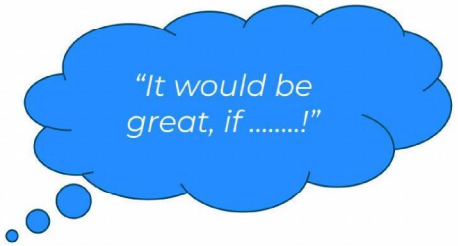
ZonMw

ME/CFS & post-COVID
research consortia: general introduction

# Collaborating on FAIR data for synergy in research

ZonMw, FAIR data workshop, Zwolle, April 9, 2024

5.1.2e Implementation lead FAIR data management
5.1.2e FAIR data & Open Access

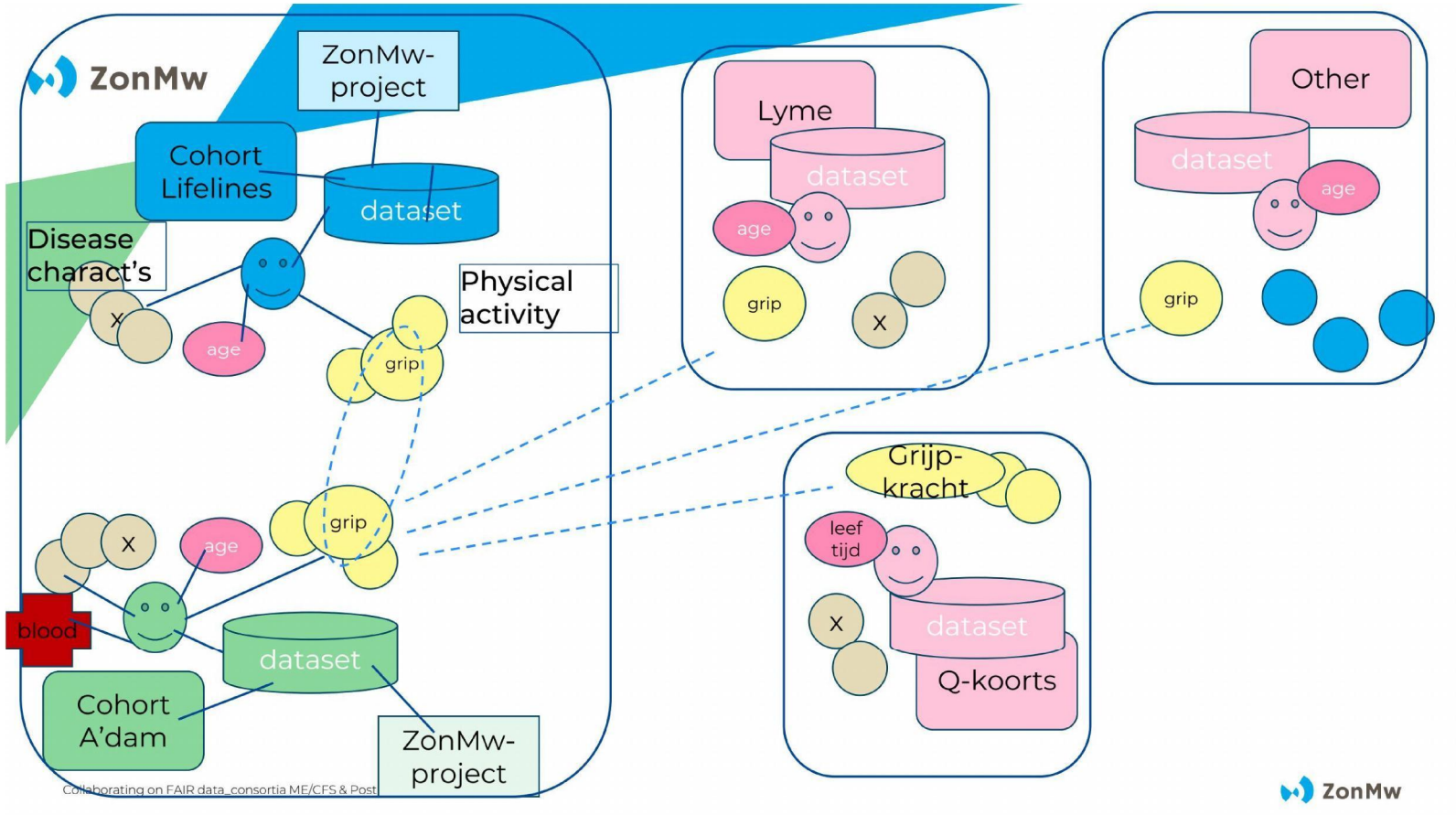



# Today on the agenda

1. **Available information > reusable for men and machines (FAIR)**
2. **The bumpy road to FAIR data**
3. **It is all about the choices you make as a research community**
4. **A closer look at data FAIRification**
6. **From DMPs to a FAIR Implementation Profile**
7. **Next steps**

**Session 1:**
**Explore DMPs** to see the standards, tools (etc) that may be useful

**Session 2:** Compare / review the input from breakout 1, and **choose what fits your consortium best**


ZonMw
ZonMw-project
Cohort Lifelines
dataset
Disease charact's
Physical activity
x
age
grip
Lyme
dataset
age
grip
x
Other
dataset
age
grip
Grijp-kracht
leef tijd
x
dataset
Q-koorts
grip
x
age
blood
dataset
Cohort A'dam
ZonMw-project
Collaborating on FAIR data_consortia ME/CFS & Post
ZonMw

ZonMw

Beware of Fake-FAIR!!

## FAIR principles

Provide **guidance** for data management:

| | |
|---|---|
| **F**indable | – vindbaar |
| **A**ccessible | – toegankelijk |
| **I**nteroperable | – uitwisselbaar |
| **R**e-usable | – herbruikbaar |

*Reusable for both **men** and **machines** (people and computers) (**F**ully-**AI**-**R**eady)*

FAIR data stay at the source (data visiting)

FAIR is 90% Metadata

**Persistent identifiers** to make metadata usable for the machine

**Box 2** | The FAIR Guiding Principles

**To be Findable:**
F1. (meta)data are assigned a globally unique and persistent
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

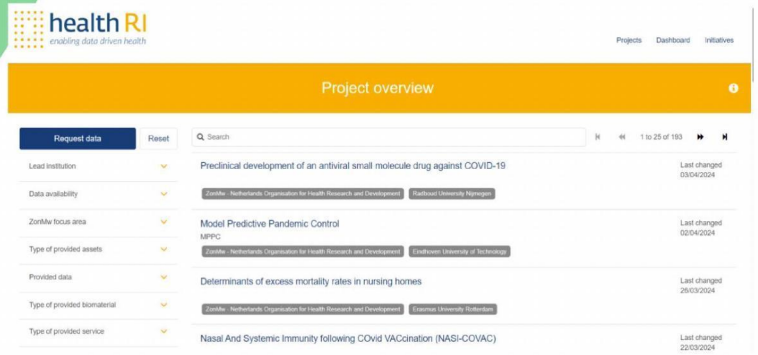
**Machine-actionable metadata**
**Technical infrastructure (accepted generic services)**
**Social decisions (domain specific agreements)**

Blue principles: researchers & data stewards

Red principles: Data stewards, FAIR data experts

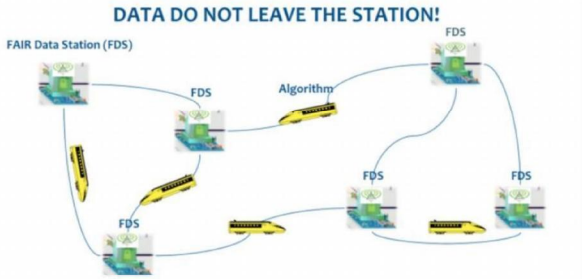
ZonMw

# *Levels of FAIRness: Examples of what you can do with it?*



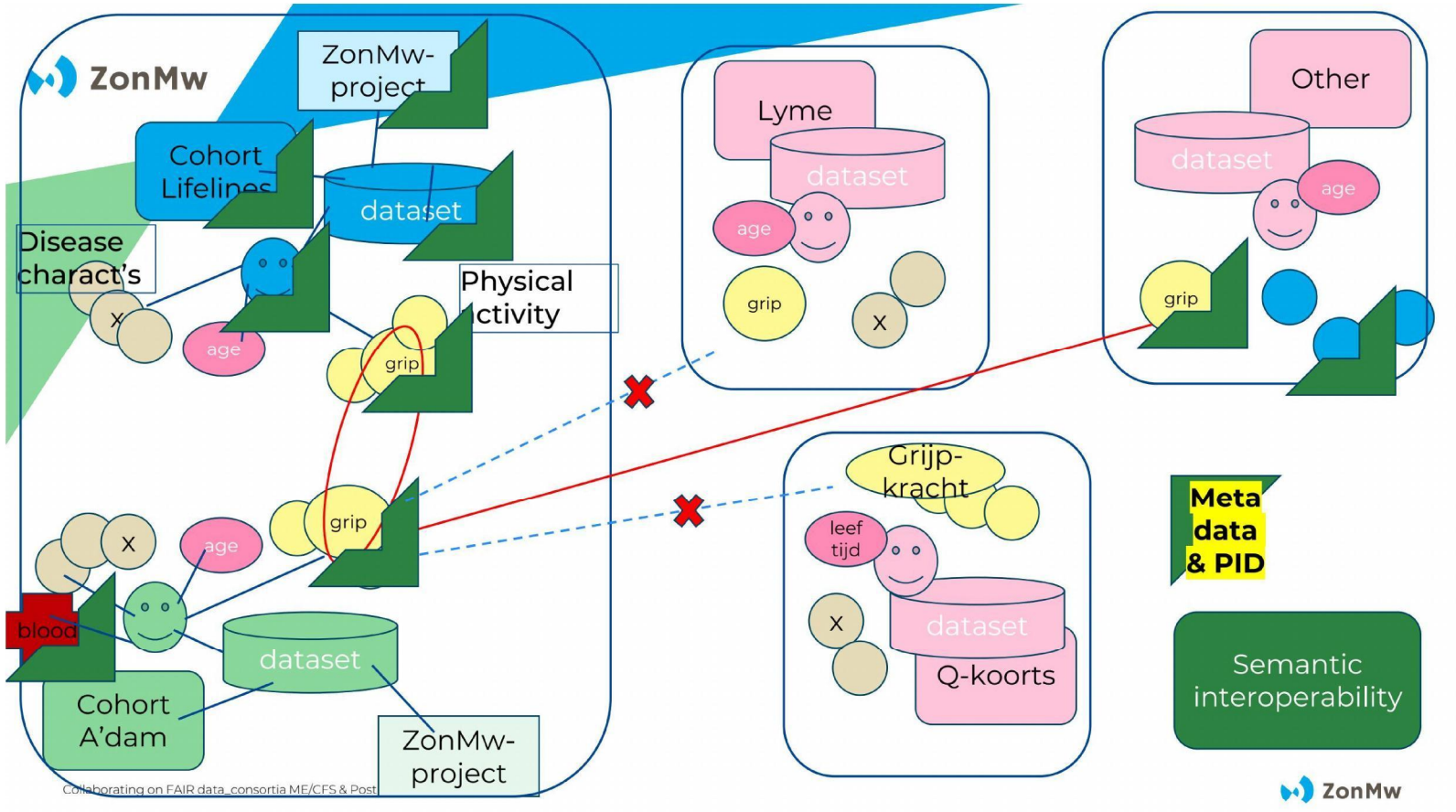

**Data portal: data not yet fully FAIR**
**Project metadata FAIR**

**FAIR = Fully-AI-Ready**




**Data visiting > query > personal health train**

ZonMw


ZonMw
ZonMw-project
Cohort Lifelines
dataset
Disease charact's
Physical activity
x
age
grip
Lyme
dataset
age
grip
x
Other
dataset
age
grip
Grijp-kracht
leef tijd
x
dataset
Q-koorts
Meta data & PID
Semantic interoperability
blood
Cohort A'dam
ZonMw-project
Collaborating on FAIR data_consortia ME/CFS & Post
ZonMw

# What steps must we take to improve reusability, interoperability?

**What scientific challenges do you want to solve?**

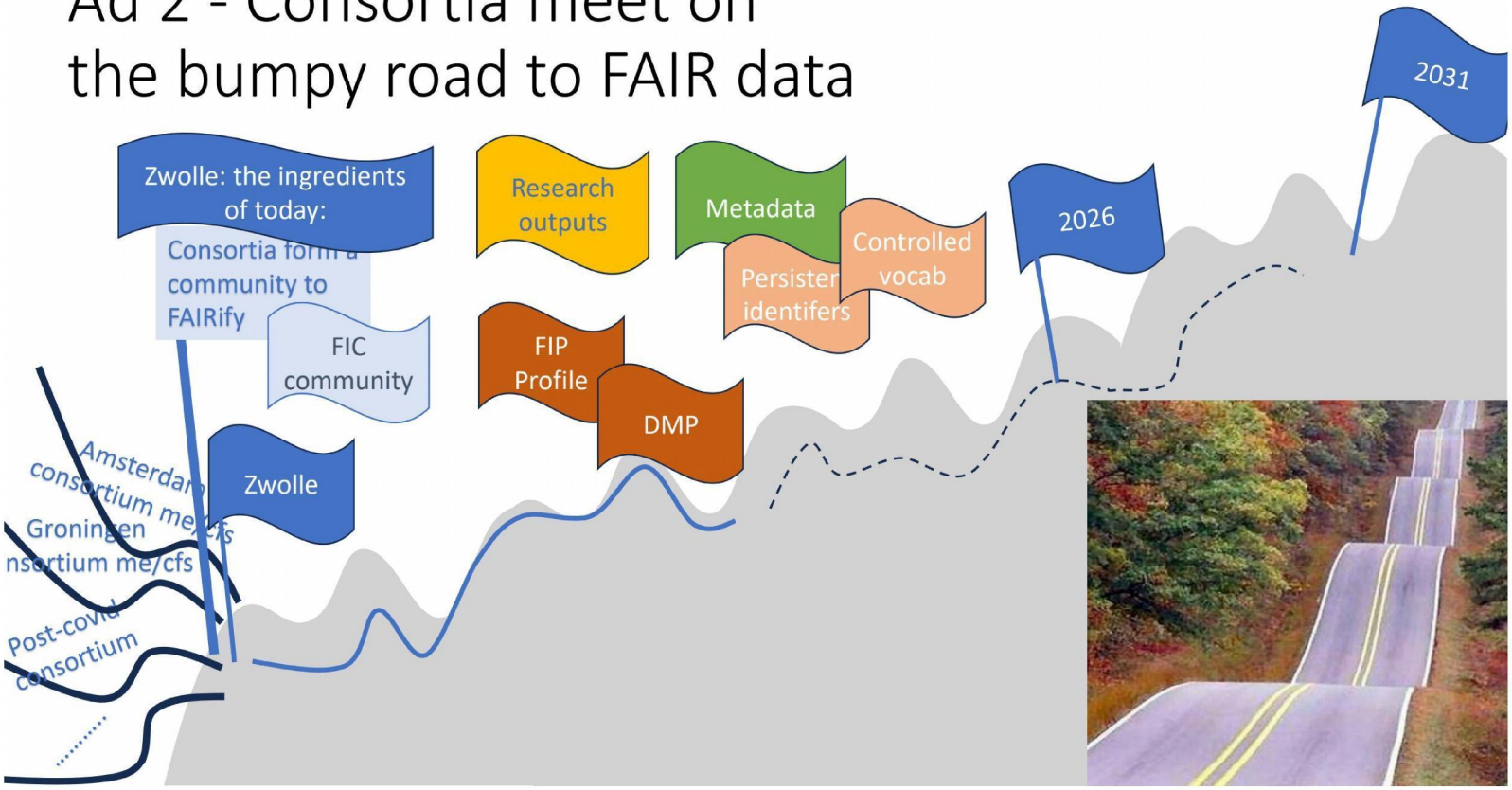
- **ME/CFS, post-covid** and its causes
- Disease characteristics, mechanisms
- Diagnosis
- Treatment
- Etc, etc, etc

The research community must choose!

**What choices are relevant for ME/CFS and Post-covid?**

- Reusable data? > red/blue principles
- Semantic interoperability? > to align topics, terms

# Ad 2 - Consortia meet on the bumpy road to FAIR data



FAIRify all research outputs! Not only datasets, also e.g. samples (biobank)

>> 5.1.2e

# Ad 4 - A closer look at data FAIR[illegible]cation

- **What are your goals for reusable / FAIR data?**
- **What do you want to FAIRify?** Datasets, sample collections, cohorts, variables…
- **Who do you need therefore?**
  - Community building, collaborating consortia >
  - **FAIR Implementation Community (FIC)**
  - Expertise from researchers, data stewards and FAIR data experts
- **How will you approach it? Community choices > today:**
  - FAIR Implementation Profile
  - Metadata

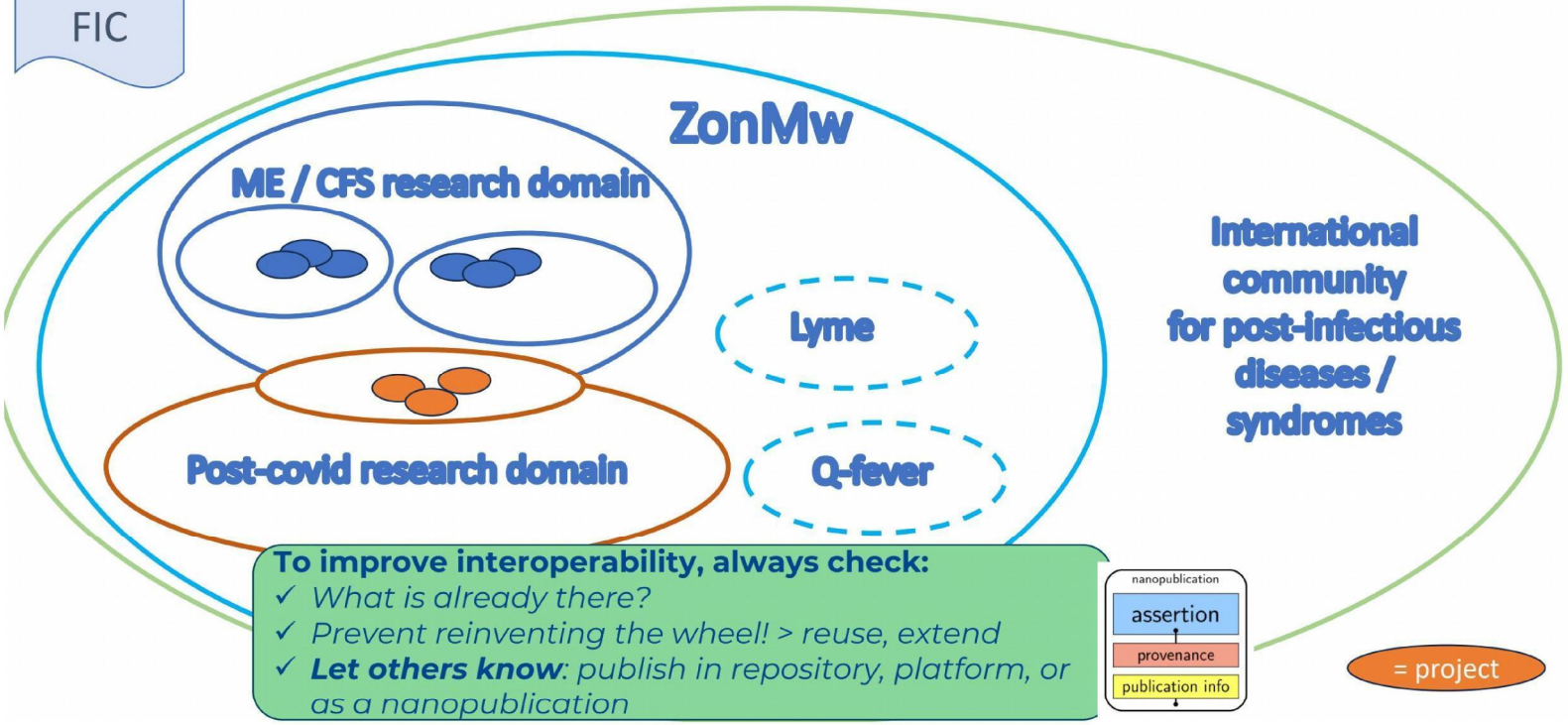
Consortia form a community to FAIRify

FIC

Consortia form a community to FAIRify

FIC

# The ME/CFS and post-covid FAIR Implementation Community: part of a larger community – including smaller ones



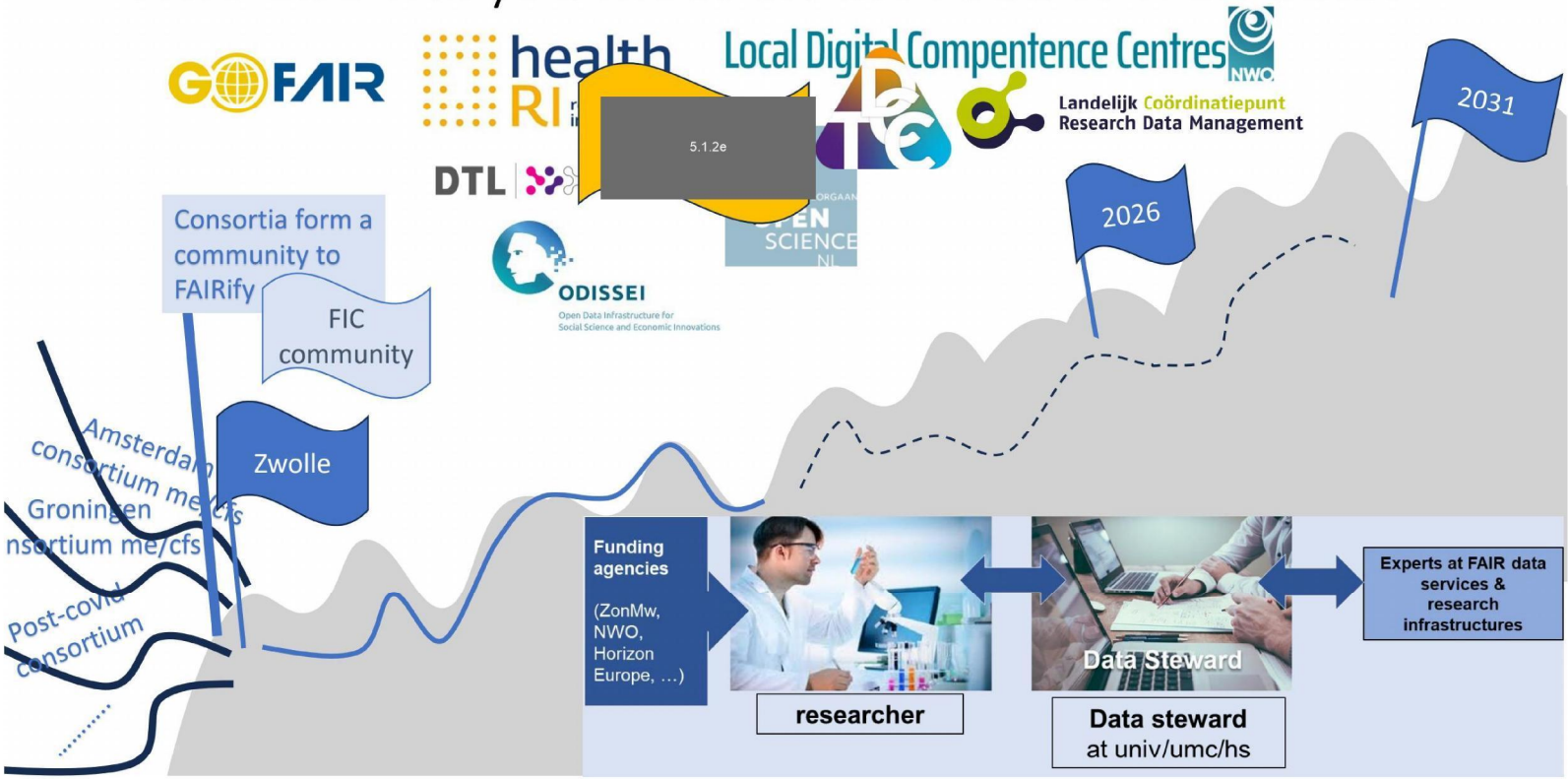

**To improve interoperability, always check:**

- ✓ *What is already there?*
- ✓ *Prevent reinventing the wheel! > reuse, extend*
- ✓ ***Let others know**: publish in repository, platform, or as a nanopublication*




= project

# Who else will you meet on the road to FAIR data?

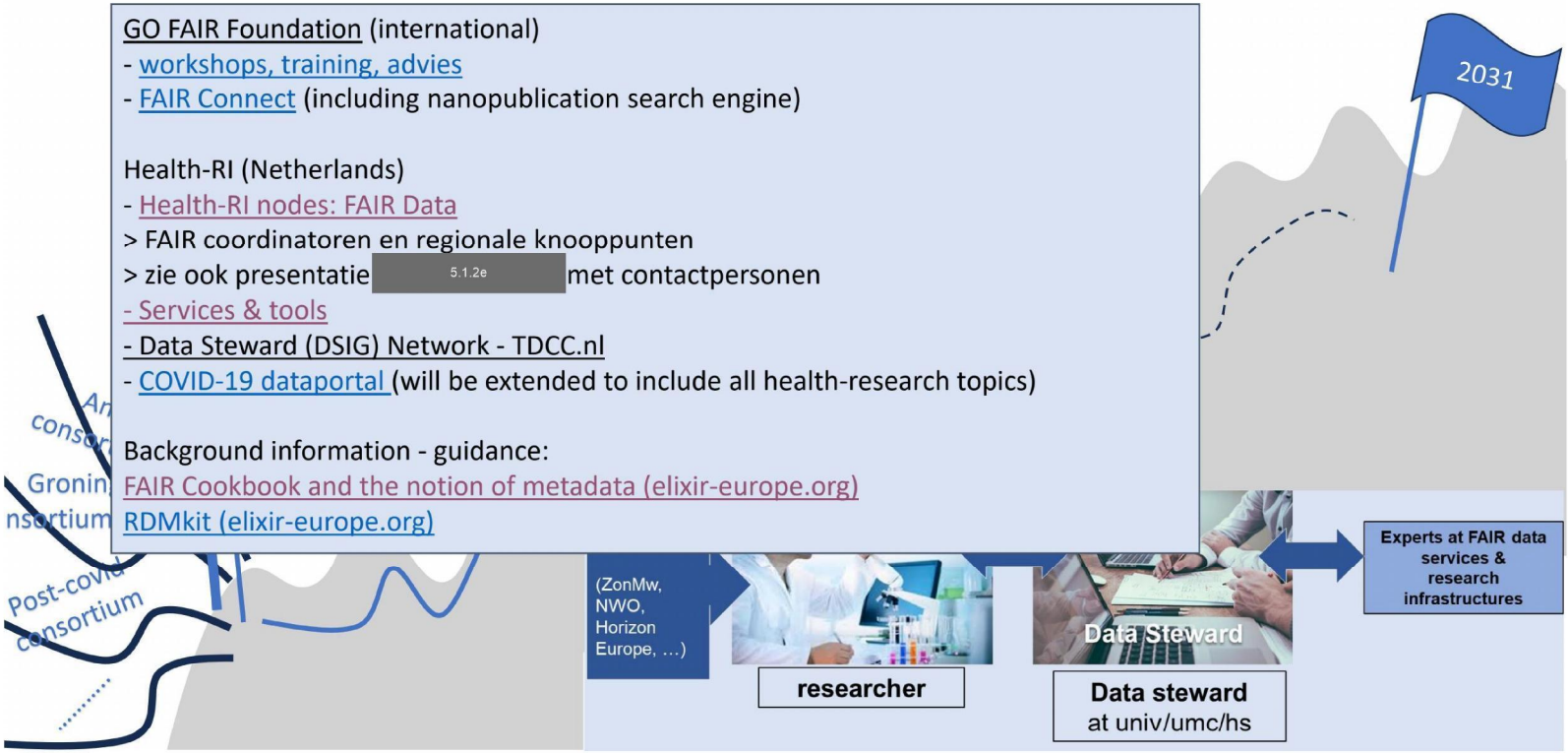


# Who will you meet on the road to FAIR data?

GO FAIR Foundation (international)
- workshops, training, advies
- FAIR Connect (including nanopublication search engine)

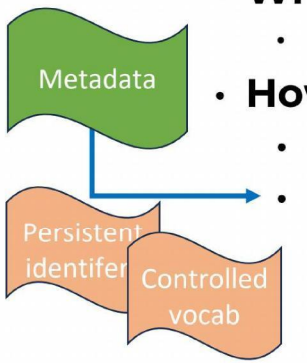
Health-RI (Netherlands)
- Health-RI nodes: FAIR Data
> FAIR coordinatoren en regionale knooppunten
> zie ook presentatie 5.1.2e met contactpersonen
- Services & tools
- Data Steward (DSIG) Network - TDCC.nl
- COVID-19 dataportal (will be extended to include all health-research topics)

Background information - guidance:
FAIR Cookbook and the notion of metadata (elixir-europe.org)
RDMkit (elixir-europe.org)



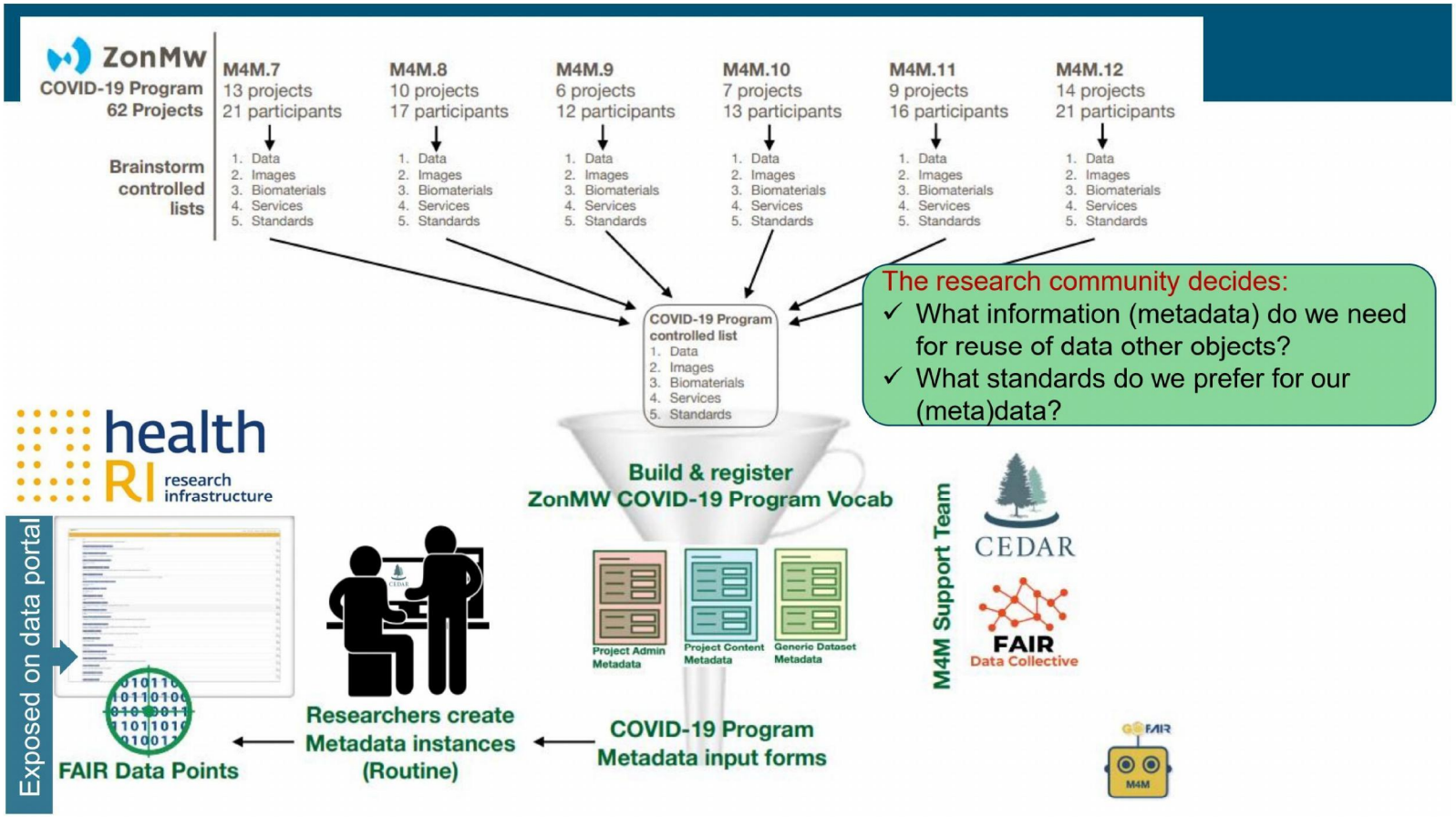
# Ad 4 - A closer look at data FAIRification

- **What are your goals for reusable / FAIR data?**
- **What do you want to FAIRify?** Datasets, sample collections, cohorts, variables...
- **Who do you need therefore?**
  - FIC & community, expertise
- **How will you approach it? Community choices > today:**
  - FAIR Implementation Profile > working sessions
  - Metadata






ZonMw


ZonMw
COVID-19 Program
62 Projects
Brainstorm controlled lists
M4M.7
13 projects
21 participants
M4M.8
10 projects
17 participants
M4M.9
6 projects
12 participants
M4M.10
7 projects
13 participants
M4M.11
9 projects
16 participants
M4M.12
14 projects
21 participants
1. Data
2. Images
3. Biomaterials
4. Services
5. Standards
COVID-19 Program controlled list
1. Data
2. Images
3. Biomaterials
4. Services
5. Standards
The research community decides:
✓ What information (metadata) do we need for reuse of data other objects?
✓ What standards do we prefer for our (meta)data?
health RI research infrastructure
Build & register
ZonMW COVID-19 Program Vocab
Project Admin Metadata
Project Content Metadata
Generic Dataset Metadata
M4M Support Team
CEDAR
FAIR
Data Collective
Exposed on data portal
FAIR Data Points
Researchers create
Metadata instances
(Routine)
COVID-19 Program
Metadata input forms
M4M

ZonMw

# Metadata Forms for ZonMw ID & AMR

**Guidelines for completing CEDAR forms**

Link to the guidelines document

**CEDAR forms** [forms to fill out, CEDAR account required]

Project Admin form

Project Content form

Collection Admin form

Collection Content form

**CEDAR forms in OpenView** [forms to view/share, no CEDAR account

Project Admin form in OpenView

Project Content form in OpenView

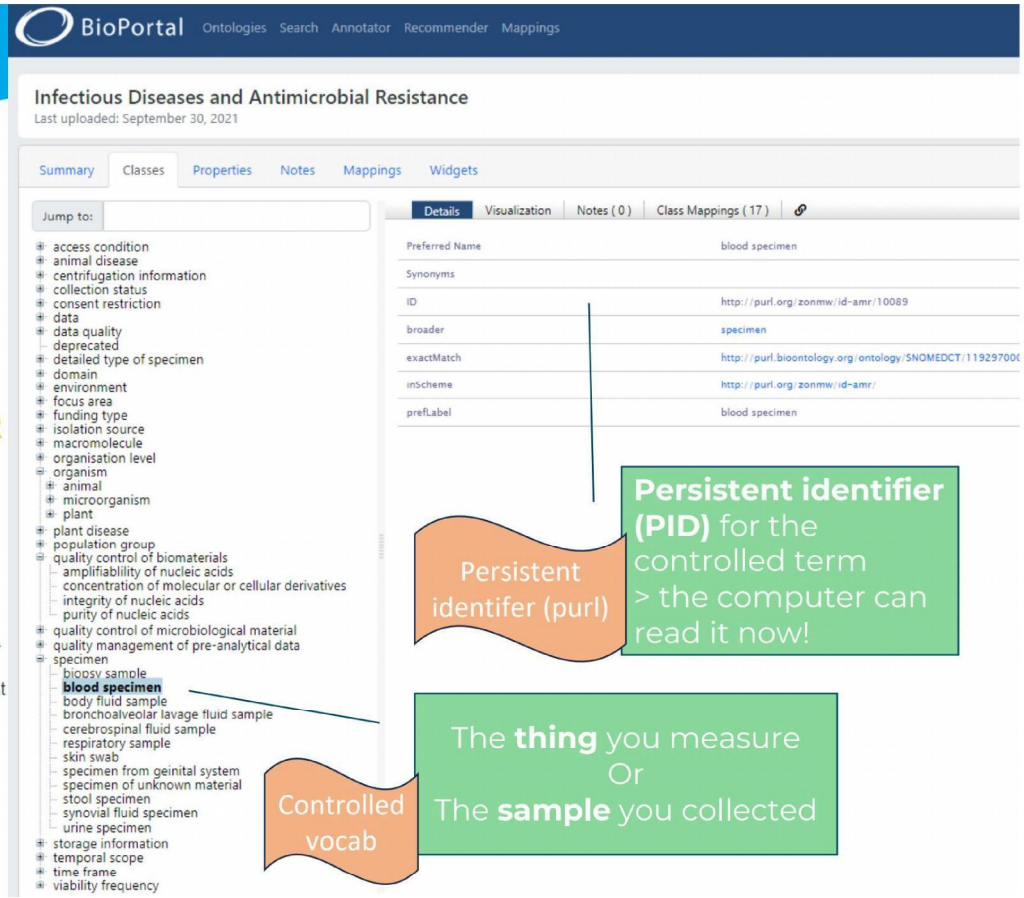
Collection Admin form in OpenView

Collection Content form in OpenView

**BioPortal vocabularies**

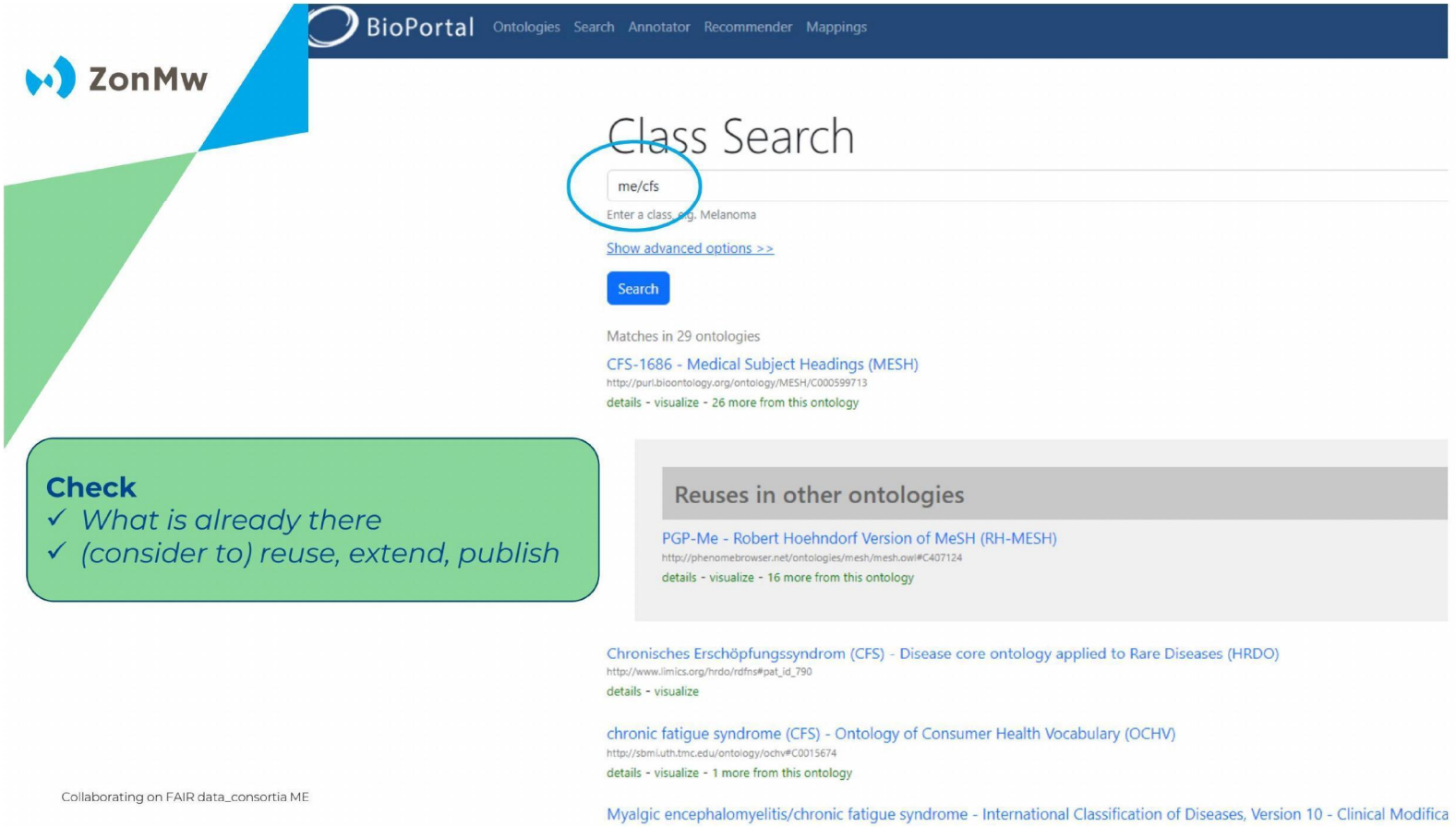
ID&AMR Vocabulary





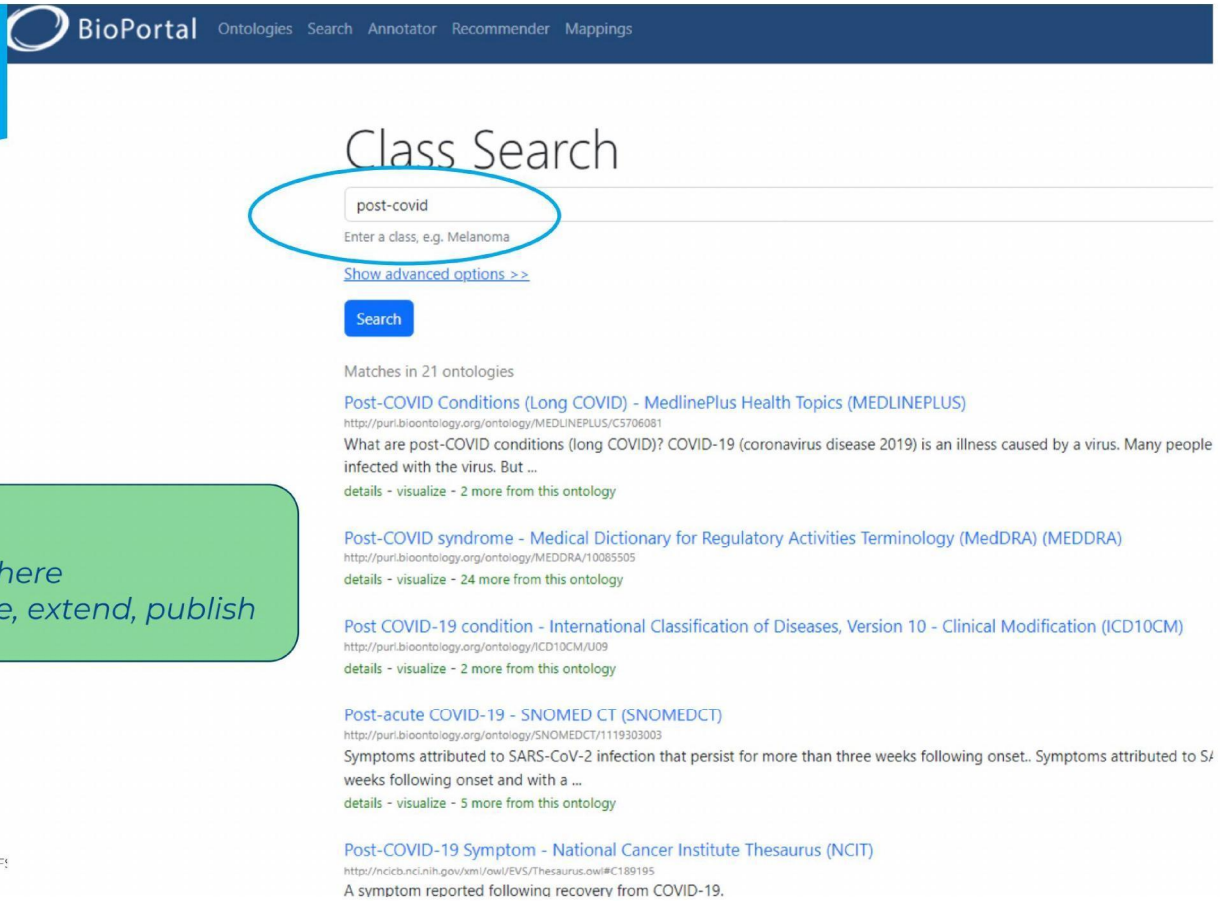
ZonMw

**Check**
- ✓ *What is already there*
- ✓ *(consider to) reuse, extend, publish*



ZonMw

**Check**
- ✓ *What is already there*
- ✓ *(consider to) reuse, extend, publish*



# Metadata for variables > lecture [redacted]

**Check**
- ✓ *What is already there* > ***for example*** [redacted] ***variables!***
- ✓ *(consider to) reuse, extend, publish*

npj | digital medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Check for updates

## Harmonization and standardization of data for a pan-European cohort on SARS- CoV-2 pandemic

[redacted]

The European project ORCHESTRA intends to create a new pan-European cohort to rapidly advance the knowledge of the effects and treatment of COVID-19. Establishing processes that facilitate the merging of heterogeneous clusters of retrospective data was an essential challenge. In addition, data from new ORCHESTRA prospective studies have to be compatible with earlier collected information to be efficiently combined. In this article, we describe how we utilized and contributed to existing standard terminologies to create consistent semantic representation of over 2500 COVID-19-related variables taken from three ORCHESTRA studies. The goal is to enable the semantic interoperability of data within the existing project studies and to create a common basis of standardized elements available for the design of new COVID-19 studies. We also identified 743 variables that were commonly used in two of the three prospective ORCHESTRA studies and can therefore be directly combined for analysis purposes. Additionally, we actively contributed to global interoperability by submitting new concept requests to the terminology Standards Development Organizations.

*npj Digital Medicine* (2022)5:75 ; https://doi.org/10.1038/s41746-022-00620-x



ZonMw


What we showed until now:
Zwolle: the ingredients of today:
Consortia form a community to FAIRify
FIC community
Zwolle
Research outputs
Metadata-templates
Controlled vocab
Persistent identifers
FAIR data point
Data portal
Nano-publications
2026
2031
Amsterdam consortium me/cfs
Groningen
nsortium me/cfs
Post-covid consortium
Time for the next speaker!