

M4M Workshops for the ZonMw COVID 19 community

October 7, 2021

FAIRification STEP4 – domain-specific metadata webinar



METADATA 4 MACHINES



Involved experts



Erik Schultes
FAIR Implementation Lead
GO FAIR Foundation

Lead and organisation of
M4M workshops



Barbara Magagna
Data Architect, Semantic Expert
Environment Agency Austria
UTwente

Process facilitation in the creation of
community vocabularies and templates



Nikola Vasiljevic
Special Consultant for Digitalization
Department of Wind Energy
DTU Orbit

Design of metadata templates and
provision of FAIRification tools

Community: COVID-19 research community

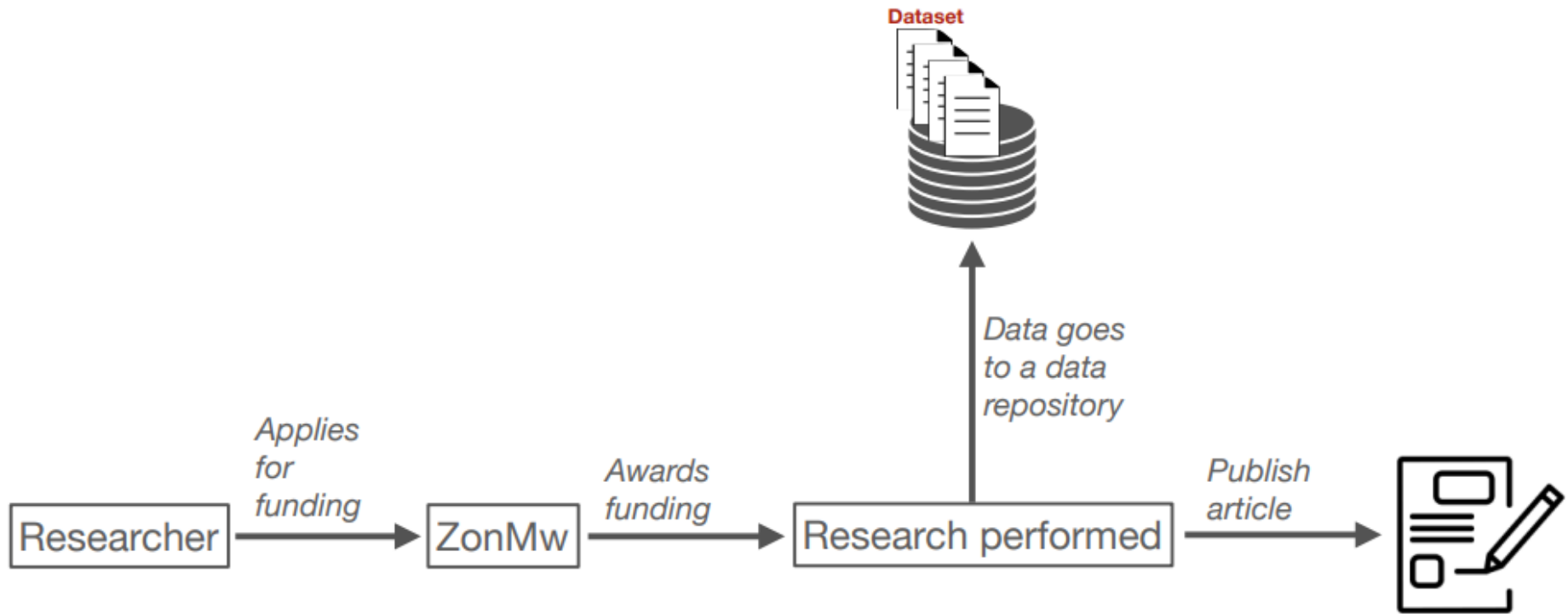
ZonMw: Dutch funder of health research and innovation projects.

-> **COVID-19 research programme** includes a series of training events for involved project data stewards in applying FAIRification approaches to improve interoperability of (meta)data in collaboration with GO FAIR, DTL and

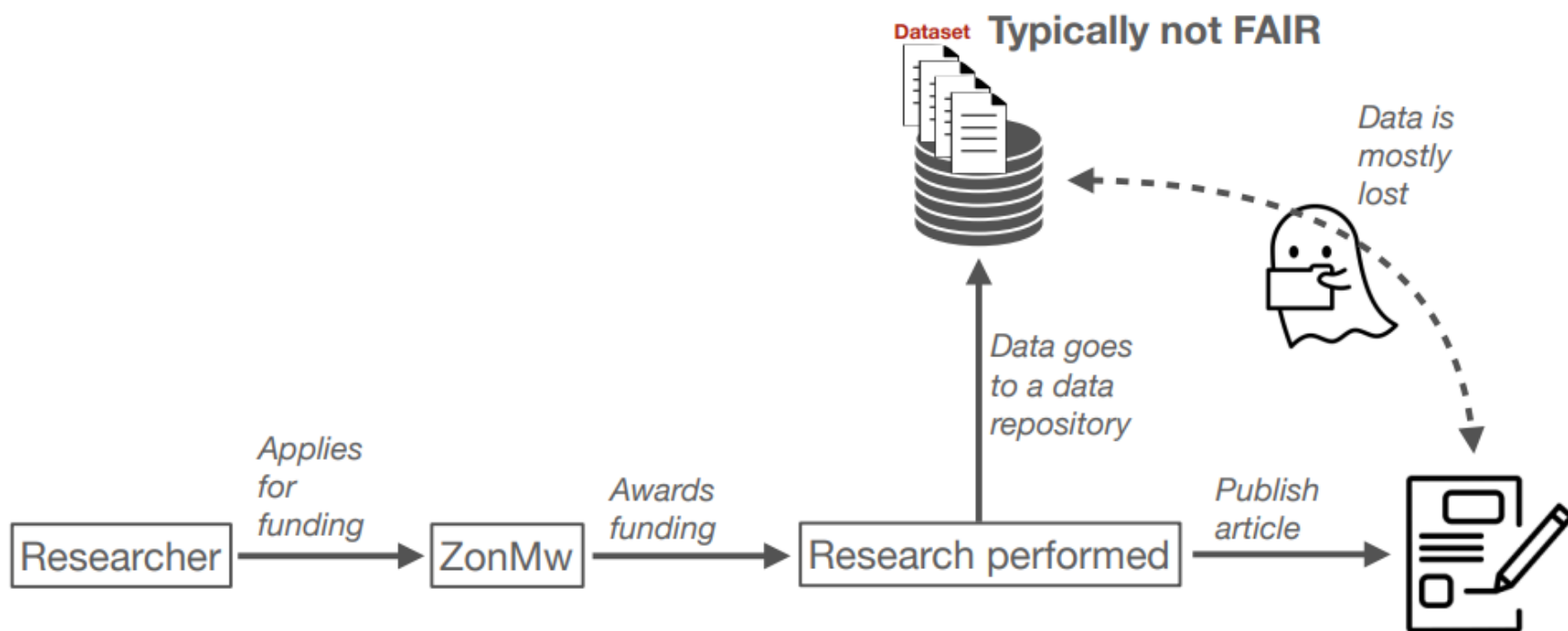
Health-RI: Dutch National Health Research Infrastructure for optimal access to knowledge, data and samples

Objective: Build a national COVID-19 observational and innovation portal with (meta)data from all more than 130 funded projects covering topics from health care to diagnostics, prognosis to immunology and socio-economic studies

Traditional Research Cycle



Traditional Research Cycle



What is needed to make data 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

FAIR is 90% machine-actionable metadata

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



Metadata for Machines

machine actionable metadata form





Metadata for Machines

Machine-actionable metadata form

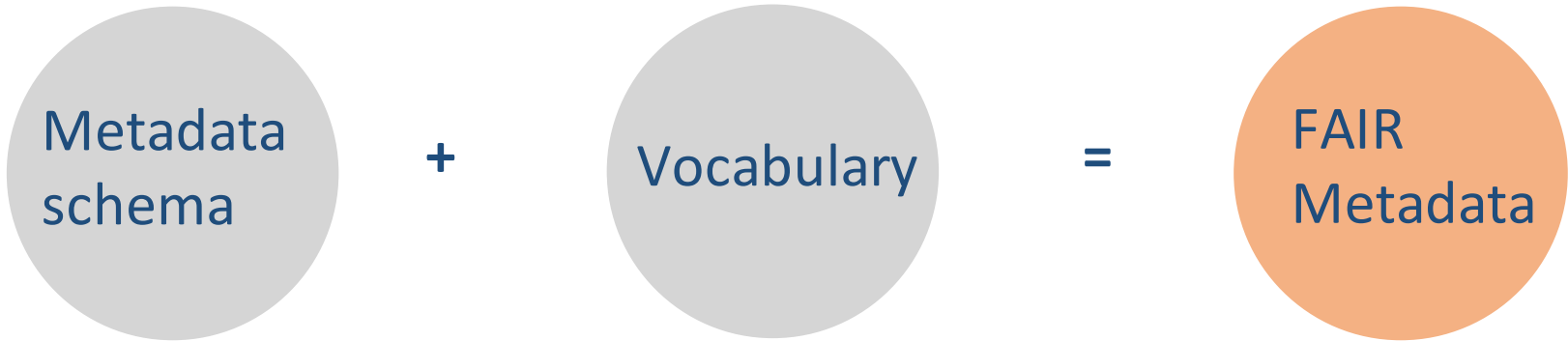
**Machine actionable metadata form =
metadata schema represented as linked data and JSON-LD**

- **LINKED DATA** builds upon standard Web technologies extending them to share information in a way that can be read automatically by machines. This enables data from different sources to be connected and queried.
- **JSON-LD** is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. **JSON-LD is an ideal data format for programming environments**, REST Web services, and unstructured databases such as Apache CouchDB and MongoDB.



Metadata for Machines

FAIR metadata





Metadata for Machines

FAIR metadata

FAIR metadata =

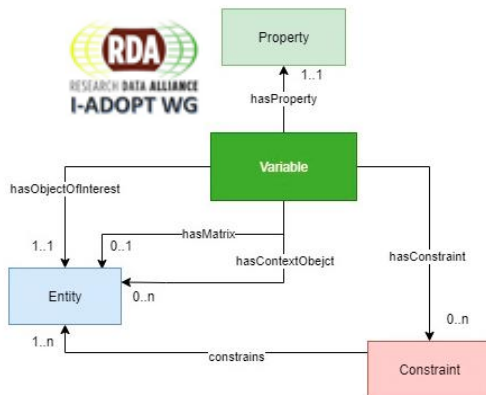
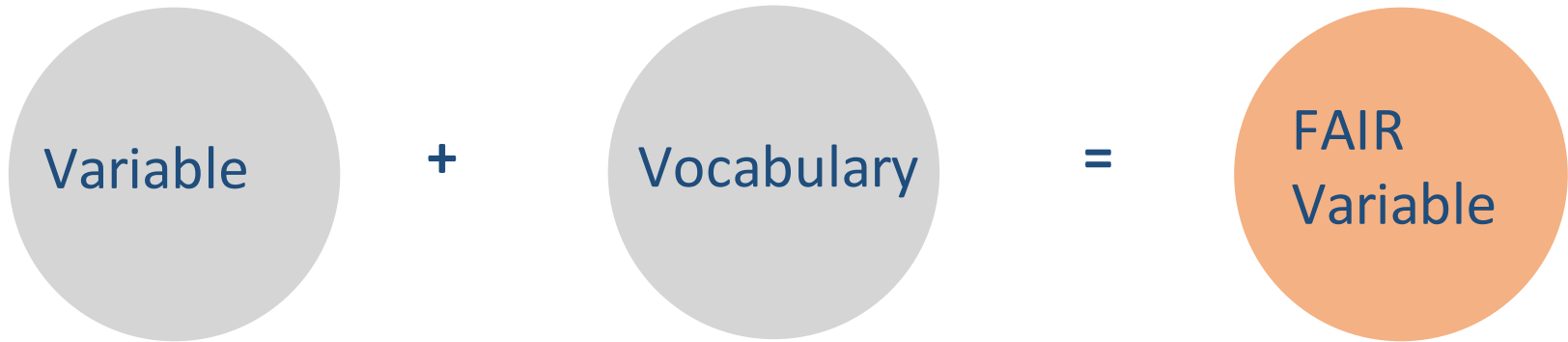
metadata using controlled vocabulary based on W3C specs

- **RDF** (Resource Data Framework) is a standard model for information (e.g. vocabularies) interchange on the Web
- **Turtle** is a common, human-readable and very compact data format for storing RDF data
- **SKOS** (Simple Knowledge Organization System) is a W3C recommendation designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or **any other type of structured controlled vocabulary.**



Metadata for Machines

semantically enhanced variable descriptions



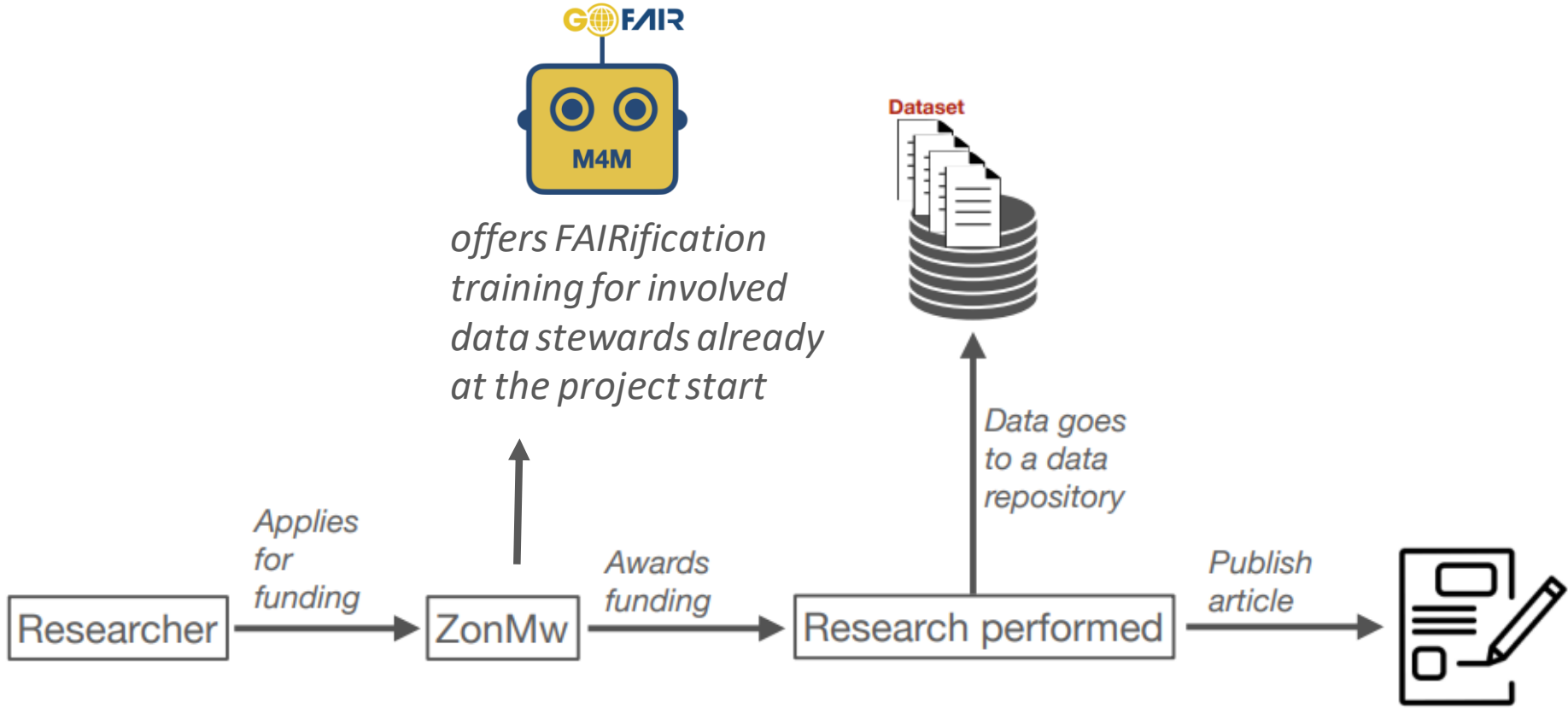


Metadata for Machines

metadata plus vocabulary plus tooling



FAIR Research Cycle



COVID-19 Program aims to make data 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

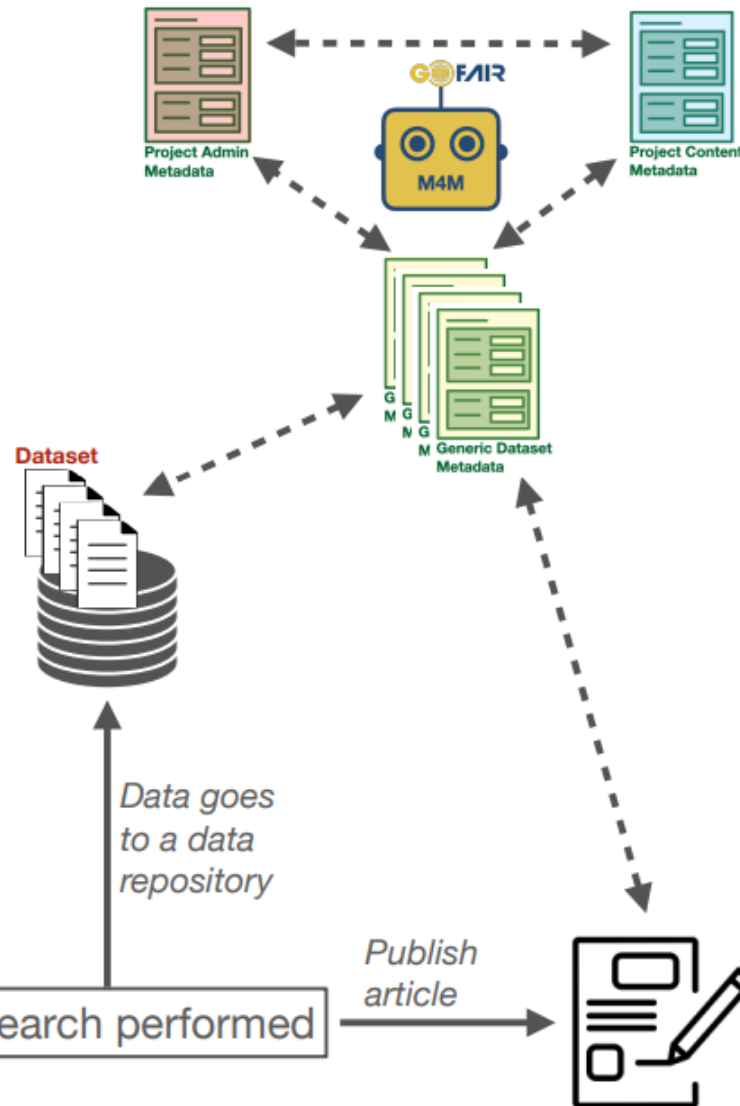
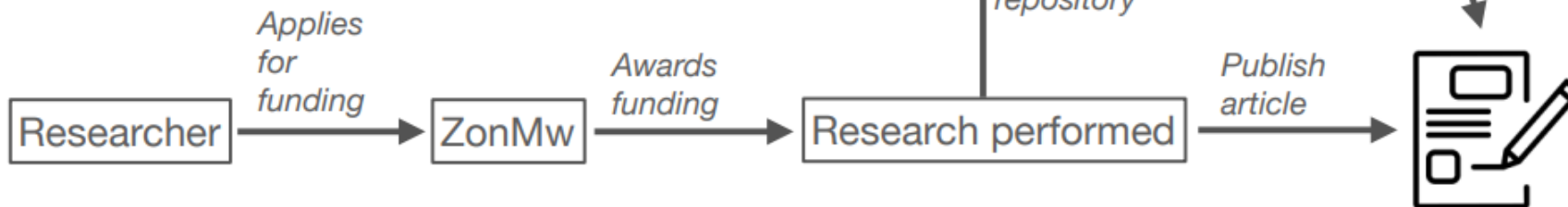
To be Reusable:

- 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

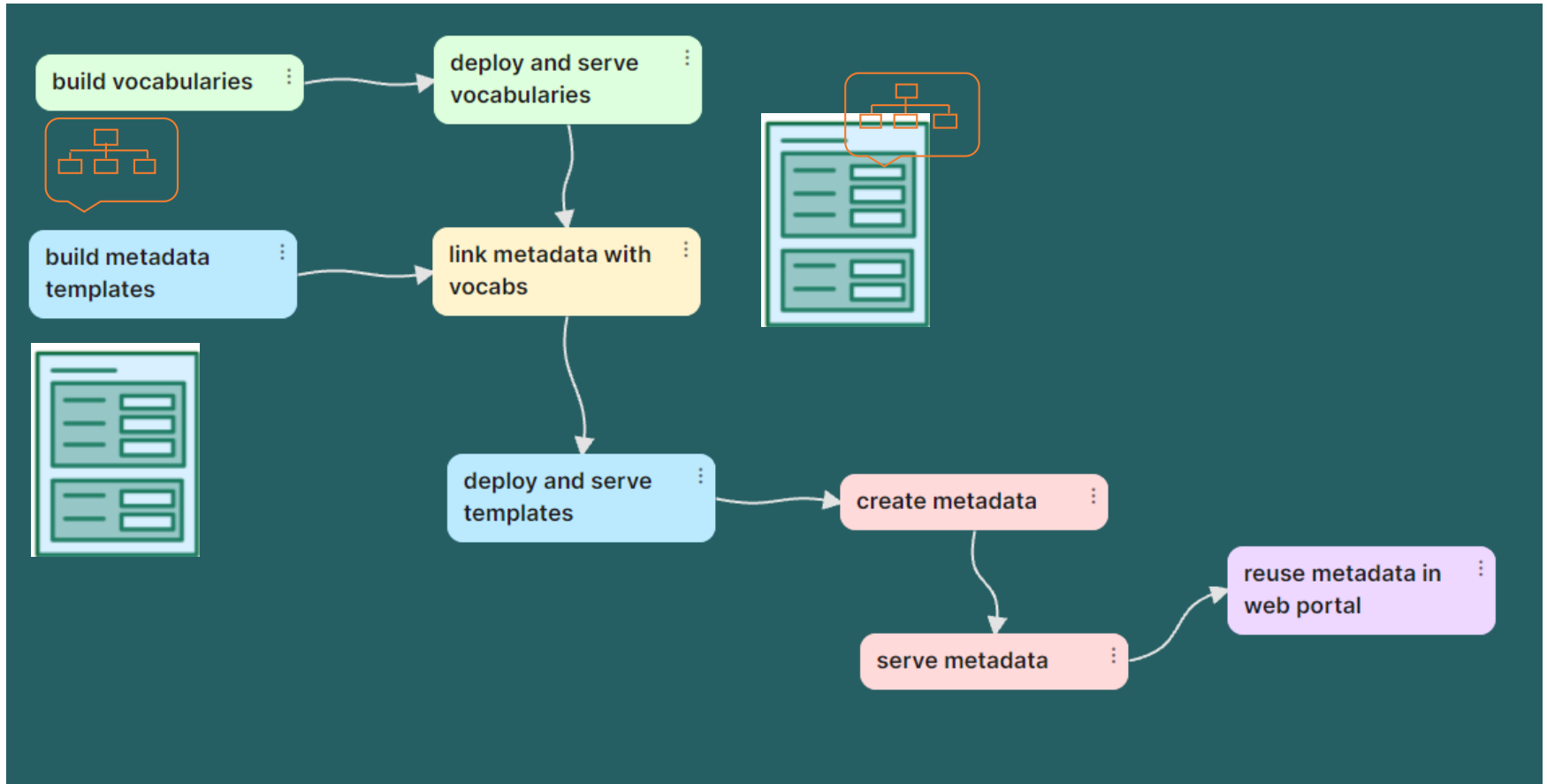
CEDAR Workbench → **Templates**



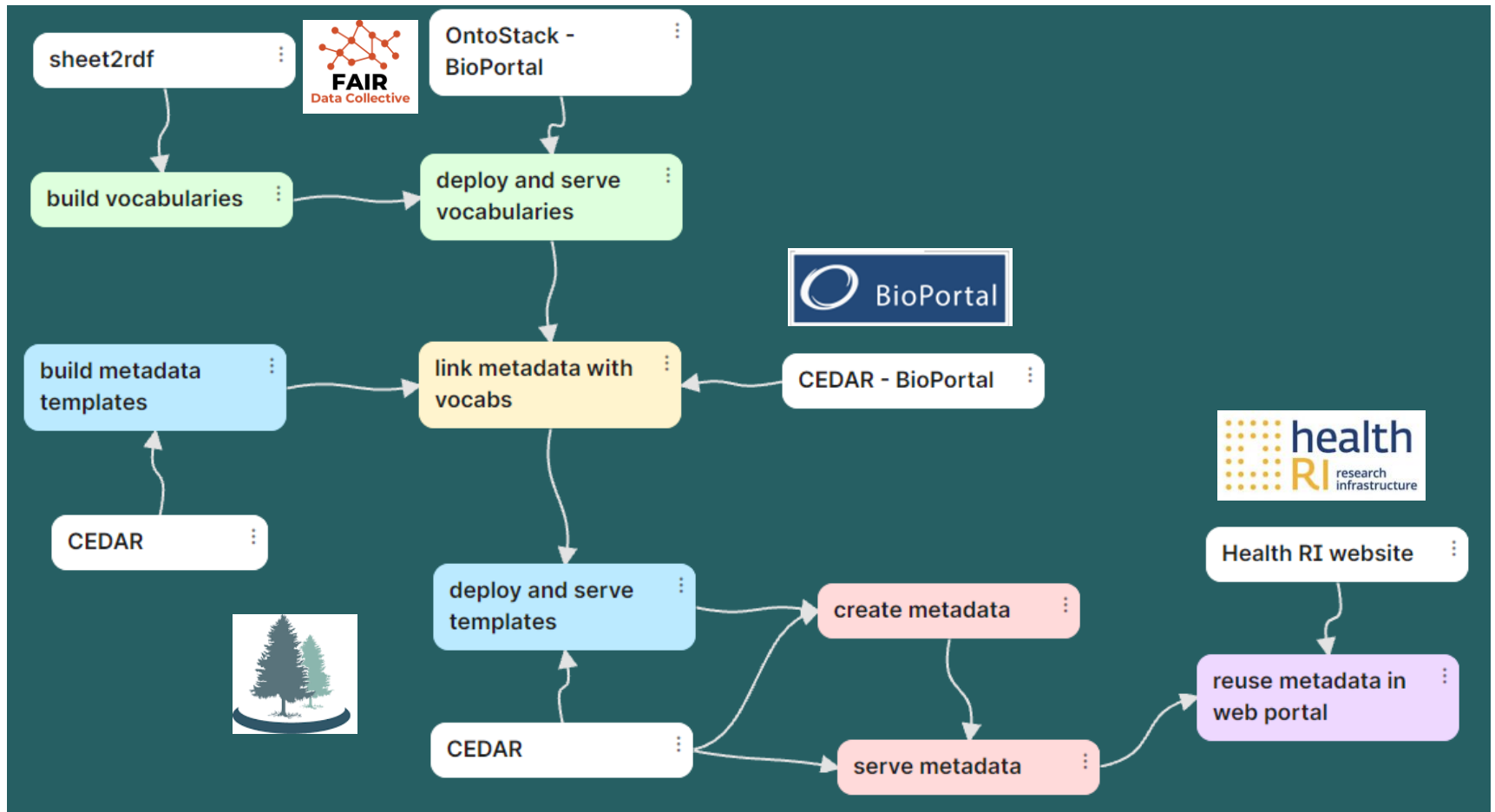
→ **Vocabulary**



FAIRification process and tools



FAIRification process and tools



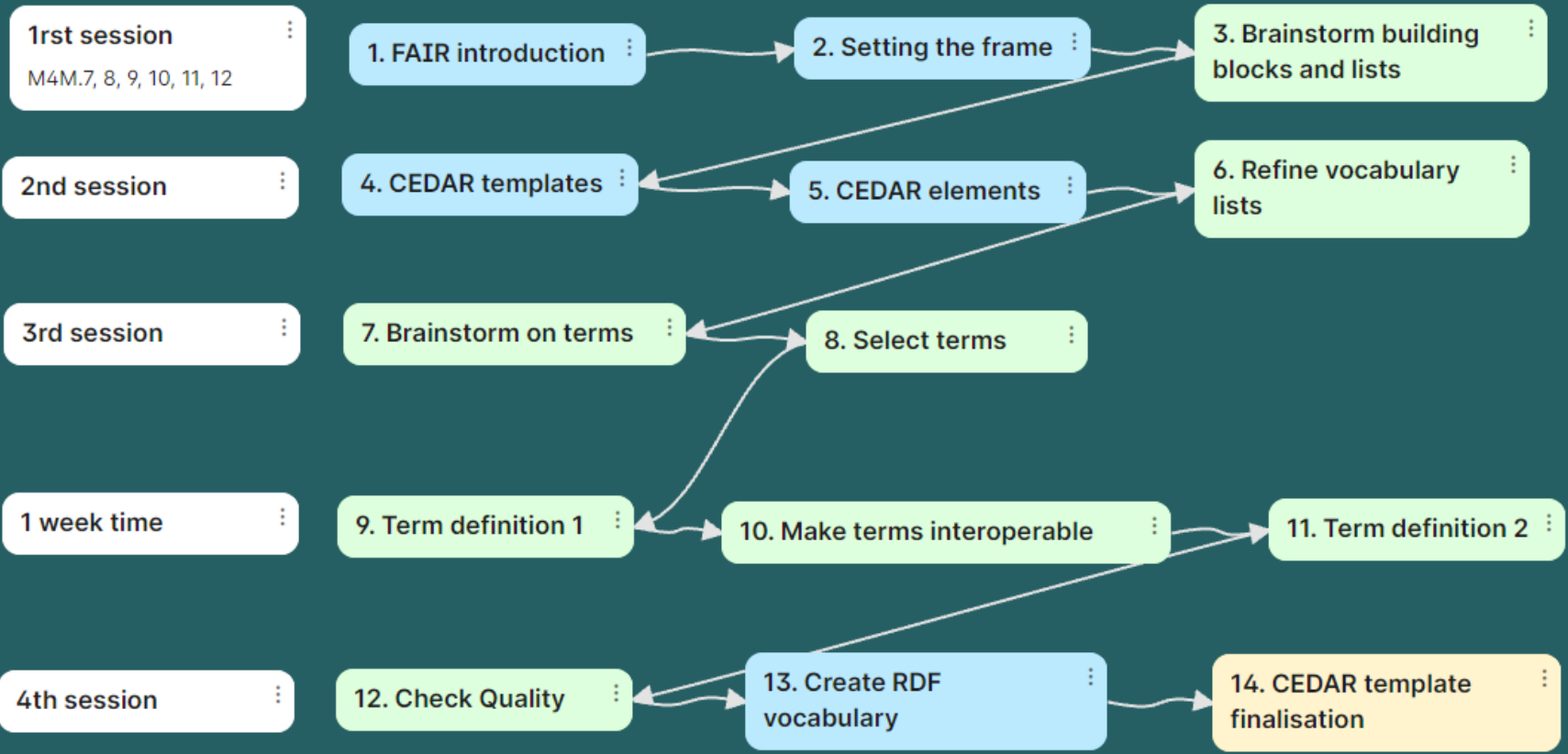
FAIRification process for COVID-19 Community

M4M workshops overview

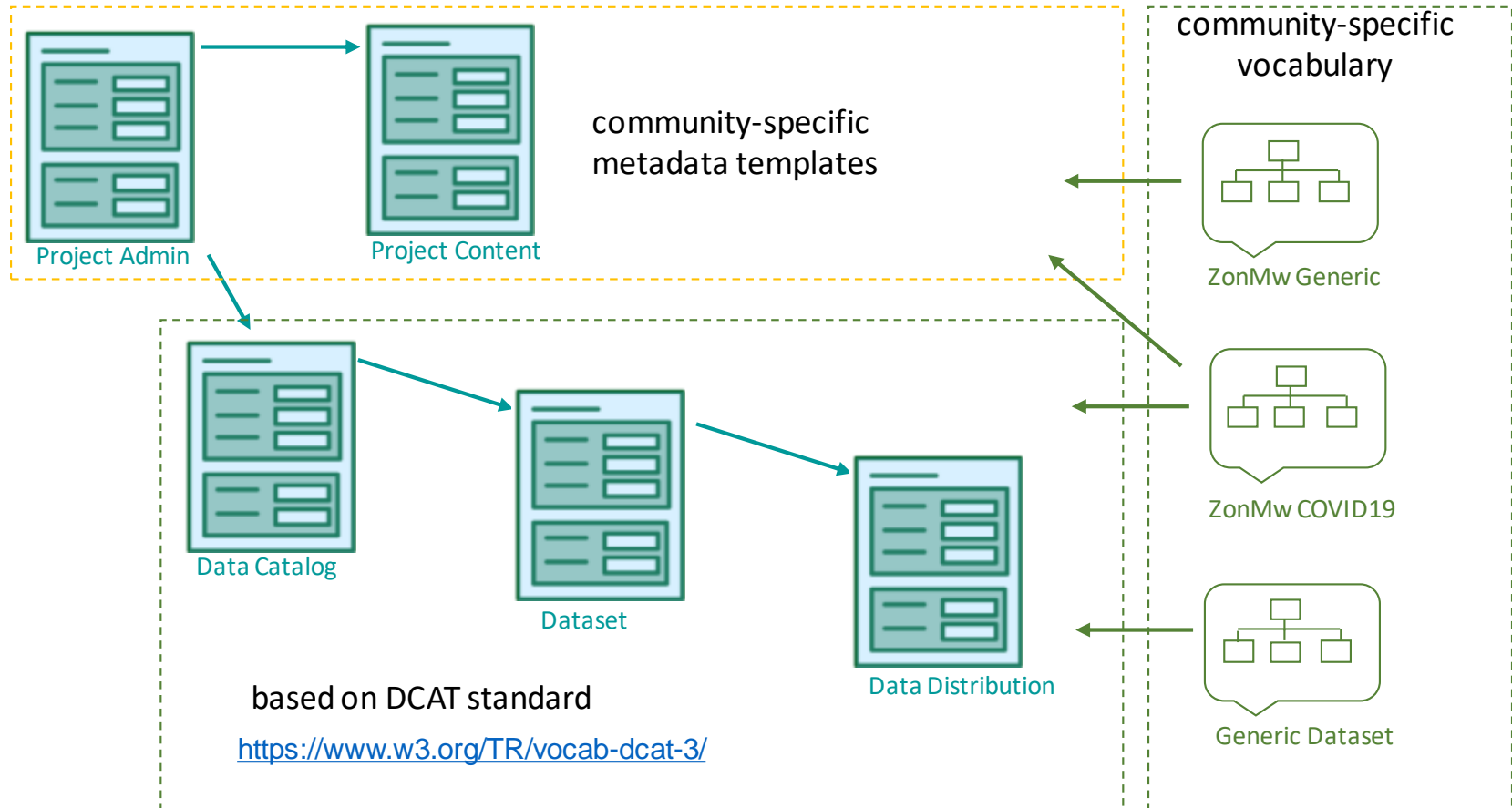
Workshop	Date	Community	Topic	Sponsor
M4M.1	October 2019	Inaugural	Setting up the concept	GO FAIR
M4M.2	January 2020	Funders	ZonMw + HRB	GO FAIR
M4M.3	January 2020	PreClinicalTrails	pre-registration form	GO FAIR
M4M.4	April-Sept 2020	VODAN Africa	Metadata for the FDP	Phillips Foundation
M4M.5	Summer 2020	AnnaEE	Climate data	DeiC
M4M.6	Summer 2020	DTU and others	Wind Energy	DeiC
M4M.7	November 2020	COVID-19 Program	Care (Treatment) / Prevention	ZonMw
M4M.8	November 2020	COVID-19 Program	Diagnostic / Testing	ZonMw
M4M.9	November 2020	COVID-19 Program	Prognosis / Risk assessments	ZonMw
M4M.10	November 2020	COVID-19 Program	Virus / Immunology / Molecular	ZonMw
M4M.11	November 2020	COVID-19 Program	Organisational / Process related	ZonMw
M4M.12	November 2020	COVID-19 Program	Socio-economic / Behavioral	ZonMw
M4M.13	February 2021	COVID-19 Program	Vocab	ZonMw
M4M.14	February 2021	COVID-19 Program	Vocab	ZonMw
M4M.15	June 2021	COVID-19 Program	Rapid M4M for datasets	ZonMw
M4M.16	June 2021	COVID-19 Program	I-ADOPT M4M for variables	ZonMw
M4M.17	June 2021	ID & AMR	R4R, COVID—>ID&AMR	ZonMw
M4M.18	Sept 2021	INCENTIVE	Influenza vaccine	EU/Horizon2020

Building templates and vocabulary for metadata

	M4M.7	M4M.8	M4M.9	M4M.10	M4M.11	M4M.12
	13 projects 21 participants	10 projects 17 participants	6 projects 12 participants	7 projects 13 participants	9 projects 16 participants	14 projects 21 participants



Metadata templates and vocabulary for projects and datasets



Project Admin Form (... funder information)

- Read & Understood
- Project Information
- Funder Information
 - Funder (1..N)

Funder Name - Multiple answers are allowed.

Please enter the funder name if not found in the drop-down list.

Funder ROR Reference

Funder Project Identifier

Funder Project Identifier*
 - Project Dates
 - (Expected) Start Date of Project*
 - (Expected) End Date of Project*
- Project Duration (in Months)*

- Organisation Information
- Participant Information
- Other

Project Content Form (... produced assets: data)

Read & Understood

Project Title (1..N)

Scope

Used Assets

Produced Assets

What assets does your project produce?*

Type of Data (1..N)

Which type of data is relevant for your project? Multiple answers are allowed.*

- audiovisual data
- diagnostic imaging
- economic data
- environmental data
- genomics data
- geographical data
- health data

Type of Biomaterial (1..N)

What type of biomaterial, if any, is relevant for your project? Multiple answers are allowed.

What organisms are the target of your analysis?

Please enter the organism if not found in the drop-down list.

If you ticked substance, what kind of substance is relevant for your project?

If you ticked specimen, what kind of specimen is relevant for your project?

If you ticked macromolecule, what kind of macromolecules is relevant for your project?

If you ticked other, which ones would you like to add?

Type of Service

Standards (1..N)

Other

Products of M4Ms

Guidelines for completing CEDAR forms

[Link to the guidelines document](#)

CEDAR forms – to be filled out

[Project Admin form](#) [CEDAR account required]

[Project Content form](#) [CEDAR account required]

[Data Catalogue form](#) [CEDAR account required]

[Dataset form](#) [CEDAR account required]

[Data Distribution form](#) [CEDAR account required]

CEDAR forms in OpenView – easy way to view/share forms

[Project Admin form in OpenView](#) [no CEDAR account required]

[Project Content form in OpenView](#) [no CEDAR account required]

[Data Catalogue form in OpenView](#) [no CEDAR account required]

[Dataset form in OpenView](#) [no CEDAR account required]

[Data Distribution form in OpenView](#) [no CEDAR account required]

BioPortal vocabularies

[ZonMw Generic Terms](#)

[Ontology for Generic Dataset Metadata Template](#)

[ZonMw COVID-19 Vocabulary](#)

→ used as resources in the Health RI [Website](#)

Health RI Portal with semantic faceted search using the provided metadata

Type of provided assets	Respective monitoring of antibody response following COVID-19 vaccination in patients with Down syndrome	01/10/2021
<input type="checkbox"/> biomaterial (physical samples)	PRIDE	
<input checked="" type="checkbox"/> data (digital resources)	University Medical Center Utrecht	
<input type="checkbox"/> services		
Provided data	Serologische surveillance van SARS-CoV-2 tijdens de 2020 pandemie onder zorgmedewerkers mét en zonder COVID-patiënten-contact in de Amsterdam Universitair Medische Centra (S3 studie)	Last changed 30/09/2021
<input type="checkbox"/> audiovisual data	S3 study	
<input checked="" type="checkbox"/> diagnostic imaging	ZonMw - Netherlands Organisation for Health Research and Development VU University Medical Center	
<input type="checkbox"/> economic data		
<input type="checkbox"/> environmental data		
<input type="checkbox"/> environmental data		
<input type="checkbox"/> genomics data		
<input type="checkbox"/> genomics data		
<input type="checkbox"/> geographical data		
<input checked="" type="checkbox"/> health data	SARSLIVA: utility of saliva in COVID-19 diagnosis - a household study	Last changed 30/09/2021
<input type="checkbox"/> none	SARSLIVA	
<input type="checkbox"/> political data	ZonMw - Netherlands Organisation for Health Research and Development Spaarne Gasthuis - Spaarne Ziekenhuis	
	Ethnicity and COVID-19: epidemiology and control measures	Last changed 30/09/2021
	ZonMw - Netherlands Organisation for Health Research and Development Amsterdam UMC	