internal
14-4-2022 (Ruduan, Kees B, Erik)



Deployment Architecture

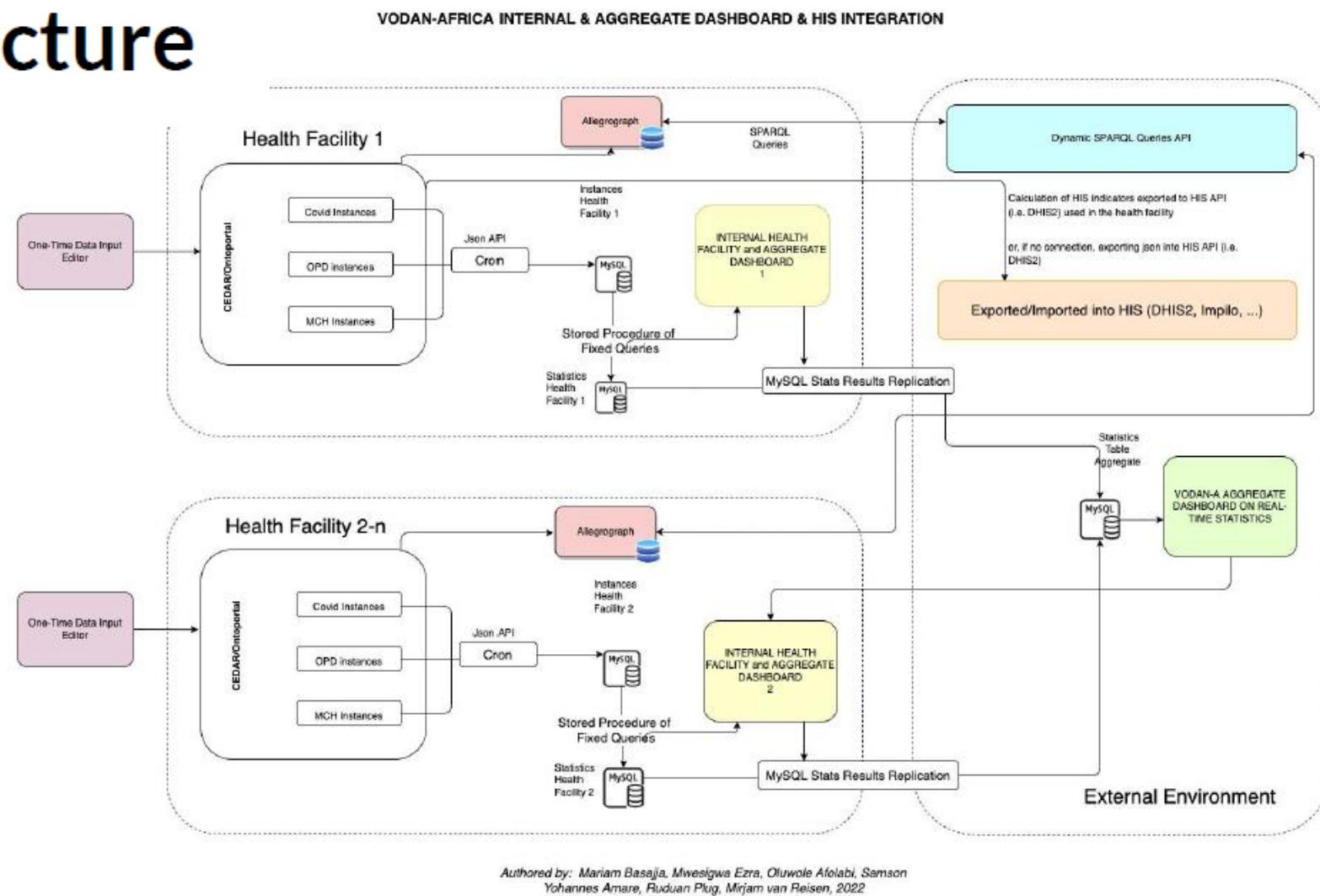
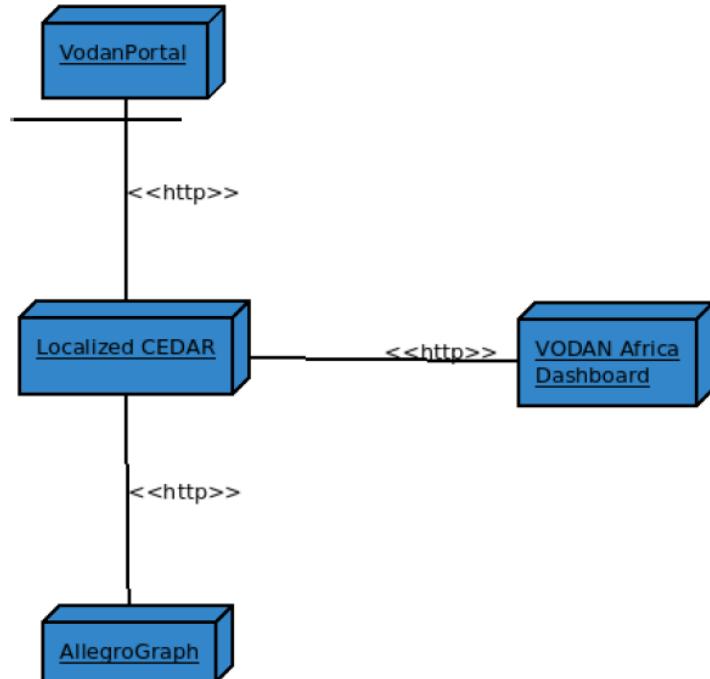


Figure 4: VODAN-Africa MVP Architecture

Mission



2022-2025: safeguarding the originally intended implementation
of the FAIR guiding principles

[Home | GO FAIR Foundation](#)

Role essentials

to avoid pseudo FAIR
the core elements are the
GFF FAIR criteria interpretation
and the hourglass

The idea of a
GFF network
of fellows, pioneers and patrons .



Fellow in residence
programm



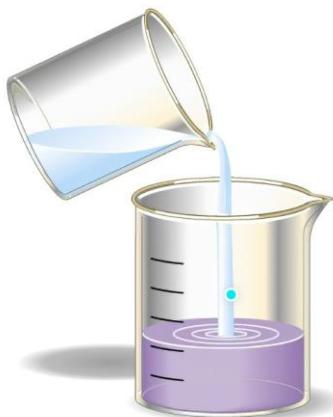
GO FAIR Initiative - GO FAIR (go-fair.org)



The GO FAIR Offices in Germany, France and The Netherlands are happy about the **commitments** of Austria, Brazil, Denmark, and the **US** to continue with GO FAIR and jointly support the community of Implementation Networks.

[GO FAIR Newsletter: January-February 2022 - GO FAIR \(\[go-fair.org\]\(http://go-fair.org\)\)](#)

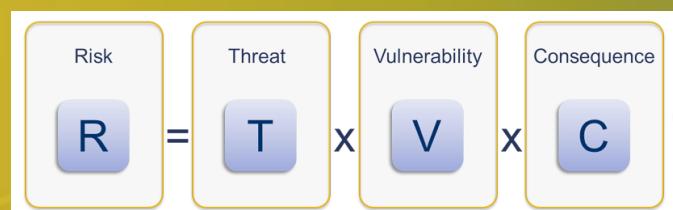
GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the [FAIR data principles](#)



Focus on certification of the FAIR Principles



Risk of dilution of the FAIR principles



Trust (less confidence)

domain data silo
(bad score Re-usability))

Topics:

- Qualification scheme's
- NEN certification programm
- FAIR implementation criteria
- Reference architecture (IFDS)
- Self assessment training, experts, FDP, FDS,..

[Home](#) » [Certification](#)[Certification process](#)[Added value](#)[Pioneer Program](#)[FAIR assessment](#)[FAIR certification for VODAN](#)

Certification

The GO FAIR Foundation (GFF) has been asked by various sectors to provide FAIR certification for FAIR-related resources. This request has come from a broad range of stakeholders and concerns the widely perceived need for independent third-party criteria and validation of resources with respect to the FAIR Principles. for technical components, domain-relevant standards, FAIR-related training, for FAIR implementation events (e.g. M4M and FIP workshops), for people (demonstrating various FAIR-related competencies) and organizations that aspire are committed to FAIR practices. It is believed that certification, when appropriately applied, can be a powerful accelerator of convergence onto wide-spread FAIR implementations.



Fellows
in residence

[VODAN: COVID-19 related projects](#)[GO FAIR IN's](#)[European Open Science Cloud](#)[Dutch Data Sharing coalition](#)

Projects

The GO FAIR Foundation participates in several projects. Currently, several COVID-19 related projects are led by the GO FAIR Foundation. The Foundation also plays an active role in multiple GO FAIR Implementation Networks, and is part of several consortia that focus on making the European Open Science Cloud (EOSC) a reality, and the GO FAIR Foundation contributes to the Dutch Data Sharing Coalition.

For projects whereby external parties are involved in the delivery the GO FAIR Foundation has defined a maximum uplift percentage for indirect costs of 20%.

[Certification process](#)[Added value](#)[Pioneer Program](#)[FAIR assessment](#)[FAIR certification for VODAN](#)

Pioneer Program

GFF has developed a "Certification Pioneer Program", whereby we have asked stakeholders and members from the community to assist, as pioneers, in defining the schemas for a number of FAIR-aspects. This pioneer-program has received support from several organizations and has delivered a number of GO FAIR endorsed elements that would assist in increasing the trust-factor and that would avoid watering down of the requirements.

See below the areas currently in progress :

Training courses

Trainers

Students (after examination)

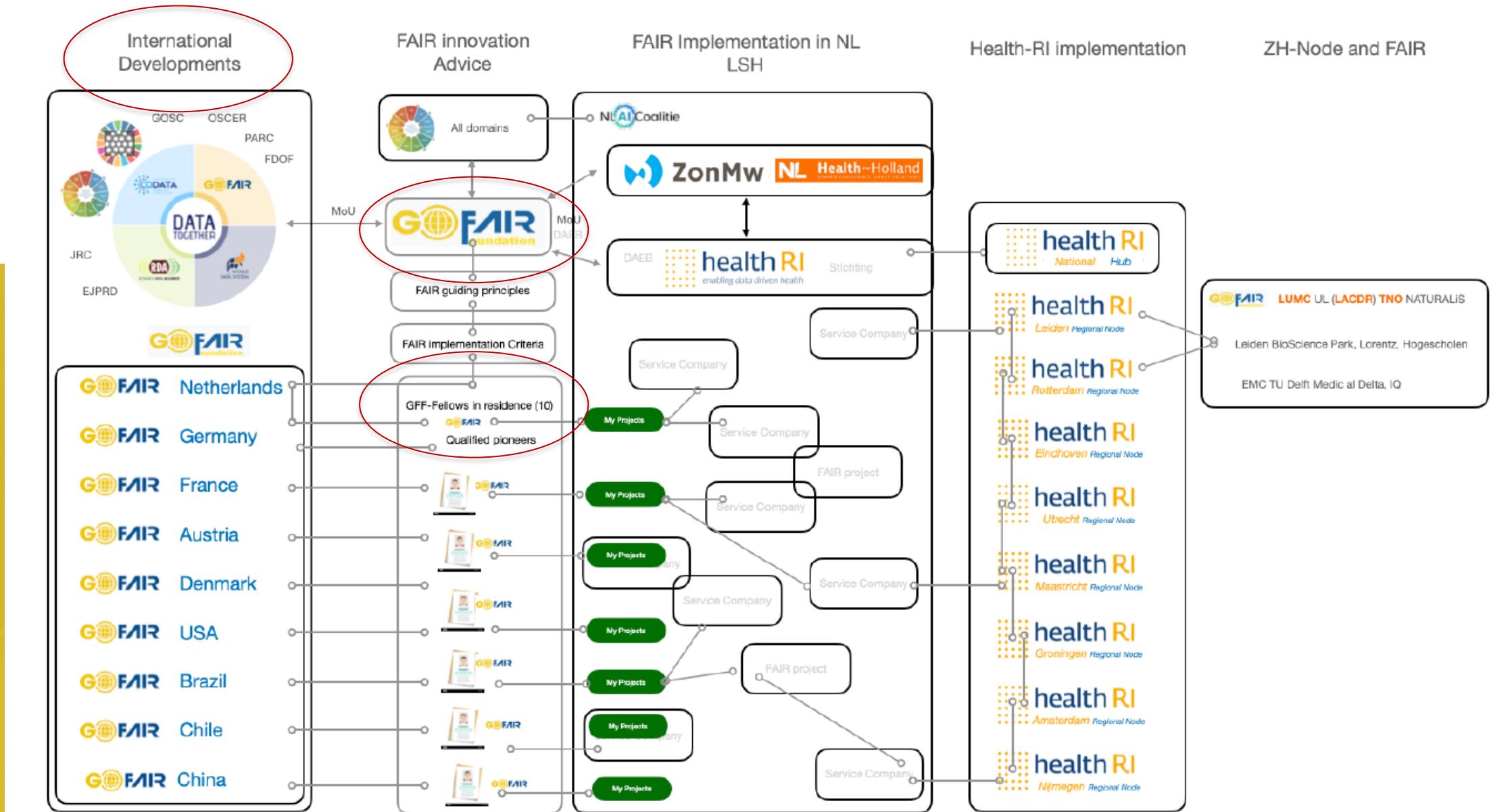
FAIR Data points

FAIR Implementation Profiles (choices made by a community)

FAIR Metadata Schemas / Templates

Organizations (Service Suppliers)

Projects (Data Exchange or Data Visiting projects)





[Leiden 2022 | Leiden2022](#)



Erik Flikkenschild,
e.flikkenschild@lumc.nl

<https://linkedin.com/in/elaflikkenschild/>
<https://orcid.org/0000-0002-7285-1651>

Further reading:

[Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research - Reisen - 2021 - Advanced Genetics - Wiley Online Library](#)

Links to this presentation:

Pdf with notes

Pdf sheets only

