

# FAIR Data for the BENEFIT community

Werkpakket 3 update

Katy Wolstencroft (UL, LIACS)

Erik Flikkenschild (LUMC, GoFair Foundation)

<https://vidensportal.deic.dk/en/FAIR>

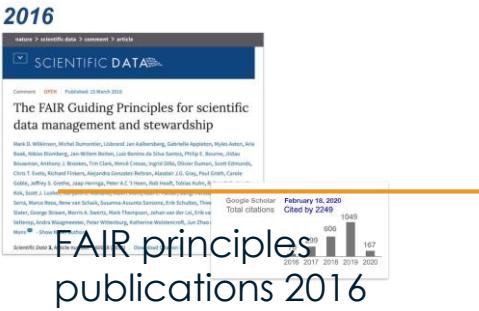
Let's talk about FAIR data...

"FAIR data?  
I do not collect any data,  
I only use what is already  
available elsewhere"

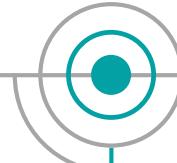


# FAIR principles

## introductie

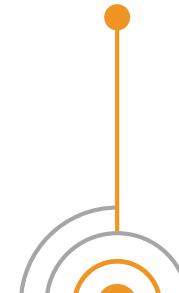


start

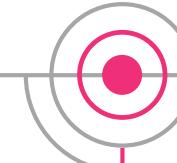


2014 Leiden, Lorentz Centre

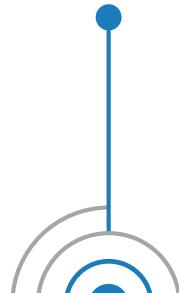
mile1



Mile2



VODAN in a box  
2020



Mile3



1<sup>st</sup> FDP data visiting  
Between continents



Historic First Data Visiting of Patient Data Stored in Residence Across Continents  
[www.vodan-totalfrica.info](http://www.vodan-totalfrica.info)  
3 mins read

Mile4

# FAIR-data

FAIR staat voor vindbaar (Findable), toegankelijk (Accessible), interoperabel (Interoperable), en herbruikbaar (Reusable). Waarom is het belangrijk dat data FAIR is?

## Findable

Metadata and data should be findable for both humans and computers

F A

## Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

I R

## Accessible

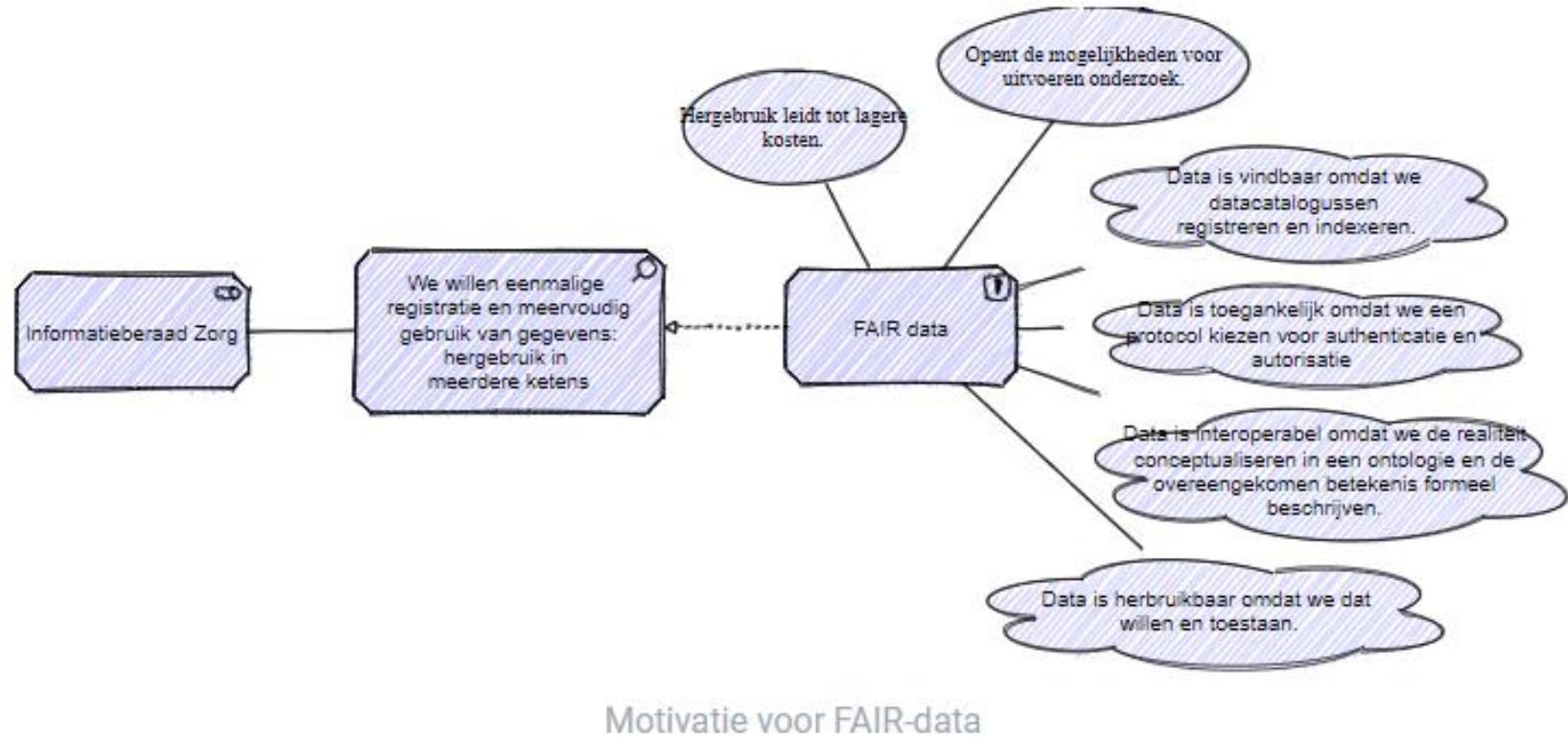
Once found, users need to know how the data can be accessed

## Reusable

The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

# FAIR data:

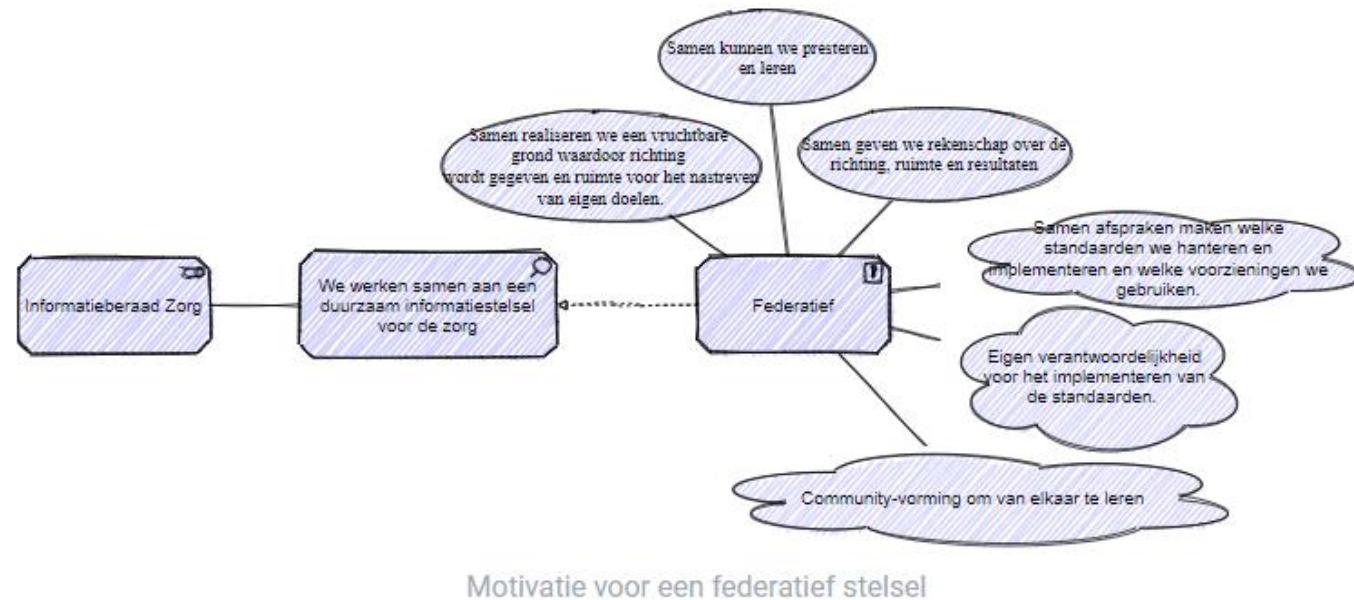
## What is the benefit for Benefit?



# Federatief

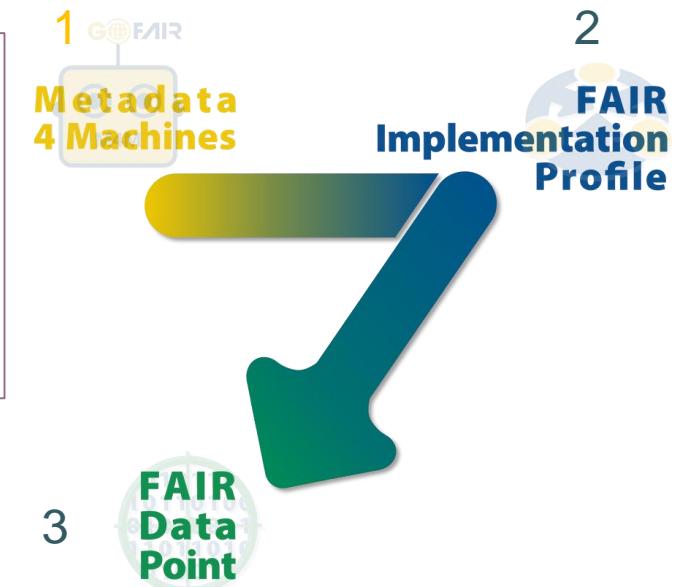
## binnen een community afspraken maken

Het Informatiestelsel Zorg heeft geen formele eigenaar. Het is van iedereen. We noemen dat een federatief stelsel. Binnen een federatief stelsel moeten we samen afspraken maken over de standaarden en de voorzieningen die we gebruiken.

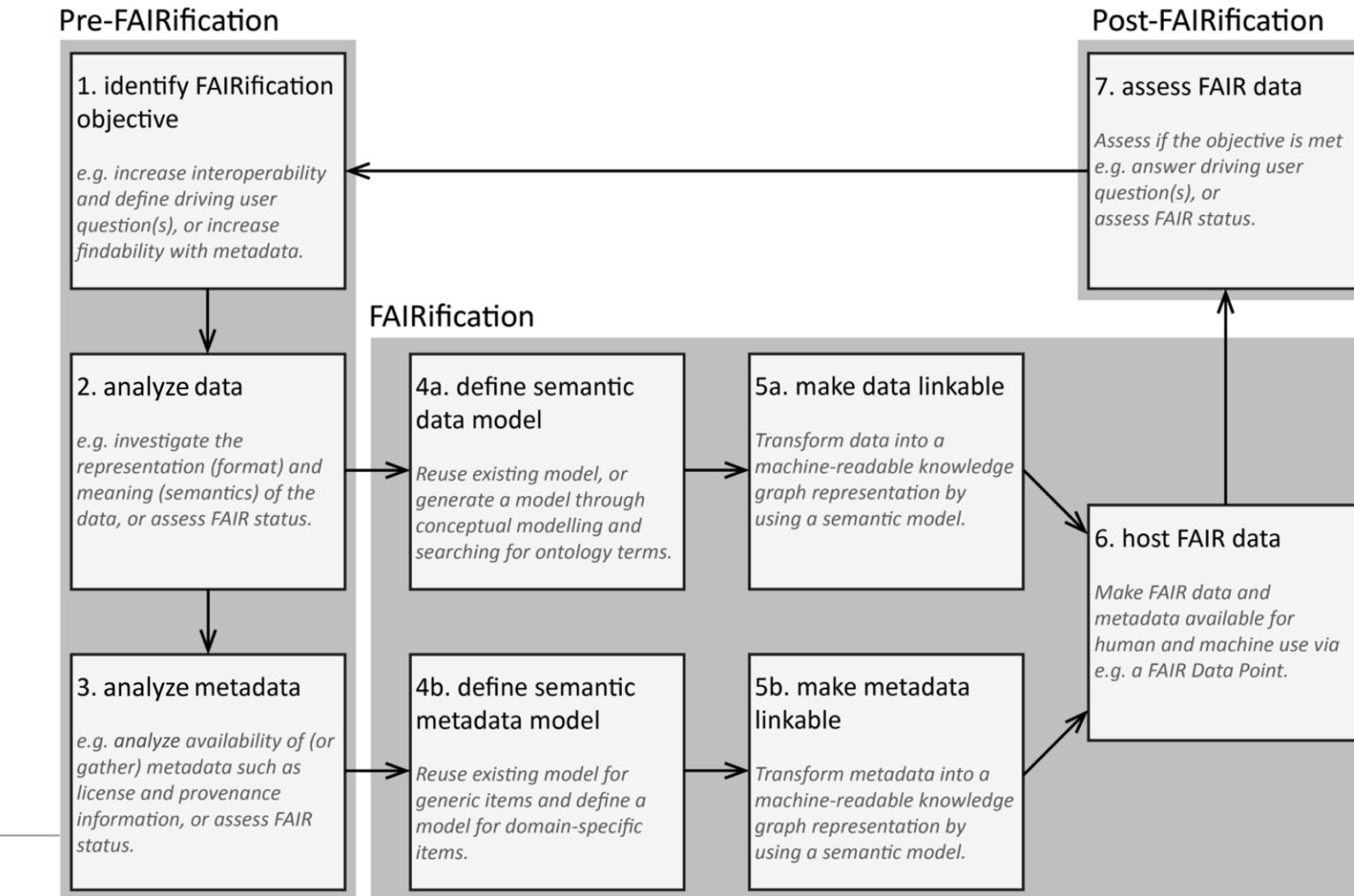


# Benefit (Community) FAIR roadmap:

1. Data FAIR maken (metadateren)
2. Afspraken maken (adopteren standaards)
3. Technische Infrastructuur realiseren

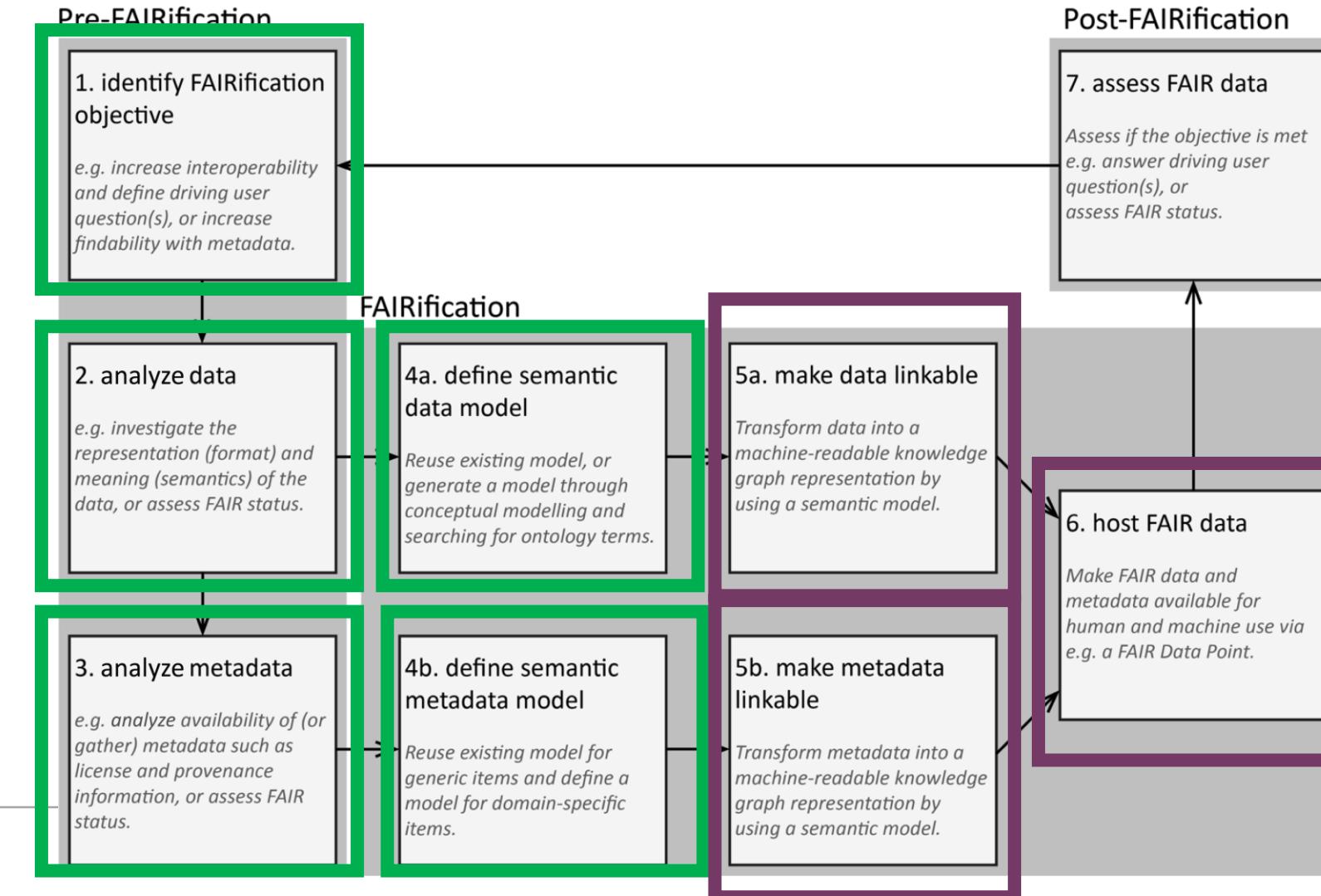


# The FAIRification Process



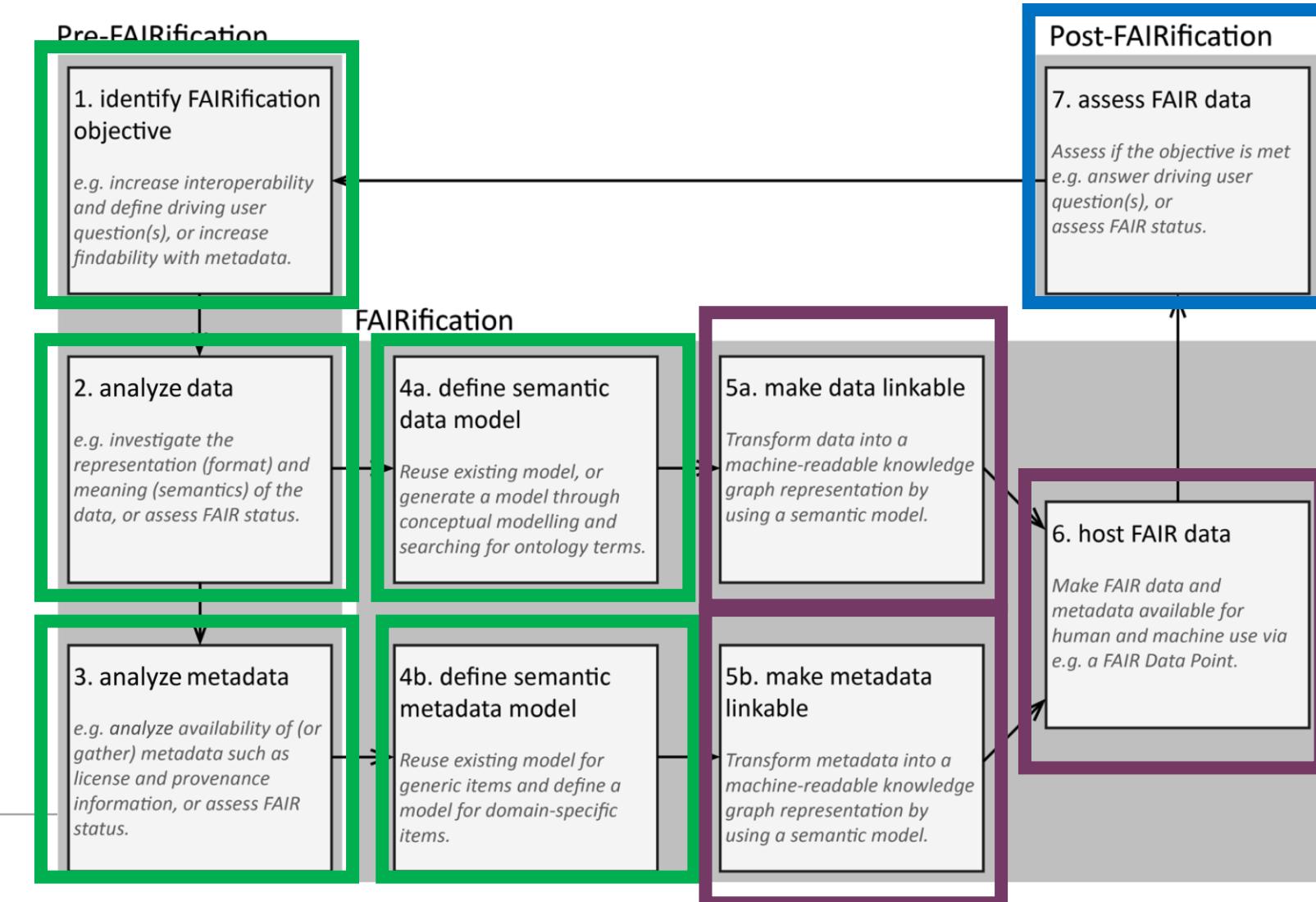
A Generic Workflow for the Data FAIRification Process, Jacobsen et al, 2020 Data Intelligence 2 (1-2): 56–65

# The FAIRification Process



- Pilot data available for controlled access in anDREa
- FAIR data point makes metadata available via FAIR Data Point

# The FAIRification Process



- Revisit research questions
- Explore links with other FDPs and other data

# Secure Environment

## VRE OP MAAT: SAMEN IN CONTROLE

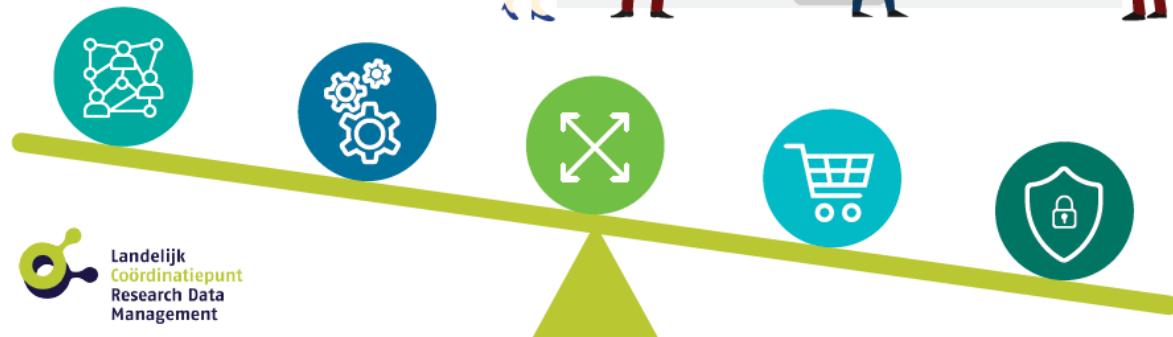
Een Virtual Research Environment (VRE) is een virtuele werkomgeving voor onderzoekers. Hoe een goede werkomgeving **het beste kan worden ingericht**, zal per onderzoeker verschillen.

In Nederland realiseren meerdere universiteiten, hogescholen en universitair medische ziekenhuizen virtuele werkomgevingen voor onderzoekers. Daarin zijn combinaties van vijf verschillende drijfveren te onderscheiden die bepalen hoe deze werkomgevingen worden gerealiseerd.

Onderzoek baant zich een weg voorbij grenzen om nieuwe kennis op te doen. Hierbij hoort de balans tussen vrijheid en controle om op nieuwe en verschillende manieren onderzoek te doen. De onderzoeksinstelling wil de onderzoeker hierin ontzorgen en tegelijk de bedrijfsmatigheid kunnen garanderen.

## OP ZOEK NAAR DE JUISTE BALANS TUSSEN:

### 5 DRIJFVEREN



## DE DYNAMIEK VAN ONDERZOEK VRAAGT OM MAATWERK

### SAMENWERKING

Vergroot de impact van onderzoek met organisatieoverschrijdende samenwerking. Dat vraagt een gedeelde omgeving waar gezamenlijke regels en afspraken worden geïmplementeerd.

### AUTOMATISER-BAARHEID

Ondersteun de efficiëntie en integriteit van onderzoek. Dat vraagt om een specifieke omgeving waarin het te volgen proces kan worden vastgelegd en herhaald.

### SCHAALBAARHEID

Vergroot de impact van onderzoek met data-intensief onderzoek. Dat vraagt om een omgeving die op de momenten dat het nodig is snel over grote rekenkracht kan beschikken.

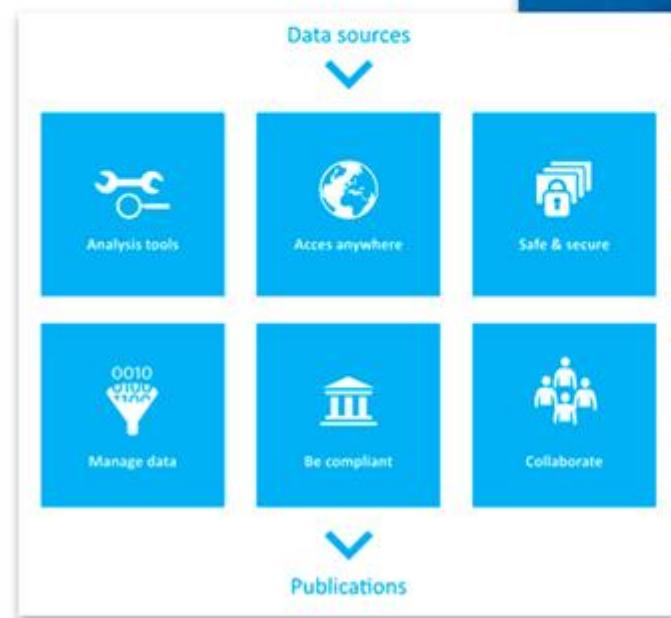
### LAAGDREMPELIGHEID

Ontzorg onderzoekers bij het realiseren van een goede onderzoeksomgeving. Dat vraagt om een omgeving die een onderzoeker zelf eenvoudig kan inrichten, gebruiken en afrekenen.

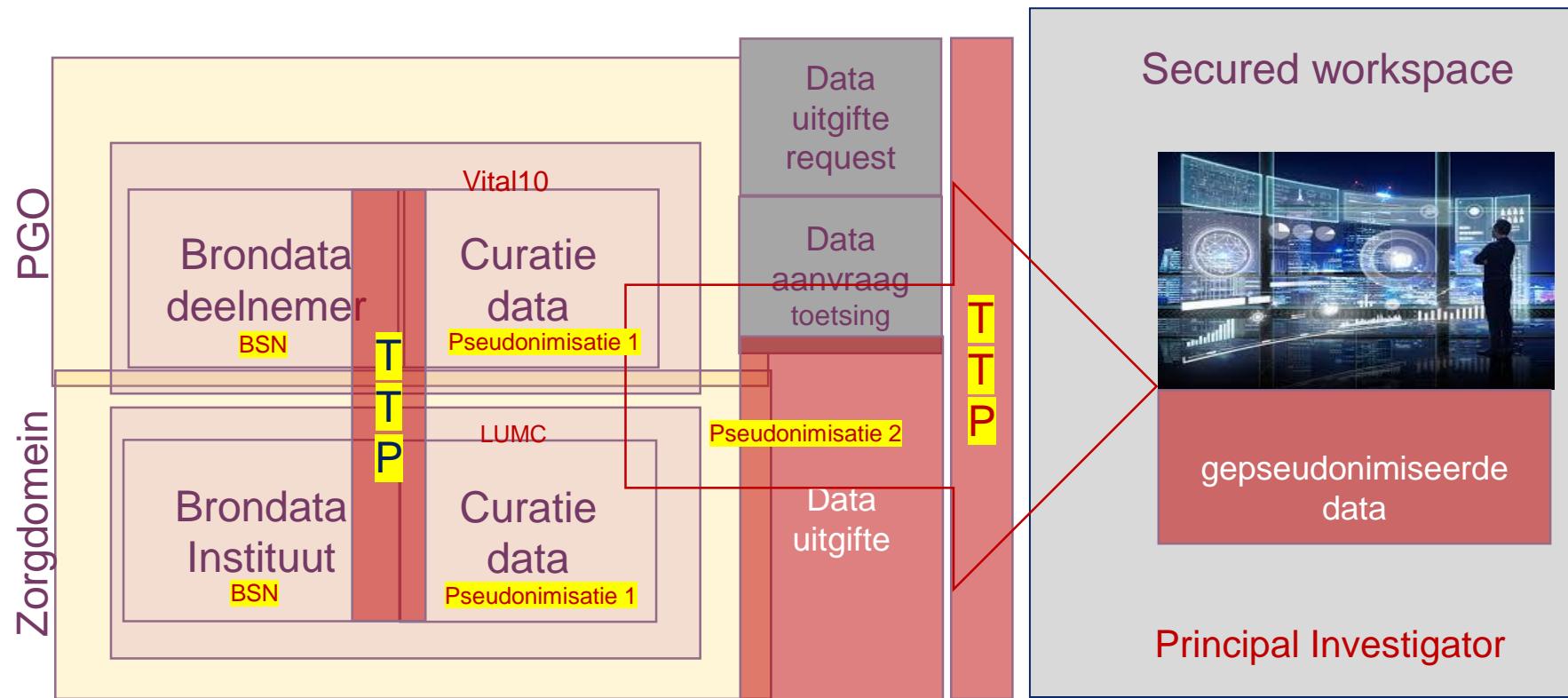
### VEILIGHEID

Voldoe aan wet- en regelgeving voor onderzoek. Dat vraagt om een omgeving waarin toegang kan worden gespecificeerd en herleidbaar is wie toegang heeft gehad tot bepaalde gegevens.

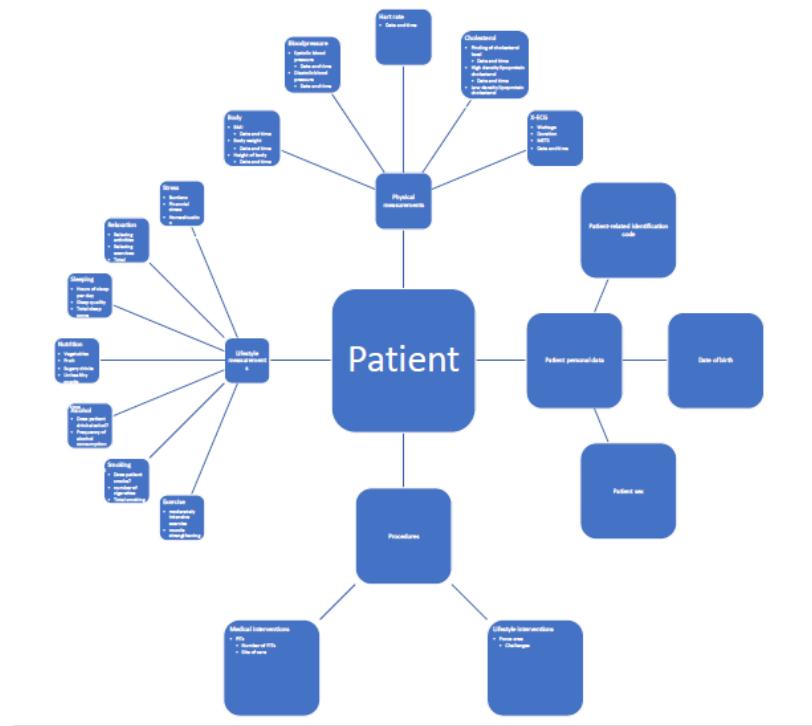
# Secure Environment



## Benefit IT architectuur (beveiligingskaders)



# Pilot Data Set

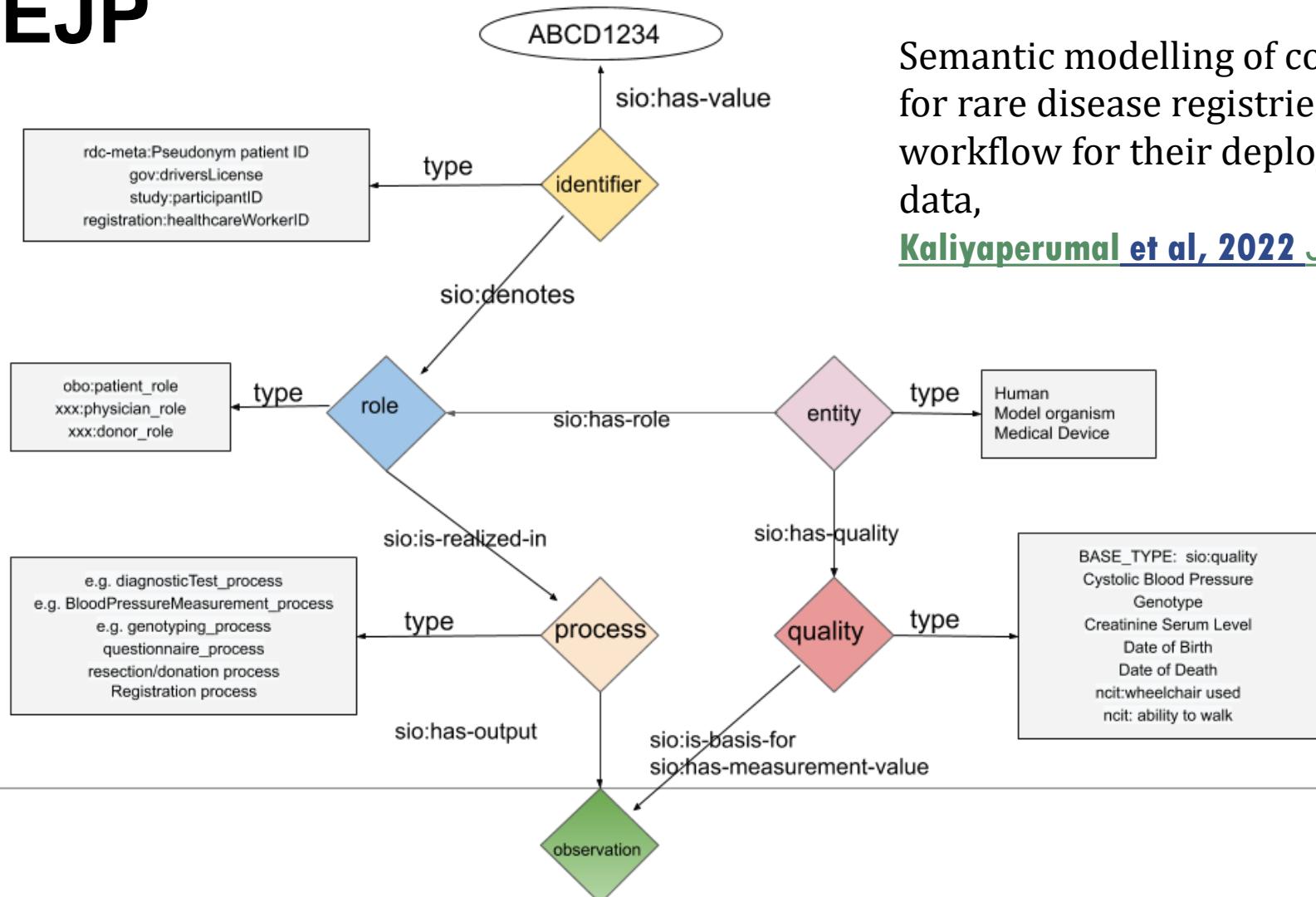


- Lifestyle measurements
- Patient personal data
- Procedures (interventions)
- Physical measurements

# Common Metadata Standards

OMOP ID	SNOMED ID	Concept Name	Class	Domain	Questionnaire Question
4141766	266741004	<a href="#"><u>Muscle strength exercise</u></a>	Procedure	Observation	Tijd spierversterkend (dagen*minuten)
4298263	386293009	<a href="#"><u>Balance exercises</u></a>	Procedure	Observation	Tijd balans (dagen*minuten)
4086504	248263006	<a href="#"><u>Duration of sleep</u></a>	Observable Entity	Observation	Uren / dag
4086501	248254009	<a href="#"><u>Quality of sleep</u></a>	Clinical Finding	Observation	Kwaliteit (range 1-4)

# Using a Common Data Model from EJP



Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data,

[\*\*Kaliyaperumal et al, 2022 J Biomed Semantics\*\*](#)

# Benefit Data in anDREa

The screenshot shows the GraphDB interface with the following details:

- Left Sidebar:** Contains icons for Import, RDF, Tabular (OntoRefine), Explore, SPARQL, Monitor, and Setup. The "Tabular (OntoRefine)" option is highlighted with a red background.
- Header:** Shows the GraphDB logo, a "FREE" badge, a search icon, and a dropdown menu for "BenefitPilot".
- Main Content:** The title "OntoRefine" is displayed with an information icon. Below it is a navigation bar with "Create Project", "Open Project" (selected, highlighted in blue), "Import Project", and "Language Settings".
- Data Table:** A table lists three projects:

	Last modified	Name	Tags	Creator	Subject	Description	Row Count
	2022-09-23 06:59 AM	fields csv					436
	2022-09-23 05:06 AM	v10 20220501 155753 entries csv					118964
	2022-09-23 04:28 AM	v10 20220501 155753 participants csv					695

# Mapping to Common Data Model

The screenshot shows the GraphDB OntoRefine interface. The left sidebar has tabs for Import, RDF, and Tabular (OntoRefine), with Tabular (OntoRefine) selected. The main area is titled "OntoRefine" with a help icon. It shows a configuration interface with tabs for Configuration (selected) and Preview. A message says "All mapping changes saved". Below are buttons for Save, Download JSON, Upload JSON, RDF, and SPARQL. A "New mapping" button is highlighted. The mapping table has columns: FieldId, Name, Label, DataType, Unit, SnomedID. The first row shows a mapping from @FieldId to a resource with IRI <IRI>, rdfs: seeAlso, and @SnomedID. The second row shows a mapping from @SnomedID to a resource with IRI <IRI>, a label 'a', and a value 'http://...OMEDCT/'. The third row shows a mapping from @DataType to a resource with IRI <IRI>, a label 'a', and a value 'sio: SIO\_000074'.

FieldId	Name	Label	DataType	Unit	SnomedID
@FieldId	<IRI>	rdfs: seeAlso	<IRI>	@SnomedID	<IRI>
@SnomedID	<IRI>	a	<IRI>	http://...OMEDCT/	<IRI>
@DataType	<IRI>	a	<IRI>	sio: SIO_000074	<IRI>



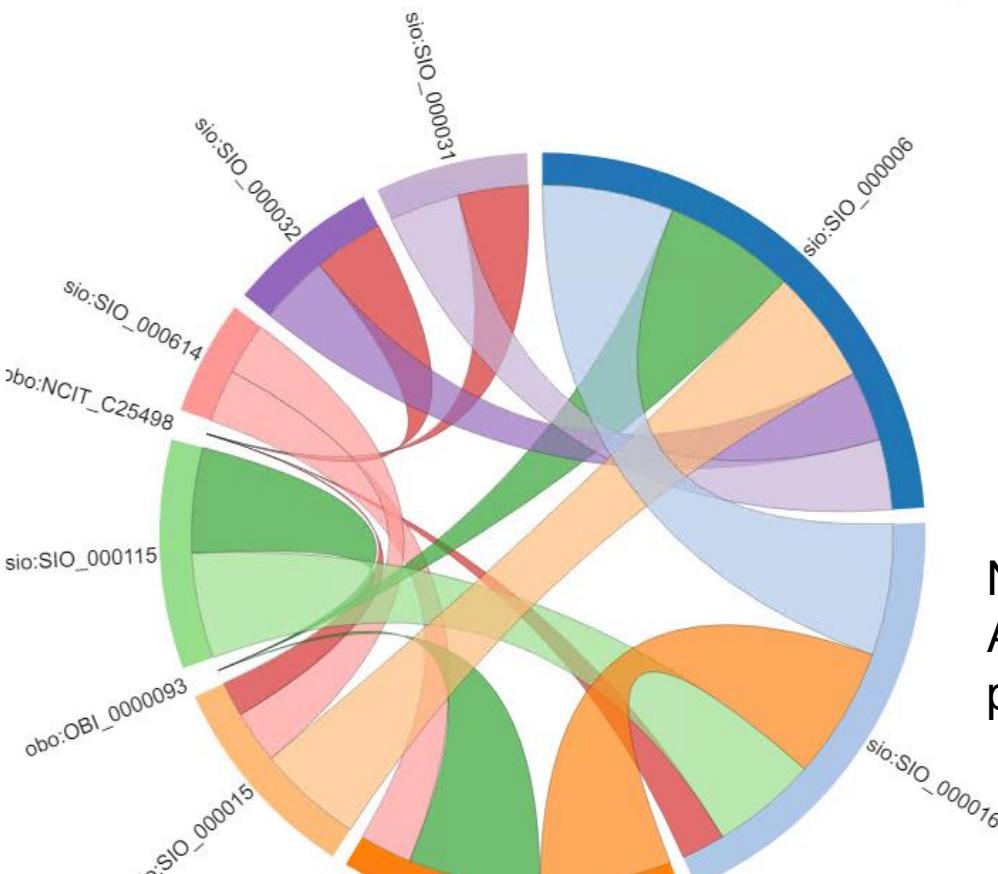
# Class relationships i

Showing the dependencies between 10 classes

 Filter classes

- All     Incoming     Outgoing

Class	Links
sio:SIO_00006	103 ↔
sio:SIO_00016	89 ↔
sio:SIO_000498	65 ↔
sio:SIO_00015	55 ↔
obo:OBI_0	42 ←



Not Findable  
Accessible in anDREa with  
permission and an account

# FAIR Data Point



**FAIR Data Point**  
Metadata for machines

Search FAIR Data Point...

Advanced

AE

[Benefit-NL FAIR Data Point](#) / BENEFIT for all: the ecosystem for he...

## [DRAFT] BENEFIT for all: the ecosystem for healthy living

The BENEFIT program aims to provide long-term support to patients with cardiovascular disease (CVD) in their own home with a healthy lifestyle. Core of the BENEFIT program is access to an advanced eHealth platform with Personal Health Portal consisting of functionality for daily goal monitoring, access to (evidence-based) lifestyle interventions, personal coaching and a reward program aimed at stimulating a wide range of health behaviours and therapy adherence. This integrated healthy living program makes it possible to tailor lifestyle interventions to the needs and preferences of end users, encouraging engagement and active use. The program also promotes self-management of a healthy lifestyle by encouraging patients to measure health indicators (such as blood pressure) themselves at home.

Owner

Publish

Edit

Settings

Delete

Metadata Issued  
26-09-2022

Metadata Modified  
26-09-2022

Conforms to

- [Catalog Profile](#)

Version  
1.0

# Finadable Through FAIR Data Point



**FAIR Data Point**

Metadata for machines

Search FAIR Data Point...

AE ▾

Advanced

## Benefit-NL FAIR Data Point

Edit

Demo FAIR Data Point Benefit-NL consortium (workpackage 3)  
<https://benefitforall.nl/benefit-consortium/>

Metadata Issued  
26-09-2022

Metadata Modified  
26-09-2022

### Catalogs

Create

#### [DRAFT] BENEFIT for all: the ecosystem for healthy living

The BENEFIT program aims to provide long-term support to patients with cardiovascular disease (CVD) in their own home with a healthy lifestyle. Core of the BENEFIT program is acces...

Issued 26-09-2022 Modified 26-09-2022

Conforms to

- [FAIR Data Point Profile](#)

Version

1.0

Language  
[English](#)

# FAIR Data Point Registry



Search FAIR Data Point...

Log in

Advanced

## FAIR Data Points

Filter:

All 220

Active 31

Inactive 55

Unreachable 98

Invalid 33

Unknown 3

Endpoint ▲ ▼

Registration ▲ ▼

Modification ▲ ▼

Status

<a href="https://directory.bbmri-eric.eu/api/fdp">https://directory.bbmri-eric.eu/api/fdp</a>	27-01-2021, 15:30:32	27-09-2022, 05:45:11	ACTIVE
<a href="https://covid19research.nl/api/fdp">https://covid19research.nl/api/fdp</a>	27-01-2021, 15:39:43	27-09-2022, 05:00:00	ACTIVE
<a href="http://rcxopenspecimenfair-fdp-client.silicolabv4.bibbox.org">http://rcxopenspecimenfair-fdp-client.silicolabv4.bibbox.org</a>	25-08-2022, 18:46:40	27-09-2022, 04:46:40	ACTIVE
<a href="https://fdp.castoredc.com">https://fdp.castoredc.com</a>	01-02-2021, 11:57:53	27-09-2022, 02:00:02	ACTIVE
<a href="http://145.38.197.183:8080">http://145.38.197.183:8080</a>	24-11-2021, 12:30:49	26-09-2022, 13:38:54	ACTIVE

# Next Steps - Evaluation

- Assess FDP and described data against research questions
    - 1.What lifestyle risk factors do CVD patients exhibit before they enter cardiac rehabilitation?
    - 2.How is the patients' lifestyle related to their health?
    3. Is there an improvement in lifestyle risk factors after cardiac rehabilitation?
-

# Next Steps - Expansion

- Benefit FAIR data paper
    - Data descriptor
    - Data sharing with secure environments
  - Extend the pilot with other data
    - Subset of Benefit data described. What is the most useful expansion?
  - Investigate interoperation with other datasets
-

# CONCLUSIE Benefit WP3

Benefit data model (FAIR, concept V 0.1)

Consortiumleden kunnen hierbij in een volgende fase aanhaken (model uitbreiden)

Workspace data & dashboards

AnDREa: *in the cloud*, controleerbare en GDPR veilige omgeving voor dataverwerking

IT infrastructuur voor FAIR data

Eerste aanzet voor aansluiting op Personal Health Train concept (voorbereiding op data visiting)

---