

The FAIR Hourglass: FAIRification – concepts and tools to create FAIR data

Erik Schultes
FAIR Implementation Lead
<http://orcid.org/0000-0001-8888-635X>
eriks@gofair.foundation

<https://www.gofair.foundation>
June 22 2022
Slides: <https://osf.io/g349c/>



What is FAIR?

Box 2 | The FAIR Guiding Principles

<https://www.nature.com/articles/sdata201618>

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

Automating F, A, I, and R

What is FAIR?

Box 2 | The FAIR Guiding Principles

<https://www.nature.com/articles/sdata201618>

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

Automating F, A, I, and R

What is FAIR?

Box 2 | The FAIR Guiding Principles

<https://www.nature.com/articles/sdata201618>

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

Automating F, A, I, and R

What is FAIR?

Box 2 | The FAIR Guiding Principles

<https://www.nature.com/articles/sdata201618>

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

Automating F, A, I, and R

What is FAIR?

Box 2 | The FAIR Guiding Principles

<https://www.nature.com/articles/sdata201618>

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:

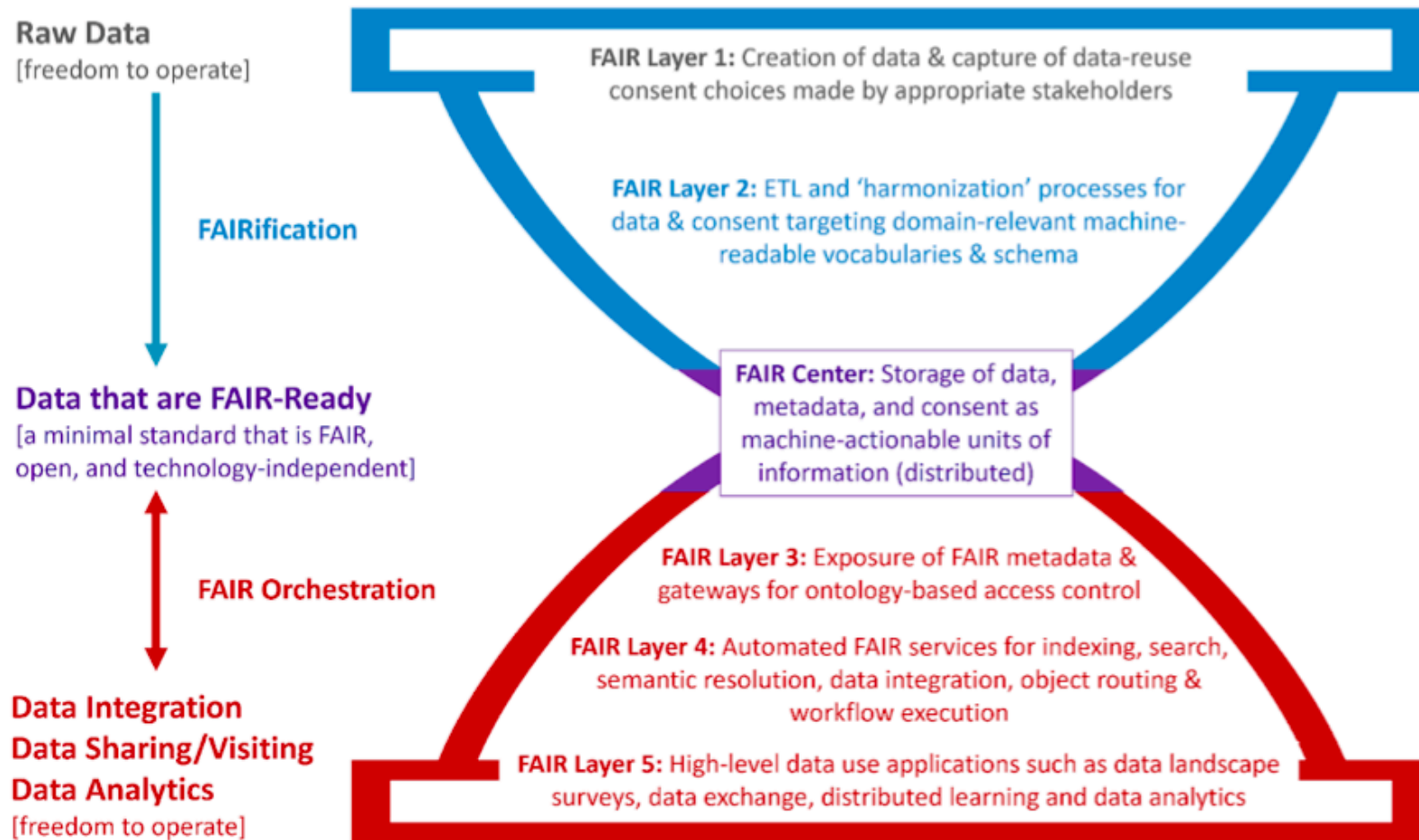
- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



Technical infrastructure (generic operations)

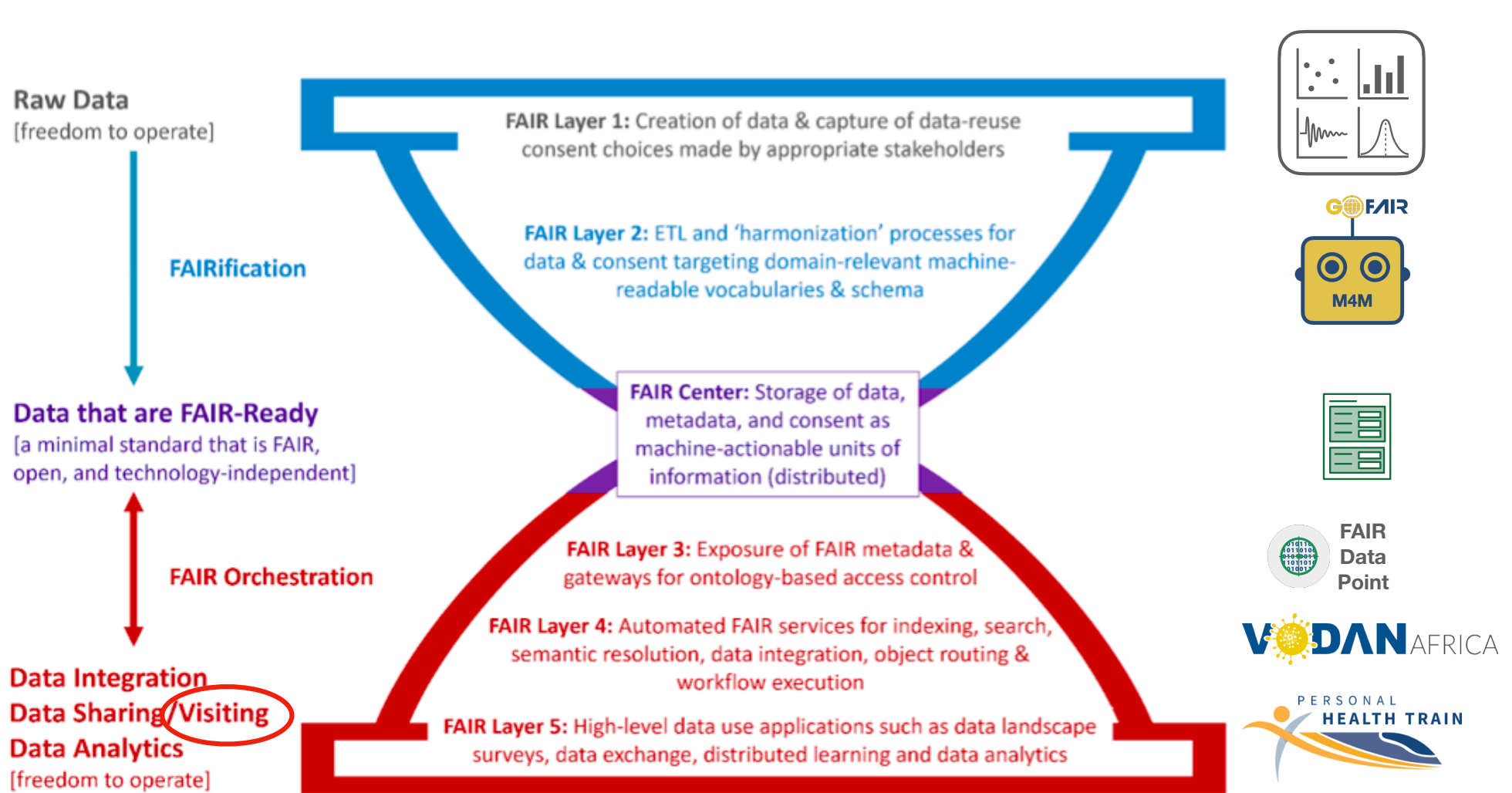
Data/metadata (domain-specific content)

FAIR Hourglass



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

FAIR Hourglass



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

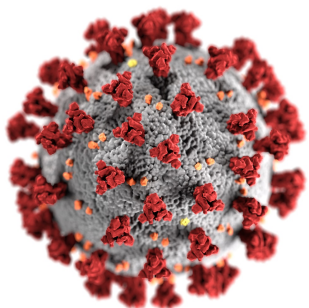
FAIR & Data Visiting

- Data Visting vs Data Sharing (sharing implies “giving away” and the loss of control over your data)
- Data Visting (Distributed Learning) avoids the “copy and transport” of vast amounts of data or data that is privacy-sensitive, by providing the **ability to analyze data at its source**, avoiding data centers and cloud environments.
- Data Visting requires mobile FAIR AI (**Trains**) and **FAIR Data Stations** where machine-actionable metadata standards are used to control access to data.



Data Visiting is emerging globally as an alternative to “data centers” and “cloud” environments. Examples:

- Dutch initiatives
 - Dutch FARM Data Train: <https://bigdata-wfsr.wur.nl/2020/09/09/fair-data-train/>
 - Dutch AI Coalition: <https://www.youtube.com/watch?v=L63ka2q4LNM>
 - Health-Holland
 - Trusted World of Cornona: <http://www.twoc.eu>
 - Citizen Centred and Controlled COVID-19 data for reuse: <https://www.health-holland.com/project/2021/2021/citizen-centred-and-controlled-covid-19-data-reuse>
- International initiatives
 - **Swiss Personalized Health Network (SPHN):** <https://www.youtube.com/watch?v=pqV0qp4oisM>
 - FAIR Data for European COVID-19 Response: <https://doi.org/10.6084/m9.figshare.c.5740430.v1>
 - VODAN Africa: <https://www.vodan-totafrica.info>



How to GO FAIR

Three-Point FAIRification Framework

[Home](#) › [How to GO FAIR](#) <https://www.go-fair.org/how-to-go-fair/>

How to GO FAIR

Since its beginning in early 2018, the GO FAIR community has been working towards implementations of the **FAIR Guiding Principles**. This collective effort has resulted in a three-point framework that formulates the essential steps towards the end goal, a global Internet of FAIR Data and Services where data are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**) for machines.

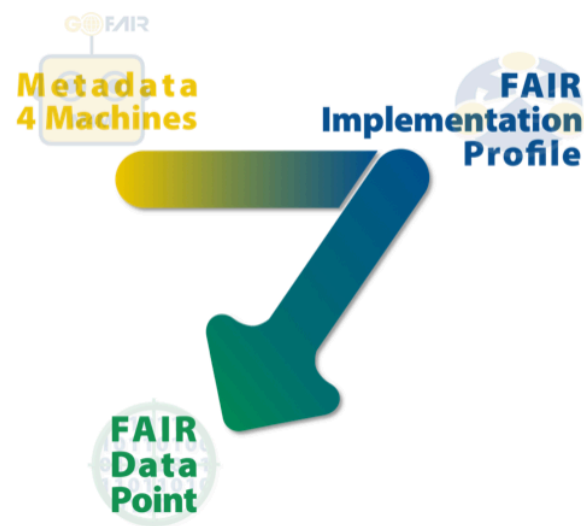


A framework guiding FAIRification

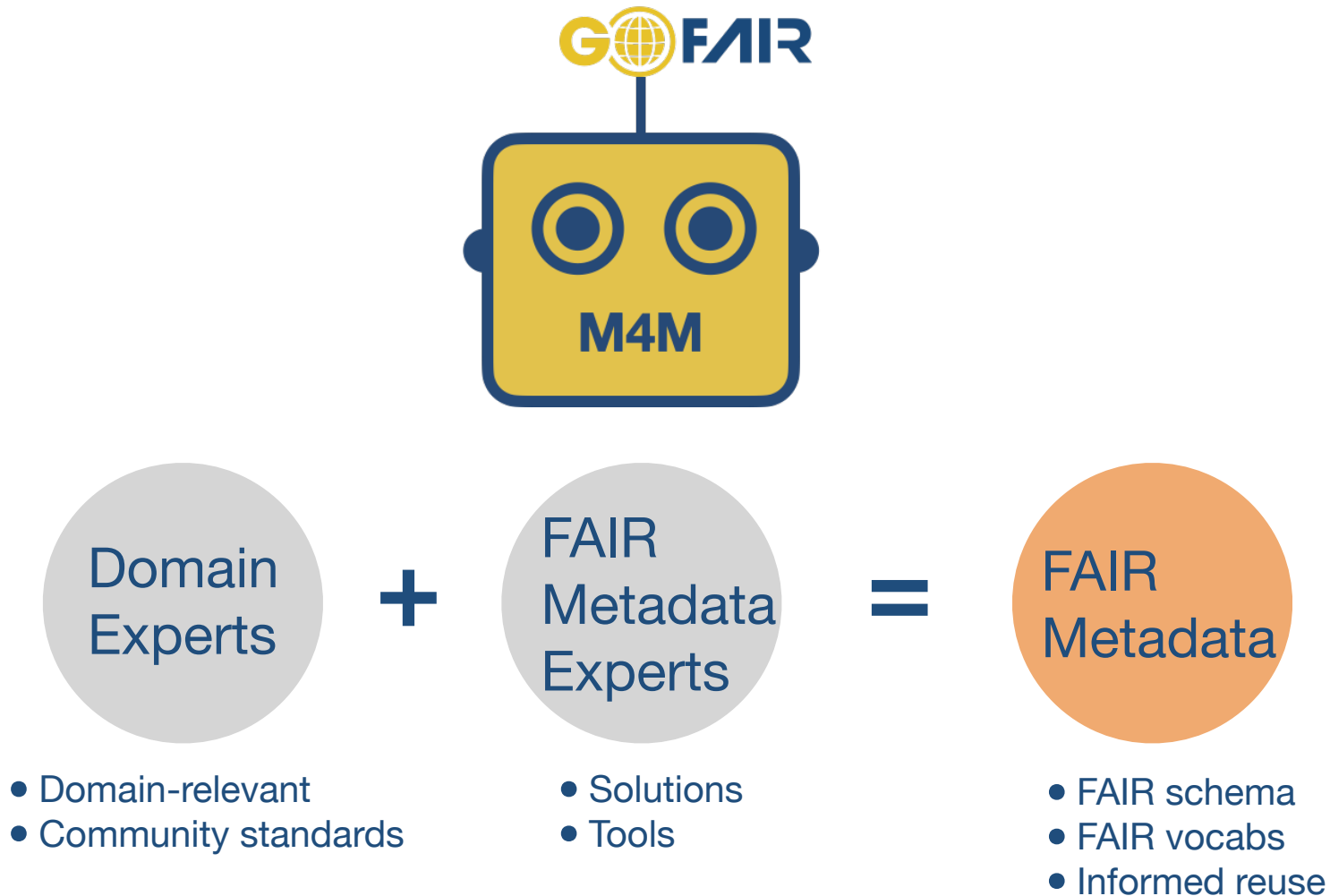
The Three-point FAIRification Framework provides practical “how to” guidance to stakeholders seeking to go FAIR.

Moreover, by following this framework, stakeholders can rest assured that their efforts toward FAIRification will be optimally coordinated with the efforts of other stakeholders in the GO FAIR community. The three-point framework maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services.

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in **Metadata for Machines** (M4M) Workshops.



Metadata for Machines Workshops



Community (decision making) is central



Metadata for Machines for ZonMw

ZonMw
COVID-19 Program
62 Projects

Brainstorm
controlled
lists

M4M.7
13 projects
21 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

M4M.8
10 projects
17 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

M4M.9
6 projects
12 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

M4M.10
7 projects
13 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

M4M.11
9 projects
16 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

M4M.12
14 projects
21 participants

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

November

**COVID-19 Program
controlled list**

1. Data
2. Images
3. Biomaterials
4. Services
5. Standards

February

Feb 9: Consolidation of lists and terms describing the project assets
Feb 17: Building the COVID-19 Program Vocabulary



<https://covid19initiatives-test.health-ri.nl/p/ProjectOverview>

March



Project Admin
Metadata



Project Content
Metadata



Generic Dataset
Metadata

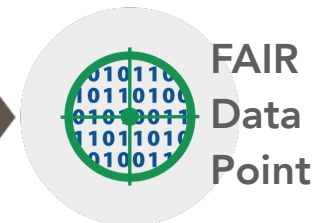


COVID
Program
Duration

**COVID-19 Program
metadata input forms**

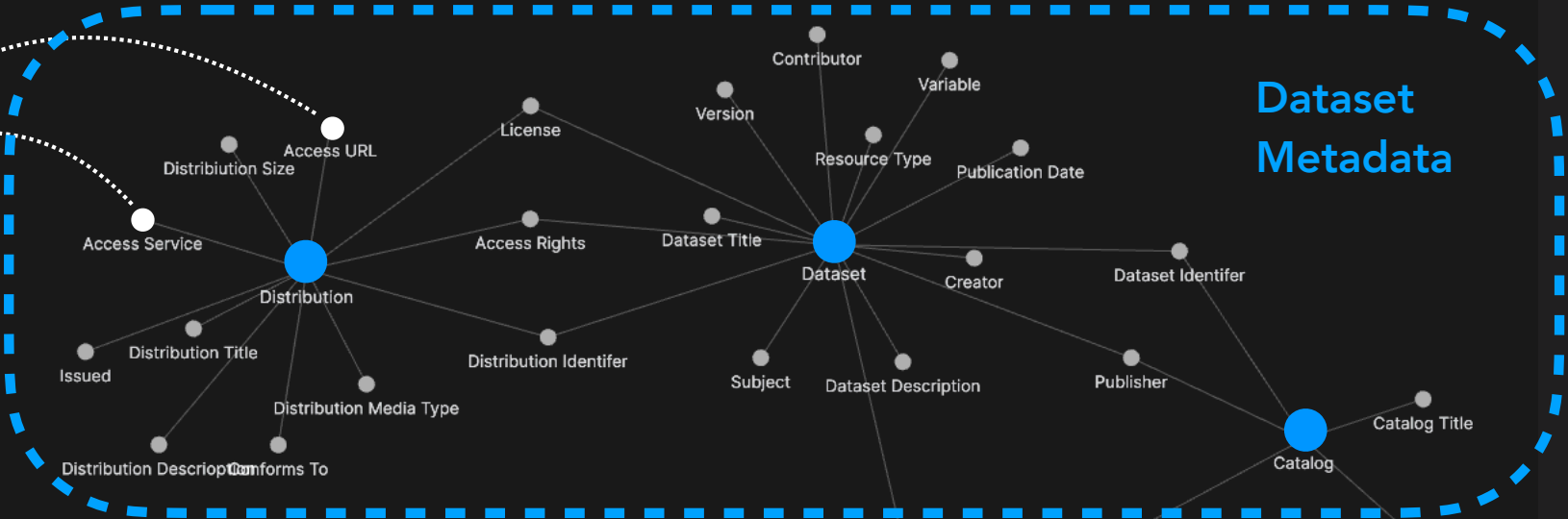
observable property	
observable property	
average physical exercise duration per day	
observed property	
duration	
observed object	
physical exercise	
matrix	
human	
interval	
per day	
modifier	
average	
unit	
s	

**Data Stewards
support researchers
to routinely create
metadata instances**

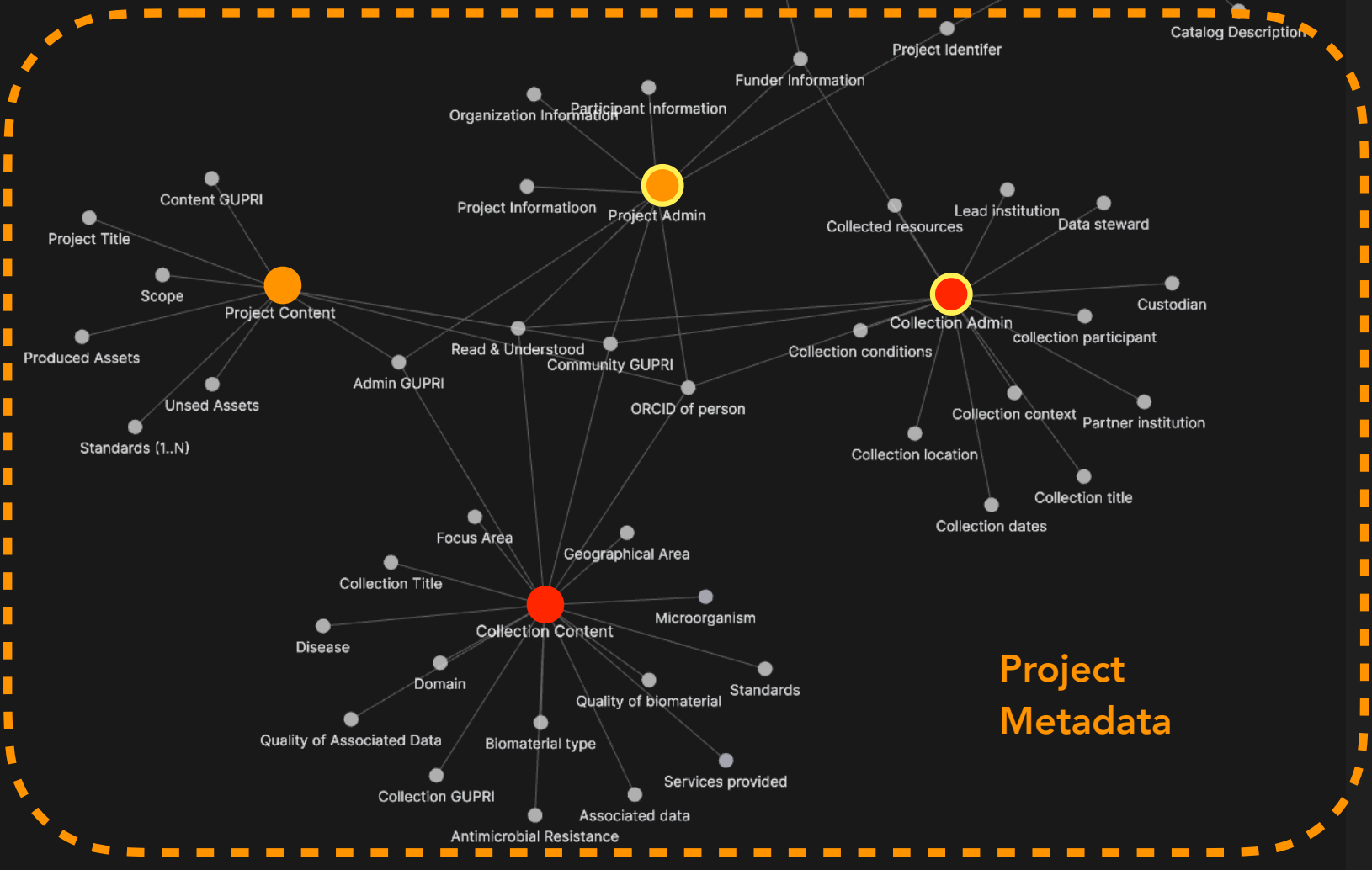




The data



Dataset Metadata



Project Metadata

FAIR Implementation Profile



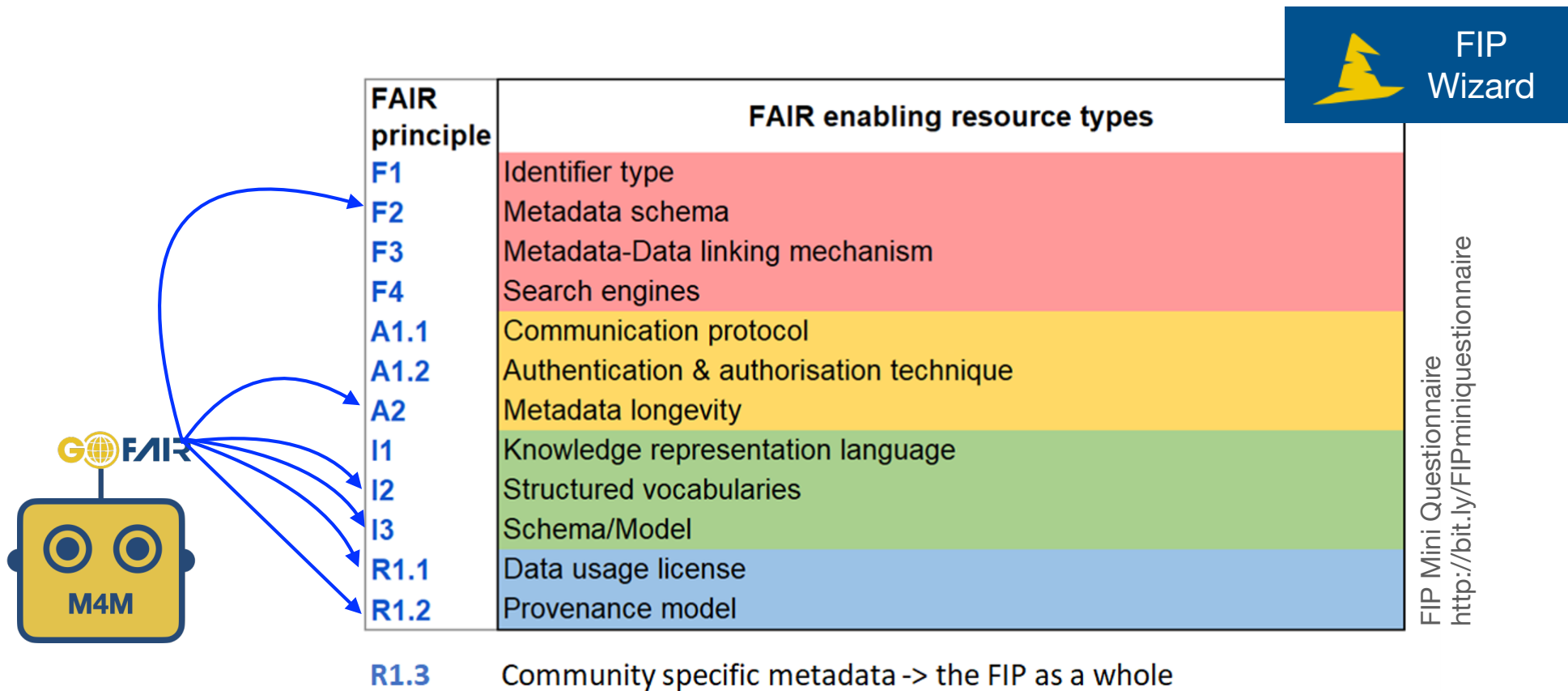
FAIR principle	FAIR enabling resource types
F1	Identifier type
F2	Metadata schema
F3	Metadata-Data linking mechanism
F4	Search engines
A1.1	Communication protocol
A1.2	Authentication & authorisation technique
A2	Metadata longevity
I1	Knowledge representation language
I2	Structured vocabularies
I3	Schema/Model
R1.1	Data usage license
R1.2	Provenance model

FIP Mini Questionnaire
<http://bit.ly/FIPminiquestionnaire>

R1.3 Community specific metadata -> the FIP as a whole

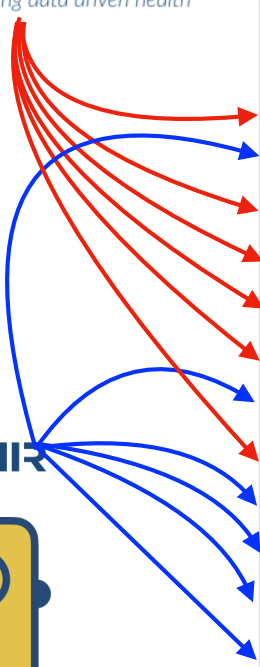
Community (decision making) is central

FAIR Implementation Profile



Community (decision making) is central

FAIR Implementation Profile



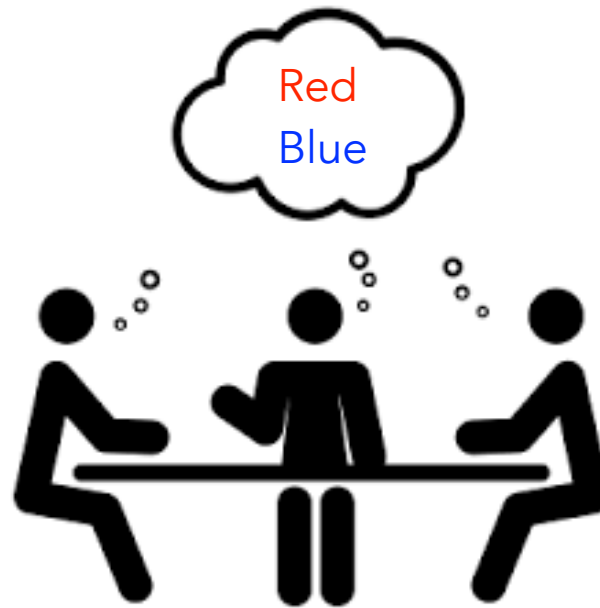
FAIR principle	FAIR enabling resource types
F1	Identifier type
F2	Metadata schema
F3	Metadata-Data linking mechanism
F4	Search engines
A1.1	Communication protocol
A1.2	Authentication & authorisation technique
A2	Metadata longevity
I1	Knowledge representation language
I2	Structured vocabularies
I3	Schema/Model
R1.1	Data usage license
R1.2	Provenance model

R1.3 Community specific metadata -> the FIP as a whole

FIP Mini Questionnaire
<http://bit.ly/FIPminiquestionnaire>

Community (decision making) is central

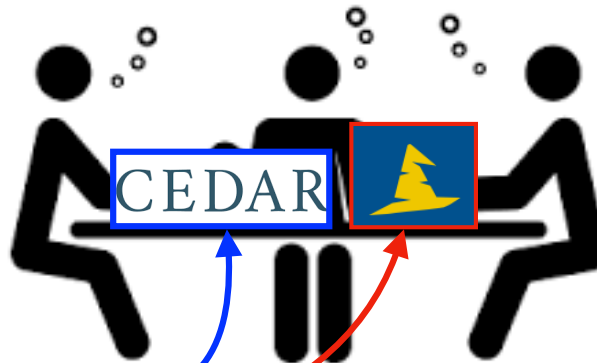
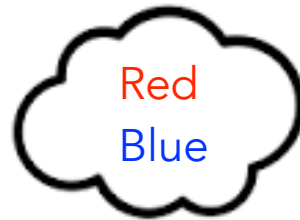
FAIR in Practice



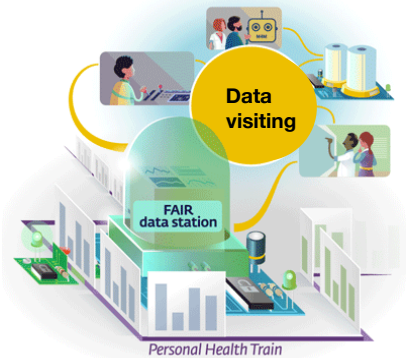
Community (decision making) is central

FAIR in Practice

- ZonMw COVID-19
- ZonMw ID/AMR
- ZonMw Dementia
- ZonMw ME/CVS
- Cancer (KWF)
- LUMC
- NFU
- Health-RI

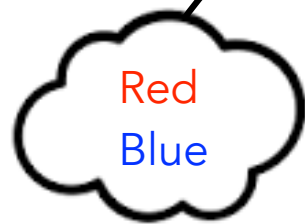


Community (decision making) is central



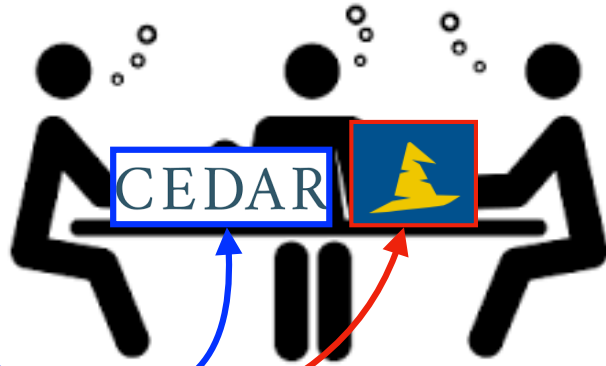
FAIR in Practice

- ZonMw COVID-19
- ZonMw ID/AMR
- ZonMw Dementia
- ZonMw ME/CVS
- Cancer (KWF)
- LUMC
- NFU
- Health-RI

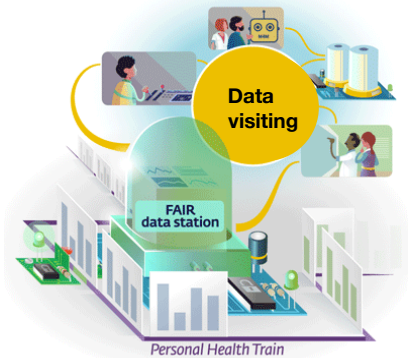


FIP

identifier type
metadata schema
metadata:Data linking mechanism
search engines
communication protocol
authentication & authorisation technique
metadata longevity
knowledge representation language
structured vocabularies
schema/Nodes
data usage license
insurance model

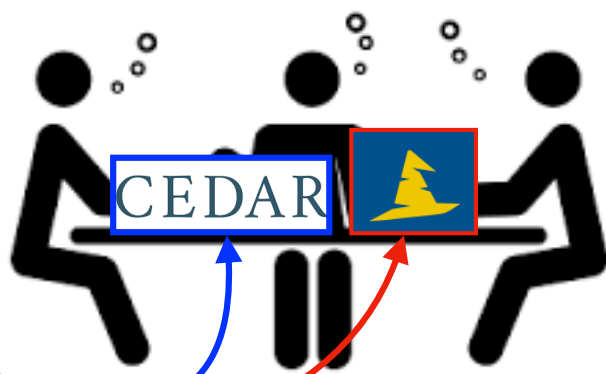
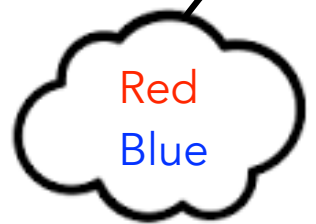


Community (decision making) is central



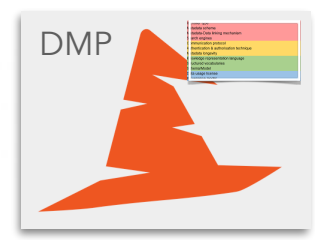
FAIR in Practice

- ZonMw COVID-19
- ZonMw ID/AMR
- ZonMw Dementia
- ZonMw ME/CVS
- Cancer (KWF)
- LUMC
- NFU
- Health-RI

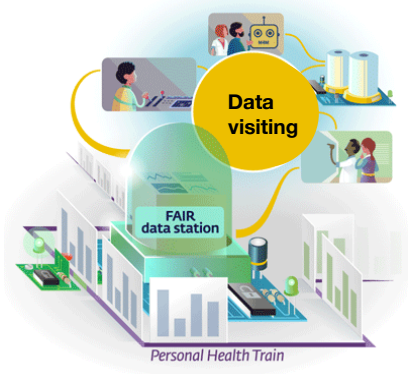


FIP

sensor type
metadata schema
metadata: Data linking mechanism
search engines
communication protocol
authentication & authorisation technique
metadata longevity
knowledge representation language
structured vocabularies
schema/Nodes
data usage license
insurance model

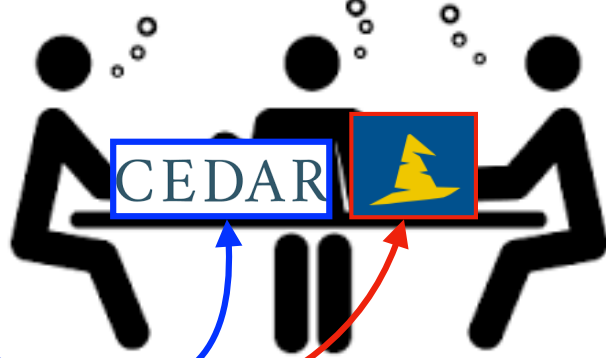
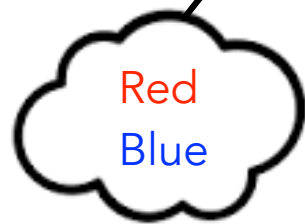


Community (decision making) is central



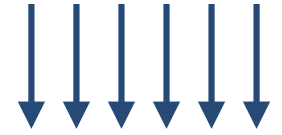
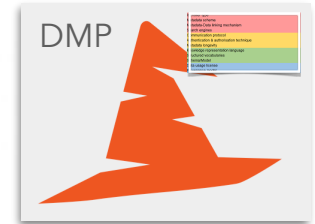
FAIR in Practice

- ZonMw COVID-19
- ZonMw ID/AMR
- ZonMw Dementia
- ZonMw ME/CVS
- Cancer (KWF)
- LUMC
- NFU
- Health-RI

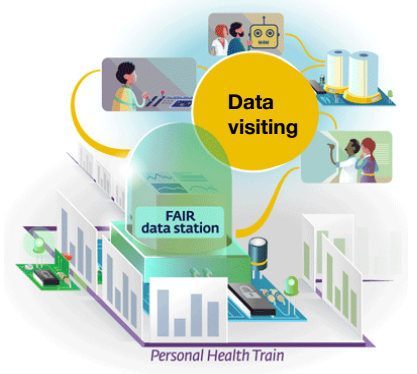


FIP

sensor type
metadata schema
metadata: Data linking mechanism
search engines
communication protocol
authentication & authorisation technique
metadata longevity
knowledge representation language
structured vocabularies
schema/models
data usage license
insurance model



Community (decision making) is central



FAIR in Practice

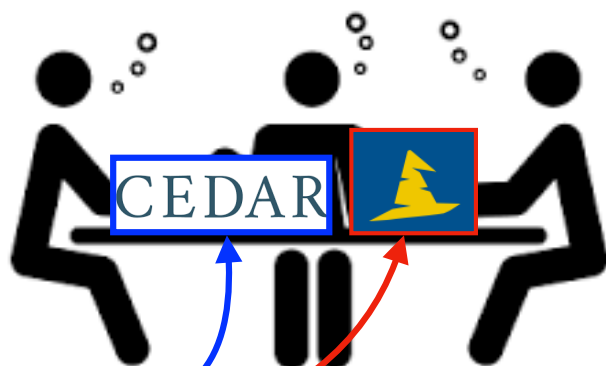
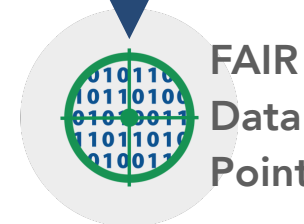
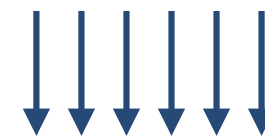
- ZonMw COVID-19
- ZonMw ID/AMR
- ZonMw Dementia
- ZonMw ME/CVS
- Cancer (KWF)
- LUMC
- NFU
- Health-RI

FAIR

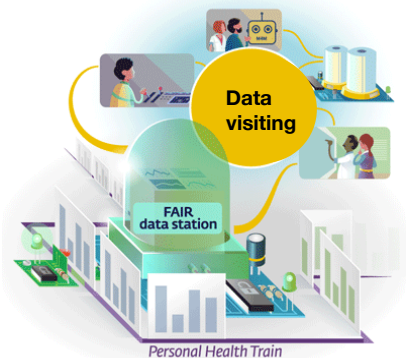


FIP

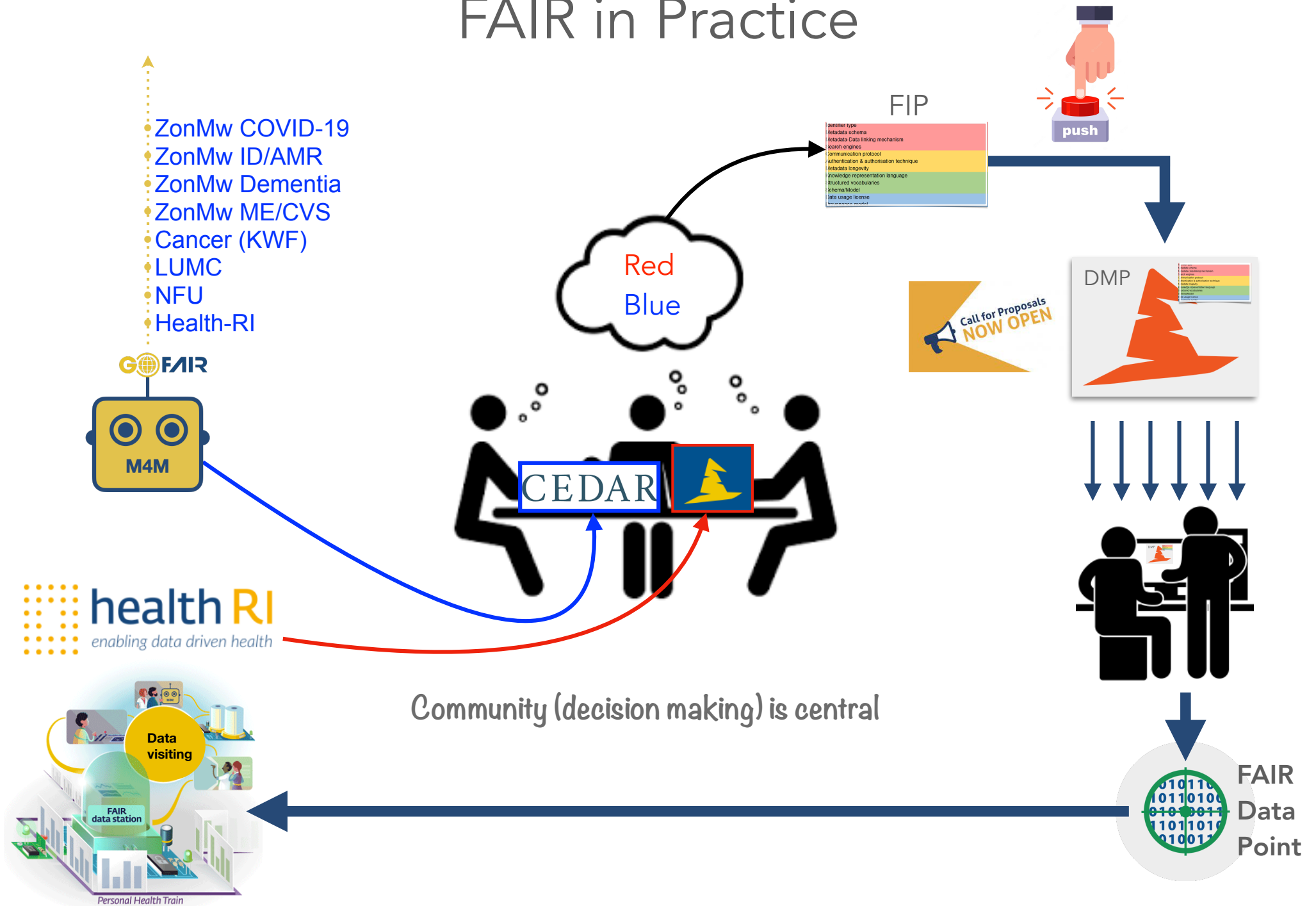
sensor type
metadata schema
metadata: Data linking mechanism
search engines
communication protocol
authentication & authorisation technique
metadata longevity
knowledge representation language
structured vocabularies
schema/Nodes
data usage license
insurance model



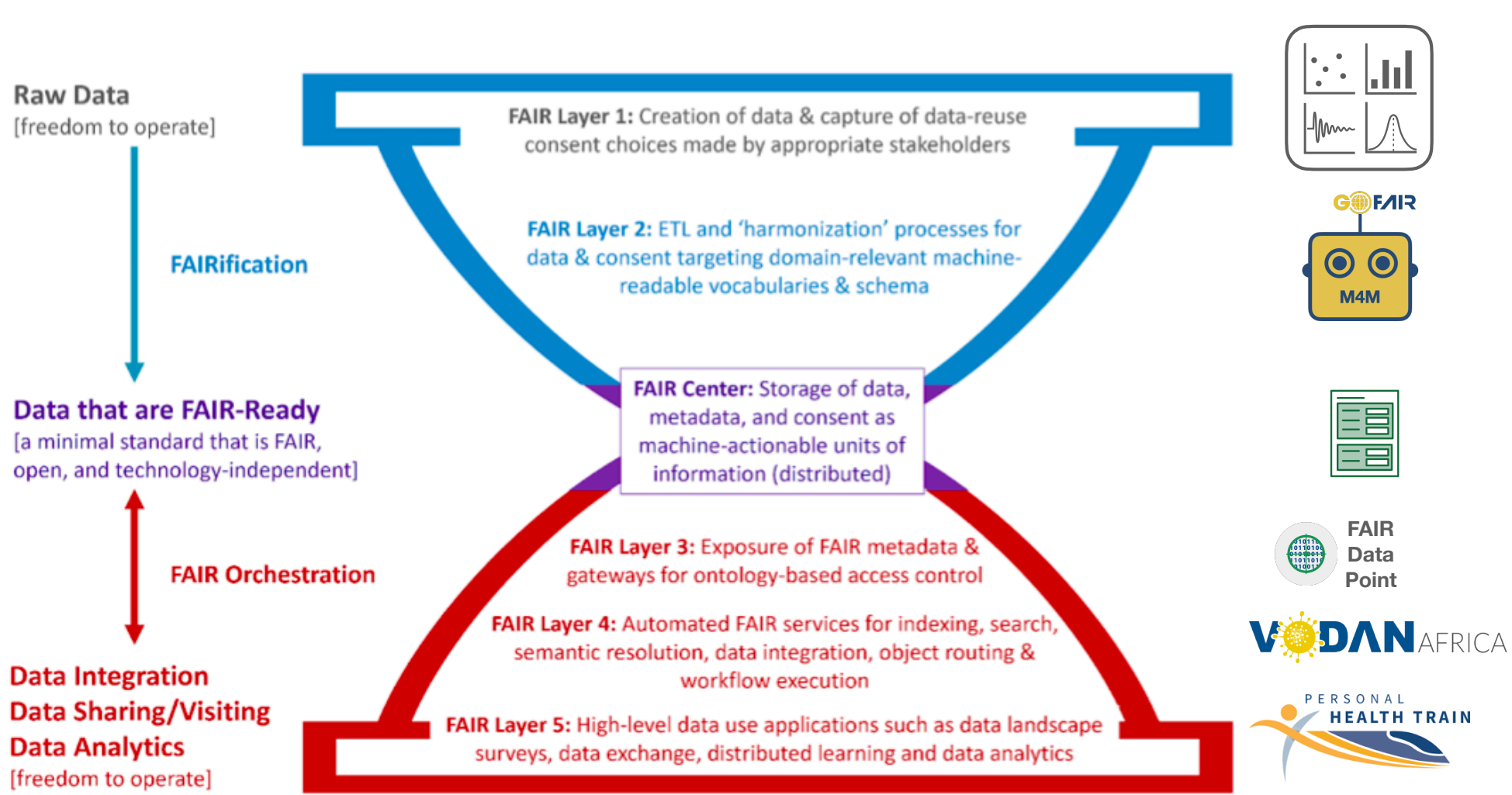
Community (decision making) is central



FAIR in Practice



FAIR Hourglass



FAIR well !

