


FAIR in times of Crisis



Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons *Scientific Data* **3**, Article number: 160018 (2016) | [Cite this article](#)**130k** Accesses | **1680** Citations | **1571** Altmetric | [Metrics](#) An [Addendum](#) to this article was published on 19 March 2019

Abstract

The machine knows what I mean

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on **enhancing the ability of machines to automatically find and use** the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

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

FAIR is partly technical

Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier;
- F2. data are described with rich metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

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;

And partly a social contract....

Findable:

- F1. (meta)data are assigned a globally unique and **persistent** identifier;
- F2. data are described with **rich** metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- F4. (meta)data are registered or indexed in a searchable resource;

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;

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;**

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;**

Virus Outbreak Data Network (VODAN)

[Home](#) › [Implementation Networks](#) › [Current Implementation Networks](#) › [Virus Outbreak Data Network \(VODAN\)](#)

The VODAN Implementation Network is one of the joint activities carried out by **CODATA**, **RDA**, **WDS**, and **GO FAIR** (Link to the [Data Together Statement](#)).

Read the full statement on [Data Together COVID-19 Appeal and Actions](#).

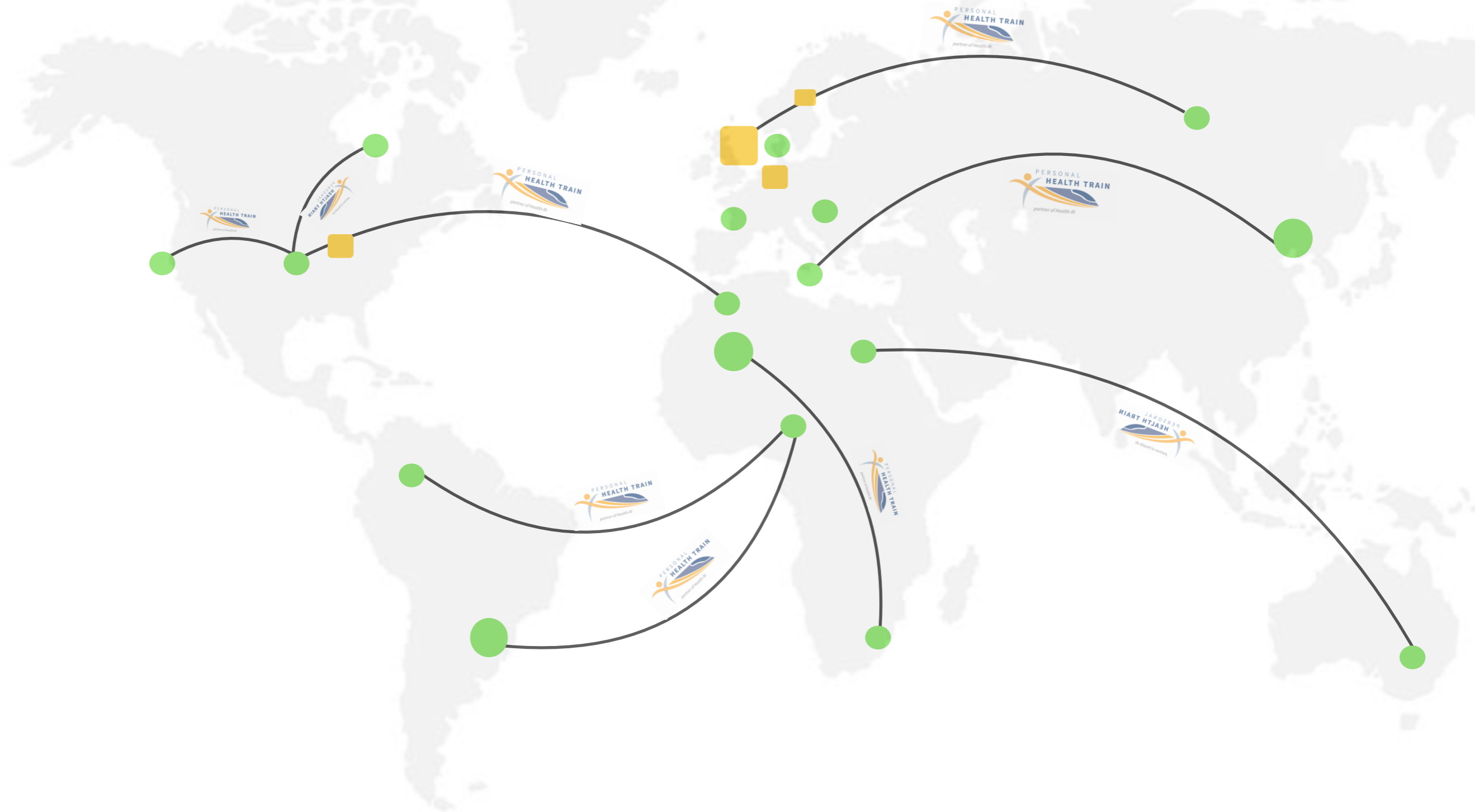
Active GO FAIR Implementation Network

The spread of the virus causing the COVID-19 outbreak is far from over. During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse. Moreover, access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. For instance, the data from the past Ebola epidemics are very difficult to find, to access, and if accessible, they are not interoperable, *let alone reusable*. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to do better and ensure that data are FAIR (in this sense also meaning **Federated**, **AI-Ready**).



The VODAN-IN approach: distributed analytics over FAIR data

<https://vimeo.com/143246458>



FAIR reference (EK) data stations



FAIR RWO data stations

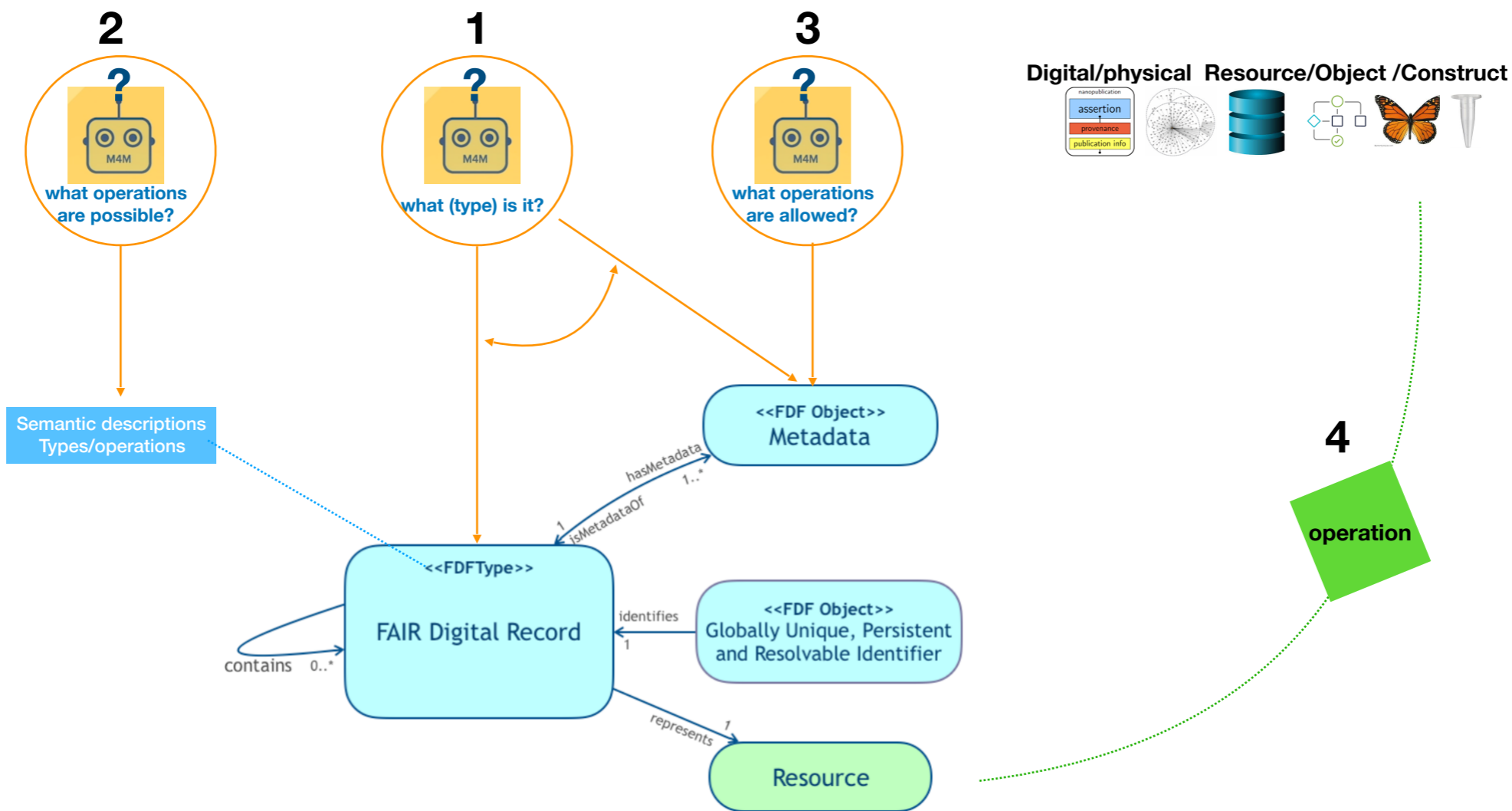


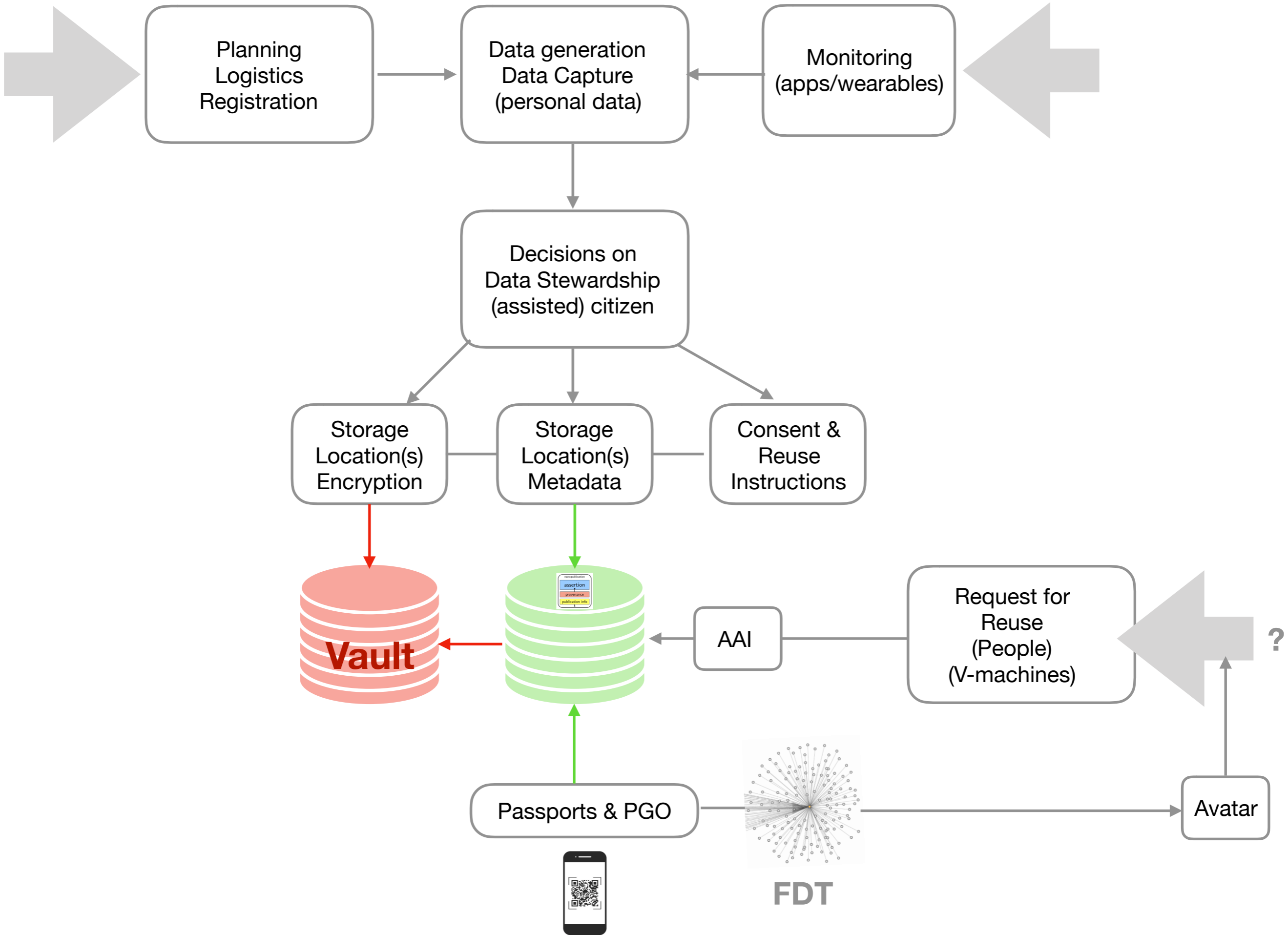
Trains - FAIR algorithms



FAIR

Machines know what it means





Planning
Logistics
Registration

Data generation
Data Capture
(personal data)

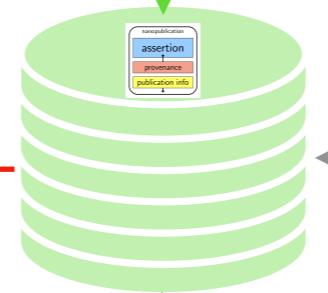
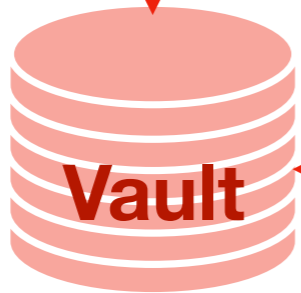
Monitoring
(apps/wearables)

Decisions on
Data Stewardship
(assisted) citizen

Storage
Location(s)
Encryption

Storage
Location(s)
Metadata

Consent &
Reuse
Instructions



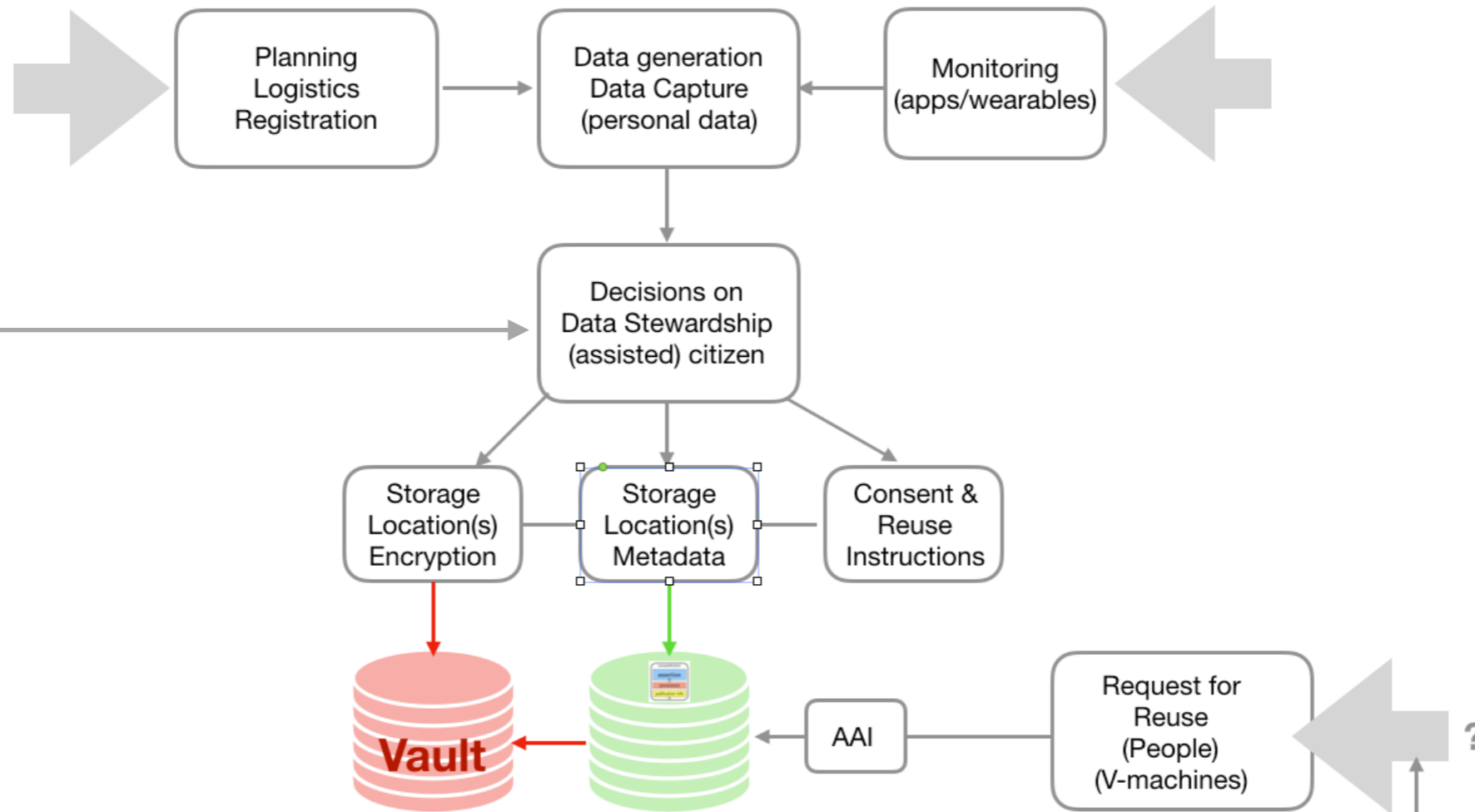
Request for
Reuse
(People)
(V-machines)

Passports & PGO



Avatar

?



Data Visiting and Distributed Research Services



Data Visiting and Distributed Research Services

Hotel Room Level

- Group rationalisation
- Query formation
- Analysis
- Discussion

Entry ('portal') Level

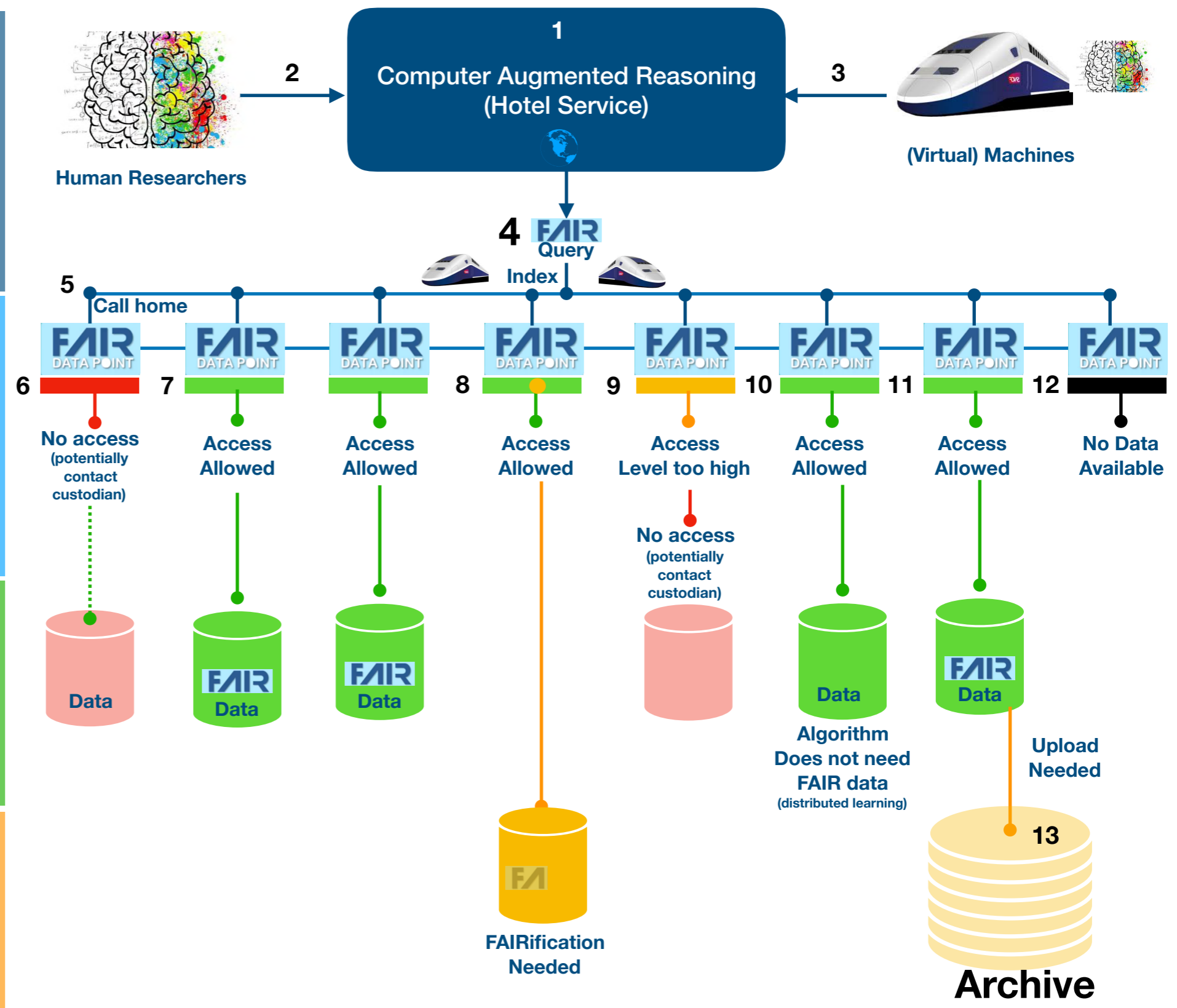
- Metadata (separate from data)
- Ontology Based Access Control
- Reporting about search results
- Data Access

(HP) re-use services

- FAIR data 'ready to use'
- Compute available
- Algorithms can run safely
- Log and provenance capture

Archiving services

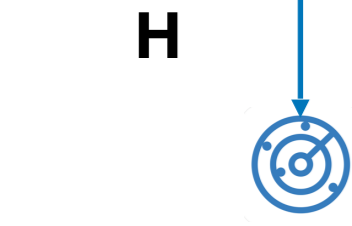
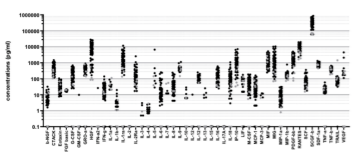
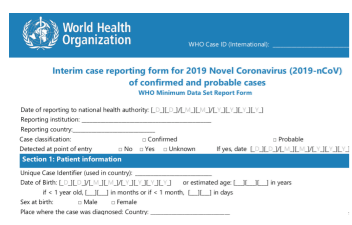
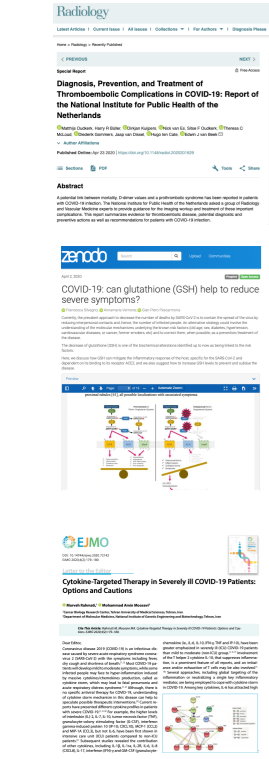
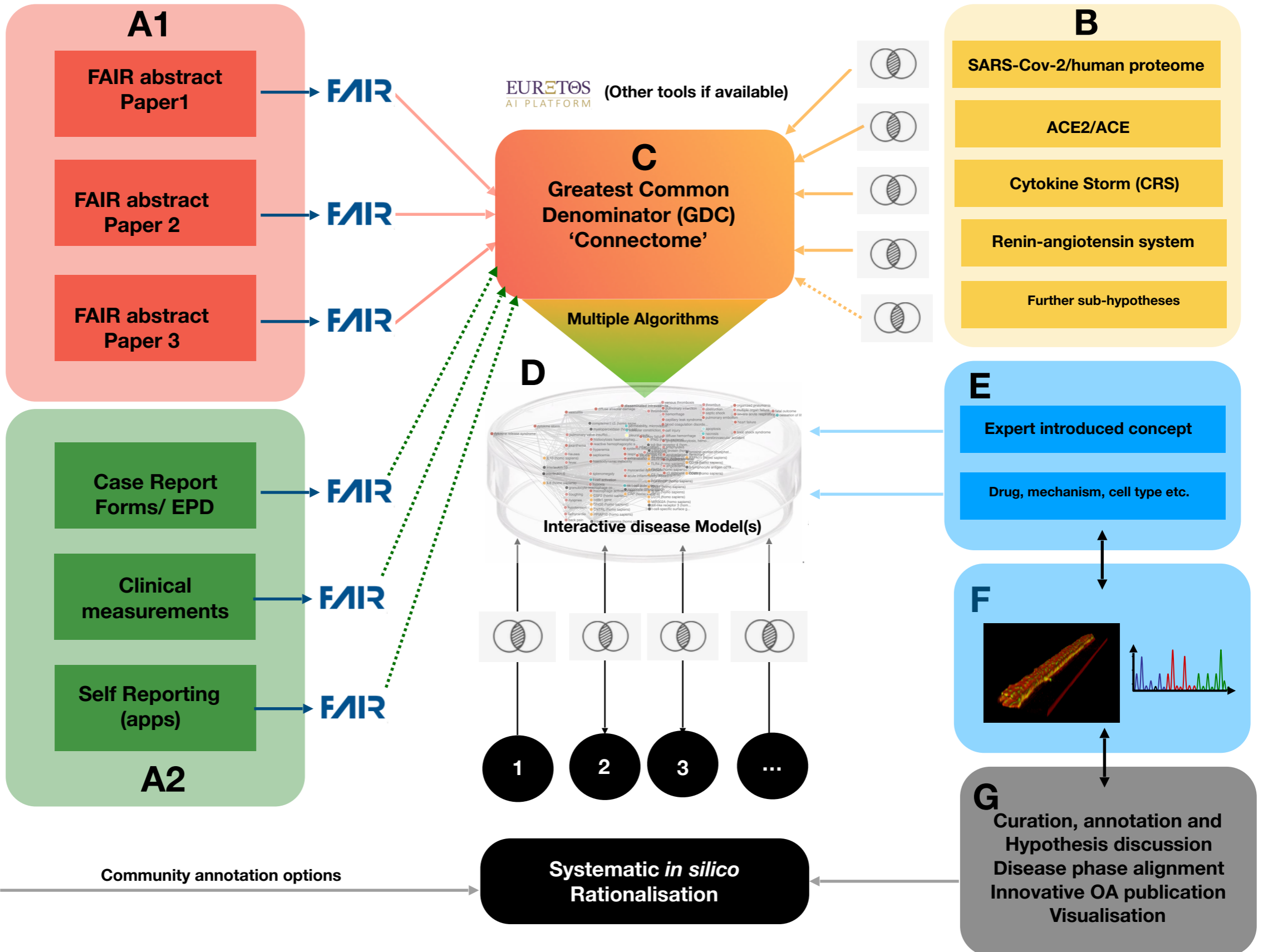
- Legacy data archived
- Upload to re-use services
- FAIR metadata may explain actual data is not (yet) FAIR

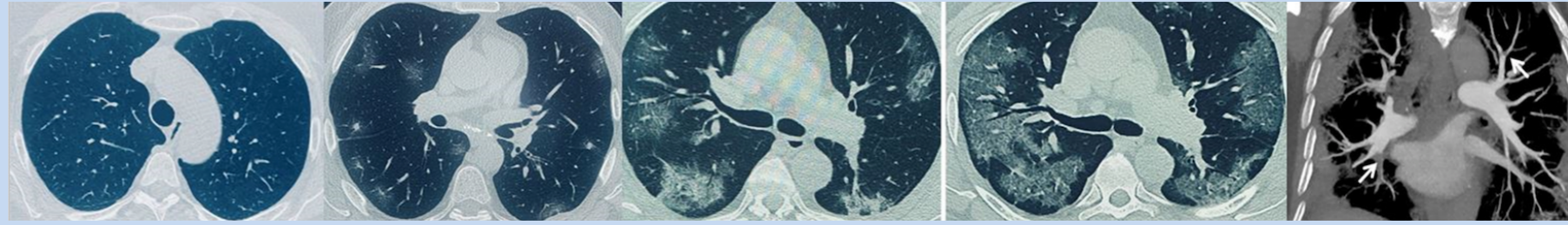


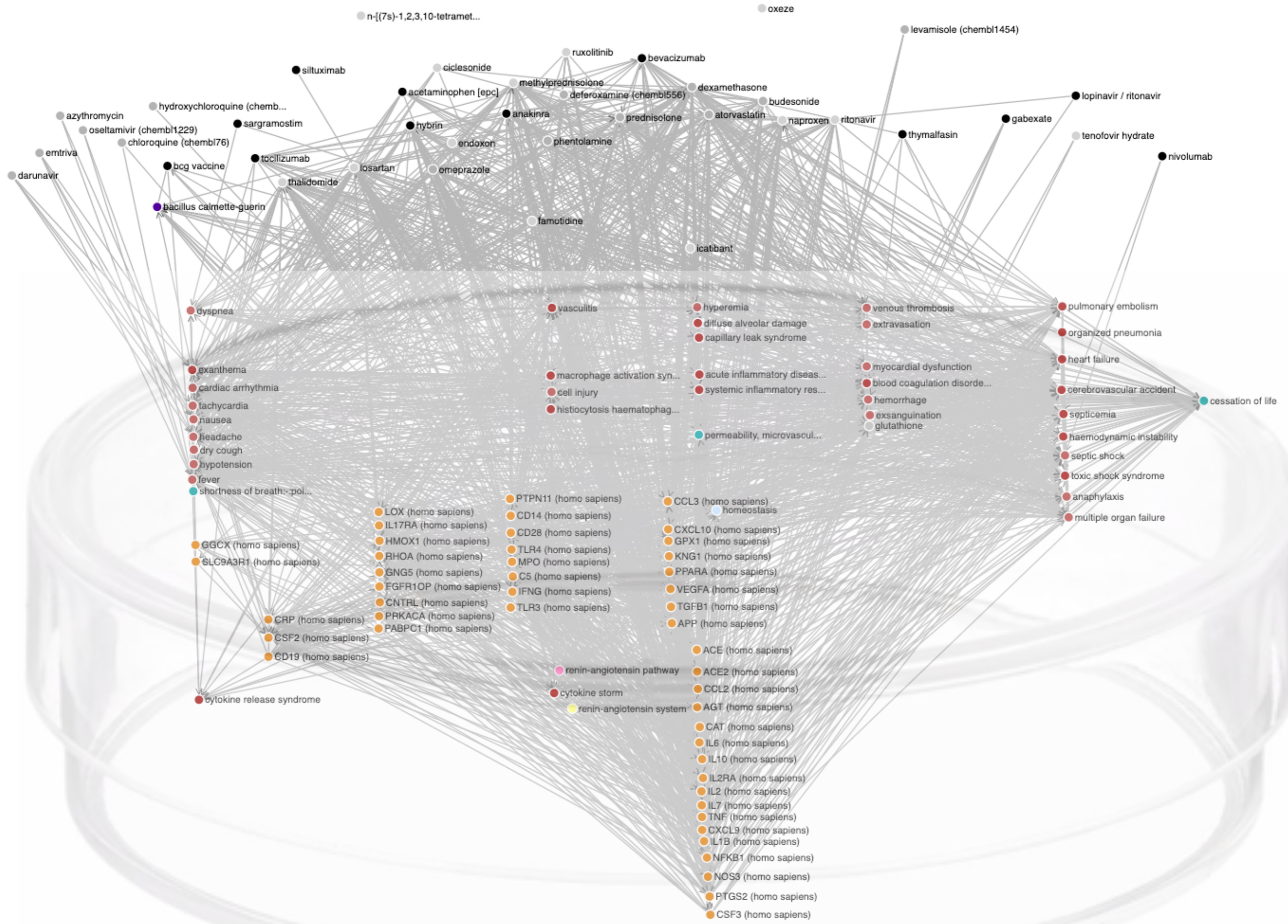
New Publications
Real World Observations
Clinical and Self reporting

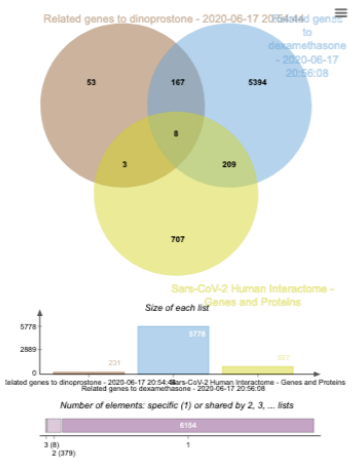
Disease Modelling Workflow (COVID-19)

AI-ready Established Knowledge
Plus selected connectors
Subhypotheses







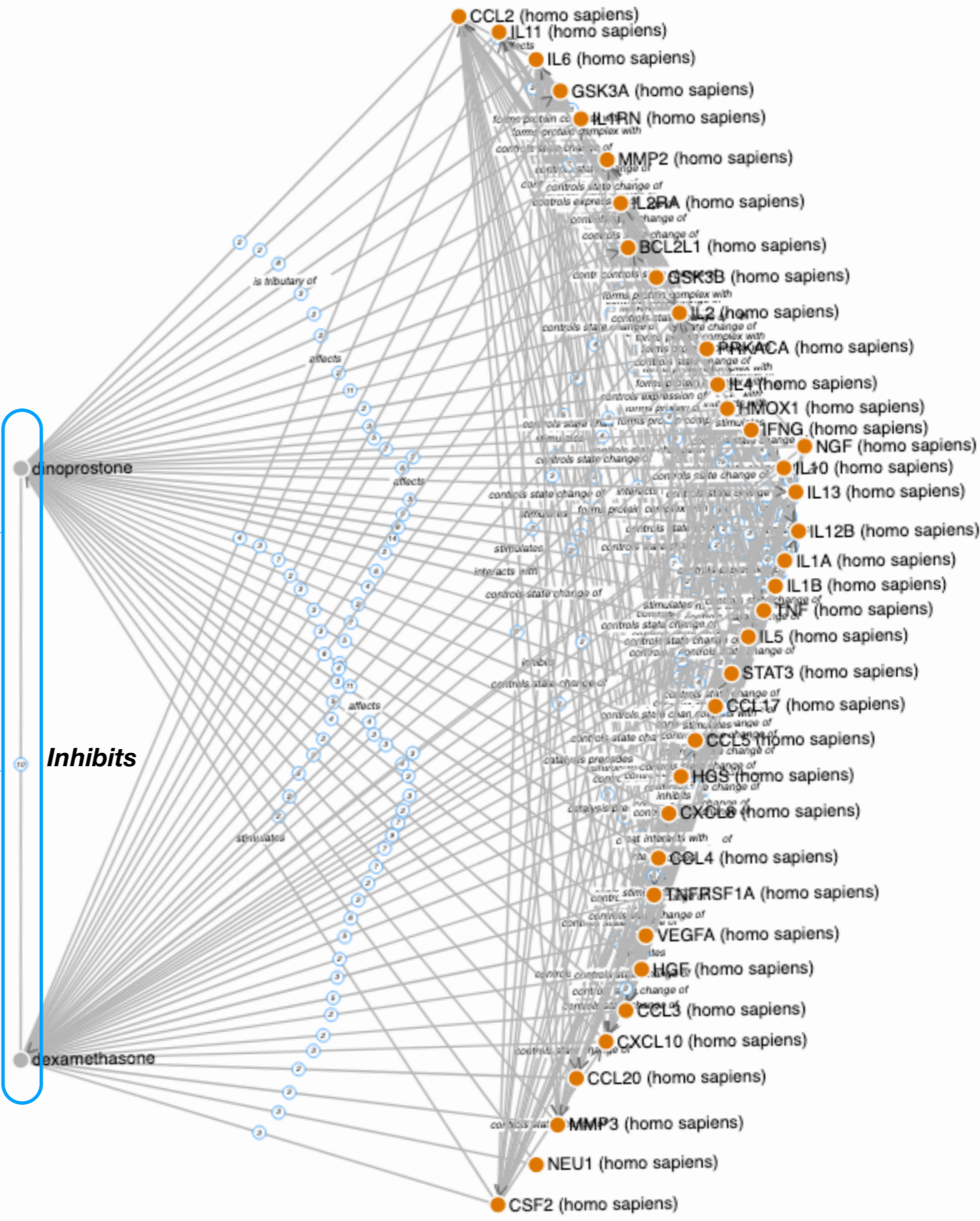
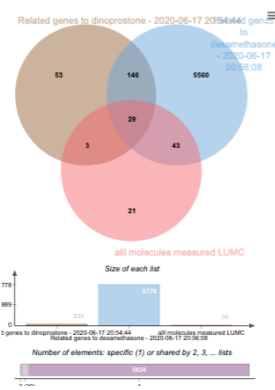


Cardinal assertion

EURETOS
AI PLATFORM

Provenance

Supporting or contesting Evidence



Workers Search available workers

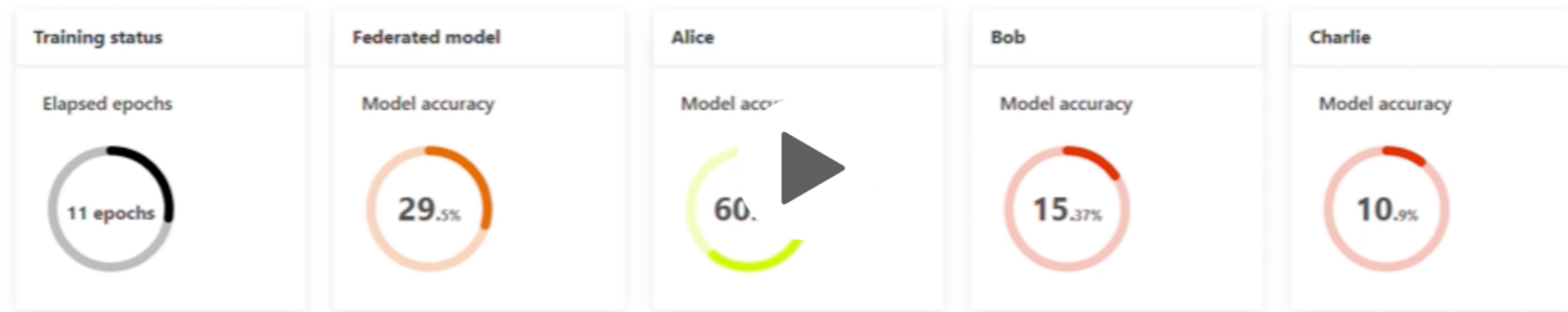
- Alice
- Bob
- Charlie

Training epochs

Federate rate every epochs

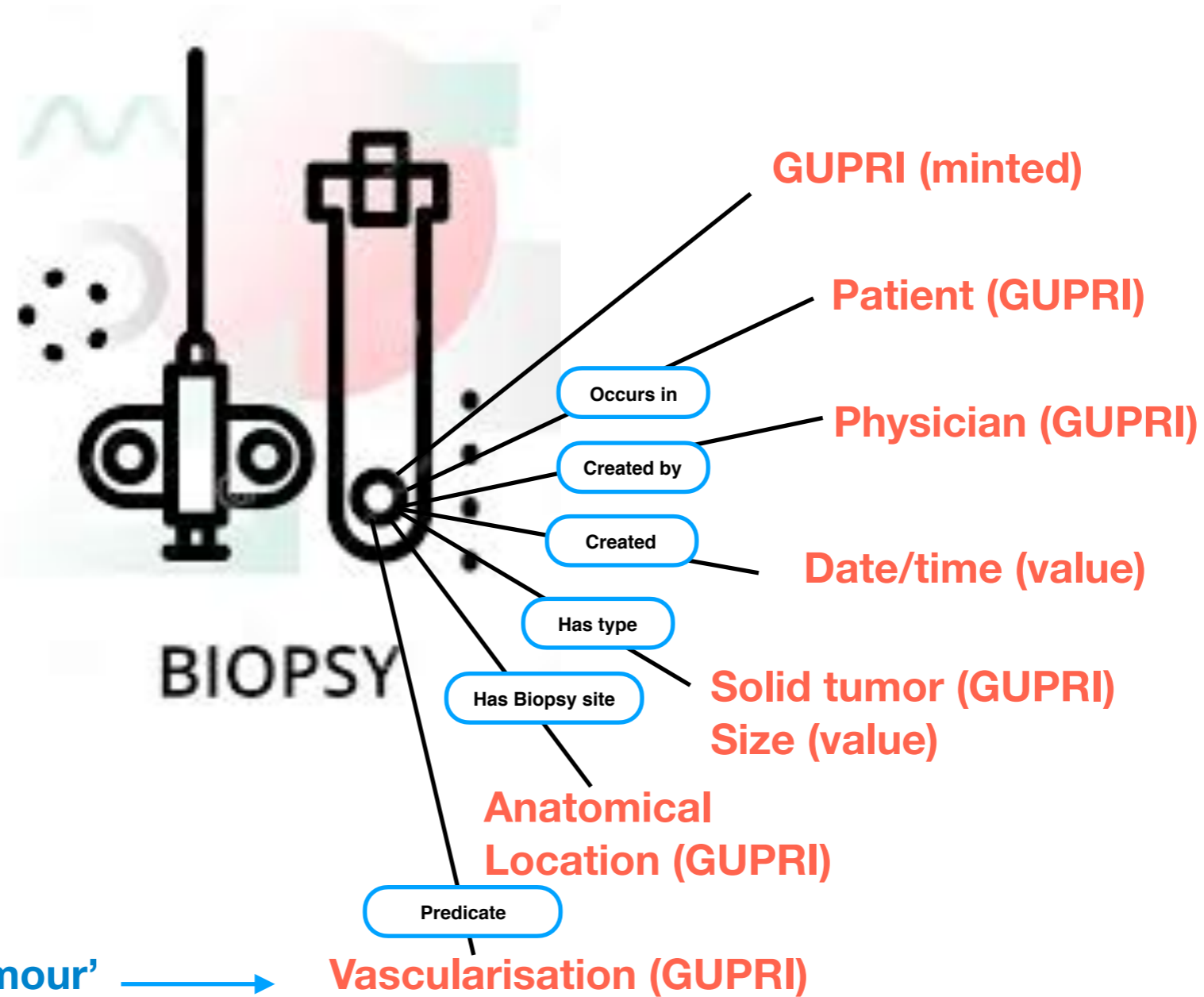
Stop federated learning

Training statistics

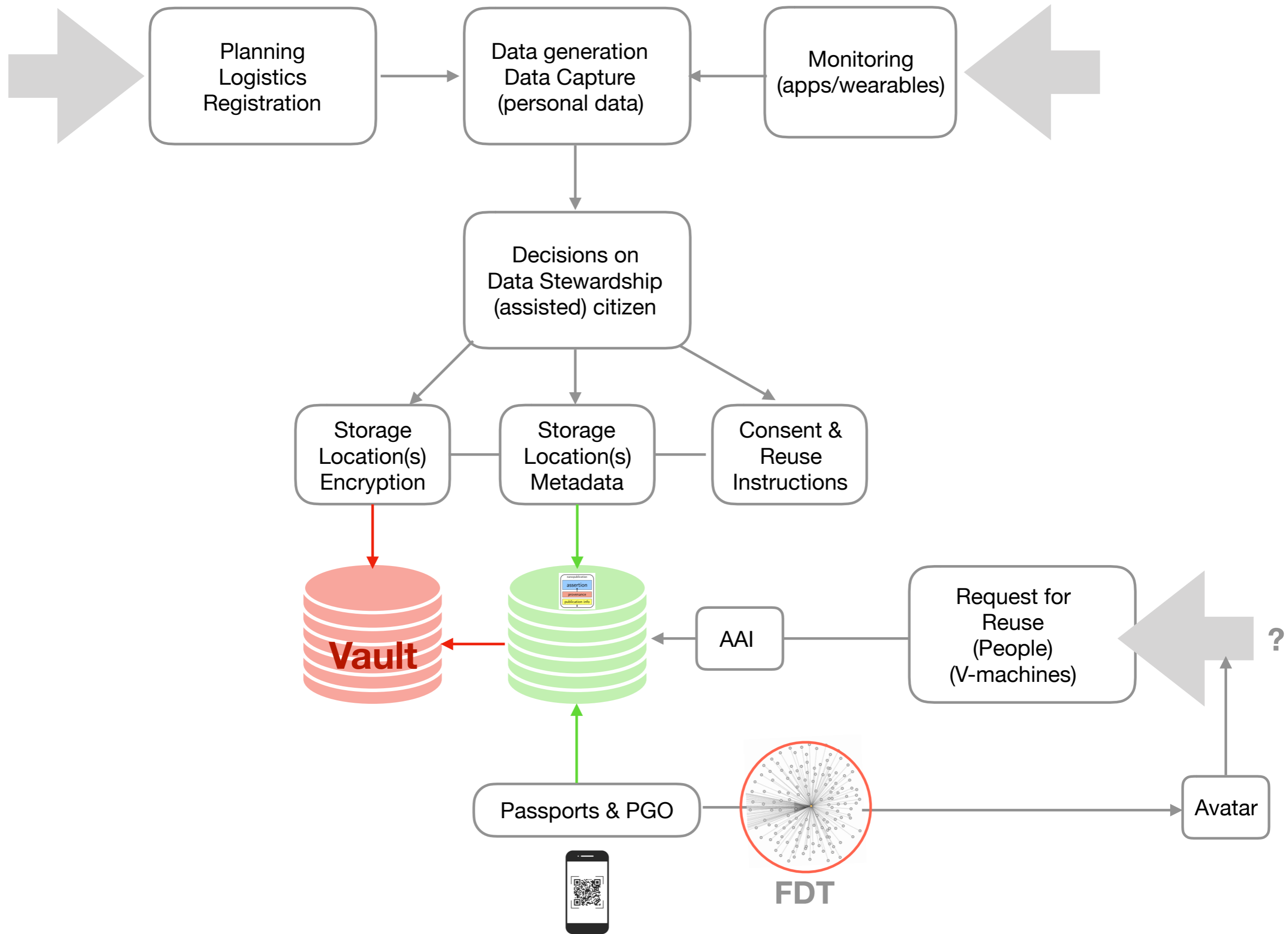


Logs ▼

FAIR Digital twins of tumors:



All CA's 'about the tumour'



Planning
Logistics
Registration

Data generation
Data Capture
(personal data)

Monitoring
(apps/wearables)

Decisions on
Data Stewardship
(assisted) citizen

Storage
Location(s)
Encryption

Storage
Location(s)
Metadata

Consent &
Reuse
Instructions

Vault

assertion
provenance
publication info

AAI

Request for
Reuse
(People)
(V-machines)

?

Avatar

Passports & PGO

FDT



Thank you for your attention



GO FAIR Foundation
Rijnsburgerweg 10
2333 AA Leiden
The Netherlands
info@gofairfoundation.org



Download PDF here