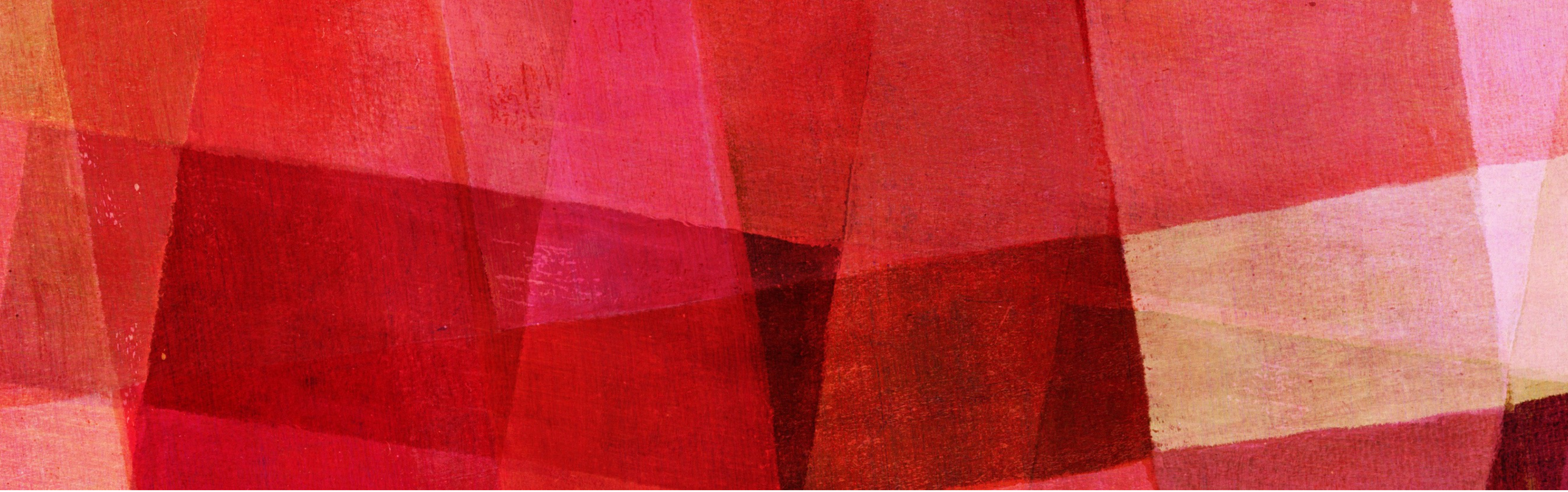


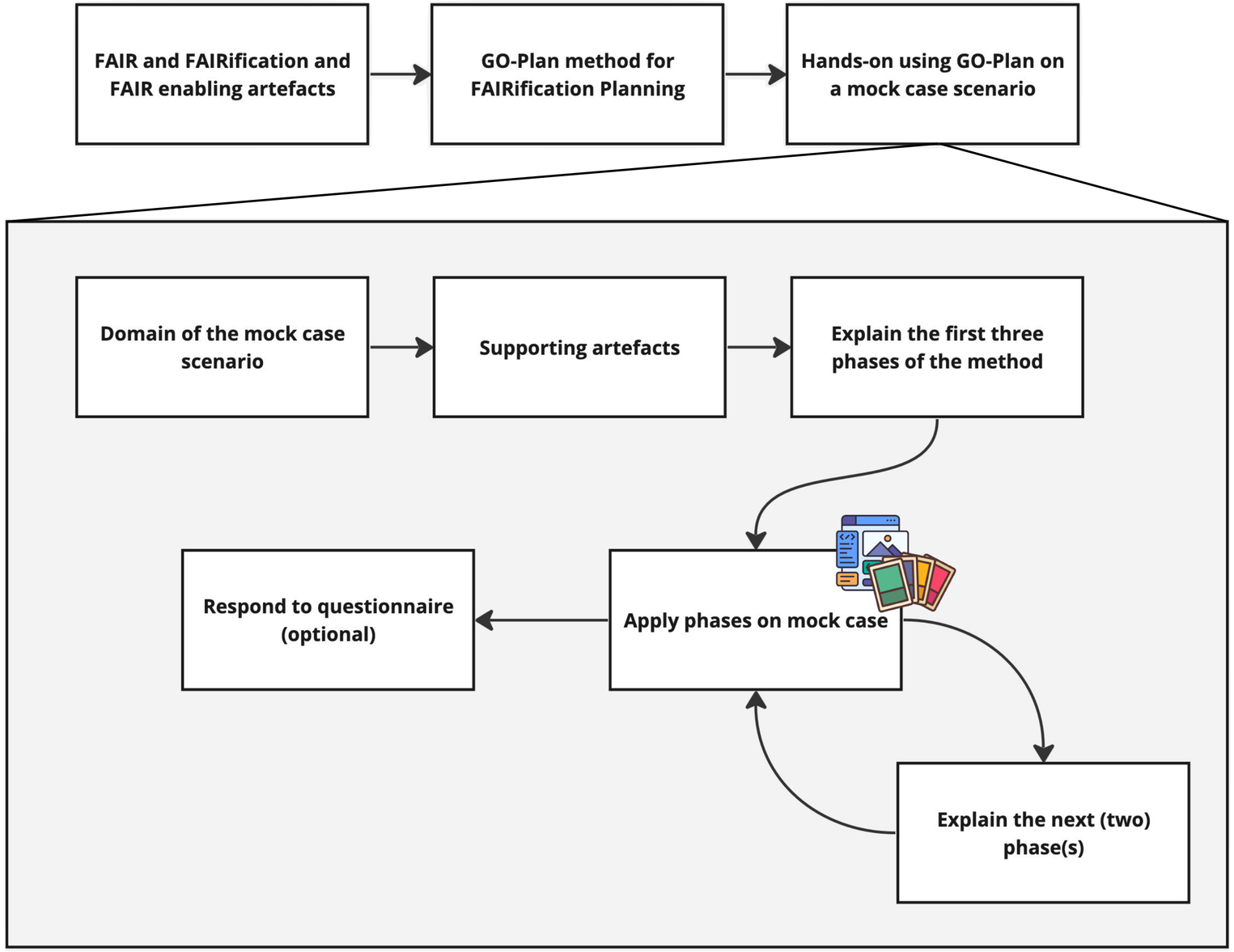
INSIGHTS IN FAIRIFICATION PLANNING

UNIVERSITY OF TWENTE.

LUMC Leids Universitair Medisch Centrum

EUROPEAN JOINT PROGRAMME RARE DISEASES

AGENDA



Template

Cards

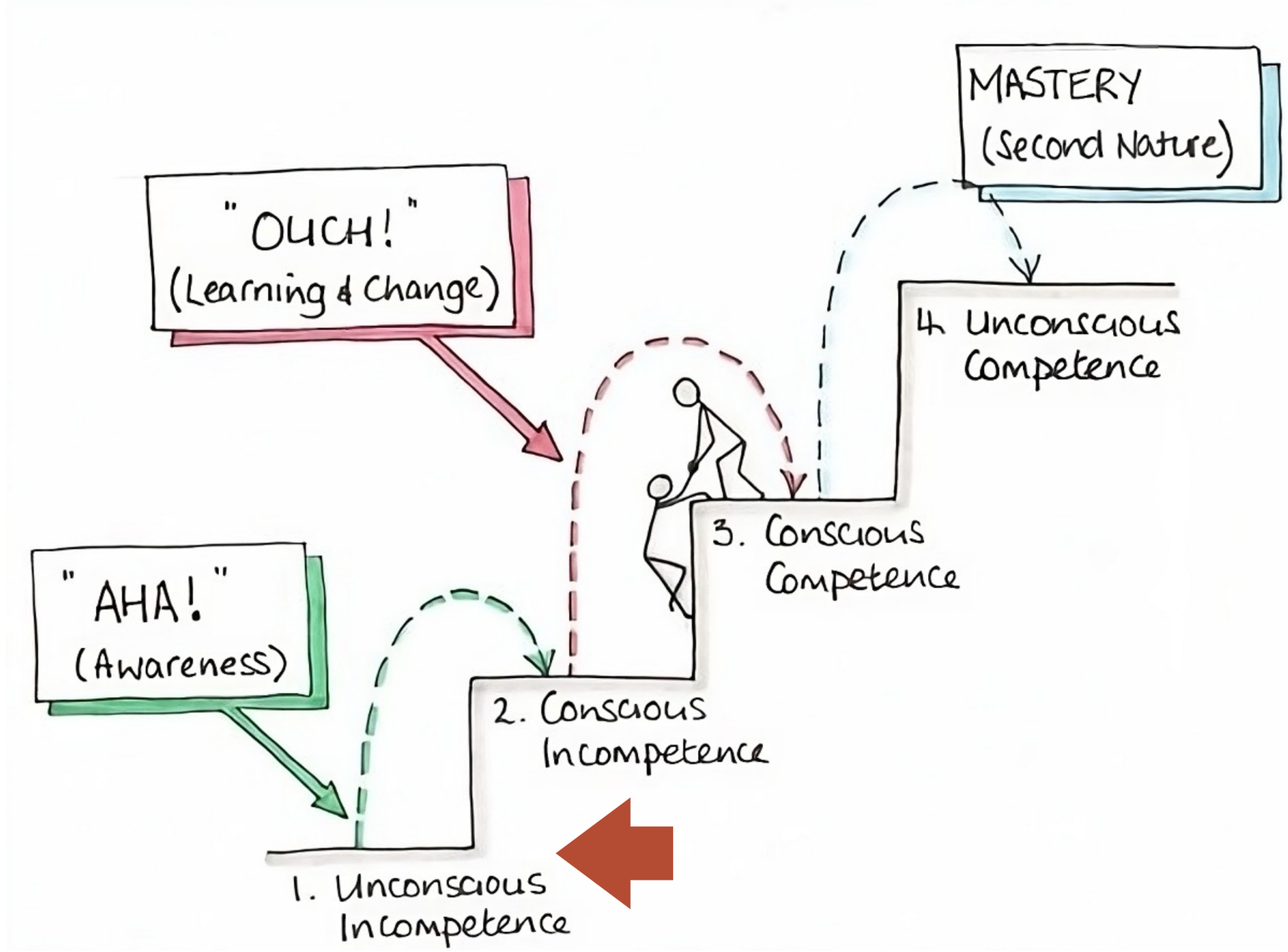
WHAT WILL YOU LEARN TODAY?

- Why FAIR and the benefits of having FAIR data
- The FAIR principles
- The process of making data FAIR: FAIRification and FAIRification workflows
- FAIR enabling software, standards, technology and other artefacts
- GO-Plan method and hands-on application

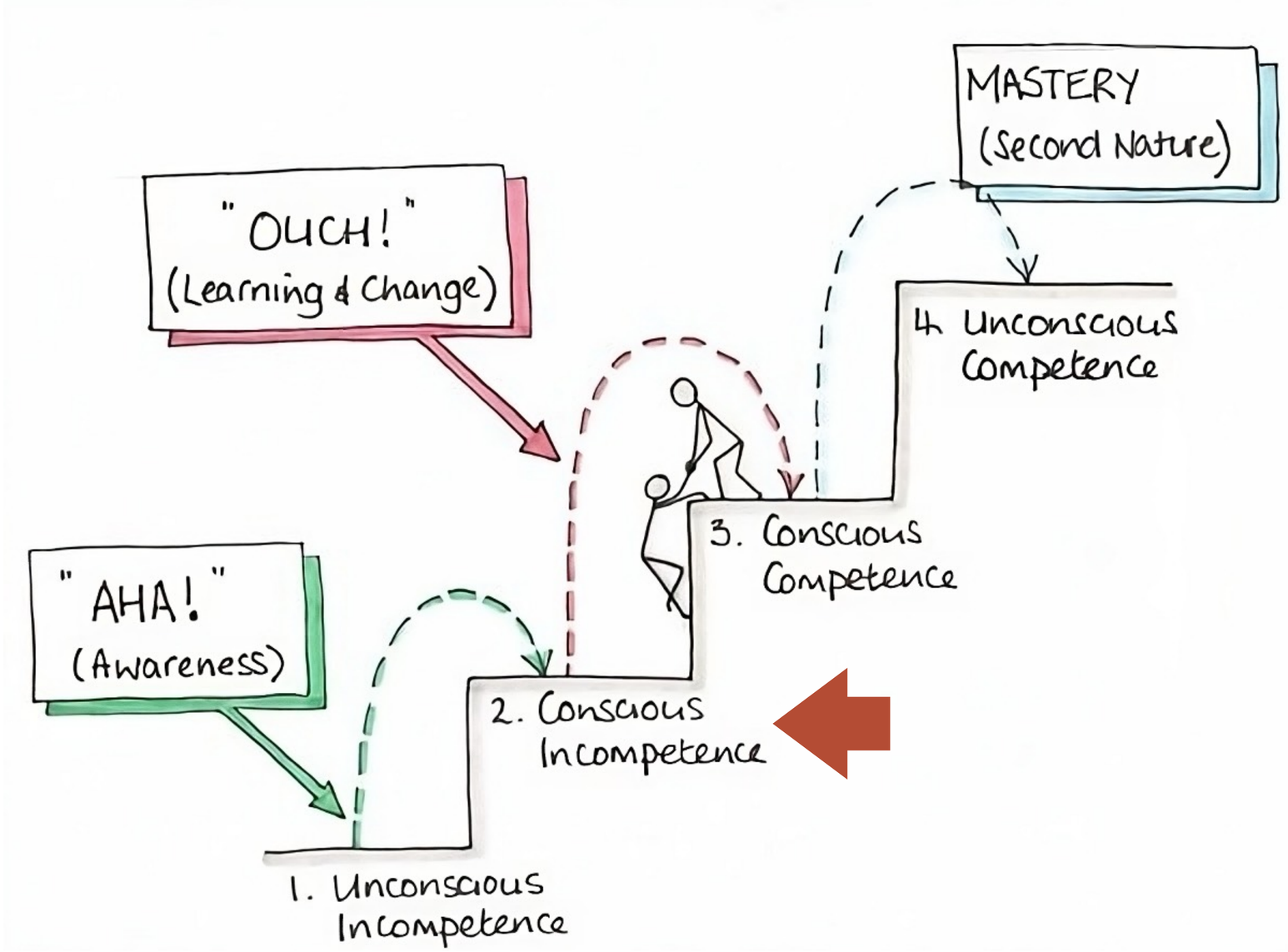
WHAT WILL YOU NOT LEARN TODAY?

- How to realise FAIR (e.g., develop code/script, transform data)
- How to use any specific software or tool
- How to create conceptual models or ontologies
- How to assess FAIRness

LEARNING GOALS



LEARNING GOALS



WHY FAIR?

- The **European Joint Programme on Rare Diseases (EJP RD)** aims at creating an effective rare diseases research ecosystem for progress, innovation and for the benefit of everyone with a rare disease.

WHY FAIR?

- The **European Joint Programme on Rare Diseases (EJP RD)** aims at creating an effective rare diseases research ecosystem for progress, innovation and for the benefit of everyone with a rare disease.

- Data cannot leave the source (legal regulations)
- Difficult to achieve Interoperability
  - Language barrier
  - Ambiguity of data terms

WHY FAIR?

EUROPEAN JOINT PROGRAMME RARE DISEASES

WHY FAIR?

EUROPEAN JOINT PROGRAMME RARE DISEASES

pacient	léčba nemocí

ασθενής	διάγνωση

WHY FAIR?

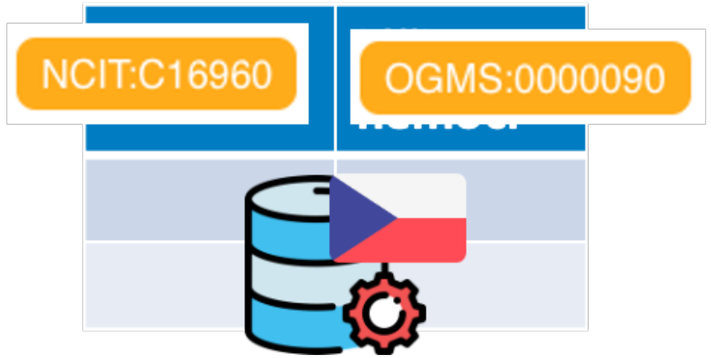
EUROPEAN JOINT PROGRAMME RARE DISEASES

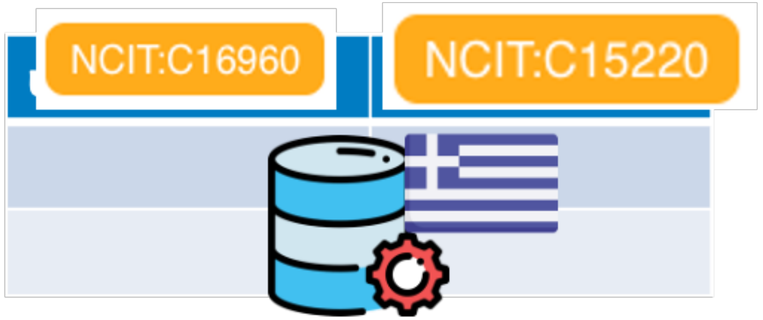
pacient	léčba nemocí

ασθενής	διάγνωση

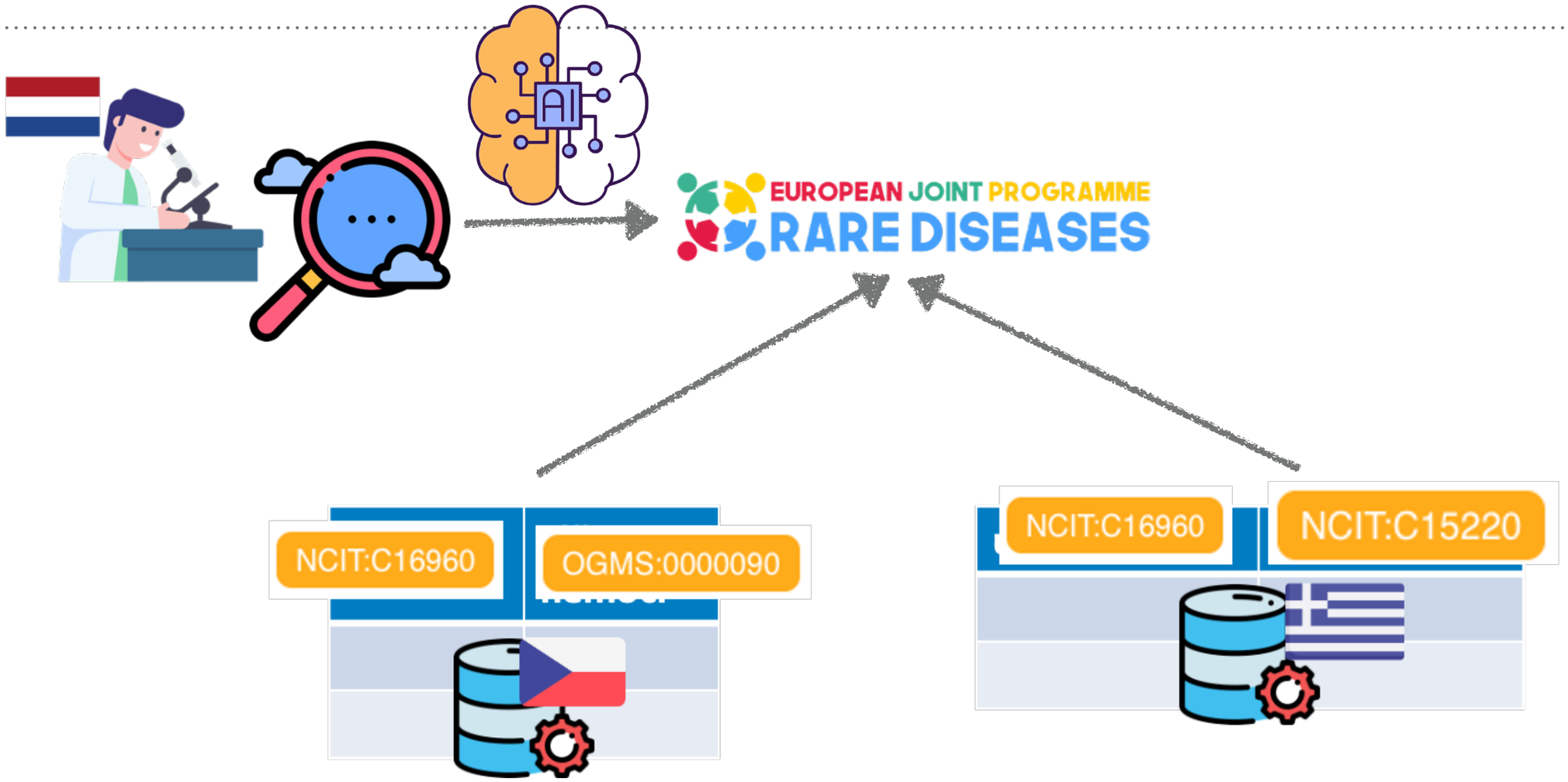
WHY FAIR?

EUROPEAN JOINT PROGRAMME RARE DISEASES

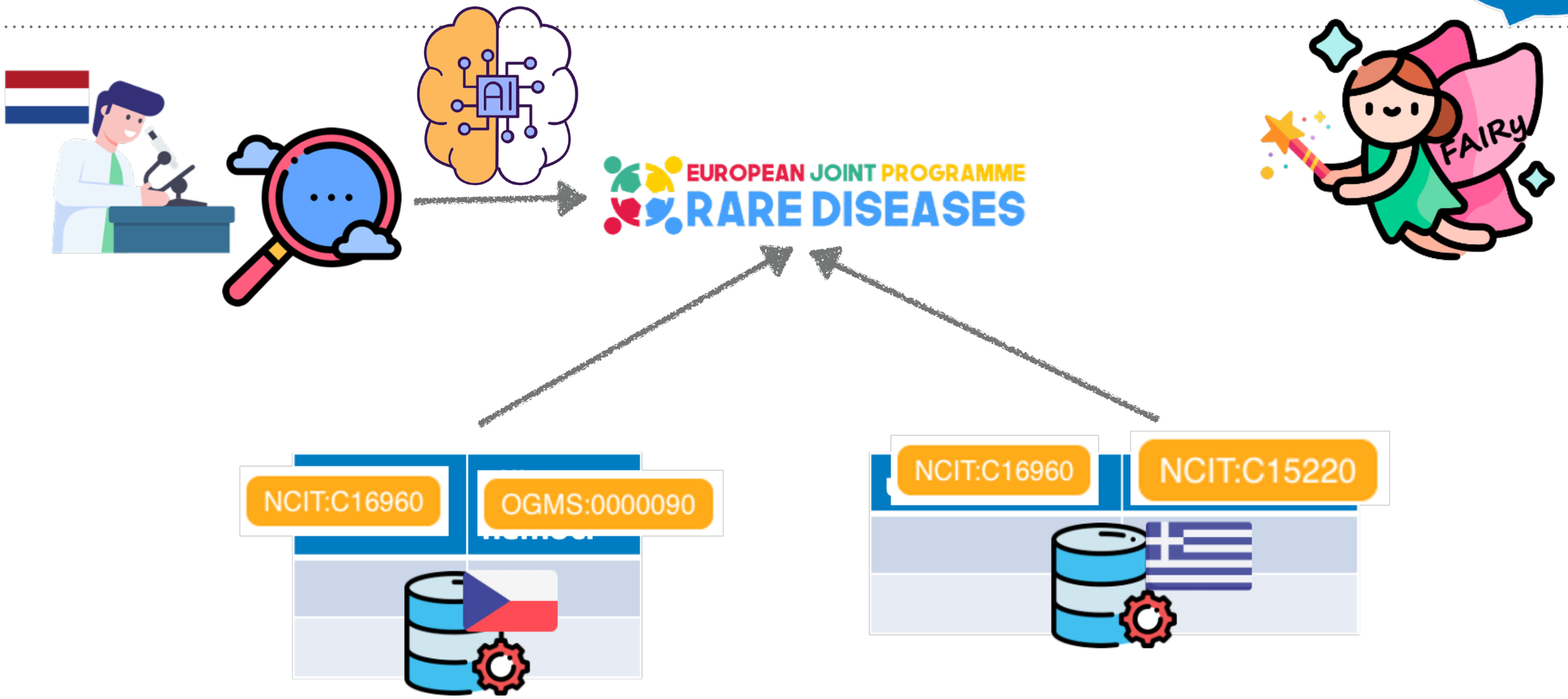




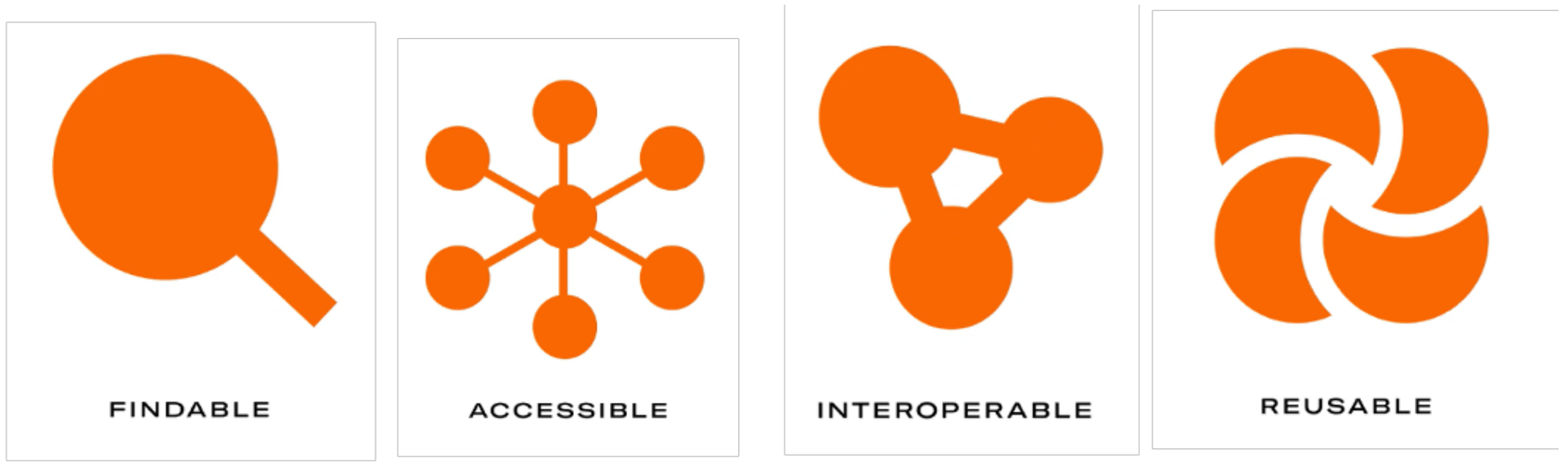
WHY FAIR?


AI
EUROPEAN JOINT PROGRAMME
RARE DISEASES
NCIT:C16960
OGMS:0000090
NCIT:C16960
NCIT:C15220

WHY FAIR?


Doei!
FAIRy
AI
EUROPEAN JOINT PROGRAMME
RARE DISEASES
NCIT:C16960
OGMS:0000090
NCIT:C16960
NCIT:C15220

THE FAIR PRINCIPLES



THE FAIR PRINCIPLES: FINDABILITY

- 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**;

THE FAIR PRINCIPLES: FINDABILITY

- 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**;

Findability

Principle	What?	What about?
F1	(Meta)data	GUPRI
F2	Data	Described by rich metadata
F3	Metadata	Identifier to data
F4	(Meta)data	Indexed by search engine

THE FAIR PRINCIPLES: ACCESSIBILITY

- 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**;

THE FAIR PRINCIPLES: ACCESSIBILITY

- 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**;

Accessibility

Principle	What?	What about?
A1	(Meta)data	Retrievable by standardised protocol
A1.1	Protocol	Free, open, universal
A1.2	Protocol	Allows for authentication and authorisation
A2	Metadata	Accessible even when data not available

THE FAIR PRINCIPLES: INTEROPERABILITY

- 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;

THE FAIR PRINCIPLES: INTEROPERABILITY

- 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;

Interoperability

Principle	What?	What about?
I1	(Meta)data	Uses a broadly applicable knowledge representation
I2	(Meta)data	Uses FAIR vocabularies
I3	(Meta)data	Qualified references to other (meta)data

THE FAIR PRINCIPLES: REUSABILITY

- 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**;

THE FAIR PRINCIPLES: REUSABILITY

- 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**;

Reusability

Principle	What?	What about?
R1	(Meta)data	Richly described
R1.1	Rich (meta)data	Includes clear usage license
R1.2	Rich (meta)data	Detailed provenance
R1.3	Rich (meta)data	Meet community standards

FAIR ENABLING ARTEFACTS

Principle	Technology Examples
Findability	Identifier: PURL, W3ID, DOI
	Searchable resource: bing, google, pubmed, community specific indexer
Accessibility	Communication protocol: http, ftp
Interoperability	Language for knowledge representation: RDF, OWL
	Vocabularies: ontologies, controlled vocabularies (bioportal, ontobee, OLS)
Reusability	Data usage license: Creative Commons
	Provenance: Data Catalog Vocabulary (DCAT)

FAIR ENABLING ARTEFACTS: FAIR DATA POINT

- A FAIR Data Point ultimately stores information about data sets, which is the definition of **metadata**.

- Features
  - Manage users
  - Manage access rights to your catalogs, datasets, and distributions
  - Human and machine readable interface for metadata

BYOD

Edit Settings Delete

Datasets | Patient Registries | Data Services

Datasets

+ Create

Epithelial-mesenchymal plasticity induced

EMT/MET cell line model to demonstrate that epithelial mesenchymal plasticity occurring in normal cells generates co-existing phenotypically and functionally divergent cell subpopu...

topic_3308 10017758 GO_0001837

Issued 18-11-2022 **Modified** 02-06-2023

Epithelial-mesenchymal plasticity induced

EMT/MET cell line model to demonstrate that epithelial mesenchymal plasticity occurring in normal cells generates co-existing phenotypically and functionally divergent cell subpopu...

topic_3308 10017758 GO_0001837

Issued 18-11-2022 **Modified** 02-06-2023

Huntington disease mRNA

Regional and cellular gene expression changes in human Huntington's disease brain

topic_3308 Orphanet_399 NCIT_C44282

Issued 16-11-2022 **Modified** 02-06-2023

Huntingtons disease transcriptomics

Conforms to
- Catalog Profile

Version
1

Language
English

License
cc-by-nc-nd3.0

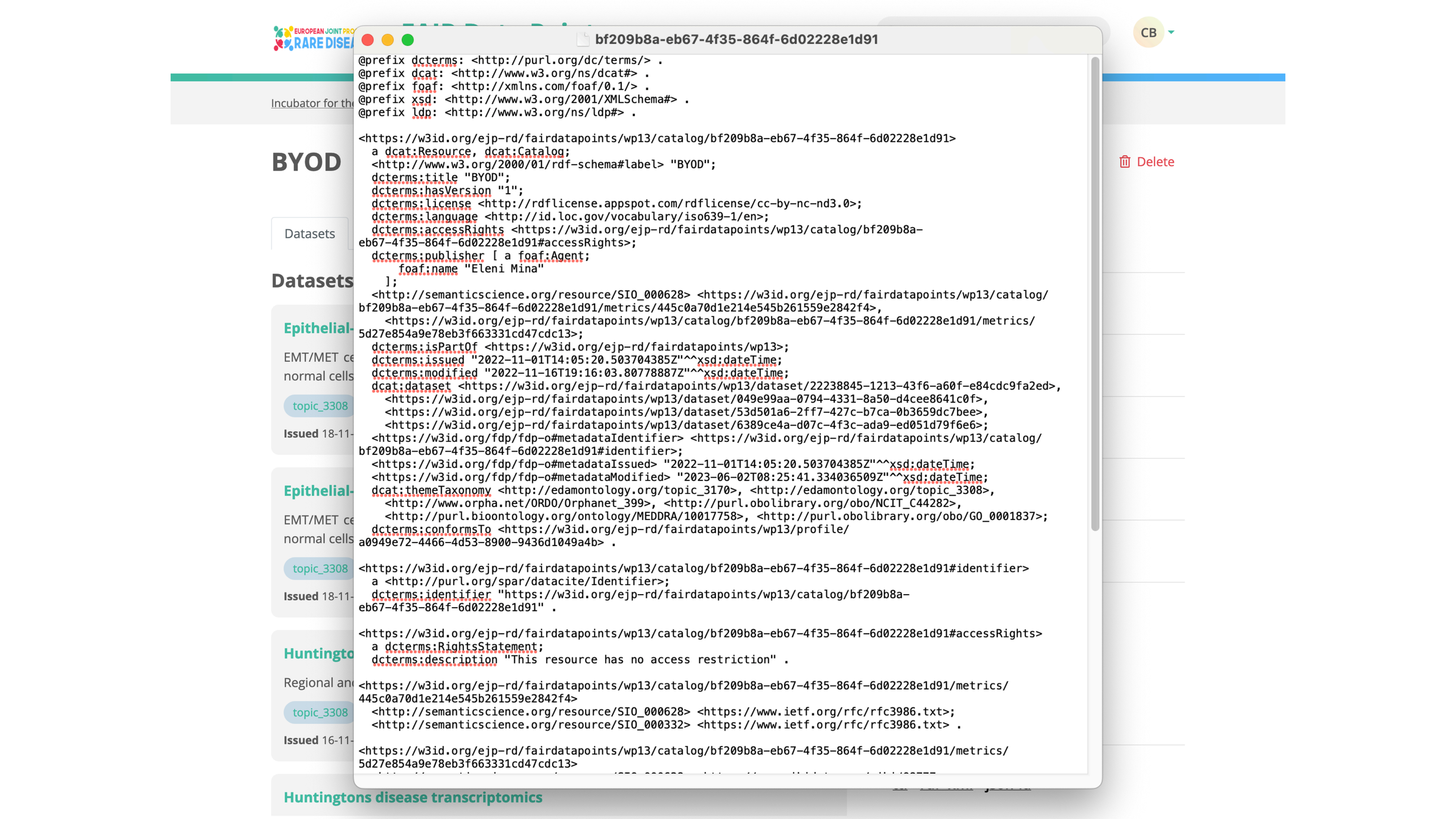
Issued
01-11-2022

Modified
16-11-2022

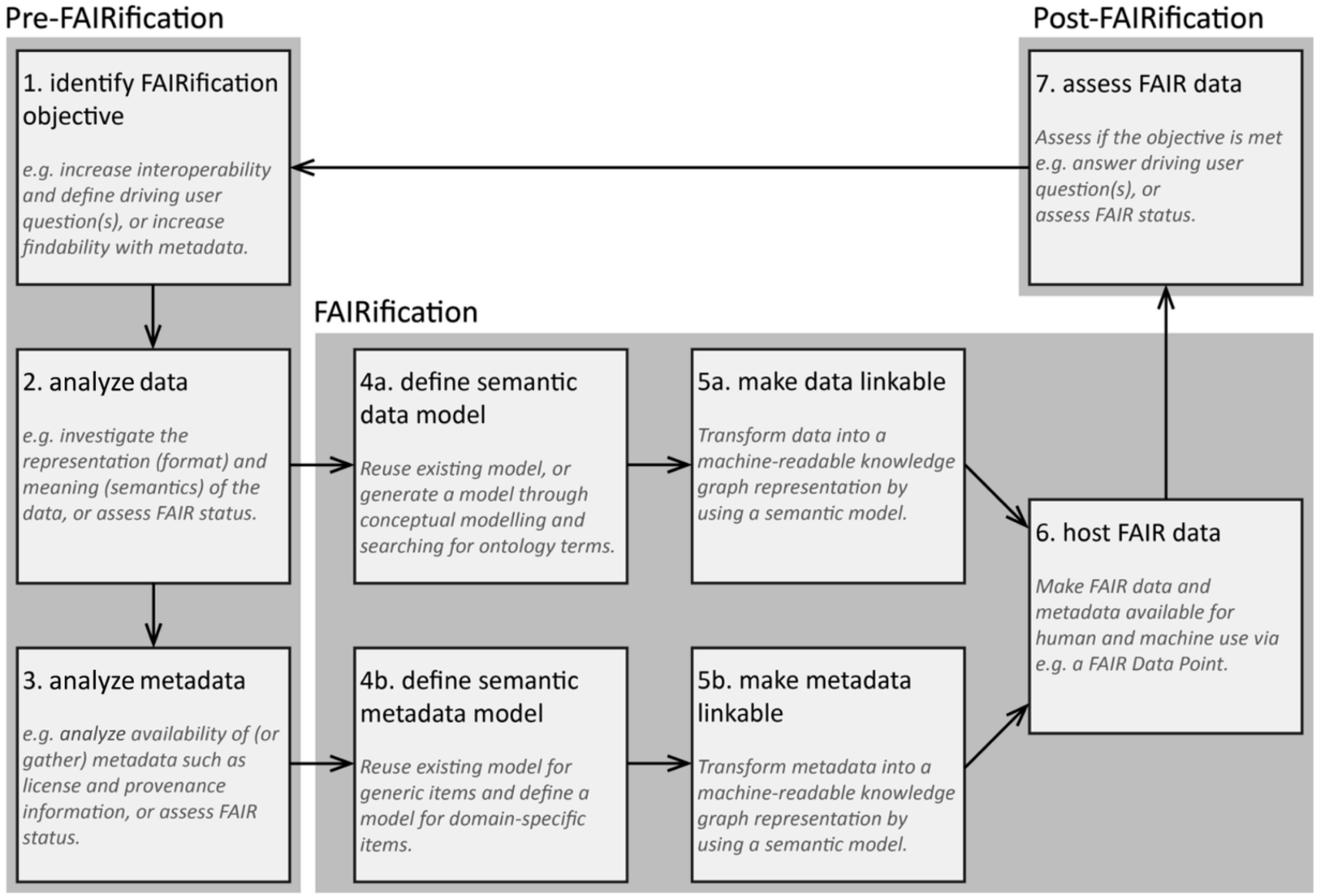
Theme taxonomy
- RNA-Seq
- Transcriptomics
- Orphanet_399

Show more

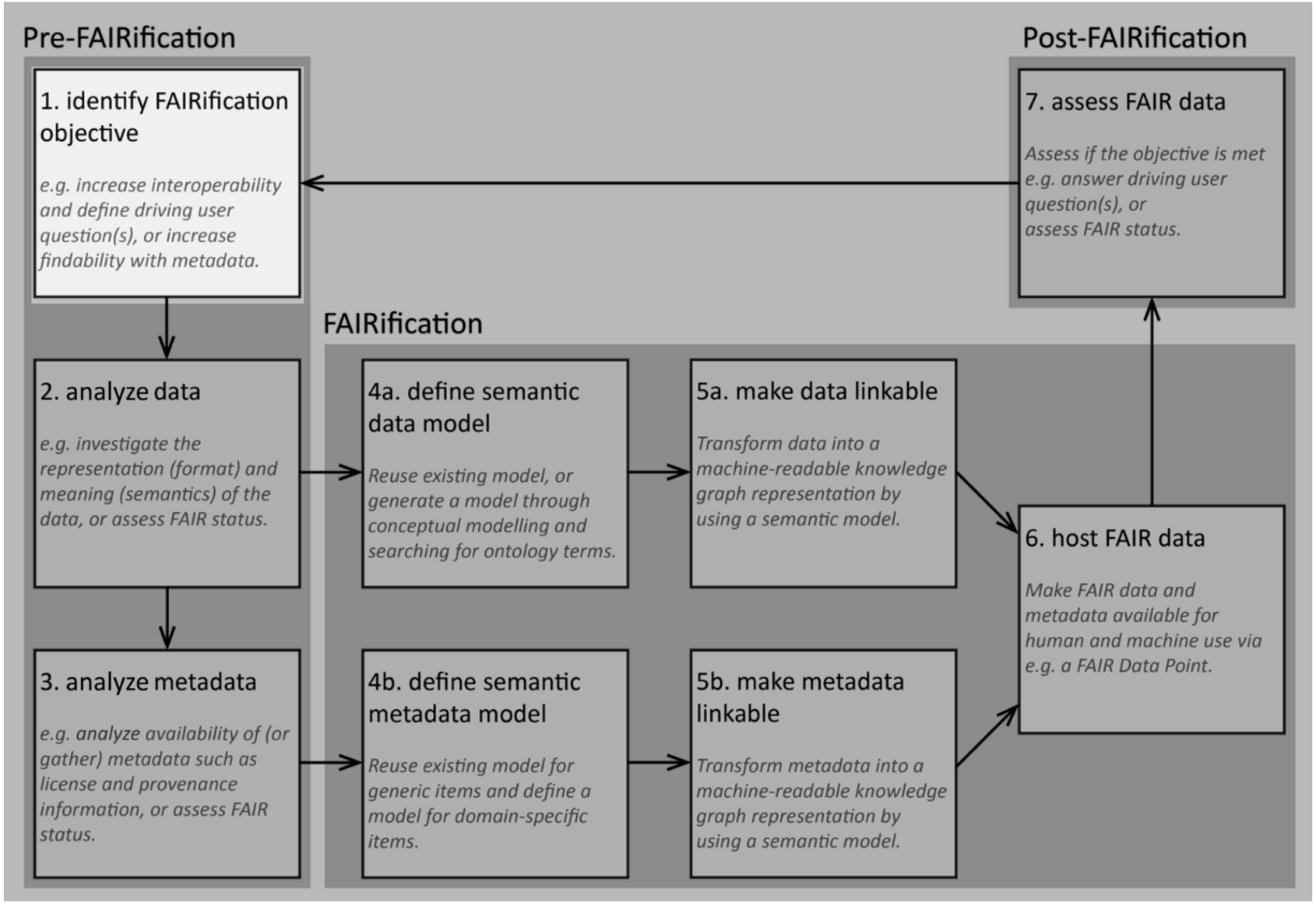
RDF metadata for machines
ttl rdf+xml json-ld


bf209b8a-eb67-4f35-864f-6d02228e1d91
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ldp: <http://www.w3.org/ns/ldp#> .
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91>
a dcat:Resource, dcat:Catalog;
<http://www.w3.org/2000/01/rdf-schema#label> "BYOD";
dcterms:title "BYOD";
dcterms:hasVersion "1";
dcterms:license <http://rdflicense.appspot.com/rdflicense/cc-by-nc-nd3.0>;
dcterms:language <http://id.loc.gov/vocabulary/iso639-1/en>;
dcterms:accessRights <https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91#accessRights>;
dcterms:publisher [a foaf:Agent;
foaf:name "Eleni Mina"
];
<http://semanticscience.org/resource/SIO_000628> <https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91/metrics/445c0a70d1e214e545b261559e2842f4>,
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91/metrics/5d27e854a9e78eb3f663331cd47cdc13>;
dcterms:isPartOf <https://w3id.org/ejp-rd/fairdatapoints/wp13>;
dcterms:issued "2022-11-01T14:05:20.503704385Z"^^xsd:dateTime;
dcterms:modified "2022-11-16T19:16:03.80778887Z"^^xsd:dateTime;
dcat:dataset <https://w3id.org/ejp-rd/fairdatapoints/wp13/dataset/22238845-1213-43f6-a60f-e84cdc9fa2ed>,
<https://w3id.org/ejp-rd/fairdatapoints/wp13/dataset/049e99aa-0794-4331-8a50-d4cee8641c0f>,
<https://w3id.org/ejp-rd/fairdatapoints/wp13/dataset/53d501a6-2ff7-427c-b7ca-0b3659dc7bee>,
<https://w3id.org/ejp-rd/fairdatapoints/wp13/dataset/6389ce4a-d07c-4f3c-ada9-ed051d79f6e6>;
<https://w3id.org/fdp/fdp-o#metadataIdentifier> <https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91#identifier>;
<https://w3id.org/fdp/fdp-o#metadataIssued> "2022-11-01T14:05:20.503704385Z"^^xsd:dateTime;
<https://w3id.org/fdp/fdp-o#metadataModified> "2023-06-02T08:25:41.334036509Z"^^xsd:dateTime;
dcat:themeTaxonomy <http://edamontology.org/topic_3170>, <http://edamontology.org/topic_3308>,
<http://www.orpha.net/ORDO/Orphanet_399>, <http://purl.obolibrary.org/obo/NCIT_C44282>,
<http://purl.bioontology.org/ontology/MEDDRA/10017758>, <http://purl.obolibrary.org/obo/GO_0001837>;
dcterms:conformsTo <https://w3id.org/ejp-rd/fairdatapoints/wp13/profile/a0949e72-4466-4d53-8900-9436d1049a4b> .
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91#identifier>
a <http://purl.org/spar/datacite/Identifier>;
dcterms:identifier "https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91" .
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91#accessRights>
a dcterms:RightsStatement;
dcterms:description "This resource has no access restriction" .
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91/metrics/445c0a70d1e214e545b261559e2842f4>
<http://semanticscience.org/resource/SIO_000628> <https://www.ietf.org/rfc/rfc3986.txt>;
<http://semanticscience.org/resource/SIO_000332> <https://www.ietf.org/rfc/rfc3986.txt> .
<https://w3id.org/ejp-rd/fairdatapoints/wp13/catalog/bf209b8a-eb67-4f35-864f-6d02228e1d91/metrics/5d27e854a9e78eb3f663331cd47cdc13>

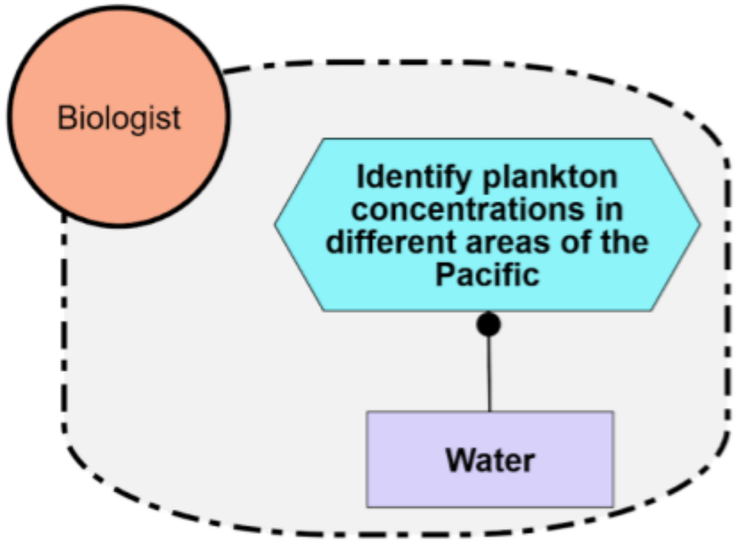
FAIRIFICATION PROCESS AND WORKFLOWS



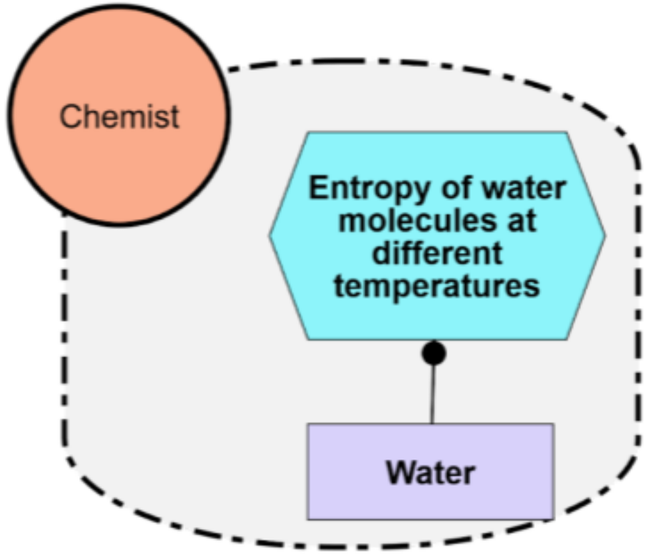
FAIRIFICATION OBJECTIVES



IMPACT OF OBJECTIVES (AKA: GOALS)

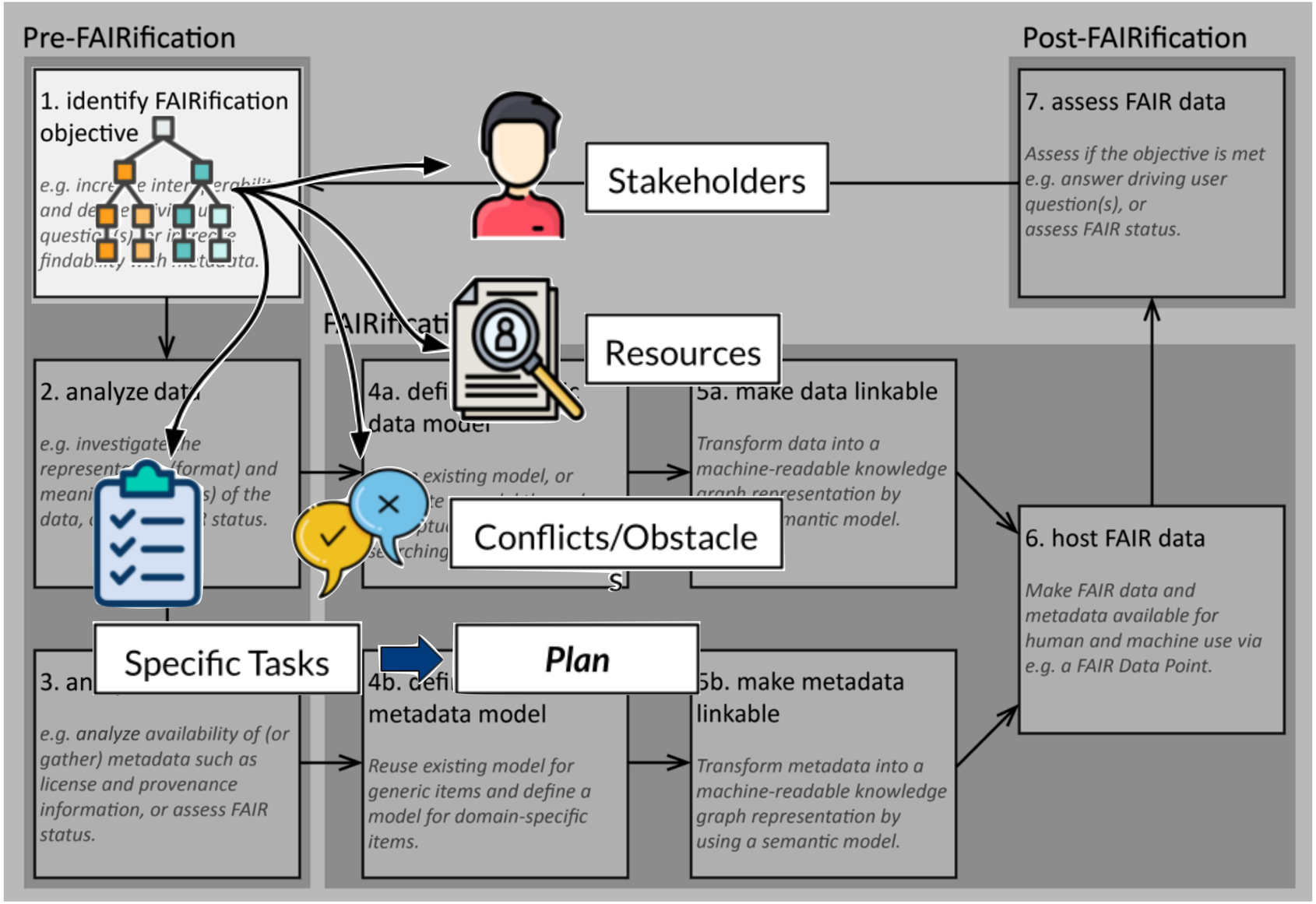


Water as a “**ocean part**”



Water as a
“**chemical compound**”

FAIRIFICATION OBJECTIVES



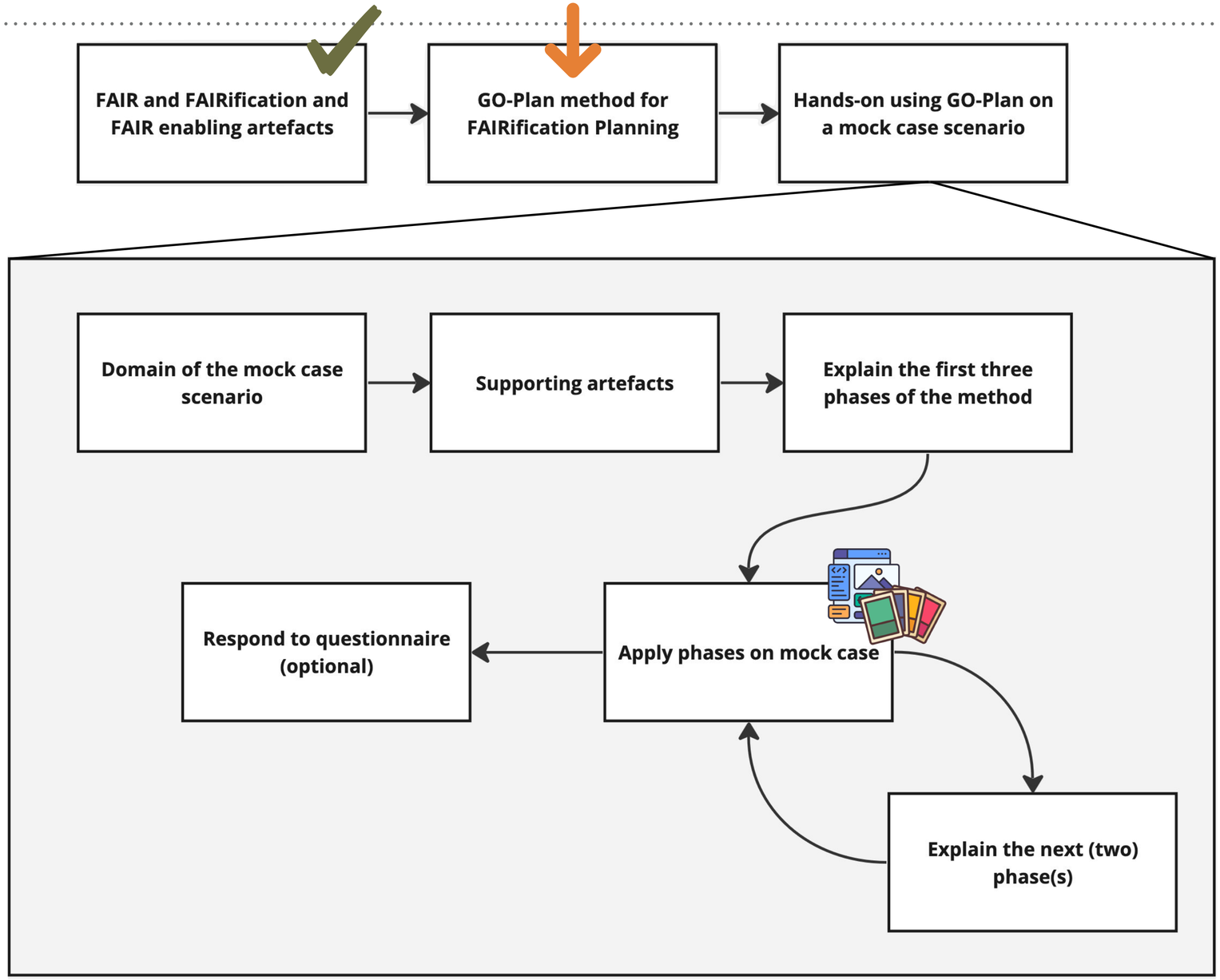
PROPOSAL: GO-PLAN - DISCLAIMER

- GO-Plan has been developed as part of my PhD work
  - Embeds experience with FAIRification projects (mine, other researchers)
    - EJP RD and European Registry Networks (ERNs)
    - LUMC and UTwente
    - Training sessions (Bring Your Own Data workshops)
    - Literature review
- Feedback about your perceptions of the method are welcome at any time

PROPOSAL: GO-PLAN – DISCLAIMER

- GO-Plan has been developed as part of my PhD work
  - Embeds experience with FAIRification projects (mine, other researchers)
    - EJP RD and European Registry Networks (ERNs)
    - LUMC
    - Training sessions (Bring Your Own Data workshops)
    - Literature review
- Feedback about your perceptions of the method are welcome at any time
- **PLEASE HELP THIS PHD STUDENT: optional questionnaire about your perceptions on using the method (10min, anonymous)**

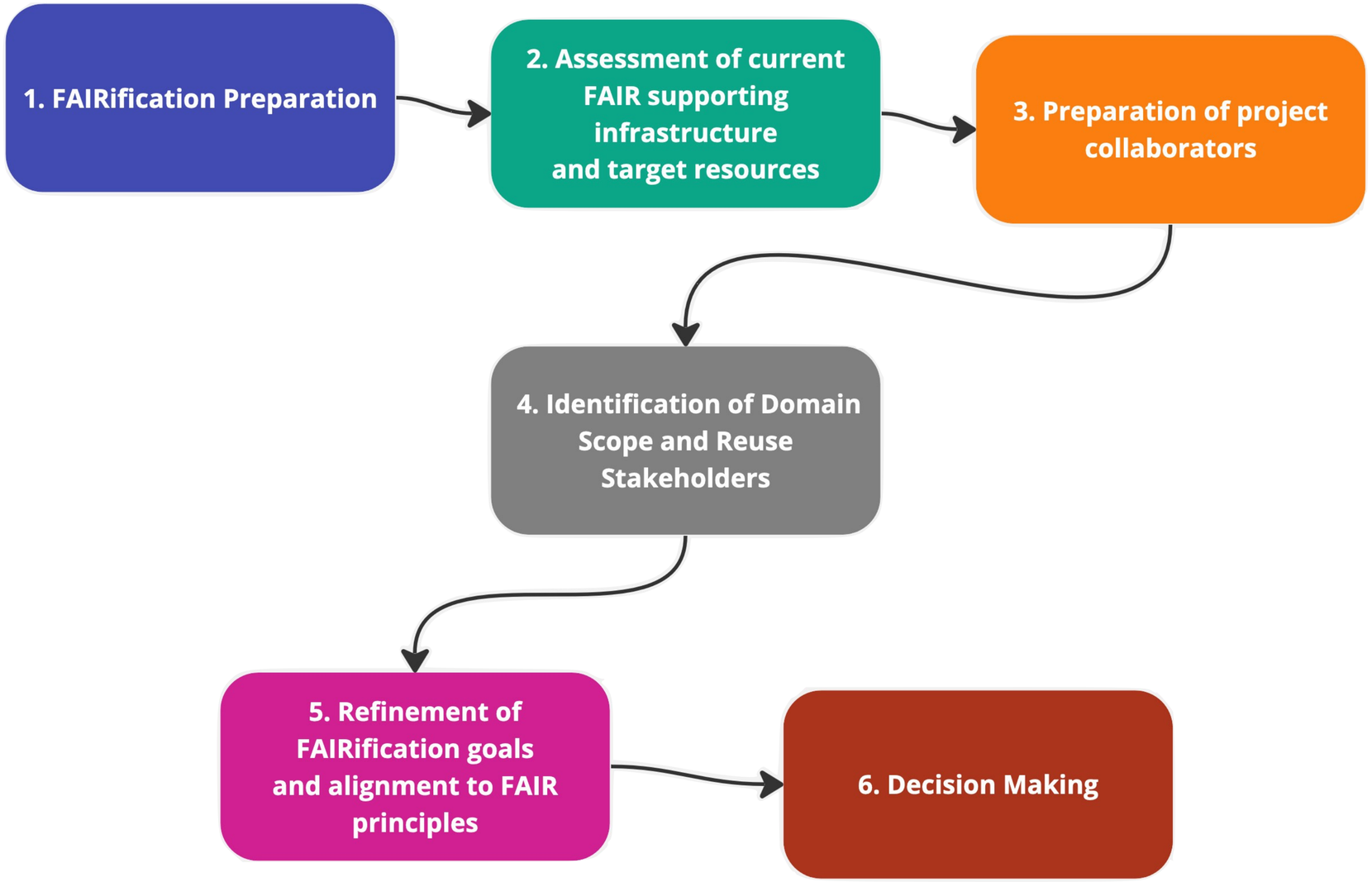
CHECKPOINT



Template

Cards

GO-PLAN: OVERVIEW



SCENARIO

- K-Woef, an animal welfare organisation, is interested in increasing dog adoption in US states with low adoption rates.
- They want to understand which dog characteristics (e.g., breed, size, sex) have higher adoption rates than others, what the differences in dog adoption rates are across the US, and what factors contribute to these differences.
- The organisation collects data from local shelters and rescue groups.
- They also make the curated data available to regional government agencies, which use the data for zoonosis control, and back to local shelters, which use the data to match dogs from the network to potential adopters in different regions.

SCENARIO

- **Problem:** Currently, K-Woef spends a lot of time and money collecting and curating data from all the local shelters and rescue organisations. Therefore, they would benefit if local shelters could make their data available in a FAIR way. The organisation has asked for your help in creating a FAIRification plan for them.

DATABASES

Dog Breeds

Dog Adoption in the US

- You will find examples of data (currently collected in a centralised way) on all breeds catalogued in the USA (dog breeds) and another on dogs available for adoption in the USA.
- These are just to give you an idea of how the data is currently collected and organised.

SUPPORTING ARTEFACTS

- Short description of the method's phases
- Template document for describing all artefacts that you produce during the method
- Links to ontologies and other repositories
- Description of the use case (scenario description)
- Link to related datasets
- Description of similar (fictitious projects)
- Link to the questionnaire

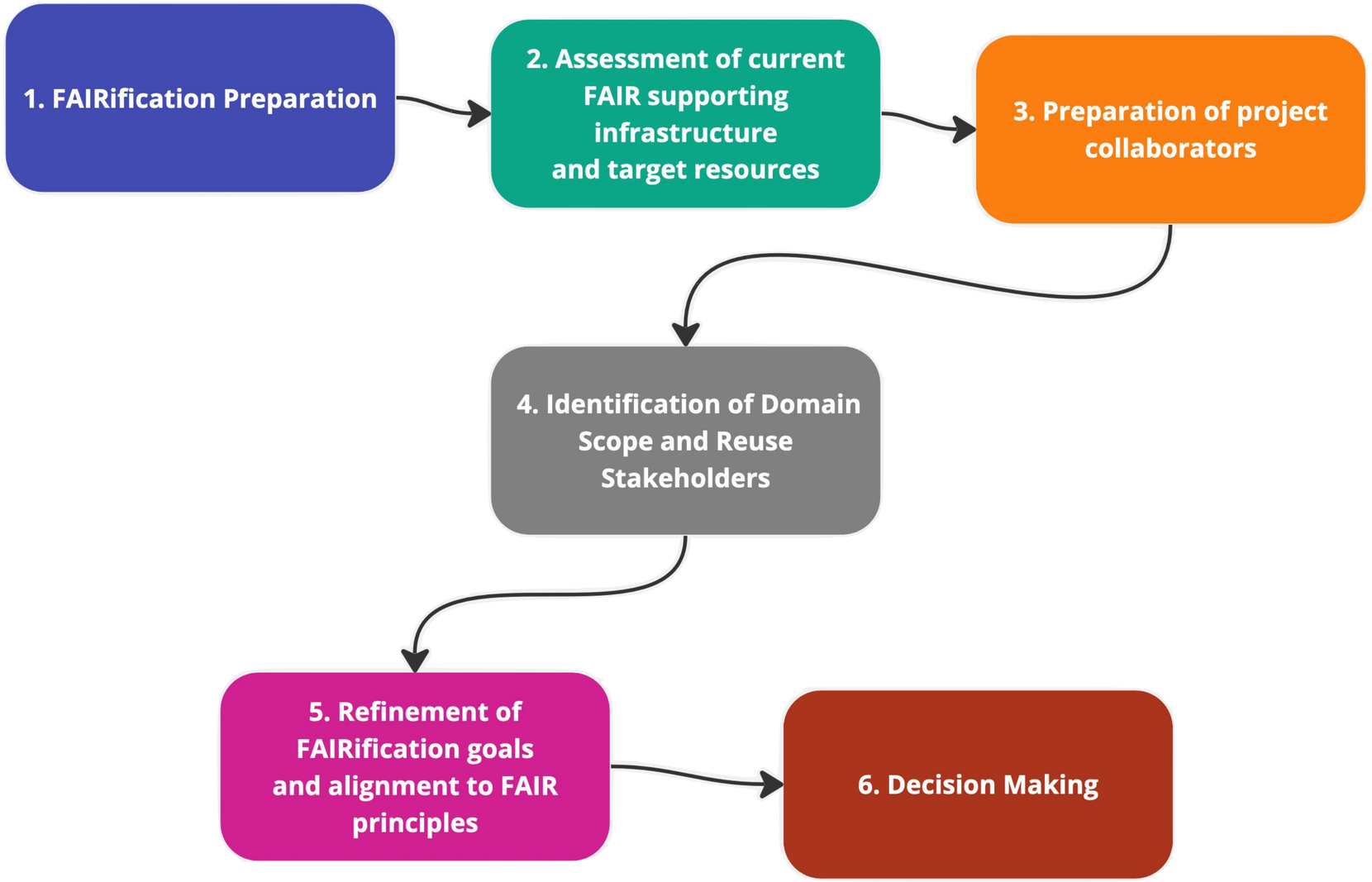
edu.nl/vcj8d

SUPPORTING ARTEFACTS

- Short description of the method's phases
- Template document for describing all artefacts that you produce during the method
- Links to ontologies and other repositories
- Description of the use case (scenario description)
- Link to related datasets
- Description of similar (fictitious projects)
- Link to the questionnaire

edu.nl/vcj8d

GO-PLAN: OVERVIEW



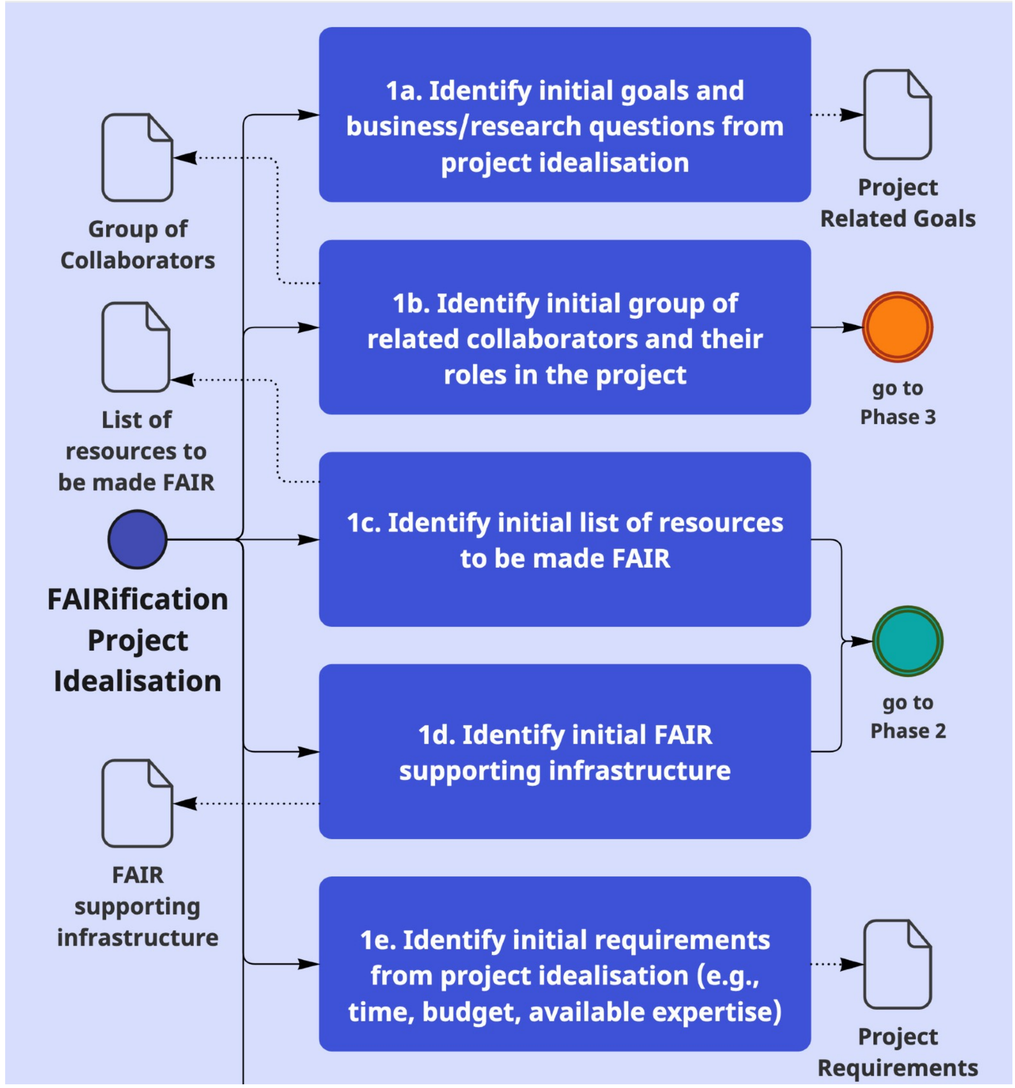
} 1st round

} 2nd round

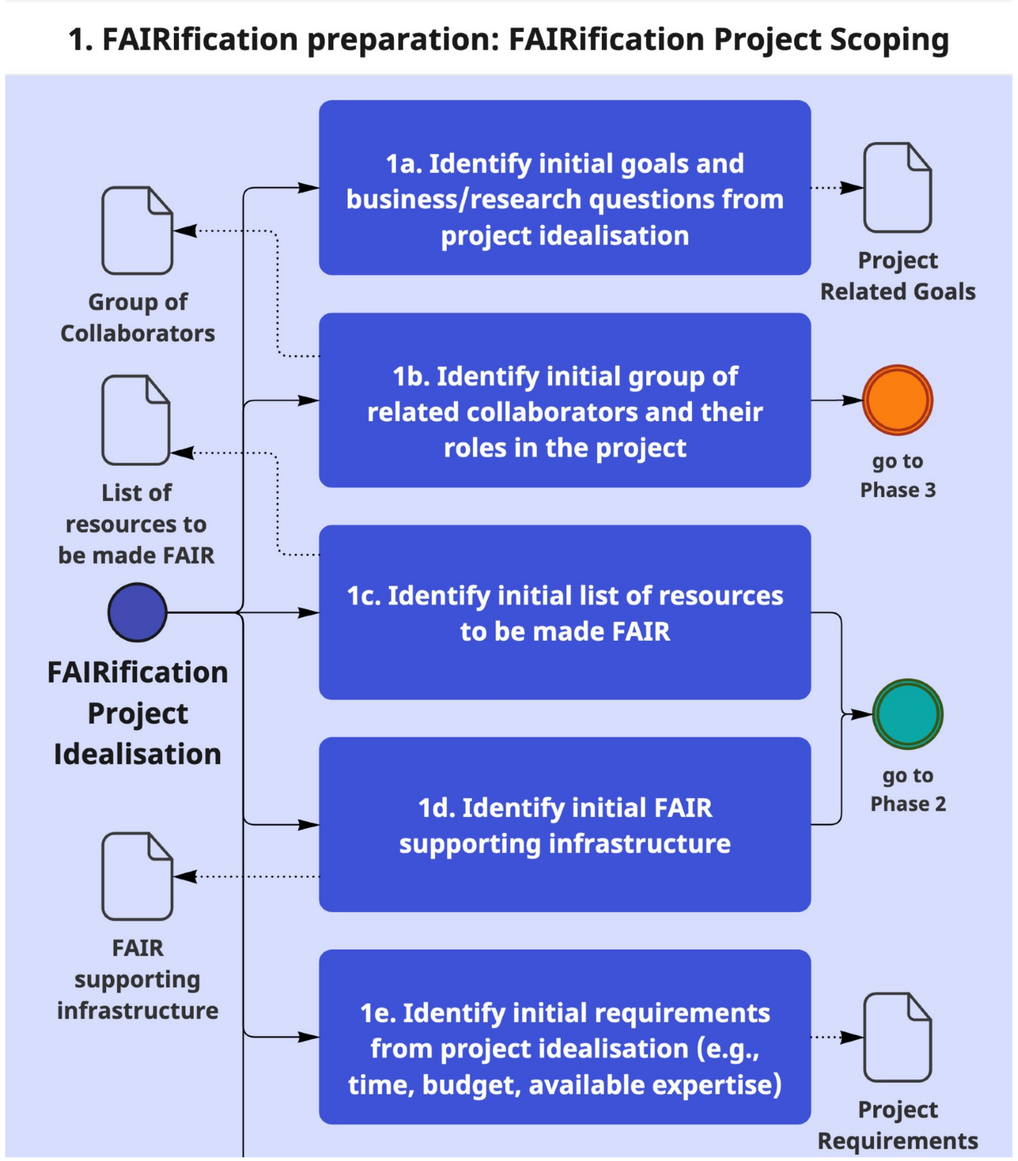
} 3rd round

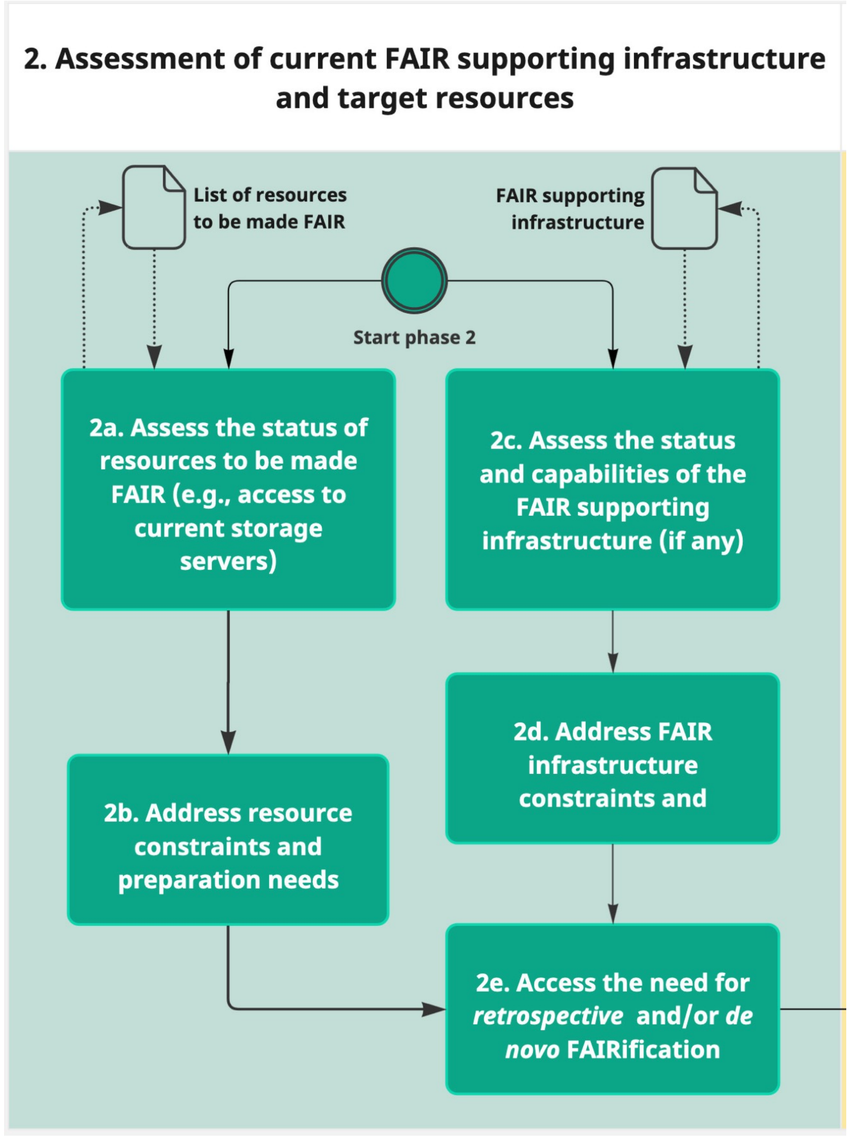
PHASES 1, 2, 3

1. FAIRification preparation: FAIRification Project Scoping

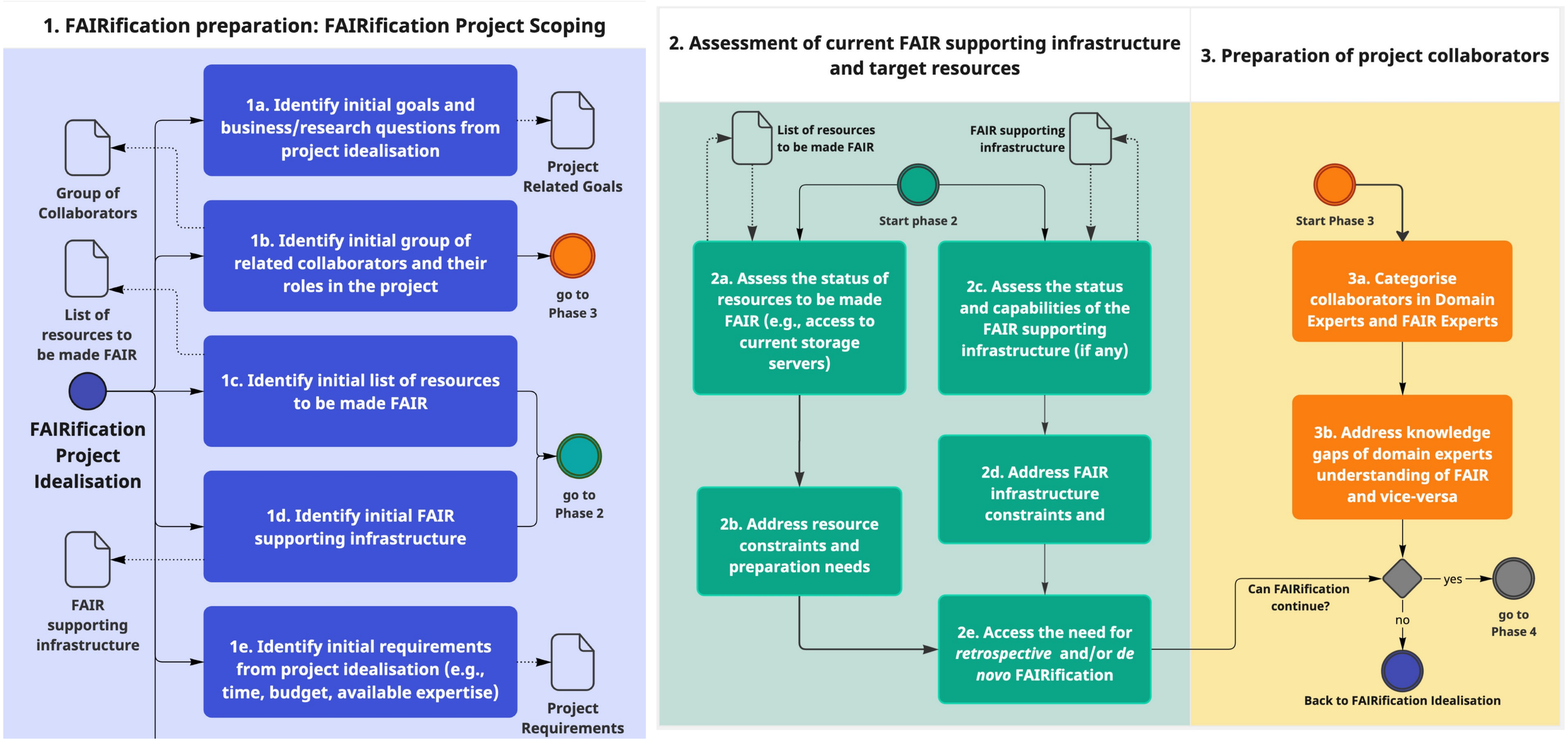


PHASES 1, 2, 3

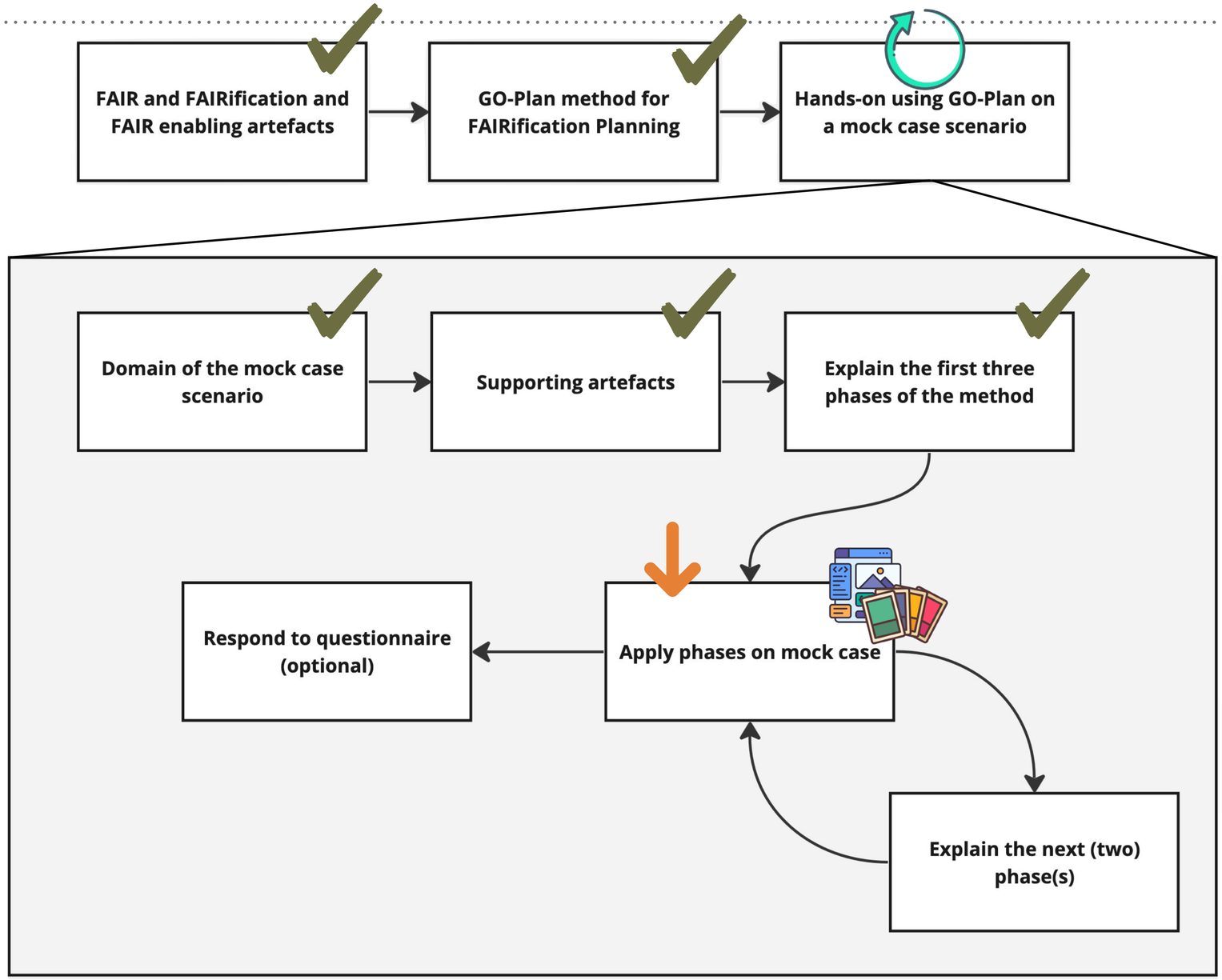

1. FAIRification preparation: FAIRification Project Scoping
1a. Identify initial goals and business/research questions from project idealisation
Project Related Goals
Group of Collaborators
1b. Identify initial group of related collaborators and their roles in the project
go to Phase 3
List of resources to be made FAIR
1c. Identify initial list of resources to be made FAIR
FAIRification Project Idealisation
go to Phase 2
1d. Identify initial FAIR supporting infrastructure
FAIR supporting infrastructure
1e. Identify initial requirements from project idealisation (e.g., time, budget, available expertise)
Project Requirements


2. Assessment of current FAIR supporting infrastructure and target resources
List of resources to be made FAIR
FAIR supporting infrastructure
Start phase 2
2a. Assess the status of resources to be made FAIR (e.g., access to current storage servers)
2c. Assess the status and capabilities of the FAIR supporting infrastructure (if any)
2b. Address resource constraints and preparation needs
2d. Address FAIR infrastructure constraints and
2e. Access the need for *retrospective* and/or *de novo* FAIRification

PHASES 1, 2, 3


1. FAIRification preparation: FAIRification Project Scoping
2. Assessment of current FAIR supporting infrastructure and target resources
3. Preparation of project collaborators
1a. Identify initial goals and business/research questions from project idealisation
Project Related Goals
Group of Collaborators
1b. Identify initial group of related collaborators and their roles in the project
go to Phase 3
List of resources to be made FAIR
1c. Identify initial list of resources to be made FAIR
FAIRification Project Idealisation
go to Phase 2
1d. Identify initial FAIR supporting infrastructure
FAIR supporting infrastructure
1e. Identify initial requirements from project idealisation (e.g., time, budget, available expertise)
Project Requirements
List of resources to be made FAIR
FAIR supporting infrastructure
Start phase 2
2a. Assess the status of resources to be made FAIR (e.g., access to current storage servers)
2c. Assess the status and capabilities of the FAIR supporting infrastructure (if any)
2b. Address resource constraints and preparation needs
2d. Address FAIR infrastructure constraints and
2e. Access the need for retrospective and/or de novo FAIRification
Start Phase 3
3a. Categorise collaborators in Domain Experts and FAIR Experts
3b. Address knowledge gaps of domain experts understanding of FAIR and vice-versa
Can FAIRification continue?
yes
go to Phase 4
no
Back to FAIRification Idealisation

AGENDA RECAP



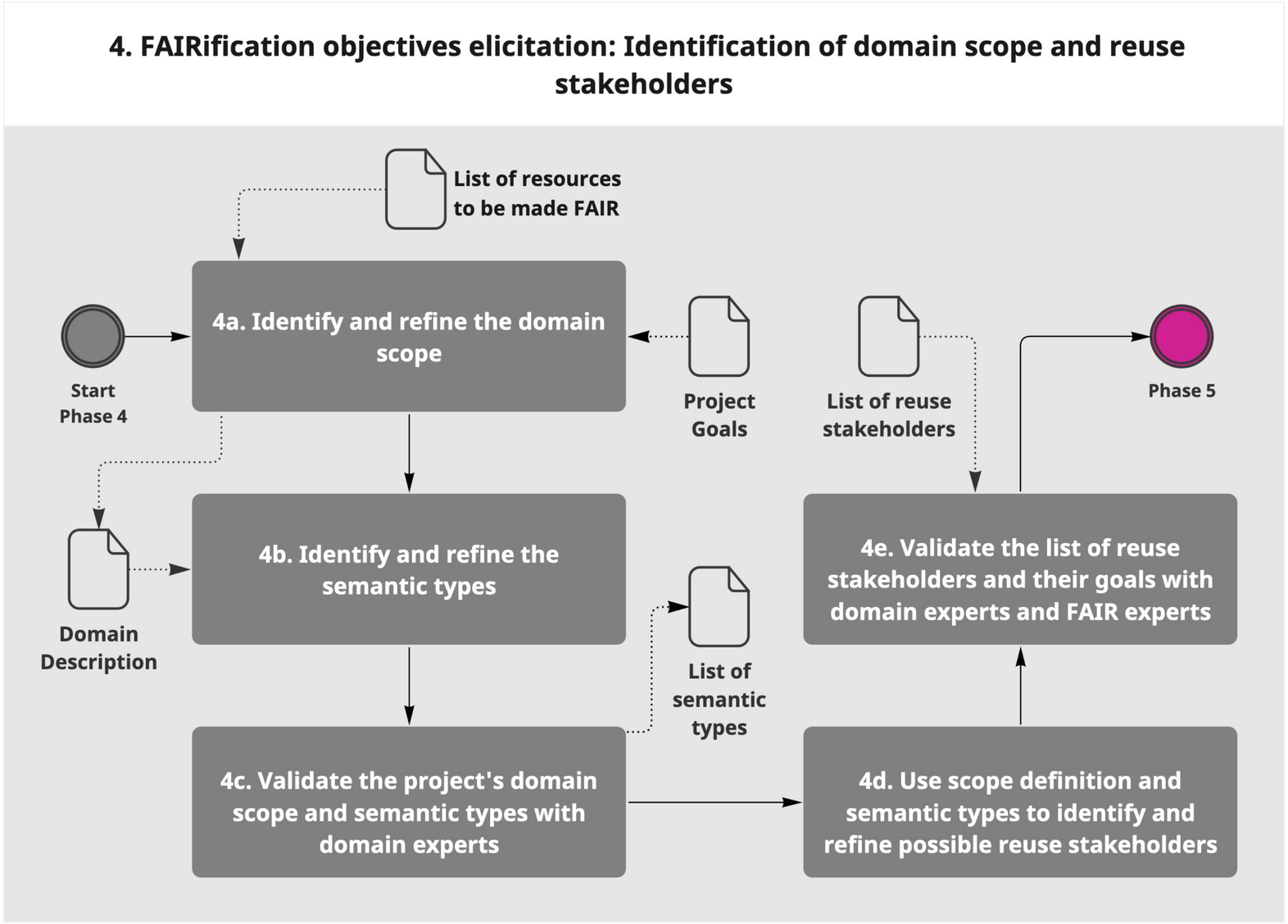
Template

Cards

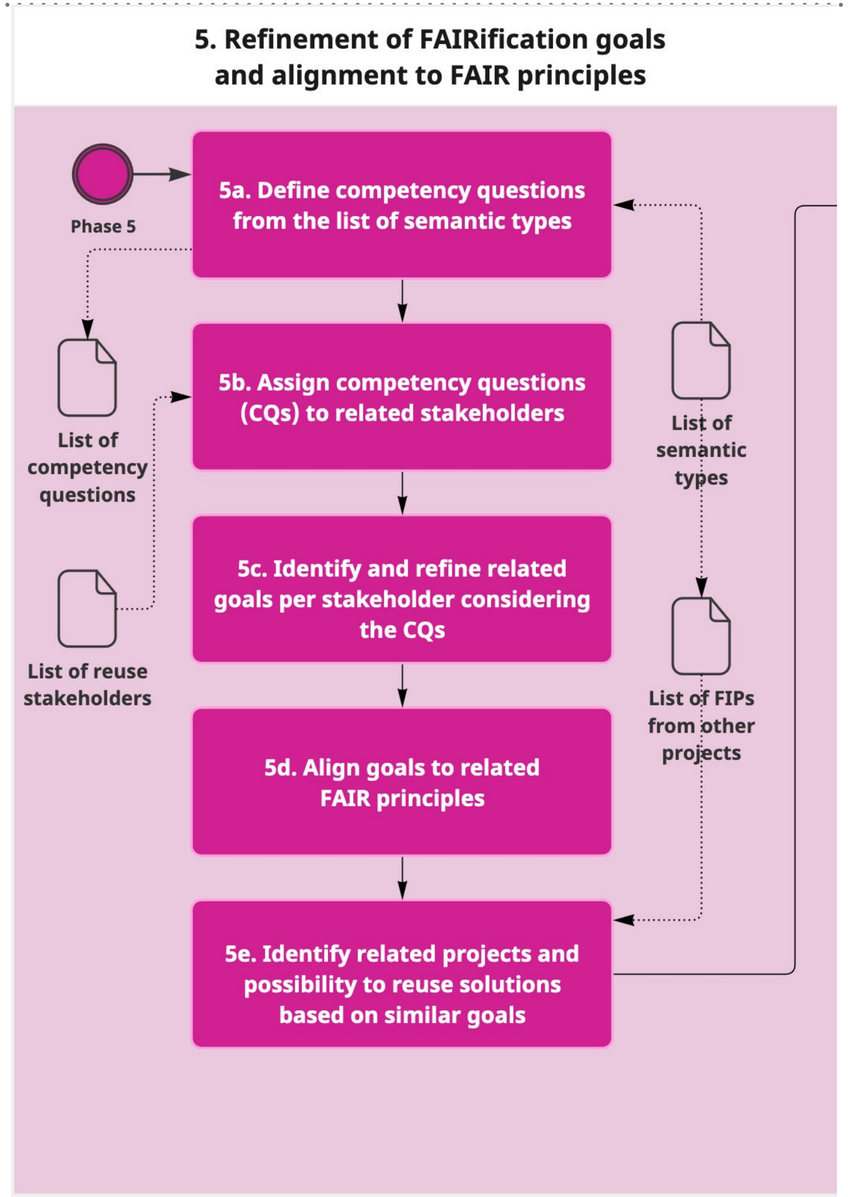
HANDS-ON

- You will apply GO-Plan on a mock case scenario, you will work in groups
- You will have 30/40 minutes to execute each phase
- Before start executing the first phase, you will blindly choose cards that describe situations about the use case (e.g., project requirements, infrastructure details)
- Each group will have a FAIR expert as a helper

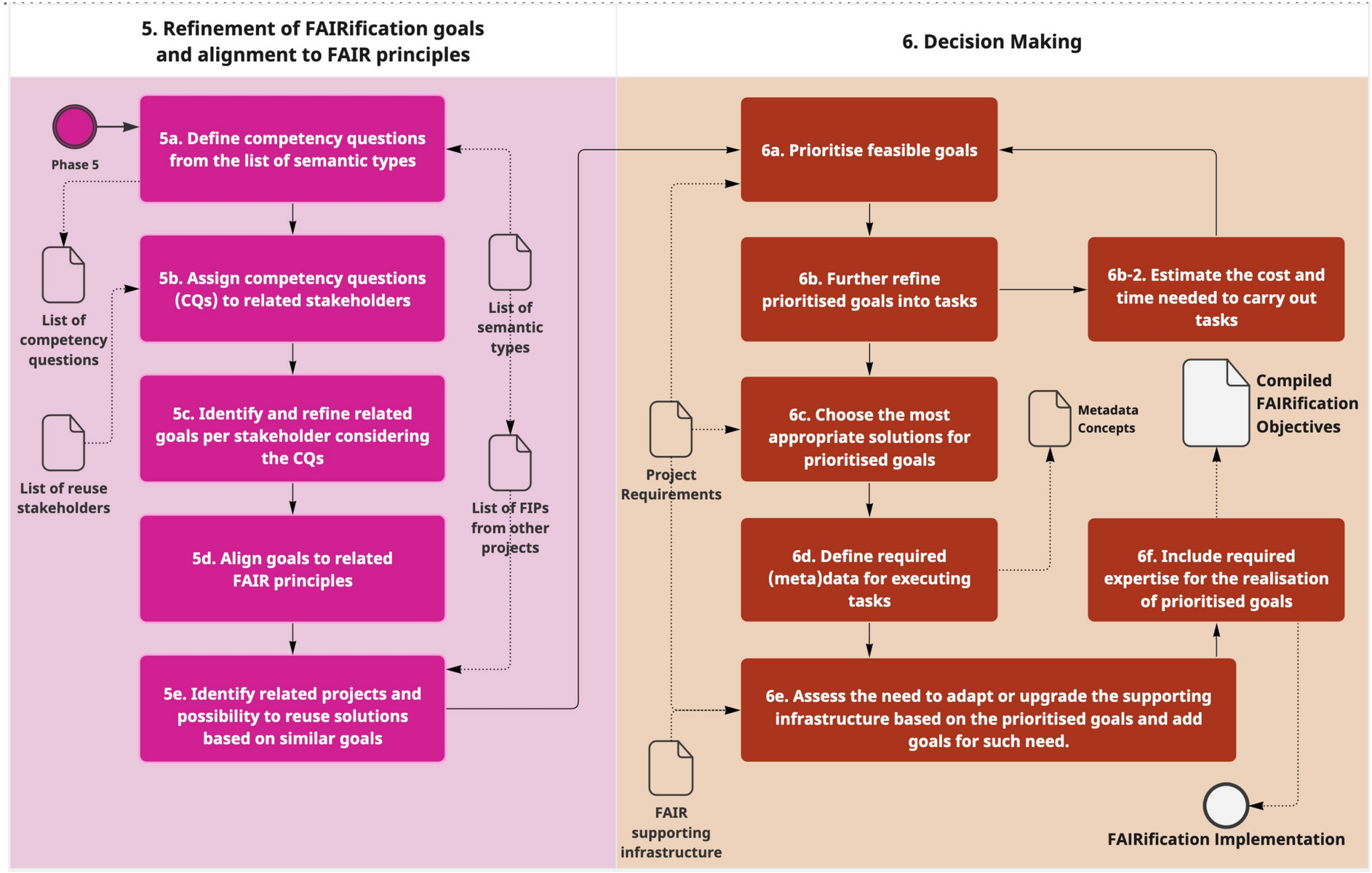
PHASE 4


4. FAIRification objectives elicitation: Identification of domain scope and reuse stakeholders
List of resources to be made FAIR
4a. Identify and refine the domain scope
Start Phase 4
Project Goals
List of reuse stakeholders
Phase 5
4b. Identify and refine the semantic types
Domain Description
4e. Validate the list of reuse stakeholders and their goals with domain experts and FAIR experts
List of semantic types
4c. Validate the project's domain scope and semantic types with domain experts
4d. Use scope definition and semantic types to identify and refine possible reuse stakeholders

PHASES 5 AND 6



PHASES 5 AND 6



QUESTIONNAIRE

UNIVERSITY OF TWENTE.

Leids Universitair Medisch Centrum

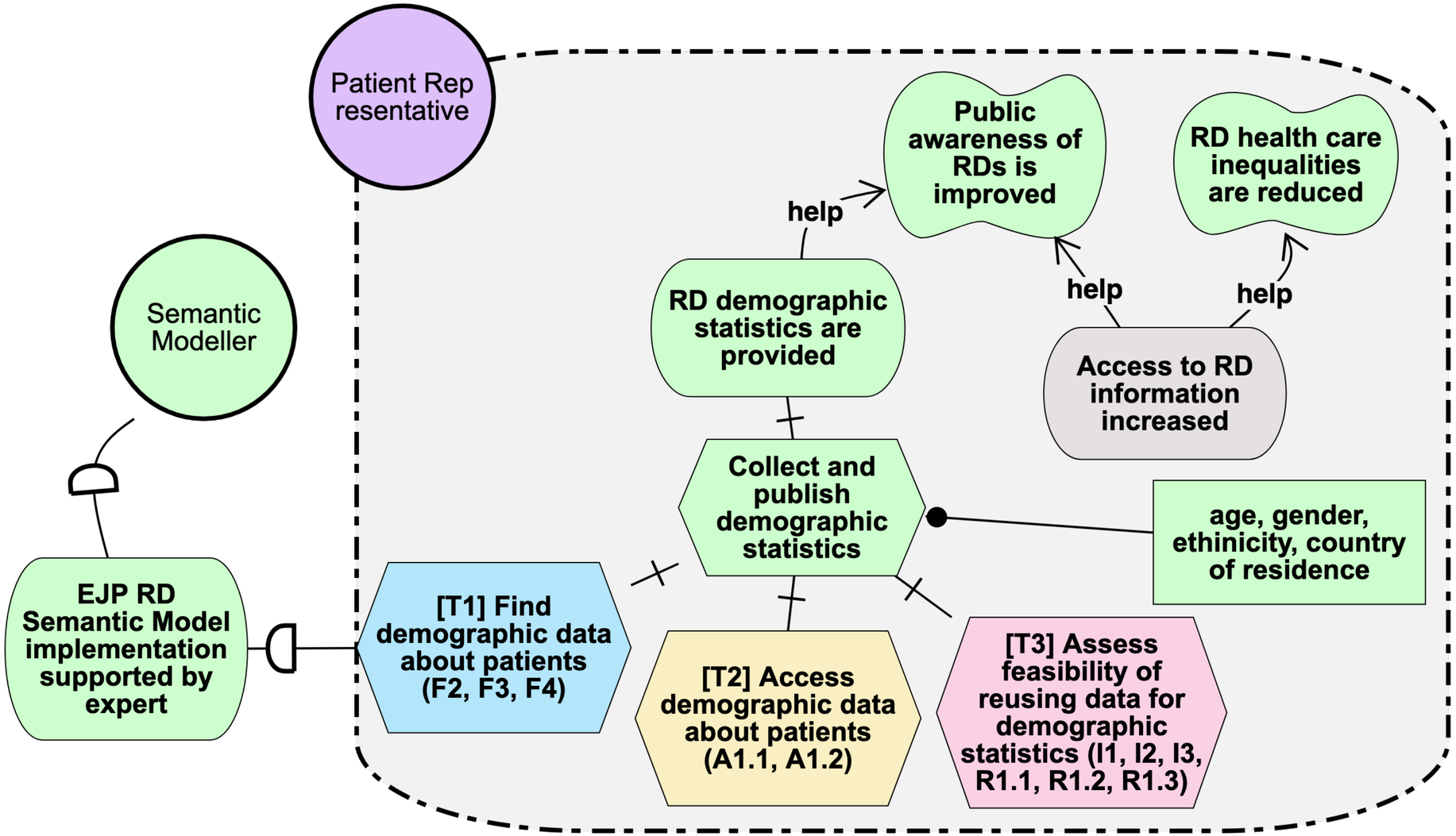
EUROPEAN JOINT PROGRAMME RARE DISEASES

edu.nl/cv9r4

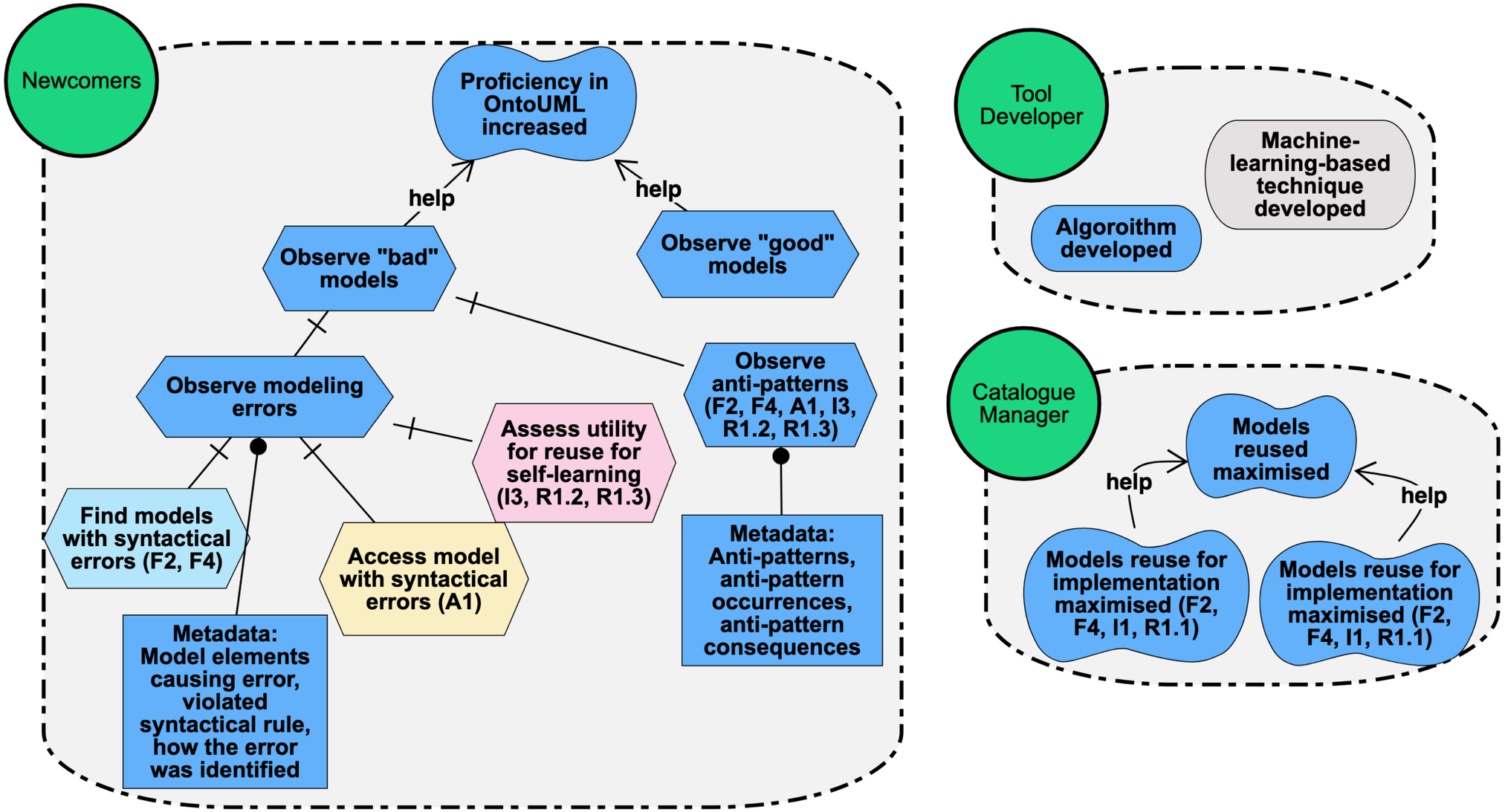
Dankjewel!

BACK-UP SLIDES

GOAL MODELLING



GOAL MODELLING



FAIR ENABLING ARTEFACTS: FAIR DATA POINT

- Which FAIR principles does the FAIR Data Point address?
  - **Findability**: publishes metadata in a machine readable format, using other FAIR vocabularies
  - **Accessibility**: uniform, open way of accessing metadata and data
    - Supports defining access conditions
  - **Interoperability**: only for metadata (if you use controlled vocabularies, ontologies..)
  - **Reusability**: supports publishing rich metadata