

GO FAIR US / EarthCube M4M Pilot Workshop

Coordinating Team:

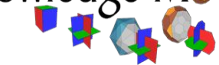
Melissa Cragin, John Graybeal,
Nancy Hoebelheinrich

8 & 9 September 2022

NOTE: This meeting will be recorded



Knowledge Motifs LLC



Mapping sensible data relationships

1 - Welcome - Workshop - Summary



Image by [Petra](#) from [Pixabay](#)

Welcome - Day 1 - Introductions

Your M4M Facilitators

Melissa Cragin



Melissa Cragin is Chief Strategist for Data Initiatives in the Research Data Services division at the San Diego Supercomputer Center (SDSC), UCSD. Prior to joining SDSC, Melissa was the Executive Director of the Midwest Big Data Hub, based at the National Center for Supercomputing Applications (NCSA) at UIUC. Previously, she served in the Office of the Assistant Director, Directorate of Biological Sciences at the National Science Foundation (NSF), and worked on data policy and public access. At SDSC, Melissa works on projects to improve data access and use, and foster the development of the national data infrastructure ecosystem and related policy, including the EarthCube Coordination Office and GO FAIR US.

John Graybeal



John Graybeal is a Technical Program Manager at the Stanford Center for Biomedical Informatics Research (BMIR), the home of Protégé and WebProtégé, BioPortal and OntoPortal, and CEDAR. He helped teach an inaugural GO FAIR M4M workshop and related training on vocabulary development. Previously he was a Principal Investigator of the Marine Metadata Interoperability Project. Most recently he has been heavily involved in metadata and harmonization efforts on behalf of the NIH Rapid Acceleration of Diagnostics RADx Data Hub, and supports the NIH Human Biomolecular Atlas Program (HuBMAP) project, as well as contributing to the Simple Standard for Sharing Ontological Mappings (SSSOM) project.

Nancy Hoebelheinrich



Nancy J. Hoebelheinrich is a digital library, archives and data repository consultant specializing in geoscience, cultural heritage, and educational data. Nancy has previously worked in academic, public and legal libraries and data repositories, and non-profit research institutes. She is currently founder and co-owner of Knowledge Motifs LLC. Nancy has been active most recently in a number of information and educational technology efforts including those of the Earth Science Information Partners (ESIP), RDA's ETHRD IG, and EarthCube. She is a member of the core GO FAIR US team leading the Ambassadors Pillar.



Welcome - Day 1

- Rules of Engagement
- Community Participation Guidelines
- Introductions
- Conceptual scope
 - Why these workshops?
 - Why you?

And, On to the show!

Common notes: <https://bit.ly/EarthCubeM4M>

Slides: <https://bit.ly/ECM4MSlides>



Welcome - Day 1

Rules of Engagement

- Participants - active engagement
 - This is an interactive workshop
 - Focus questions on process and mechanics
 - Contribute to Common Doc (capture questions and feedback)
- Observers - this is a passive role - you are welcome to shift status now
 - Mics and Video off throughout workshop
 - Chat is not available, but organizers welcome email afterward

Common notes: <https://bit.ly/EarthCubeM4M>



Welcome - Day 1

Community Participation Guidelines

Expectation of collegial behavior and kind interactions.

SDSC and EarthCube are committed to providing harassment-free environments for our events and collaborative workspaces. Harassment in any form is unacceptable.

To address concerns during the event, contact Melissa - mcragin@sdsc.edu

If you believe you have experienced or are experiencing unacceptable behavior (Section 3), please submit a report [HERE](#).



Welcome - Day 1

Introductions

- Kerstin Lehnert, Lamont-Doherty Earth Observatory
- Steve Richard, US Geoscience Information Network (USGIN)
- Carina Bennett, University of Arizona, OSIRIS REx
- Taylor Johnson, University of Arizona, SAMIS
- Joshua Kantarges, OSIRIS REx

Workshop Background & Motivation

- Present advanced tools and processes to help EarthCube members manage metadata
- Advertise these techniques to the wider science community
- Evaluate the value of these approaches for EarthCube science metadata, and
 - to consider potential uptake in more communities in the U.S.

Conceptual Scope

Why you?

- Opportunity to build future-looking developments for global open science
- Opportunity for AstroMat
- Your Background/Expertise
- Relation to work that your community has already started

UA-DSD-XXX
Effective Date: TBD
Revision – DRAFT 0.1

Origins Spectral Interpretation **Resource**
Identification Security-Regolith Explorer
(OSIRIS-REx) Project – Sample Analysis Phase

OSIRIS-REx Sample Analysis Image Format
Data Standard Document

UA-DSD-XXX, Rev. DRAFT 0.1

April 1, 2022
Updated 2022-06-21



[UA-DSD OSIRIS-REx-SA Image Format DRAFT](#)

Expectations for our Work Together

- If nothing else, today is about controlled vocabulary and schema
- Implementation of the FAIR Guiding Principles is in early stages.
- In the M4M workshop, we are often building an airplane while flying it.
- Consider yourself co-developers!



<https://www.youtube.com/watch?v=L2zqTYgcpfg>

Courtesy of Erik Schultes

Agenda – Day One, Sept 8, 2022

#	Duration	Agenda Item
	4'00" total	Build controlled vocabularies
1	0:10	Welcome: Self-introductions of presenters and participants Workshop: Conceptual scope, motivations and practicalities Summary: Short introduction about the workshop
2	0:30 GO	Introduction to Automating FAIR using linked, machine-actionable (meta)data
3	0.45 GO	How to build machine-actionable controlled vocabularies
	0.15	Break
4	0:15 GO	How to build domain-specific controlled vocabularies
4A	1:30 0:30+ each	Exercises (your choice!) in building domain-specific controlled vocabularies
4B	0:30	4A: Pick one or more vocabularies to improve
4C	0:50	4B: Create a new vocabulary (sheet) for your domain
	0:20	4C: Set up a workflow with GitHub & BioPortal
		Pose / discuss advanced vocabulary topics
5	0.15 GO	Round-off and what happens next? Summary: A summary of the first day of M4M and a prelude to day 2 of M4M.

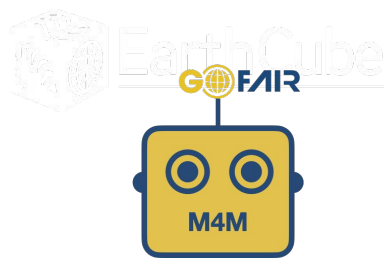
Agenda - Day Two, Sept 9, 2022

#	Duration	Agenda Item
	4' total	Build a machine-actionable metadata template
0	0:10 GO	Review and issues from Day 1
1	0:45 GO	How to build machine-actionable metadata templates
2	0:20 GO 0:40 GO	Step 1: Make a copy of existing template (e.g., Generic Dataset Metadata Template) Step 2: Create metadata fields and assign controlled vocabularies
	0:15	Break
3	0:30 GO 0:20 GO 0:10 GO	Step 3: Add created or new fields to metadata elements and add RDF properties Step 4: Structure new or copied template by composing and replacing elements Step 5: Make machine-actionable metadata by filling out the form
4	0:20 GO	FAIR Orchestration and Integration with other systems
5	0:15 GO	Questions and Discussion
6	0:15	Round-off and what happens next? Summary: What we did during M4M and discussion of next steps.

2 - Introduction to Automating FAIR using linked, machine-actionable (meta)data



Image by [Peter H](#) from [Pixabay](#)



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.



M4M workshops overview

WorkshopDate	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	PreClinicalTrials	pre-registration form	GO FAIR
M4M.4	April-Sept 2020	VODAN Africa	Metadata for the FDP	Philips 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 – Recordings	ZonMw
M4M.9	November 2020	COVID-19 Program	Prognosis / Risk assessments	ZonMw
M4M.10	November 2020	COVID-19 Program	Virus / Immunology / Molecular – Recordings	ZonMw
M4M.11	November 2020	COVID-19 Program	Organisational / Process related – Recordings	ZonMw
M4M.12	November 2020	COVID-19 Program	Socio-economic / Behavioral – Recordings	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 – Recordings	EUHorizon2020
M4M.19	December 2021	NICEST2	Climate data	EOSC Nordic



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.



<https://www.gofairfoundation.org/m4m/>



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

Total citations Cited by 7099
June 19 2022



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)

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)

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)

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

ORCID
DOI

Separate but linked
data & metadata

Search engine

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

FAIR versus Open
See R1.1

See F3

To be Interoperable:

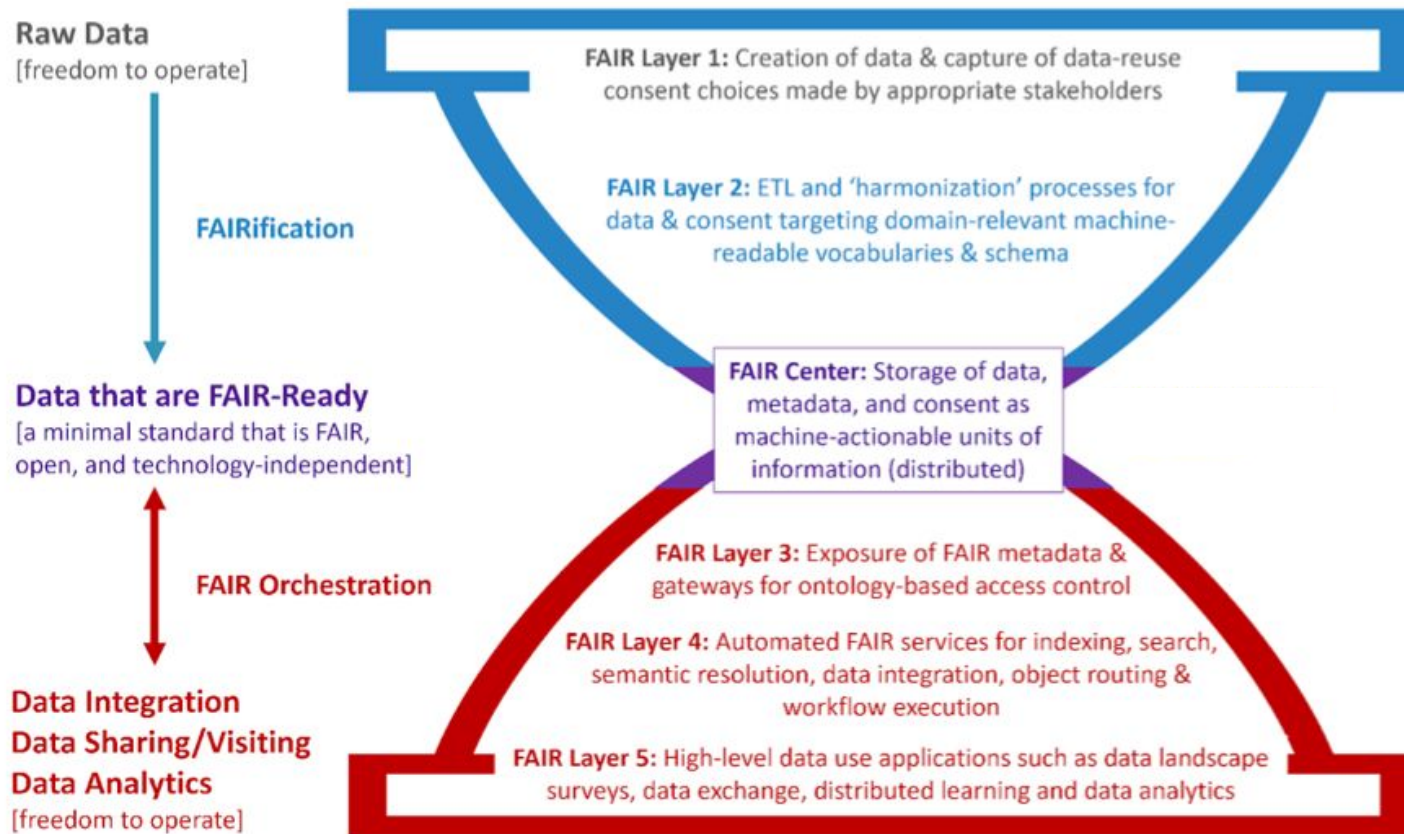
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

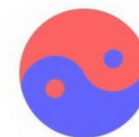
- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

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

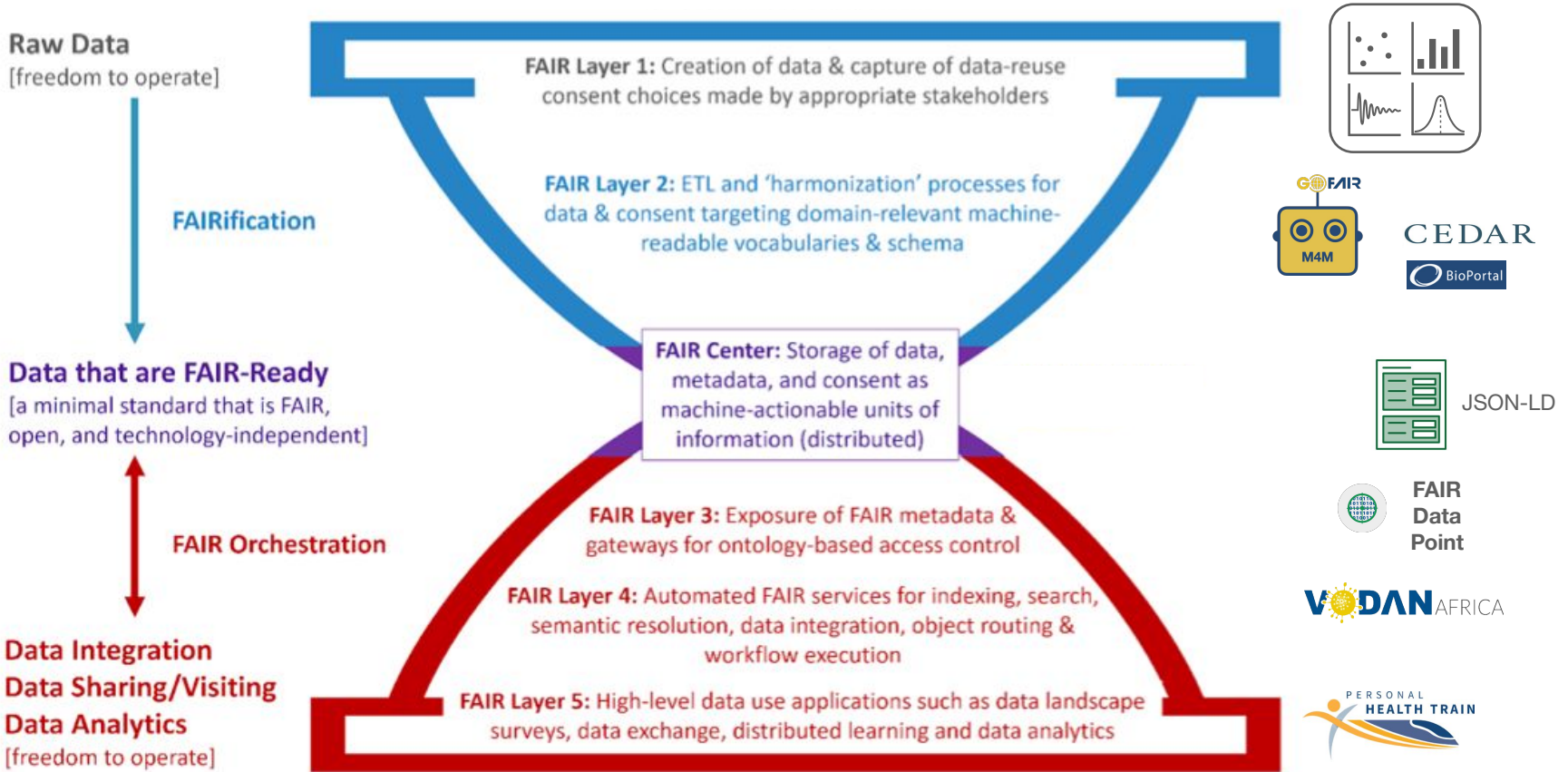
FAIR Hourglass



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



FAIR Hourglass



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



M4M FAIRification roadmap

Courtesy of Barbara Magagna



M₄M FAIRification roadmap



M4M FAIRification roadmap

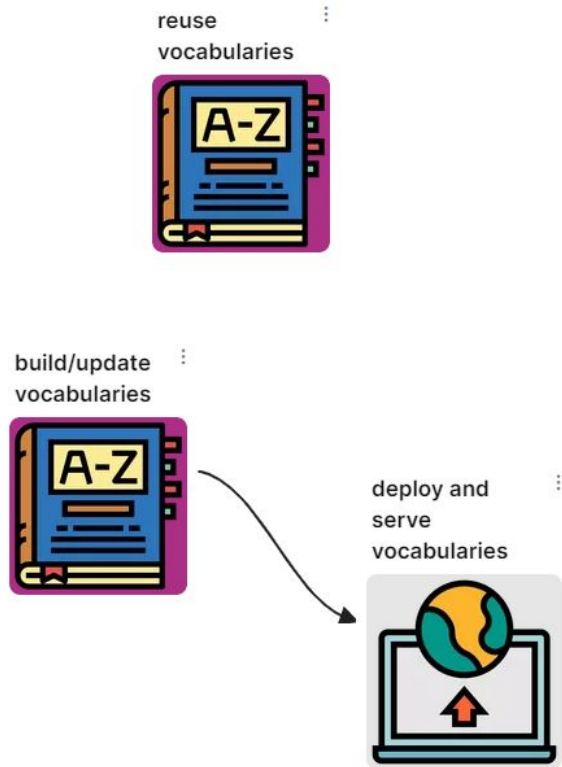
reuse
vocabularies



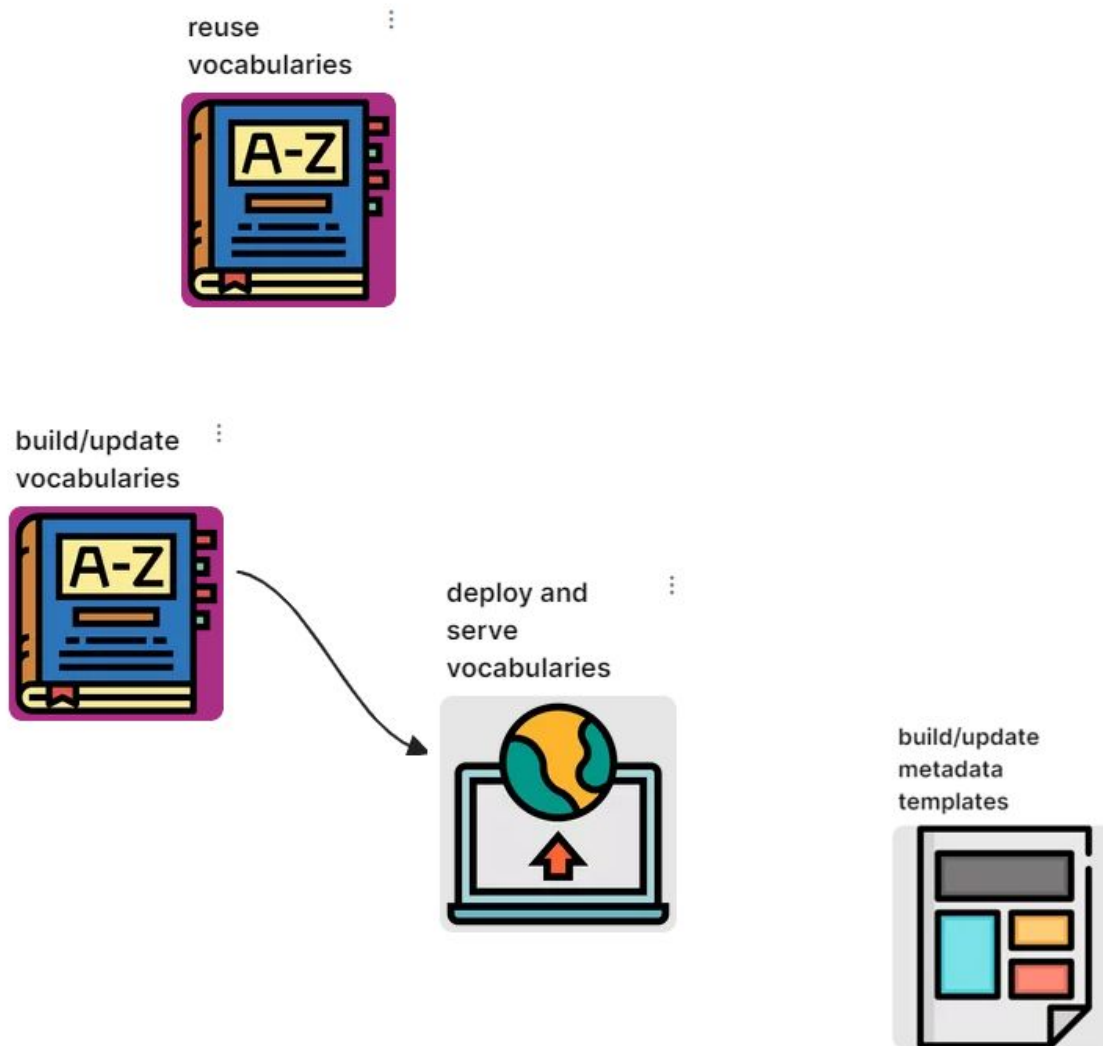
build/update
vocabularies



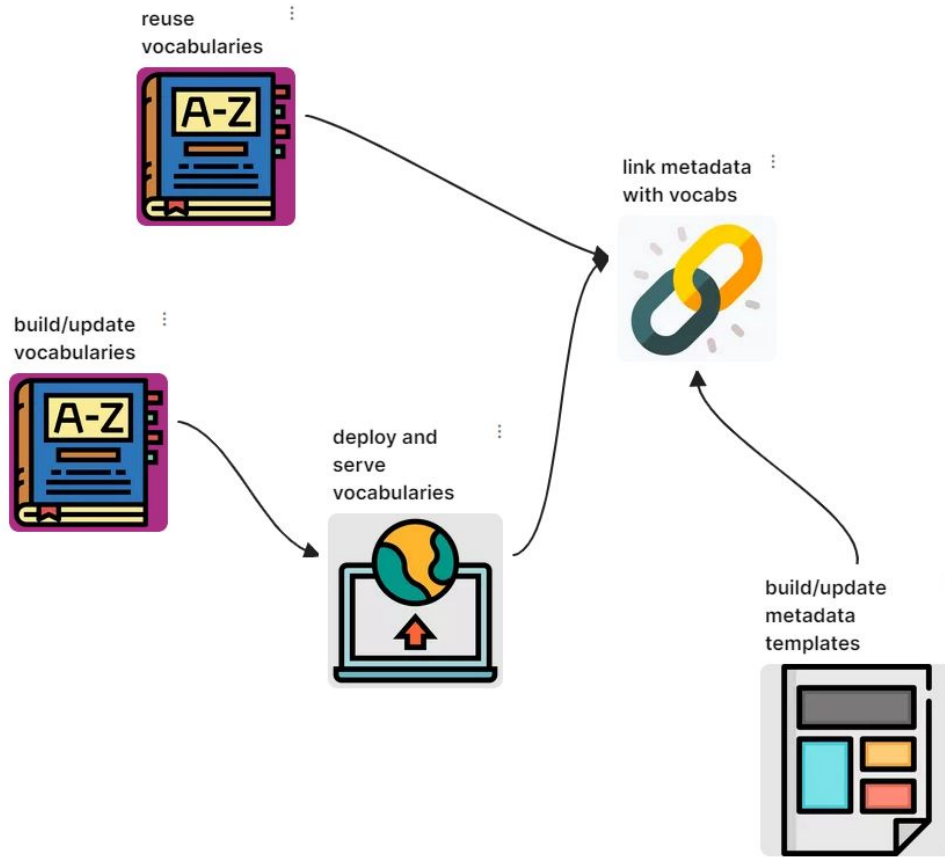
M₄M FAIRification roadmap



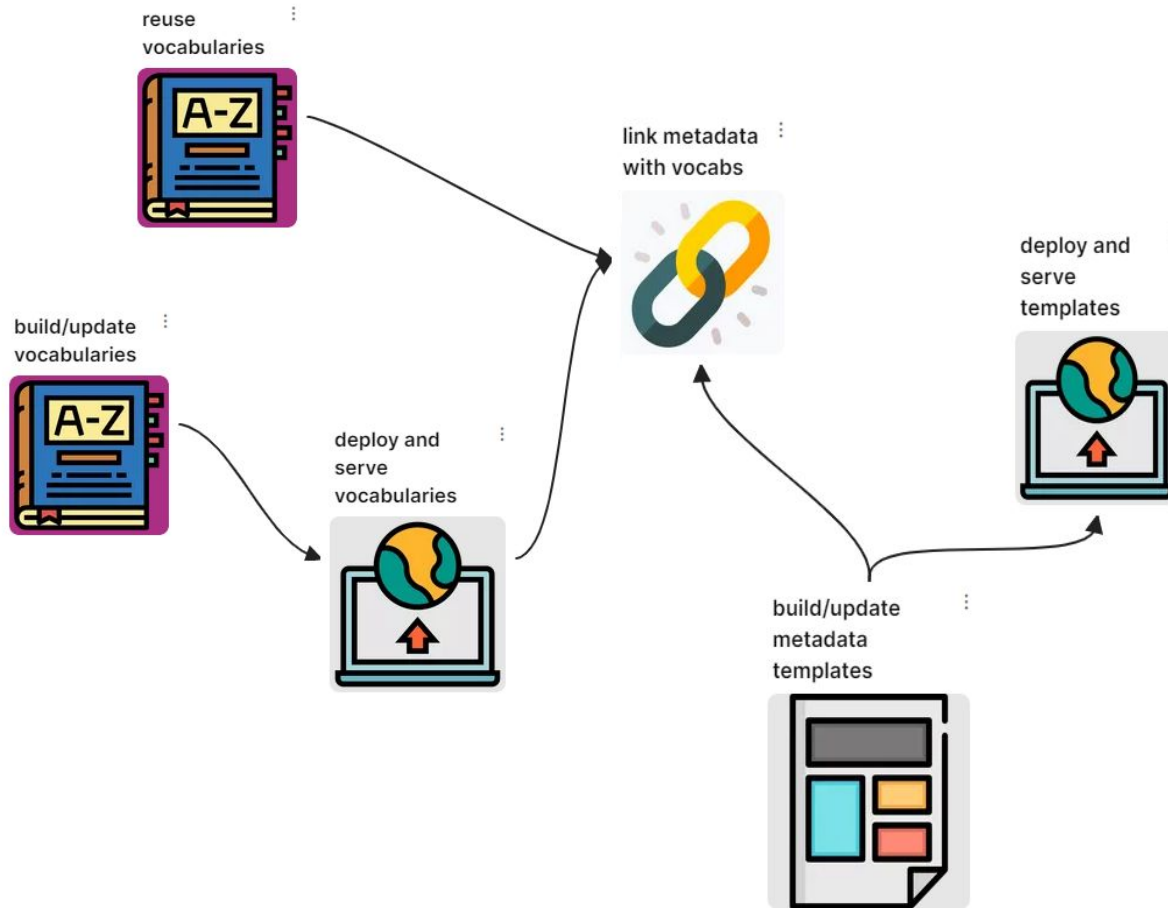
M₄M FAIRification roadmap



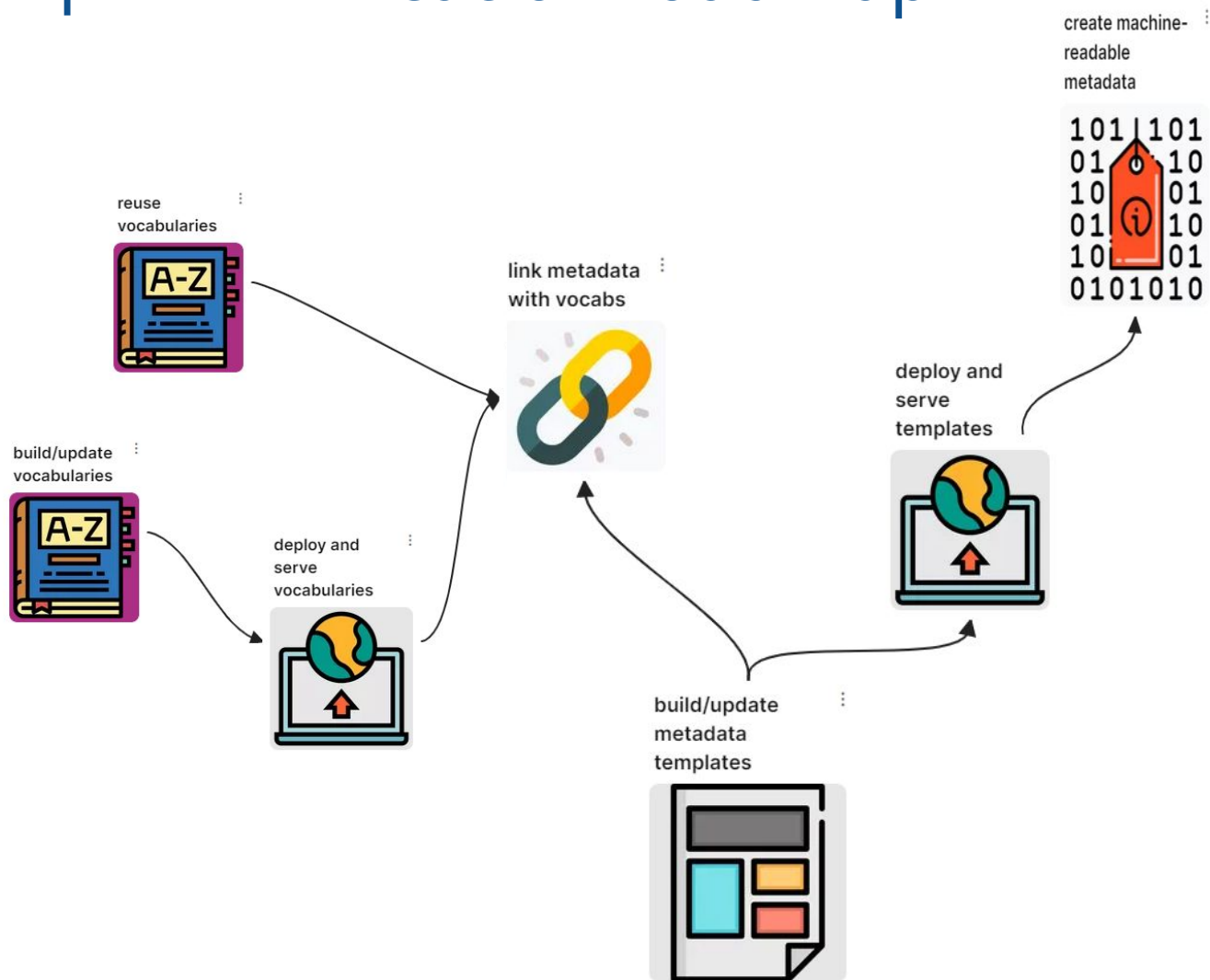
M₄M FAIRification roadmap



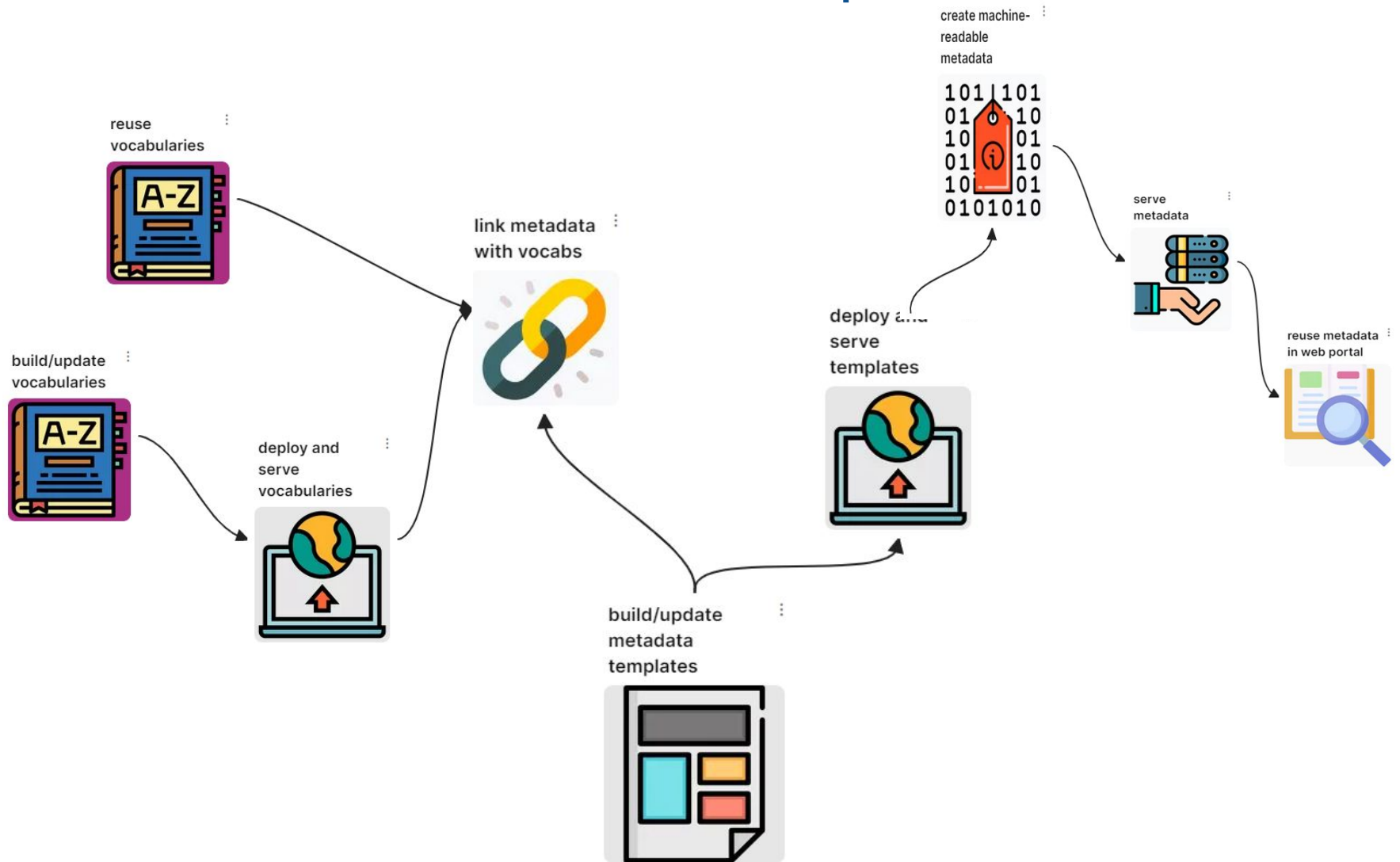
M₄M FAIRification roadmap



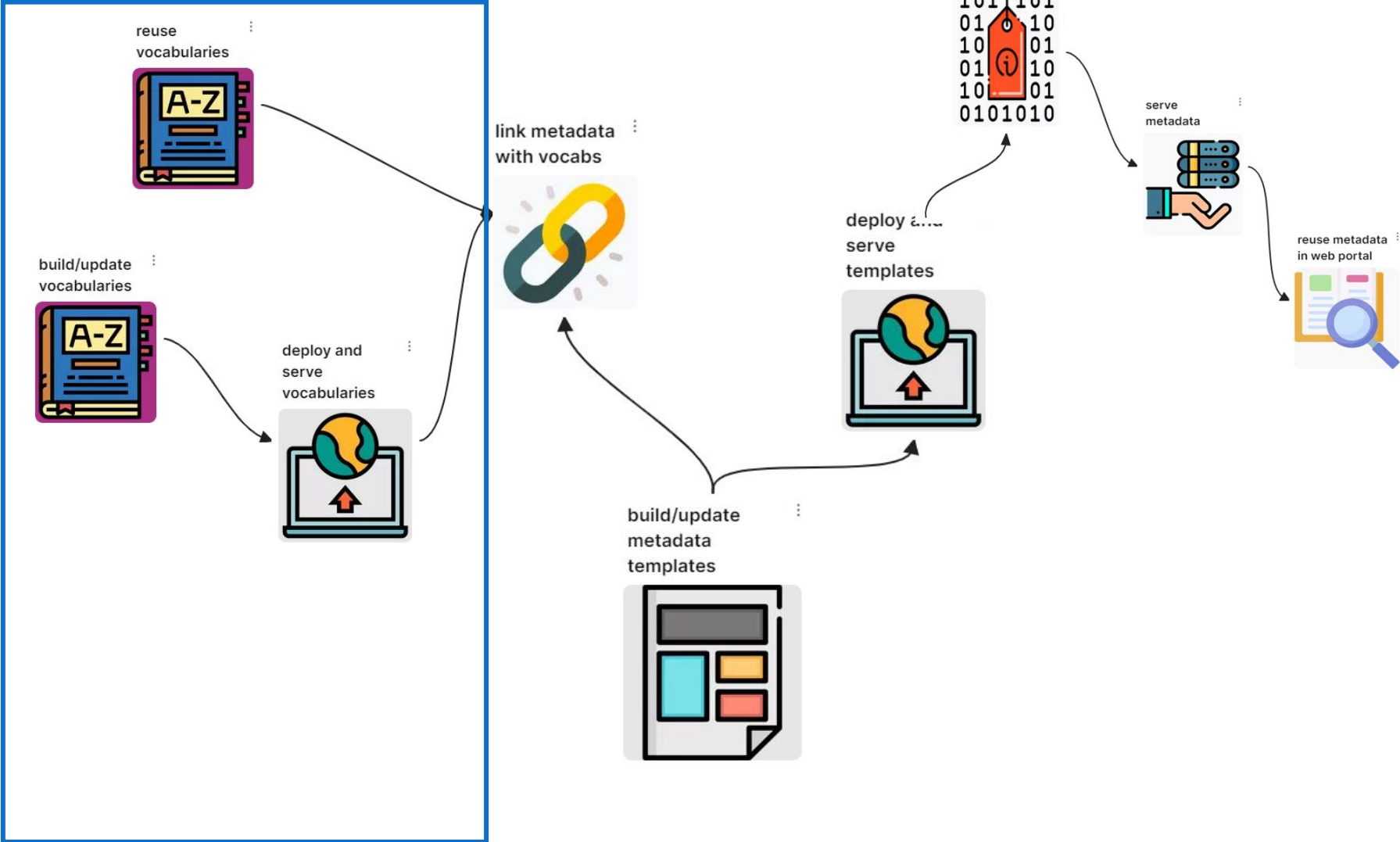
M4M FAIRification roadmap



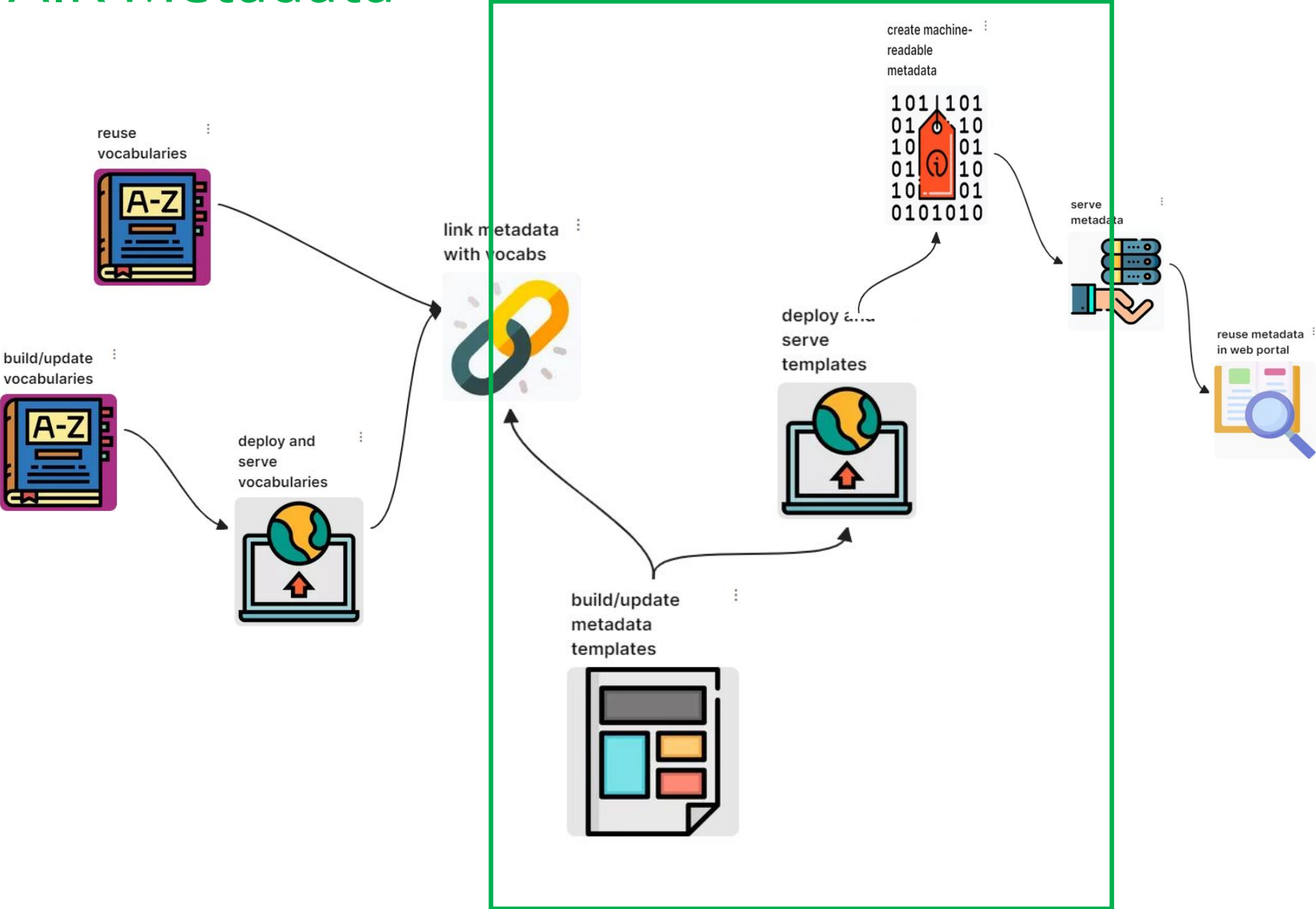
M₄M FAIRification roadmap



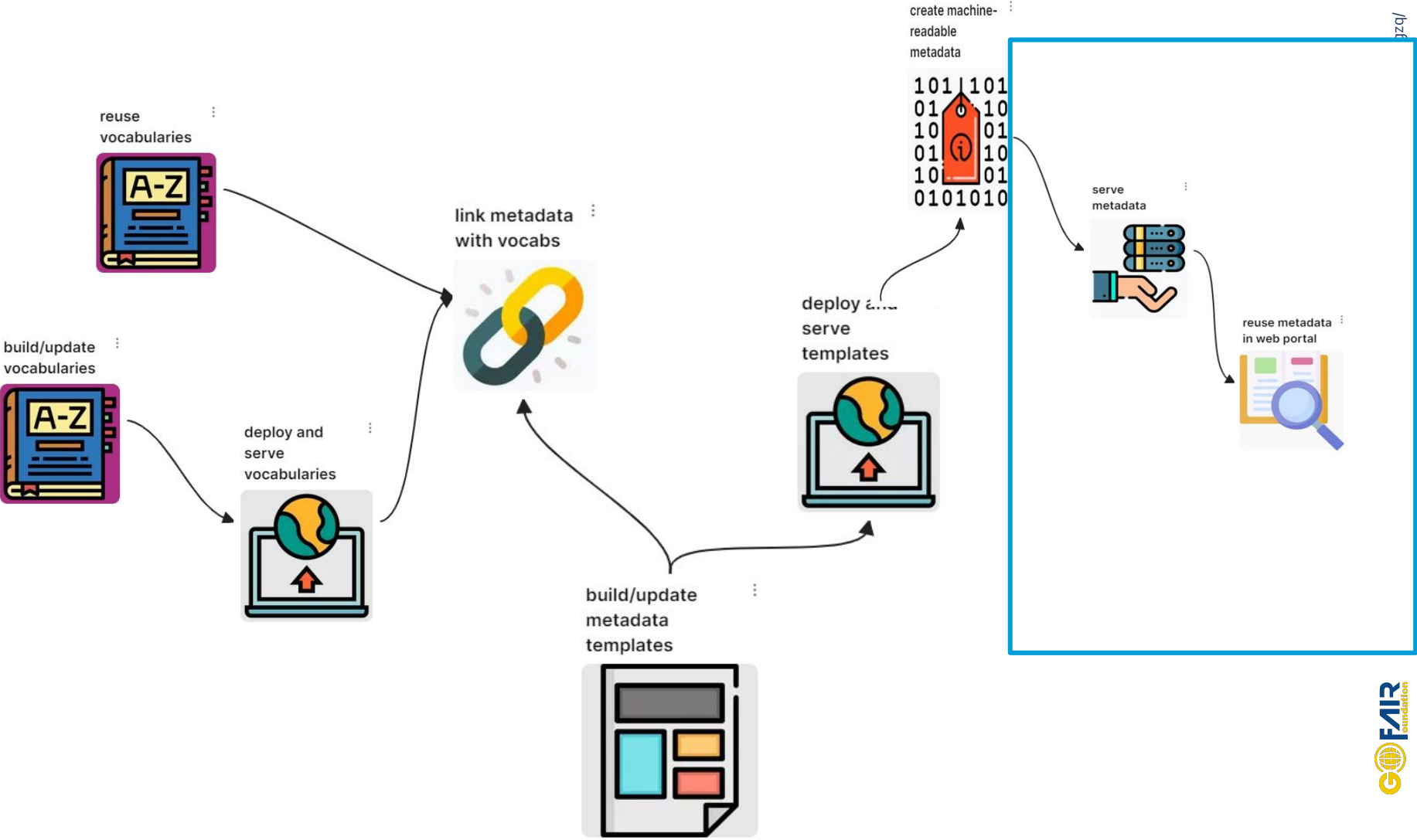
FAIR Vocabularies



FAIR Metadata



FAIR Orchestration



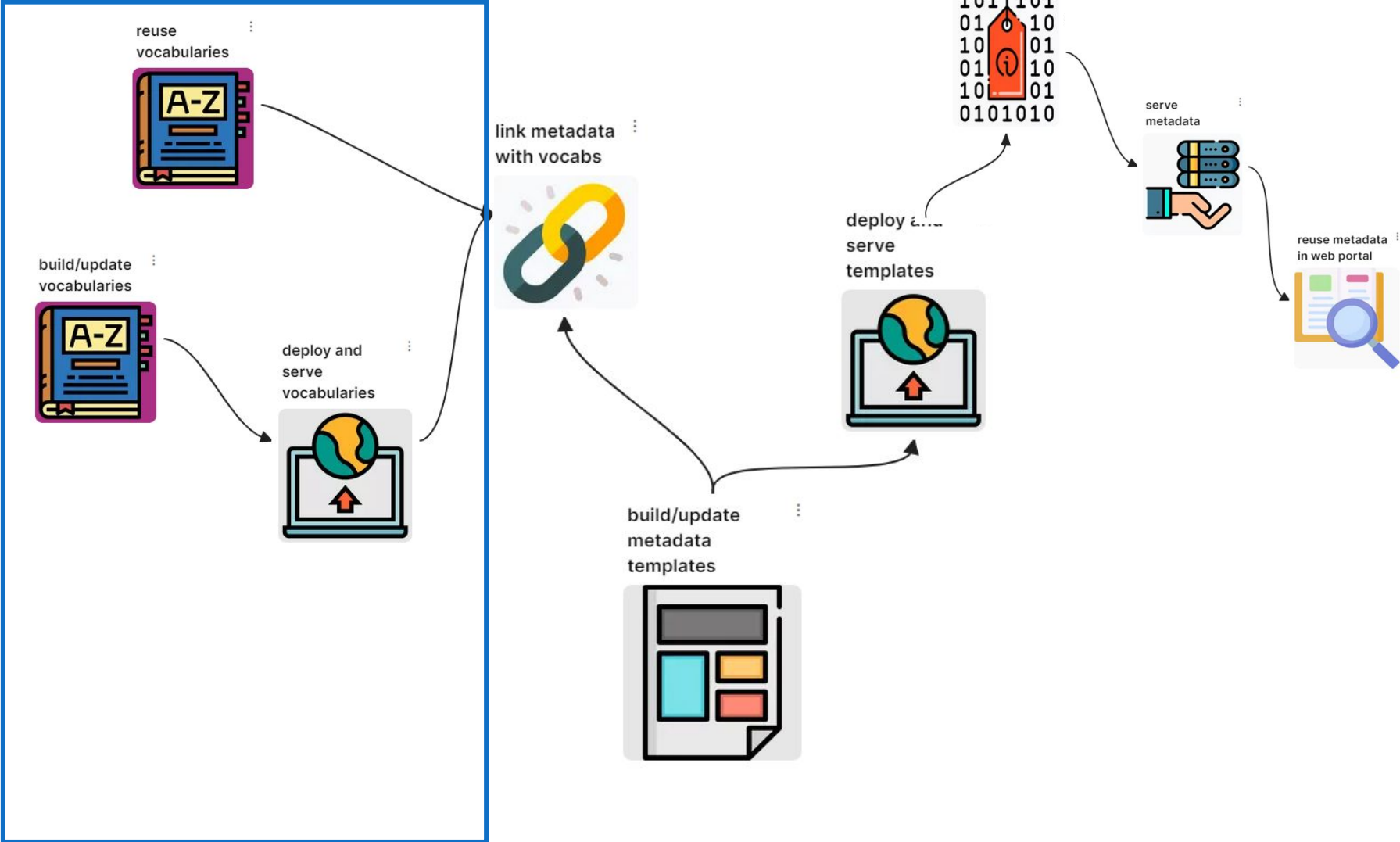
g2q/

FAIR vocabularies

Courtesy of Barbara Magagna



FAIR Vocabularies



What are key characteristics of a FAIR vocabulary?

- FINDABLE
- ACCESSIBLE
- INTEROPERABLE
- REUSABLE

Starting Point: One level higher

Basic Vocabulary Creation

an overview with references and recipes

John Graybeal for VODAN Africa
1 February 2021
22 February 2021



<https://docs.google.com/presentation/d/1vr9jhhkWPfCosjtxw0fy8rTn5u8Ba-qPyqB8R-e8k53o>

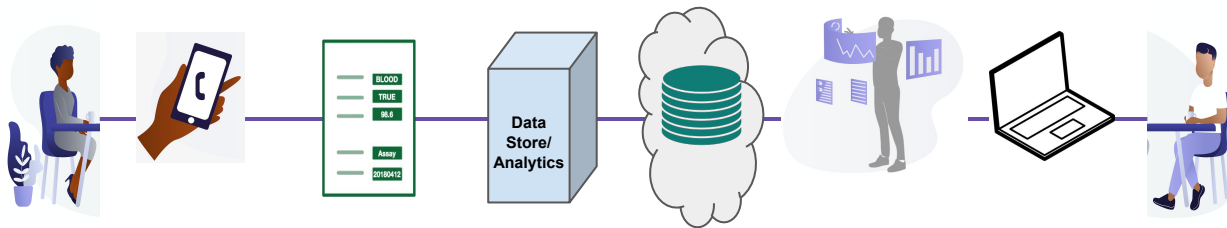
What do we want out of vocabularies?

1. We want to talk to each other ...



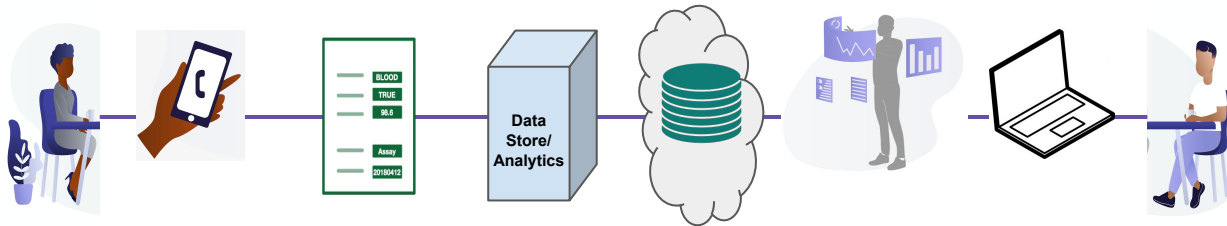
What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers



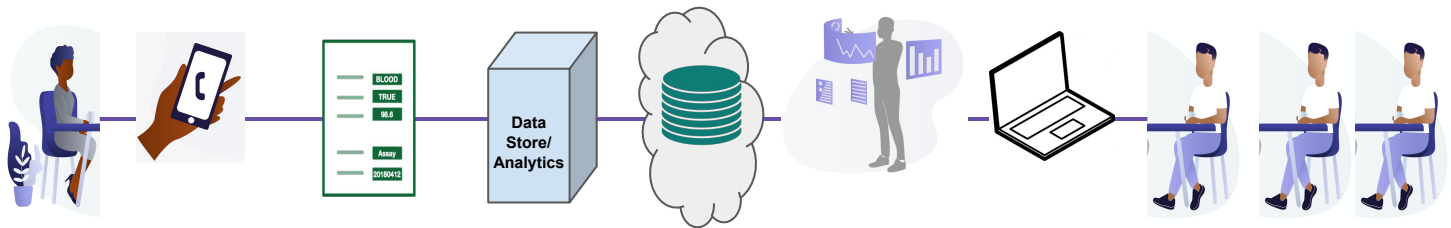
What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers
 - a. Computers are much more precise (picky) than human brains



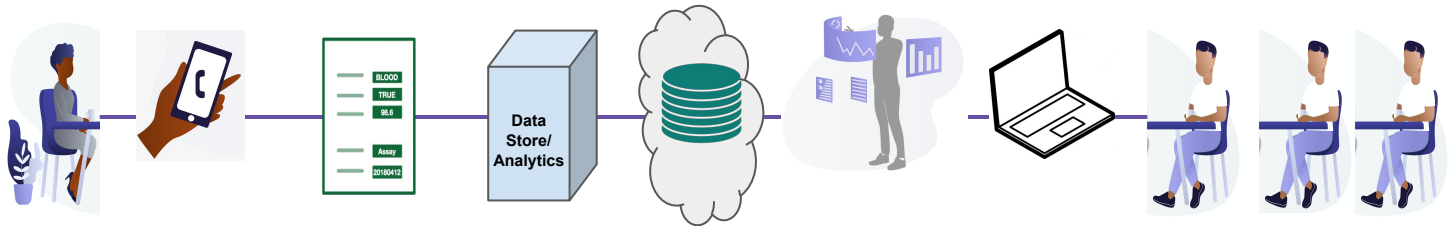
What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers
 - a. Computers are much more precise (picky) than human brains
 - b. What we 'say' often lasts way longer, and spreads more, than we imagine



What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers
 - a. Computers are much more precise (picky) than human brains
 - b. What we 'say' often lasts way longer, and spread wider, than we imagine



2. We want to make it easier to use software and data "later"



What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers
 - a. Computers are much more precise (picky) than human brains
 - b. What we 'say' often lasts way longer, and spread wider, than we imagine



2. We want to make it easier to use our work "later"
 - a. Thousands of file formats + data layout specifications for vocabularies
 - b. One set of these—the W3C semantic web specifications—is the clear 'best' and overall winner
 - c. Another of these—PDF—is a disastrous loser (not computable)



What do we want out of vocabularies?

1. We want to talk to each other ... via data in our computers
 - a. Computers are much more precise (picky) than human brains
 - b. What we 'say' often lasts way longer, and spread wider, than we imagine

2. We want to make it easier to use our work "later"
 - a. Thousands of file formats + data layout specifications for vocabularies
 - b. One set of these—the W3C semantic web specifications—is the clear 'best' and overall winner
 - c. Another of these—PDF—is a disastrous loser (not computable)



So it's not about **you**, it's about your legacy

Because let's face it, it would be a lot easier to just "use *your* words."



So it's not about **you**, it's about your legacy

Because let's face it, it would be a lot easier to just "use *your* words."

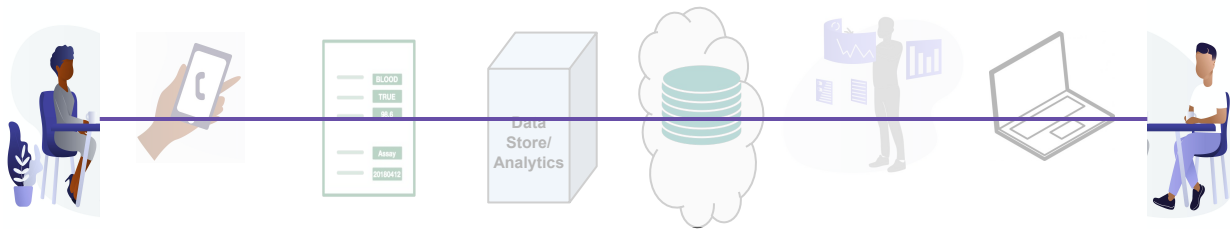
And also the legacy of government agencies worldwide...

Because large governmental agencies tend to be 'late adopters'

Whether we're talking about the U.N., U.S. and its agencies, or any other big organization, it's almost always very hard to re-use their vocabularies.



So our motivation is to 'talk' to each other using computers



How does that translate to our work with vocabularies?



Why build and use controlled vocabularies?

- To know if terms in different datasets mean the same thing
- To annotate data using concepts with agreed meaning
- To use vocabularies that are standardized and 'computable'
- To use vocabularies and concepts from your community
 - Shared meaning (and even spelling!)
 - Reflecting the community's consensus
 - Enabling community collaboration
 - Receiving ongoing love and care
 - To save time creating our own vocabulary list
 - And, to help us be FAIR

What are key technical characteristics of a FAIR vocabulary?

- **FINDABLE:** the vocabulary, and also each of the concepts, must have a GUPRI registered (indexed) in a community service
 - Globally Unique
 - Persistent
 - Resolvable
 - Identifier
- **ACCESSIBLE:** retrievable on the web, downloadable in different formats
- **INTEROPERABLE:** encoded in standard representations, mapped to other resources
- **REUSABLE:** open license, described with metadata, governed

Principles



List of main principles for **vocabularies** to follow

A vocabulary should have:

1. A registered location where it is served to the public
2. An openly accessible location where it is maintained and versioned
3. A well-defined IRI path (ideally matching the registered location)
4. Metadata describing it as thoroughly as is feasible, ideally using Metadata for Ontology Description and Publication (MOD)
5. If SKOS, all the best practices described for Bioportal SKOS Support (<https://www.bioontology.org/wiki/SKOSSupport>)
6. A published and followed process for keeping the vocabulary up-to-date
7. An open license
8. A well-defined community-accepted method to deprecate terms that change



List of main principles for vocabulary **terms**

Vocabulary terms should have:

1. Unique identifying fragment (may be codes or term labels)
2. Labels
3. Definitions
4. Unique IRIs that are resolvable
5. If derived from another source, the source should be referenced as `dcterms:source`
6. Mapping to a `skos:broader` Concept, unless it is a root term
7. As much as feasible, the annotations referenced in Rule 6 of *10 Simple Rules for Making a Vocabulary FAIR*

(<https://arxiv.org/abs/2012.02325>)



Rule #6 (*10 Simple Rules for Making a Vocabulary FAIR*):

Add these annotations to your vocabulary and terms:

1. **Labels and synonyms:** Encode term names and synonyms as `skos:prefLabel` or `skos:altLabel`, respectively.
2. **Textual definitions:** Textual definitions are encoded as `skos:definition`
3. **Codes and symbols:** Codes and symbols are encoded as `skos:notation`
4. **Notes or comments for clarifications:** Comments can be encoded using `skos:note`. Clarifications on usage can be recorded using `skos:scopeNote`
5. **Per-item metadata, if available:** Use standard elements such as `dcterms:creator`, `dcterms:created`, `dcterms:modified`, `dcterms:source`, `rdfs:seeAlso`.
6. **Relationships between terms:** If other non-specific relationships between items in the vocabulary are provided in the source document, they may be encoded using `skos:related`.
7. **Subsets:** If subsets or other term groupings are present in the source, encode each as a `skos:Collection`. Collections may be nested. Concepts may be members of more than one collection.
8. **Definition of the whole vocabulary.** The complete vocabulary should be encoded as a `skos:ConceptScheme`, with every `skos:Concept` having a `skos:inScheme` relationships to the scheme, or the top items in broader/narrower chains having a `skos:topConceptOf` relationship to the concept scheme.
9. **Language tags:** Use language tags for the term labels in multilingual vocabularies (and/or to describe the base language of the whole vocabulary).



List of main principles for creating a **new** vocabulary

1. There is no existing open vocabulary that fully suits your needs
2. You can not build upon an existing vocabulary by adding terms to it
3. It is impractical to reuse the vocabulary as is (e.g., it is not structured, or its identifiers do not resolve)
4. The existing vocabulary is no longer maintained



3 - How to build machine-actionable controlled vocabularies



Image by [Pavlo](#) from [Pixabay](#)

Technical solutions for vocabularies

Courtesy of Barbara Magagna



What are key technical characteristics of a FAIR vocabulary?

- **FINDABLE:** the vocabulary, and also each of the concepts, must have a GUPRI registered (indexed) in a community service
 - Globally Unique
 - Persistent
 - Resolvable
 - Identifier
- **ACCESSIBLE:** retrievable from the web, downloadable in different formats
- **INTEROPERABLE:** encoded in standard representations, mapped to other resources
- **REUSABLE:** open license, described with metadata, governed

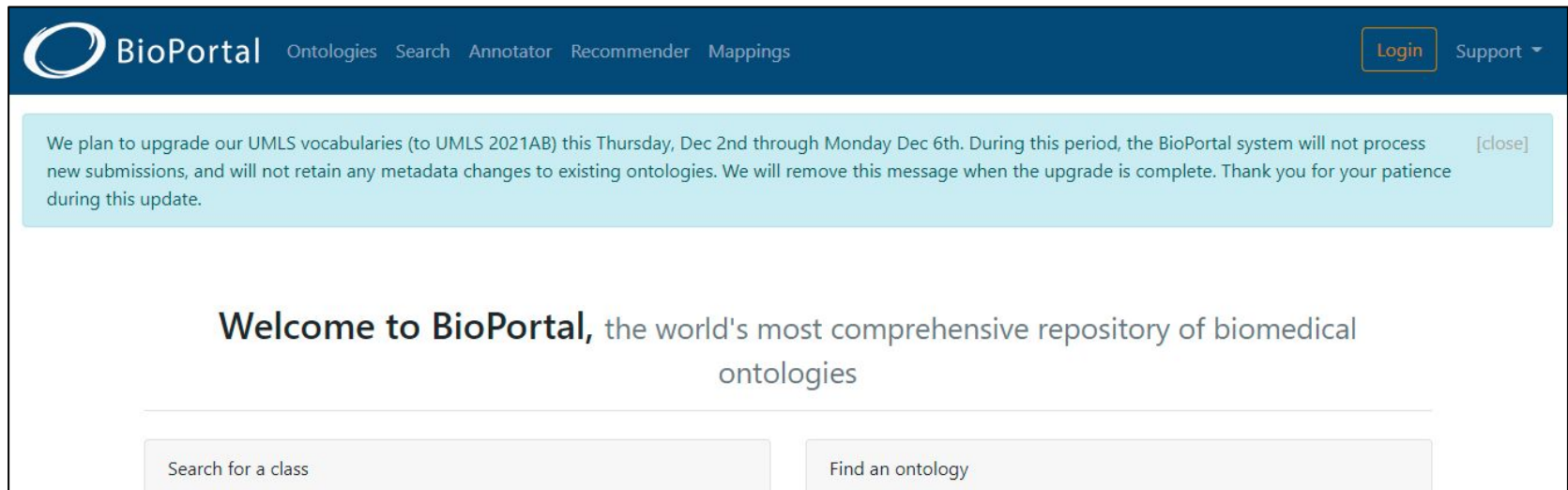
Technical solutions for building vocabularies

- **FINDABLE**

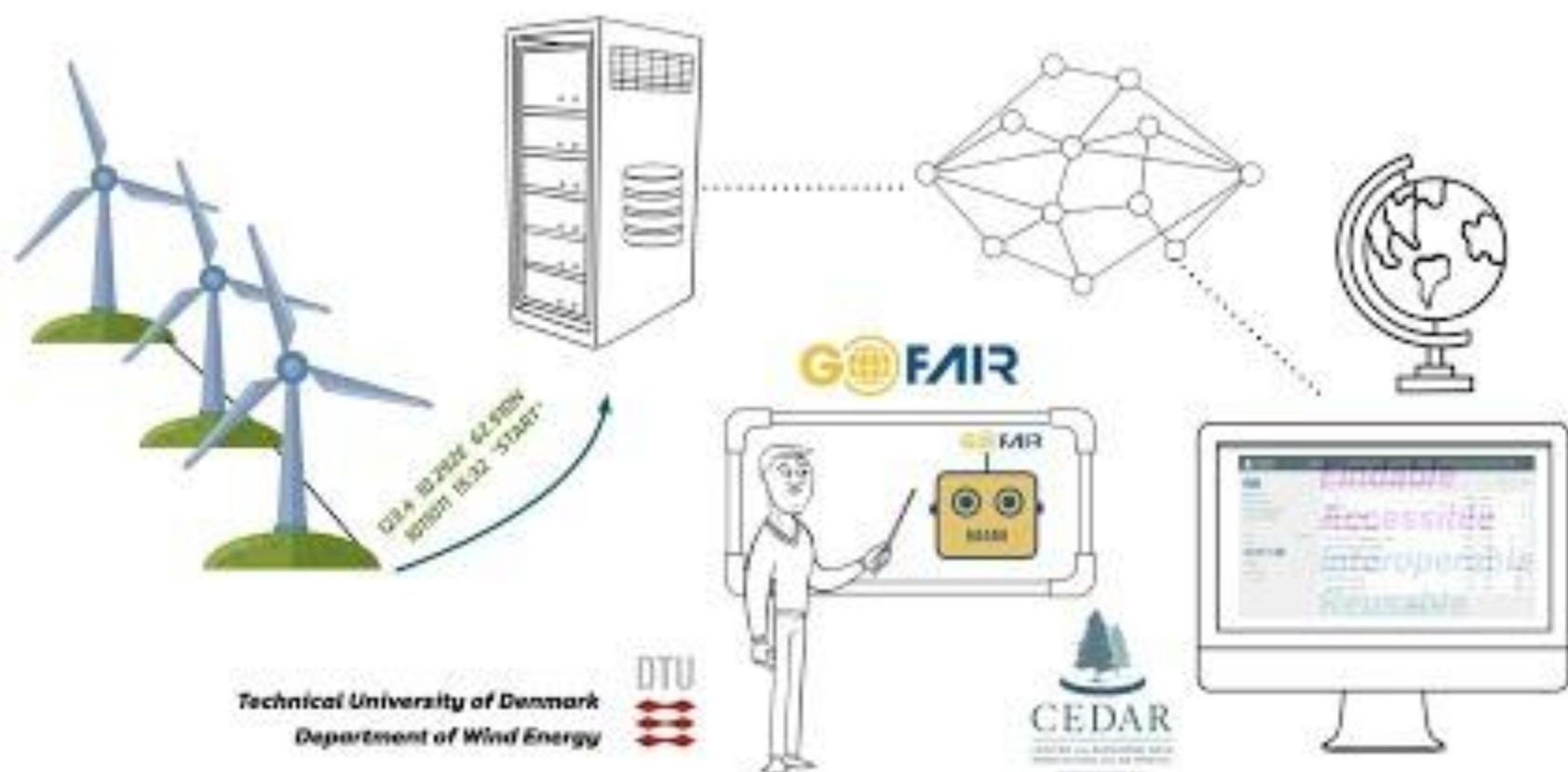
purl.org – assign persistent URLs to controlled vocabulary terms

BioPortal – ontology registration and indexing service

-> <https://bioportal.bioontology.org/>



The screenshot shows the BioPortal website interface. At the top, there is a dark blue navigation bar with the BioPortal logo on the left and links for "Ontologies", "Search", "Annotator", "Recommender", and "Mappings" in the center. On the right side of the navigation bar, there are "Login" and "Support" buttons. Below the navigation bar, a light blue banner contains a message about a system upgrade: "We plan to upgrade our UMLS vocabularies (to UMLS 2021AB) this Thursday, Dec 2nd through Monday Dec 6th. During this period, the BioPortal system will not process new submissions, and will not retain any metadata changes to existing ontologies. We will remove this message when the upgrade is complete. Thank you for your patience during this update." A "[close]" link is visible at the end of the message. Below the banner, the main content area features the text "Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies". At the bottom of the page, there are two search input fields: "Search for a class" and "Find an ontology".



excel2rdf and sheet2rdf: Template Content



ConceptScheme URI	http://ontology.deic.org/cv/vocab-name/			
PREFIX	vocab-name	http://ontology.deic.org/cv/vocab-name/		
PREFIX	pav	http://purl.org/pav/		
PREFIX	dct	http://purl.org/dc/terms/		
PREFIX	owl	http://www.w3.org/2002/07/owl#		
PREFIX	xsd	http://www.w3.org/2001/XMLSchema#		
PREFIX	skos	http://www.w3.org/2004/02/skos/core#		
dct:title				
dct:description				
dct:creator				
dct:rights				
pav:version				
pav:createdOn				
pav:lastUpdatedOn				
Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	skos:definition@en	skos:broader(separator=",")
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				
vocab-name:				

Vocabulary template structure



ConceptScheme URI	http://ontology.deic.org/cv/vocab-name/		General setup
PREFIX	vocab-name	http://ontology.deic.org/cv/vocab-name/	
PREFIX	pav	http://purl.org/pav/	
PREFIX	dct	http://purl.org/dc/terms/	
PREFIX	owl	http://www.w3.org/2002/07/owl#	
PREFIX	xsd	http://www.w3.org/2001/XMLSchema#	
PREFIX	skos	http://www.w3.org/2004/02/skos/core#	
dct:title			Controlled vocabulary metadata
dct:description			
dct:creator			
dct:rights			
pav:version			
pav:createdOn			
pav:lastUpdatedOn			
Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	Definition of terms and properties
vocab-name:			
vocab-name:			
vocab-name:			
vocab-name:			
vocab-name:			
vocab-name:			
vocab-name:			

Vocabulary template – general setup and prefixes



ConceptScheme URI	http://ontology.deic.org/cv/vocab-name/			
PREFIX	vocab-name	http://ontology.deic.org/cv/vocab-name/		
PREFIX	pav	http://purl.org/pav/		
PREFIX	dct	http://purl.org/dc/terms/		
PREFIX	owl	http://www.w3.org/2002/07/owl#		
PREFIX	xsd	http://www.w3.org/2001/XMLSchema#		
PREFIX	skos	http://www.w3.org/2004/02/skos/core#		

- **ConceptScheme URI:** This is a base URL for all your controlled terms (aka concepts in SKOS).
In this example:
<http://ontology.deic.org/cv/vocab-name/>
- **skos , dct, pav, xsd, and owl** are generic controlled vocabularies (ontologies) which contain properties that are used to define terms and metadata of a domain specific controlled vocabulary :
skos – **Simple Knowledge Organization System**
dct – **Dublin Core Terms**
pav – **Provenance Authoring and Versioning**
owl – **Web Ontology Language**
xsd - **XML Schema Definition**
- **PREFIX:** A prefixed name is turned into an URL by concatenating the URL associated with the prefix.
For example:
vocab-name will be turned into <http://ontology.deic.org/cv/vocab-name/>
pav will be turned into <http://purl.org/pav/>
dct will be turned into <http://purl.org/dc/terms/>

...

Vocabulary template – general setup metadata



dct:title	Insert values here		
dct:description			
dct:creator			
dct:rights			
pav:version			
pav:createdOn			
pav:lastUpdatedOn			

- It is recommended to re-use general ontologies to provide metadata about a controlled vocabulary, e.g.:
 - ❑ **dct:title** – controlled vocabulary title
 - ❑ **dct:creator** – ORCID ID of the vocabulary creator (full URL!), repeat this row as many time as there are vocabulary creators
 - ❑ **dct:rights** – usage license of the vocabulary, preferably chose license from <https://spdx.org/licenses/> and place a full URL of it (e.g., <http://spdx.org/licenses/CC0-1.0> for CC0-1.0 license)
 - ❑ **pav:version** – version of the controlled vocabulary, e.g. in a form of *majorChange.minorChange.bugFix*
 - ❑ **pav:createdOn** – initial datetime of the vocabulary creation in ISO 8601 format (include time zone!)
 - ❑ **pav:lastUpdatedOn** – datetime of the last vocabulary update in ISO 8601 format (include time zone!)

Vocabulary template – general setup versioning



dct:title			
dct:description			
dct:creator			
dct:rights			
pav:version			
pav:createdOn			
pav:lastUpdatedOn			

Insert values here

- Every time you update your control vocabulary update:

pav:lastUpdatedOn

and

pav:version

small modifications → change last digit → 0.1.0

minor modifications → change middle digit → 0.1.0

major modifications → change first digit → 0.1.0

- *Small, minor and major up to you to decide, but small changes could be spelling corrections, minor could be addition of new terms, major could be re-arrangement of hierarchy or deprecation/substitution of terms.*

Vocabulary template – SKOS property prefLabel



Identifier	skos:prefLabel@en
vocab-name:	
vocab-name:	
vocab-name:	

- **@en** part in **skos:prefLabel@en** indicates in what language the preferred label is given (en = English).
- To change preferred label language tag find appropriate two-letter tag in:
<https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>
- **You can have preferred labels in multiple language, but you can only have one preferred label per language tag!!!**
- **Terms must have unique preferred labels, i.e. two different terms cannot have same preferred labels!**

- The preferred label (short) for the term
- Unique within vocabulary

Vocabulary template – SKOS property altLabel



Identifier	skos:prefLabel@en	skos:altLabel(separator=",")
vocab-name:		
vocab-name:		
vocab-name:		

- **skos:altLabel** is an alternative label for the term you are defining. Unlike **skos:prefLabel** you can have both multiple alternative labels in multiple languages.
- To enter multiple **altLabel** in column simply separate them with “,” (comma), that’s why we have **(separator=“,”)** in **skos:altLabel(separator=“,”)**
- If you prefer other ways of separating values in this column change this value accordingly.

- Alternate label(s) (synonyms)

Vocabulary template – SKOS property definition



Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	skos:definition@en
vocab-name:			
vocab-name:			
vocab-name:			

- **skos:definition** column is used to define your terms.
- Always define your terms! If you have trouble defining it yourself, get a domain expert(s) to help you.
- Similar to **skos:prefLabel**, you can have definitions in multiple languages, but there can be only one definition per language tag!
- Change the current language tag (**en**) or add additional **skos:definition** columns associate with different language tags as you like/need.

- Description of the term (not redundantly)
- Like previous properties, supports languages

Vocabulary template – SKOS property broader



Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	skos:definition@en	skos:broader(separator=",")
vocab-name:				
vocab-name:				
vocab-name:				

- **skos:broader** column is used to express hierarchy in your controlled vocabulary in cases when your controlled vocabulary is not a simple flat list of terms but has more of a tree structure (see [taxonomies](#)).
- Specifically, if your term is under some broader concept then you should put **Identifier** of that broader concept in this column.
- In case your term has several broader concepts, simply separate their corresponding **Identifiers** with comma.
- If a term is a top concept (i.e., does not have broader concepts) leave this column empty.

- Points to broader terms ('up' hierarchy)
- Use broader terms exact Identifier name

Vocabulary template – SKOS property broader



(Separate multiple values with commas.)

- **dct:source** – IRI reference to the source material that you used to define the term
- **skos:closeMatch** – use Identifier of closely related term from this or another vocabulary
- **skos:exactMatch** – use Identifier of conceptually identical term from this or another vocabulary
- **dct:creator** – use ORCID ID of term creator
- **dct:contributor** – use ORCID ID of term contributor

excel2rdf and sheet2rdf: Workflows

A screenshot of a GitHub repository page for the workflow "Sheet2RDF". The page shows the workflow configuration file "sheet2rdf.yml" and a list of 11 workflow runs. The first run is highlighted, showing it was manually run by "mabablue". A "Run workflow" button is visible, and a dropdown menu is open, showing options to run the workflow from the "main" branch.

Search or jump to... Pull requests Issues Marketplace Explore

fair-data-collective / M4M19-subjects-vocabulary Public Watch 0 Star 0 Fork

<> Code Issues Pull requests **Actions** Projects Wiki Security Insights Settings

Workflows [New workflow](#)

All workflows

Sheet2RDF

Sheet2RDF
sheet2rdf.yml

Filter workflow runs

11 workflow runs Event Status Branch Actor

This workflow has a workflow_dispatch event trigger. Run workflow

✓ Sheet2RDF
Sheet2RDF #11: Manually run by mabablue

Use workflow from
Branch: main
Run workflow

excel2rdf and *sheet2rdf*: Workflows



- Automatic workflows in GitHub
- Executed by means of GitHub actions
- Contains underlying shell, python and java programs which:
 - convert the google or Excel sheet to the machine-readable controlled vocabulary
 - Test the derived controlled vocabulary
 - Commit the results and test logs to the GitHub repository
 - Makes the vocabulary available to OntoStack and/or BioPortal to be served to humans and machines

excel2rdf and *sheet2rdf*: Demonstration

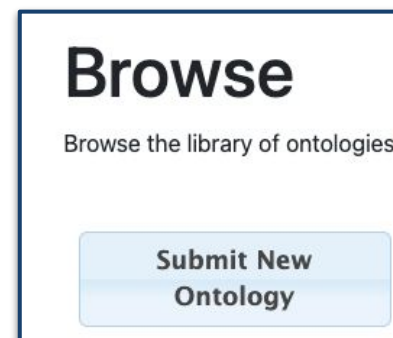


- Modify the Google or Excel spreadsheet
- If Excel, check it in to the GitHub repo
- Wait for Google's workflow to kick in (5 min)
- If BioPortal, log in to re-submit ontology (or wait for overnight automated discovery)
- View your updates in published repositories
- In case of problems, email support@bioontology.org (public list)

excel2rdf and *sheet2rdf*: Doing this yourself



- Fork the appropriate repo from <https://github.com/fair-data-collective>
- If Google, populate a sheet with template
- Configure workflow for your spreadsheet source (Google) or name (Excel)
- If sheet2rdf with OntoStack, configure that
- Set up BioPortal new ontology to read from your binary product in GitHub (example)



Break

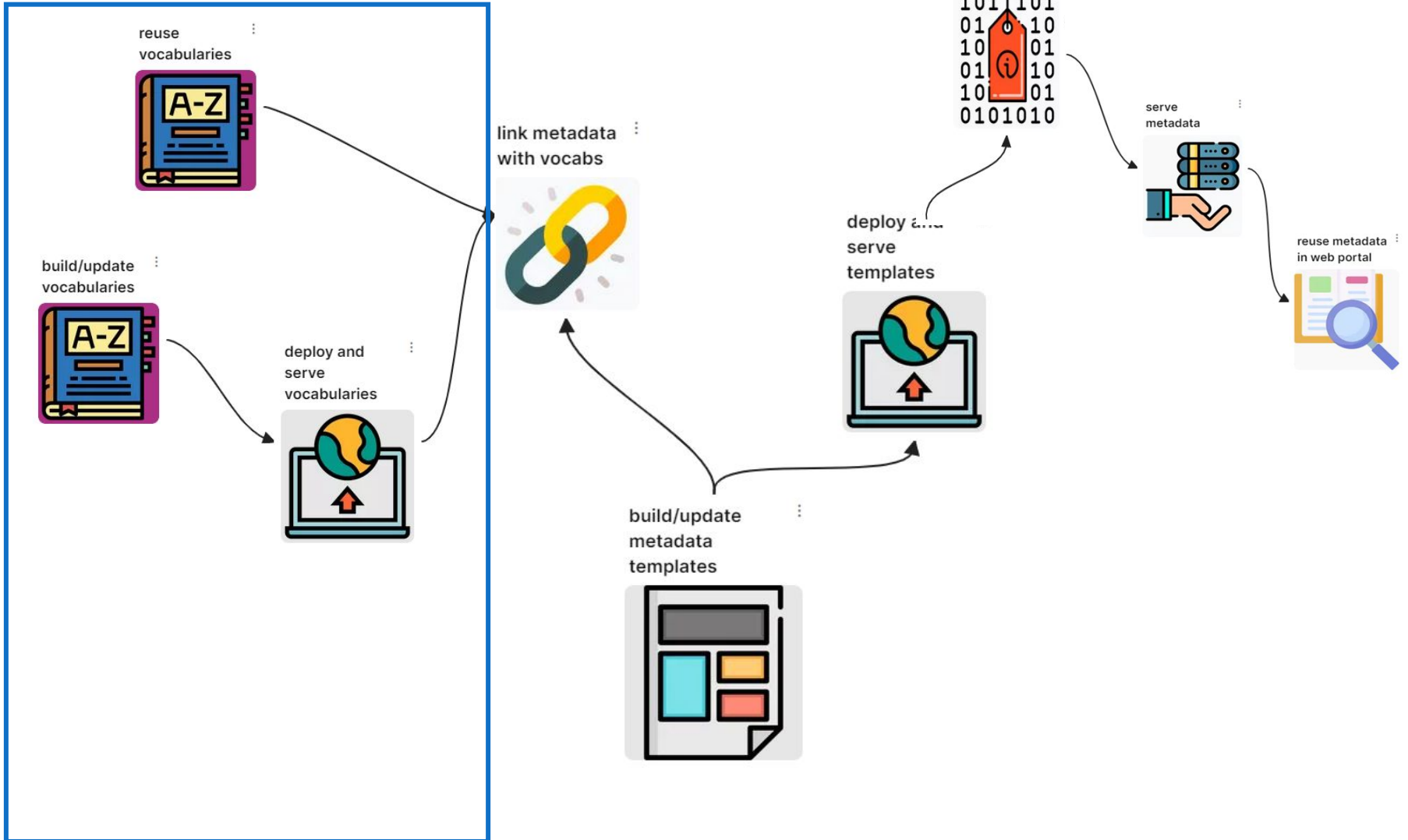


4 - Building domain-specific controlled vocabularies

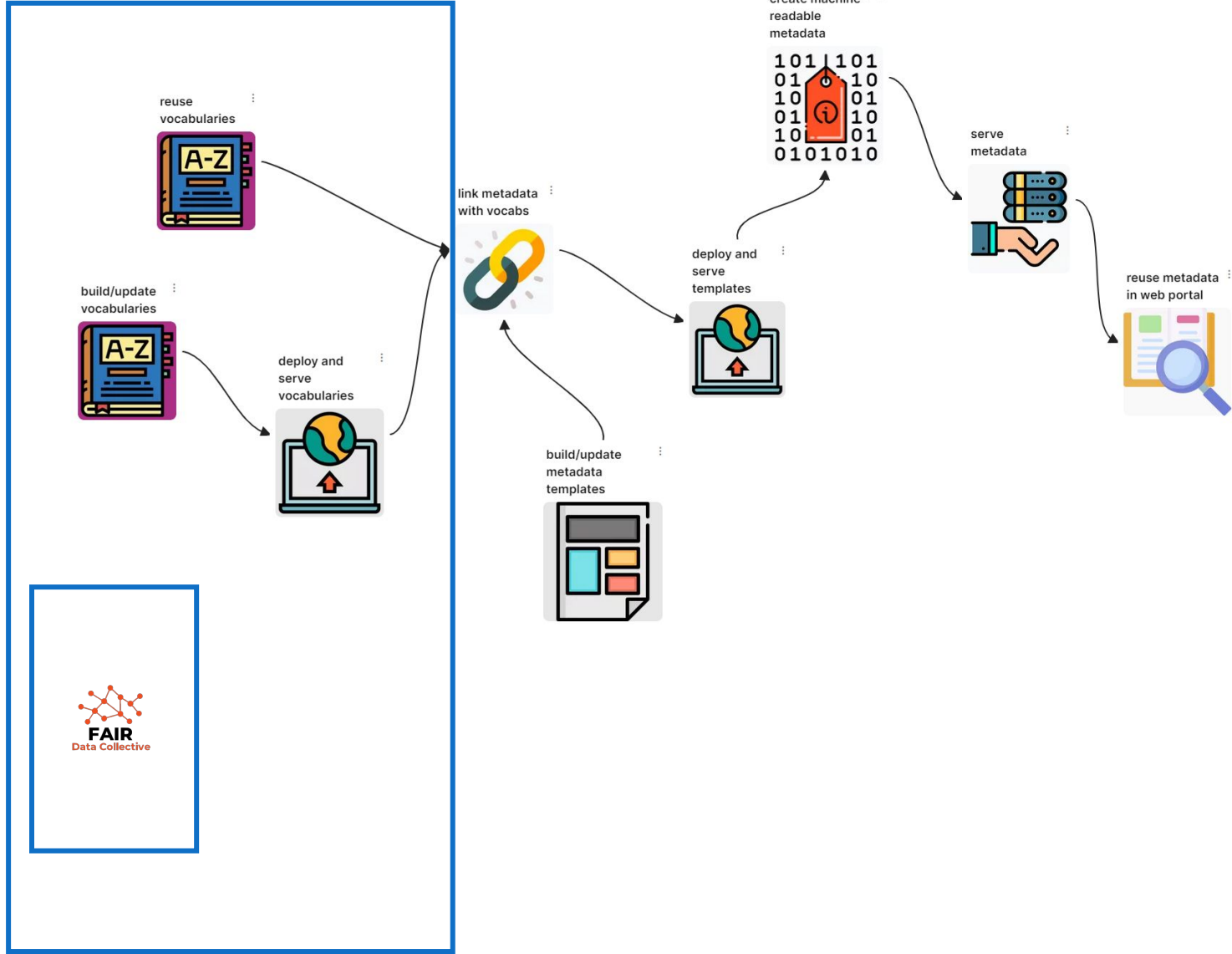


Image by [Bernard FLEURANDEAU](#) from [Pixabay](#)

Create Vocabularies



Create vocabularies



Domain-specific vocabularies

Let's look at example vocabularies in a domain

- Vocabularies you already started work on: ✓
- Vocabularies you might want to include: ?
- How to find other relevant vocabularies: 🔍

There are some questions and notes we can bring up for you to consider as you're working on exercises.

These are examples applicable to one set of attendees, but you may find useful lessons (and can bring your own terms to the table for the exercises).

Domain specific vocabularies – ✓

- SAMIS-JSC CURATION INTERFACE CONTROL DOCUMENT (ICD) - DRAFT v2
 - 8.2.1.3 **sampletype** (powder, aggregate, chip, particle, liquid, fibsection, atomprobetip, microtomeslices, residue)
 - 8.2.1.3 **containertype** (aluminumcontainer, glassvial, teflonbag, polycarbonatecontainer, aluminumwrapping, customcontainer)
 - 8.2.2.4 **sampletype** (thinSection, thickSection, pottedButt, pressedGrainsEpoxy, pressedGrainsMetal)
 - 8.2.2.4 **containertype** (same as 8.2.1.3? always?)
- UA-DSD_OSIRIS-REx-SA_Image_Format_DRAFT
 - 4.2.1.1 **dataProductType** (optical microscope image, secondary electron image, backscatter electron image, x-ray maps, high-resolution cathodoluminescence image, bright or dark field images, diffraction pattern image, raster ion image, SCAPS ion image, nanoindentation topographic image, microindentation topographic image, QRIS single channel image, QRIS multichannel image)
 - 4.2.2 **pixelLocation** (upperLeft, upperRight, lowerLeft, lowerRight, center)

Domain specific vocabularies – ?

- SAMIS-JSC CURATION INTERFACE CONTROL DOCUMENT (ICD) - DRAFT v2
 - 8.2.1.3 peclassification (?)
 - Glossary of Terms
 - Table of Abbreviations
- UA-DSD_OSIRIS-REx-SA_Image_Format_DRAFT
 - 4.2.1.1 bitDepth (1, 8, 16, 32)
- Instruments, Labs, and Data Product Leads Spreadsheet
- DataDictionary metadata fields
 - columns: productEntity, domainEntity, instrument, Meaning, Keyword, Format, Units
 - other tabs: [products, BaseDataTypes, DataContentTypes, productContent, ...]

Domain specific vocabularies –

Searching in Bioportal yields ...

- 'Image Type' => DICOM Controlled Terminology
- Recommender (powder, aggregate, chip, particle, liquid, fib section, atom probe tip, microtome slices, residue) => NCIT (6/9)
- MatPortal yields DEB ontology

This is a bit of an open exploration exercise, to potentially discover useful content.

ESIP's [Community Ontology Repository](#) is another possible source of earth science vocabularies.

Or check out this [crowd-sourced repository list](#)!

Exercise: Our Plan

- You will have a choice of exercises.
- You can stay in main session, or be in a breakout room.
- We have (something more than 90 minutes) all told.
- Nancy and I will be floating and monitoring chat.
- After we answer any questions, first breakouts start.
- We'll come back to plenary after 30 minutes & report.
 - Show your product for team.
 - Identify all issues and questions.
- After reports, return to breakouts for TBD minutes.
 - They can be same breakout or another one.
- Again we'll come back to plenary for discussion.
- Finally, we'll discuss advanced issues for 20 minutes.

Exercise: Choices

You can pick one to two areas to work on:

- 0:40+ ● Pick one or more vocabularies to improve 4A
- 0:40 ● Create a new vocabulary for your domain 4B
- 1:00 ● Set up a workflow with GitHub & BioPortal 4C

Note: We want to use your work in our metadata templates tomorrow. So, think of terms used as *values*.

4A: Pick one or more vocabularies to improve

1. (optional) Make a copy of the Workshop Vocabulary
2. Using the information that starts at [Vocabulary Template Structure](#), modify your vocabulary to include more details
3. Add additional root (top-level) terms to create more vocabulary groups
4. Convert the vocabulary to SKOS regularly to confirm it still validates (can do manually)
5. Update the Workshop Vocabulary at the end of your work today (will load into Bioportal)

4B: Create a new vocabulary for your domain

1. Make a copy of the excel2rdf or Google spreadsheet
2. Update the metadata to describe your product (per slides start–end)
3. Create the new vocabulary with a parent term and some sub-terms (slides start–end)
4. Frequently convert the vocabulary to SKOS to confirm it still validates (can do manually)

4C: Set up a workflow with GitHub & BioPortal

1. Set up GitHub repo per excel2rdf and sheet2rdf:
Doing this yourself
2. Make a minor change in your vocabulary and confirm a new TTL file is generated
3. Go to BioPortal and create a new ontology
 - a. See John for naming advice and questions
 - b. Make ontology private for now
 - c. Submit ontology, visit after a few minutes
4. Discuss further details with John (clearing caches, updating submission).

Closing: Pose/discuss advanced vocabulary topics

- How to blend existing vocabularies with yours
- Identifiers: use codes or meaningful labels?
- What about versioning of terms and vocabs?

Capture the questions and discussion points in the workshop minutes.

Want to learn more? Check out this fairly long curated list of SKOS-focused resources.

Your new products

Summary - Day 1

Welcome - Day 2



Image by [R. E. Beck](#) from [Pixabay](#)

- Any comments, questions from Day 1 activities?
 - Need(ed) any more/other background information?
- Common notes: <https://bit.ly/EarthCubeM4M>
- Slides: <https://bit.ly/ECM4MSlides>



Agenda - Day Two, Sept 9, 2022

#	Duration	Agenda Item
	4' total	Build a machine-actionable metadata template
0	0:10 GO	Review and issues from Day 1
1	0:45 GO	How to build machine-actionable metadata templates
2	0:20 GO 0:40 GO	Step 1: Make a copy of existing template (e.g., Generic Dataset Metadata Template) Step 2: Create metadata fields and assign controlled vocabularies
	0:15	Break (~12:55pm)
3	0:30 GO 0:20 GO 0:10 GO	Step 3: Add created or new fields to metadata elements and add RDF properties Step 4: Structure new or copied template by composing and replacing elements Step 5: Make machine-actionable metadata by filling out the form
4	0:20 GO	FAIR Orchestration and Integration with other systems
5	0:15 GO	Questions and Discussion
6	0:15	Round-off and what happens next? Summary: What we did during M4M and discussion of next steps.

Review and Issues from Day 1

What we covered:

- FAIR & who can best do what
- 3 Point FAIRification Framework
- Automating FAIR for vocabularies using CEDAR and BioPortal
- Reviewed a curated list of SKOS-focused resources
- Yesterday's video recording - please don't share

What we did:

- Updated & uploaded draft SAMIS vocab
 - (correcting naming issues for identifiers & added definitions)
- Created new vocab for OSIRIS-ReX Analytic Methods & uploaded to BioPortal

1 - How to build machine-actionable metadata templates

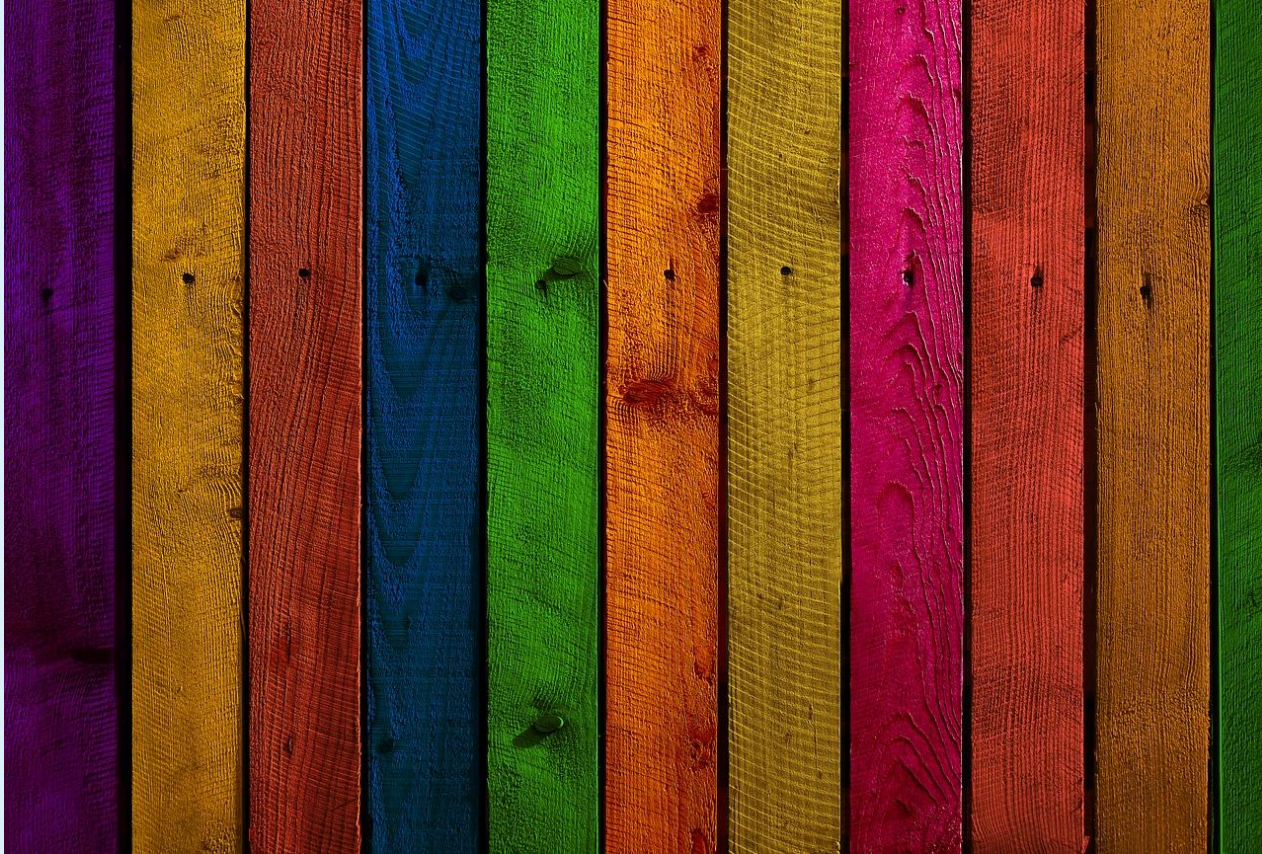


Image by [Pavlo](#) from [Pixabay](#)

3 Levels

1. Making metadata FAIR(er)—what's that about?
2. Using standards to make metadata FAIRer—the DCAT specifications
3. Using CEDAR to make FAIRer metadata

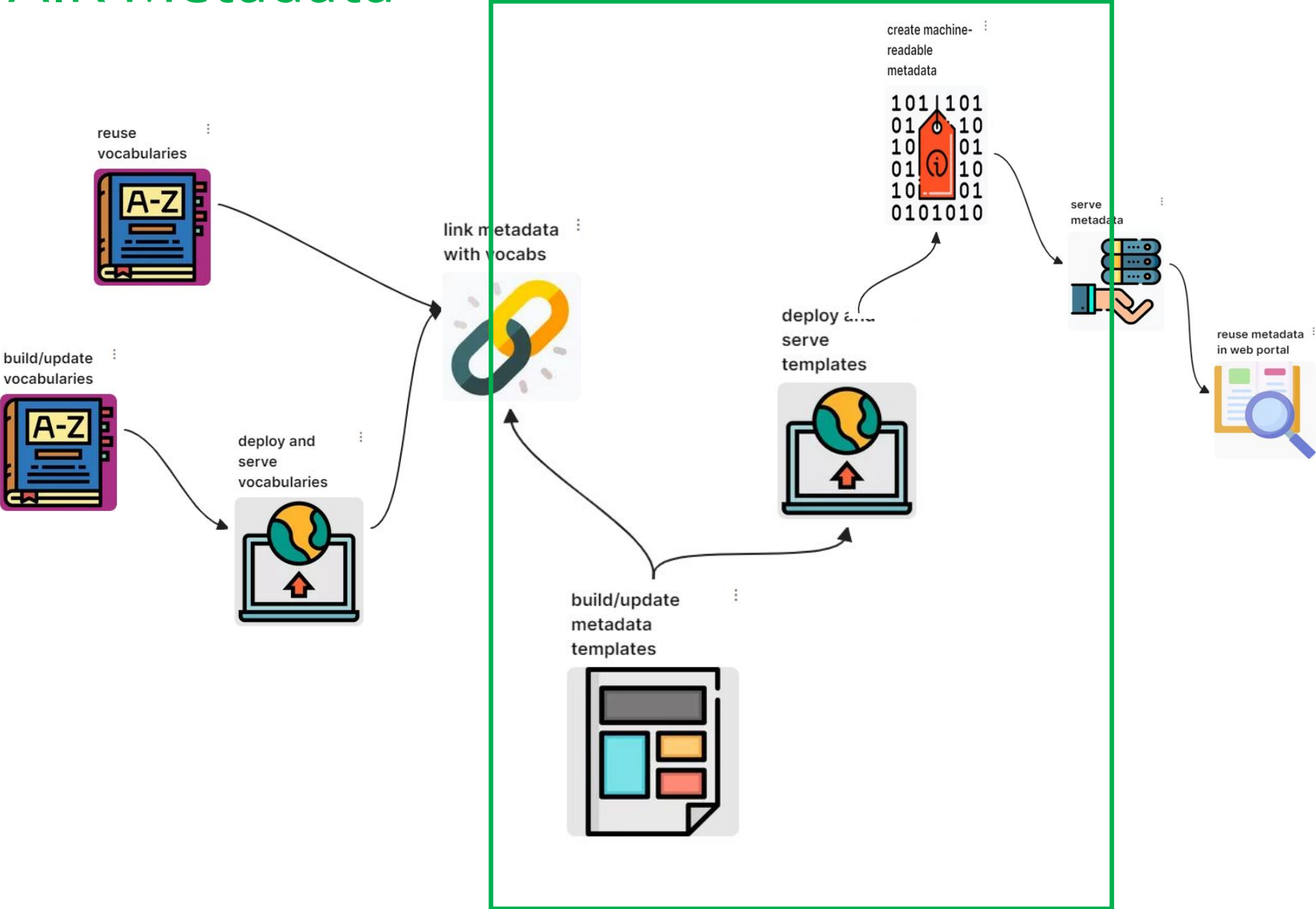
FAIR metadata

Courtesy of Barbara Magagna



(Updated by GO FAIR US)

FAIR Metadata



Towards FAIRer Metadata

Metadata as unstructured text description file in same directory



Towards FAIRer Metadata

Metadata as tabular information (property/value) in same directory

Property	Value
Title	Data XYZ
Creator	Barbara
Publication date	08-12-2021
License	CC-BY-NC-ND
Resource Type	Dataset

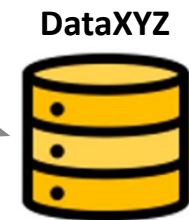


Towards FAIRer Metadata

Metadata as tabular information with use of controlled term as the value

Property	Value
Title	Data XYZ
Creator	Barbara
Publication date	08-12-2021
License	CC-BY-NC-ND
Resource Type	Dataset

Resource Type
Audiovisual
Collection
Data paper
Data stream
Dataset



Towards FAIRer Metadata

Metadata as tabular information with use of FAIR vocabulary as the value

Property	Value
Title	Data XYZ
Creator	Barbara
Publication date	08-12-2021
License	CC-BY-NC-ND
Resource Type	http://vocab.fairdatacollective.org/gdmt/Dataset



DataXYZ

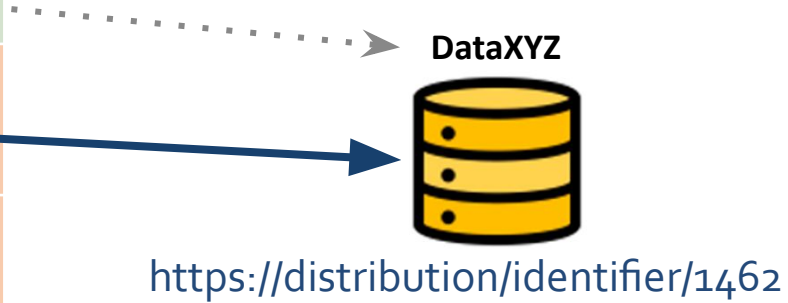
ID	prefLabel
http://vocab.fairdatacollective.org/gdmt/ResourceType	Resource Type Category
http://vocab.fairdatacollective.org/gdmt/Audiovisual	Audiovisual
http://vocab.fairdatacollective.org/gdmt/Dataset	Dataset



Towards FAIRer Metadata

Metadata as tabular information with use of identifiers (GUPRIs) for the values

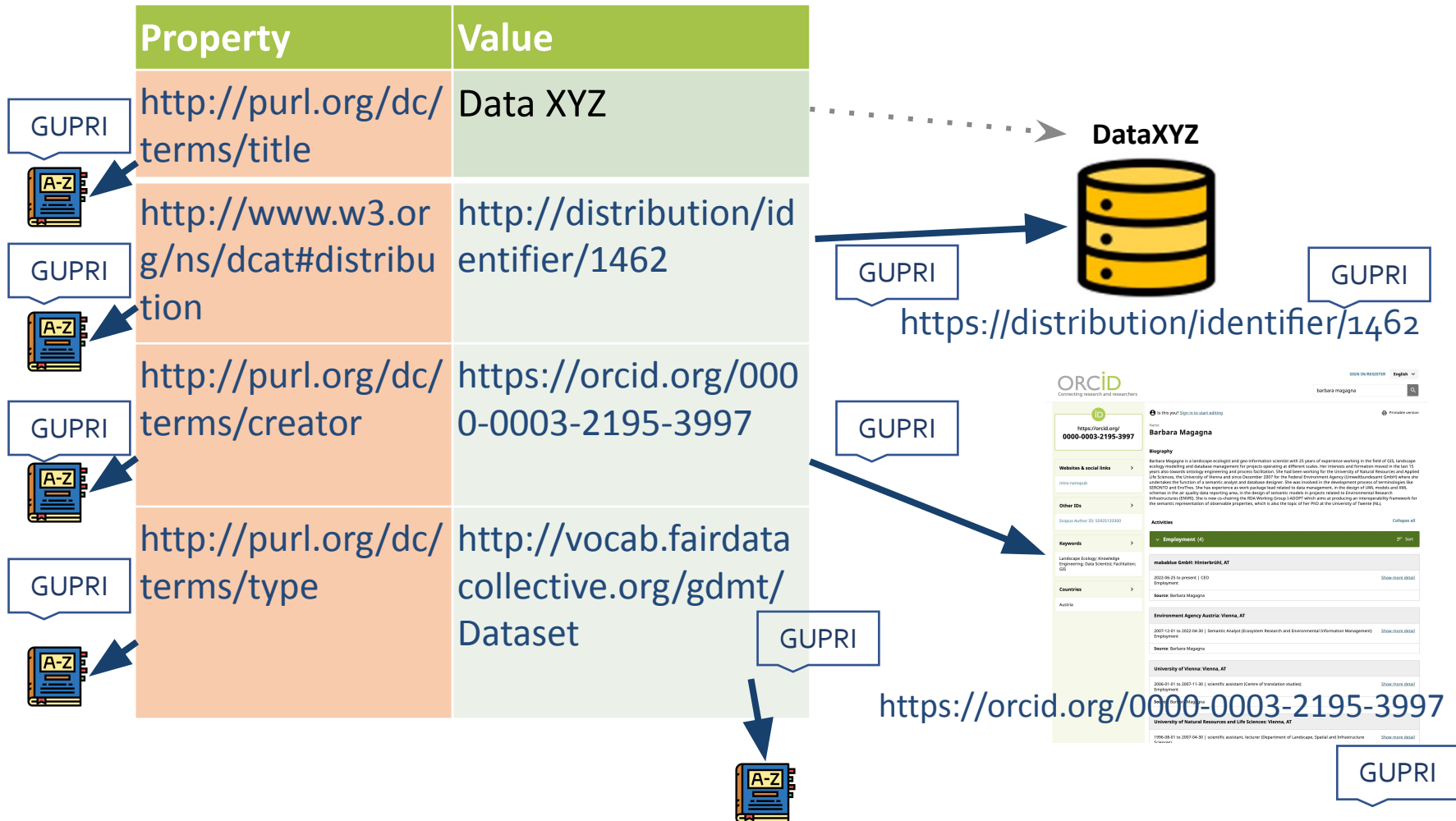
Property	Value
Title	Data XYZ
Distribution identifier	https://distribution/identifier/1462
Creator	https://orcid.org/0000-0003-2195-3997
Resource Type	http://vocab.fairdatacollective.org/gdmt/Dataset



<https://orcid.org/0000-0003-2195-3997>

Towards FAIRer Metadata

Metadata as tabular information with identifiers (IRIs) for properties



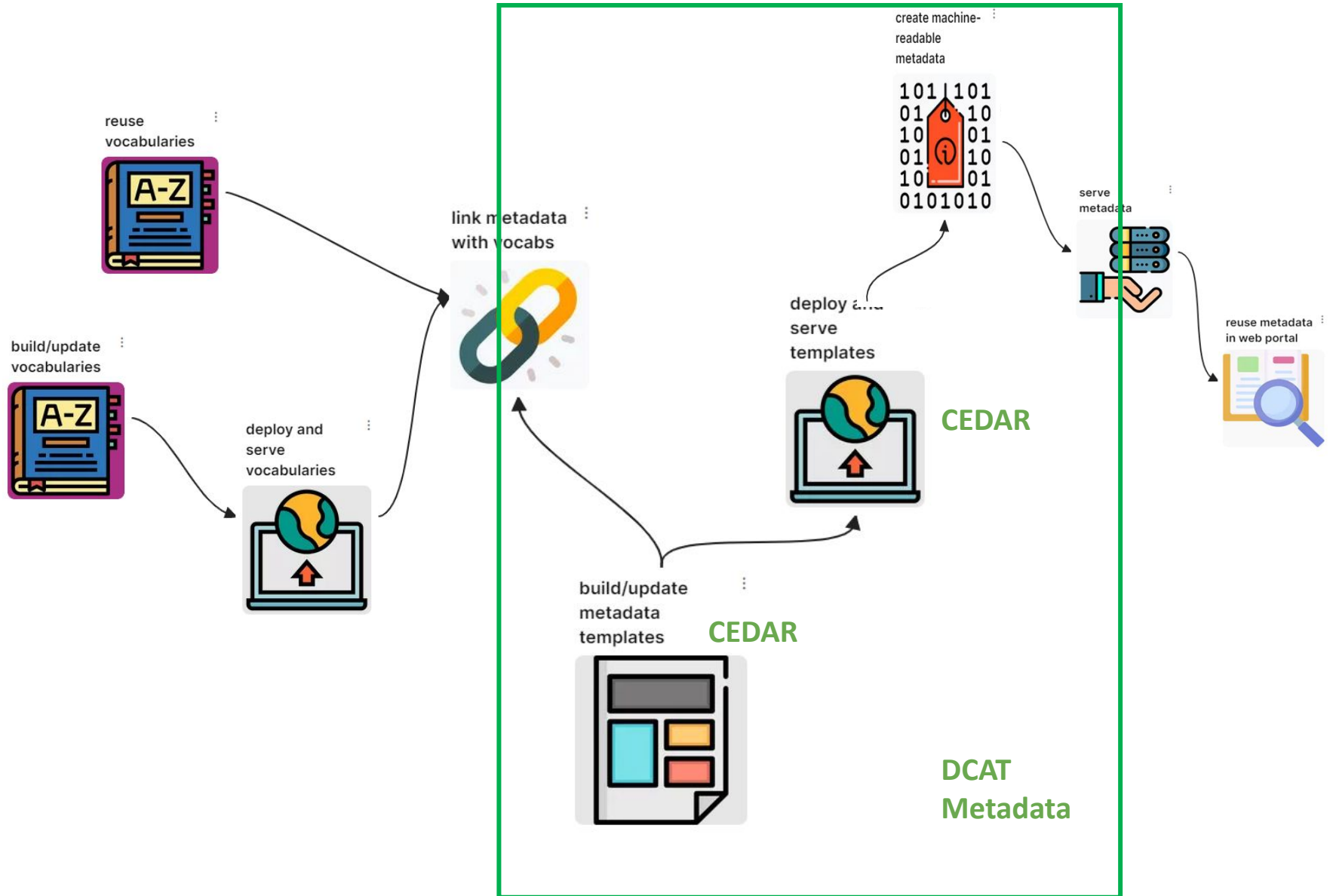
DCAT Specification

Courtesy of Barbara Magagna



(Updated by GO FAIR US)

Technical Solutions: FAIR Metadata



Metadata Schema – DCAT Vocabulary

Data Catalog Vocabulary (DCAT) - Version 2

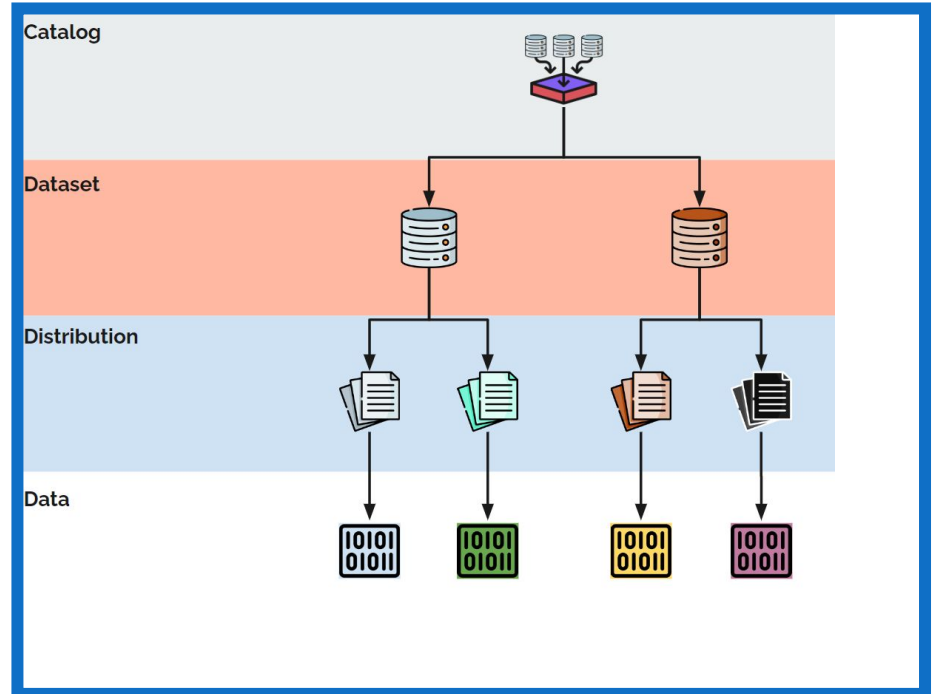
W3C Recommendation 04 February 2020



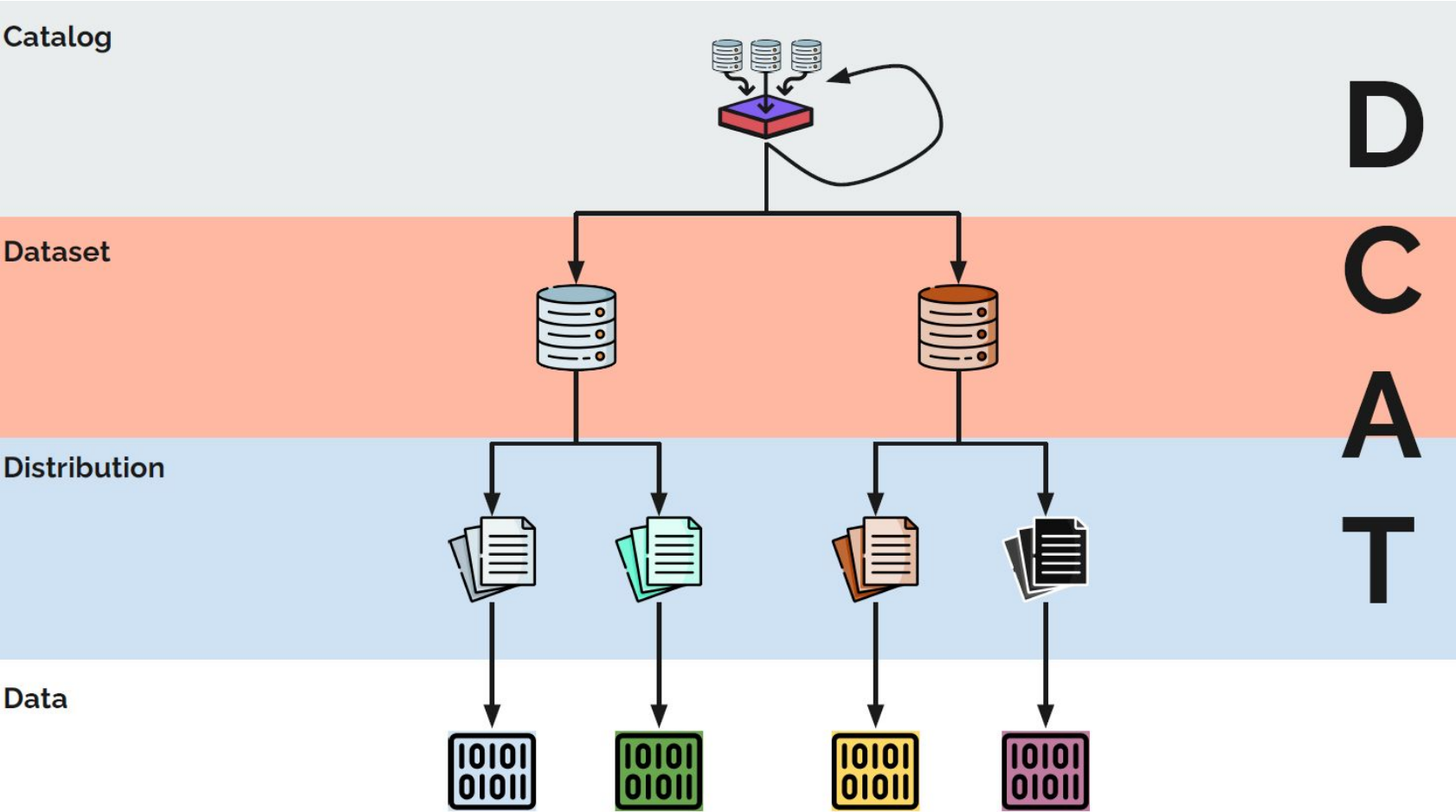
<https://www.w3.org/TR/vocab-dcat-2/>

DCAT Application Profile for data portals in Europe

DCAT-AP 2.1.0



DCAT Schema

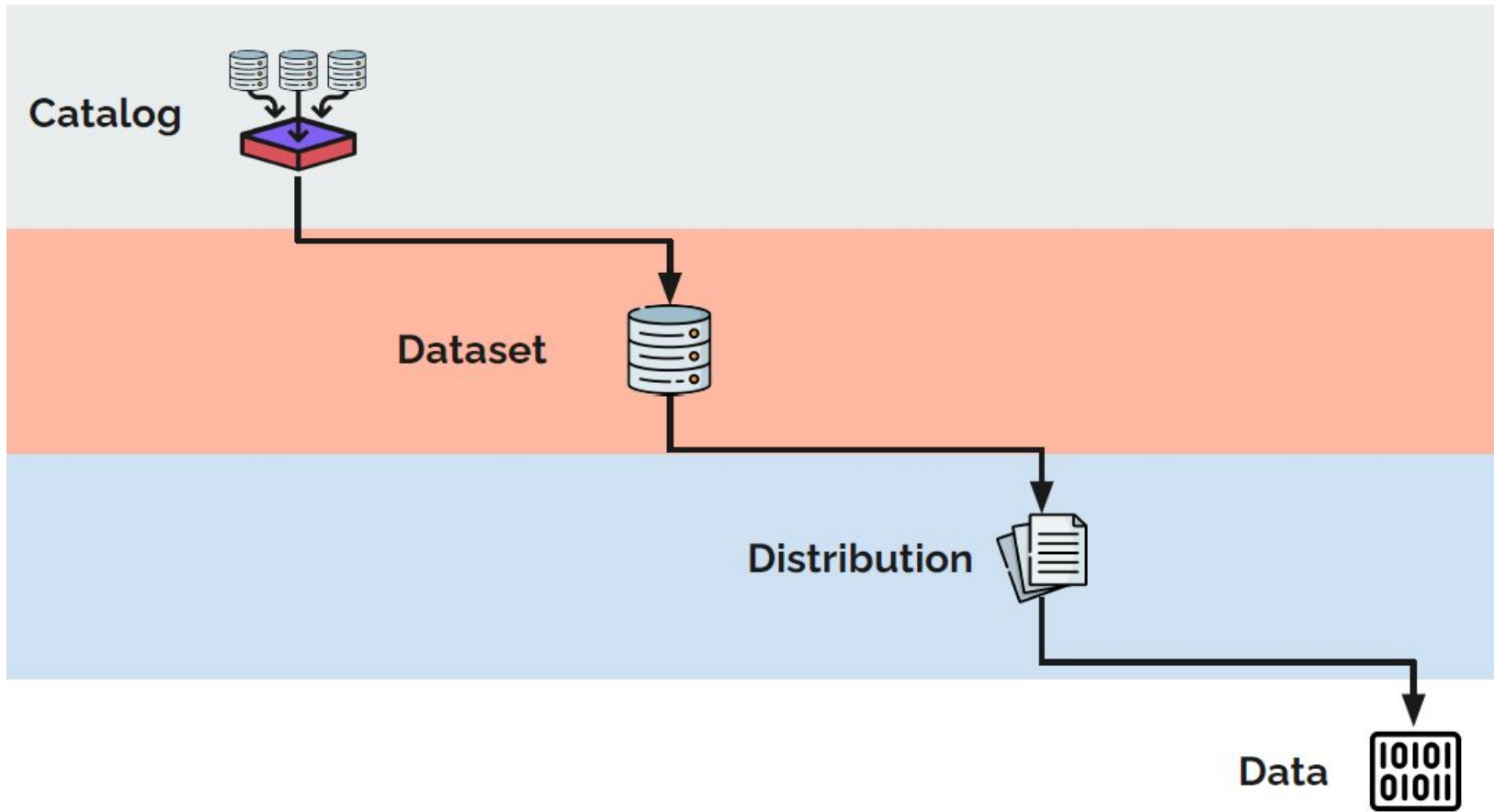


Distribution metadata fields

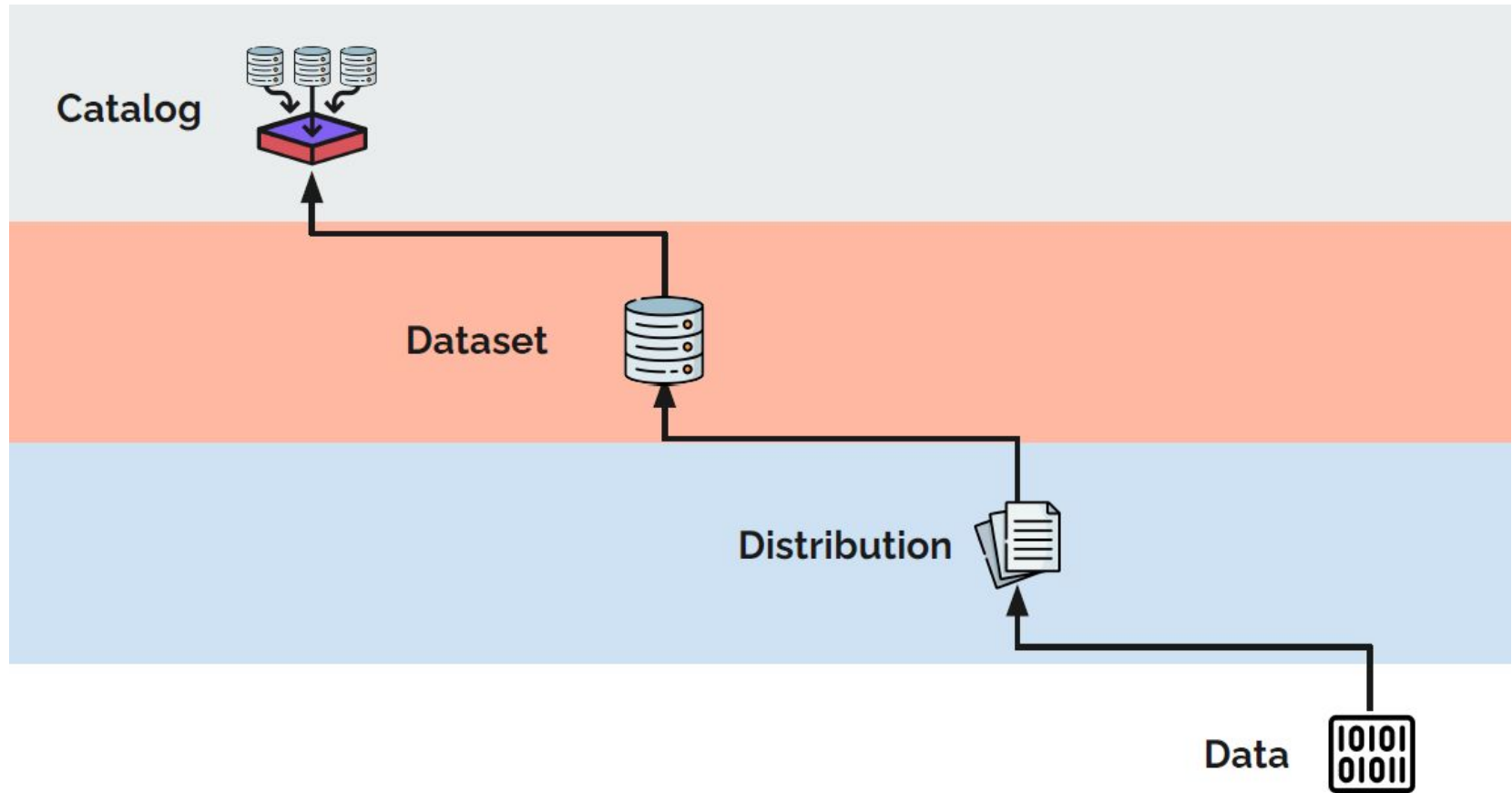
- Title
- Description
- Media Type
- ...
- Access URL
(landing page, API doc, direct download links, SPARQL endpoints...)

This is a solution for FAIR Principle F3

Data user perspective



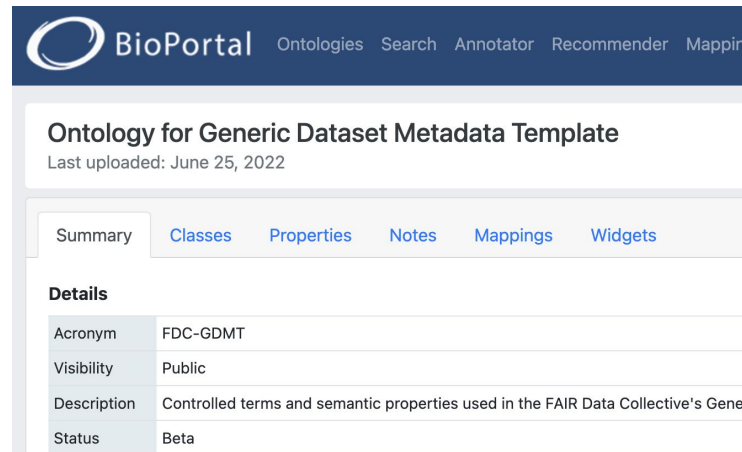
Data owner perspective



CEDAR, DCAT, and GDMT

CEDAR has a template (or two) based on DCAT
 We'll use Generic Dataset Metadata Template

- Created for these workshops
- Uses a DCAT-based vocabulary: FDC-GDMT
- Think of this as one instantiation of DCAT
- Based on an EU Application Profile (DCAT-AP 2.1.0)



BioPortal Ontologies Search Annotator Recommender Mapping

Ontology for Generic Dataset Metadata Template

Last uploaded: June 25, 2022

Summary **Classes** Properties Notes Mappings Widgets

Details

Acronym	FDC-GDMT
Visibility	Public
Description	Controlled terms and semantic properties used in the FAIR Data Collective's Gene
Status	Beta

View

Generic Dataset Metadata Template (GDMT) ⓘ

- > Resource Type ⓘ
- > Dataset Identifier ⓘ
- > Title (1 of N) ⓘ
- > Description (1 of N) ⓘ
- > Language ⓘ
- > Version ⓘ
- > Subjects and Keywords ⓘ
- > Creator (1 of N) ⓘ
- > Contributor (1 of N) ⓘ
- > Rights ⓘ
- > Date (1 of N) ⓘ
- > Funding (1 of N) ⓘ
- > Related Resources (1 of N) ⓘ

Metadata Schema – DCAT Vocabulary

Data Catalog Vocabulary (DCAT) - Version 2

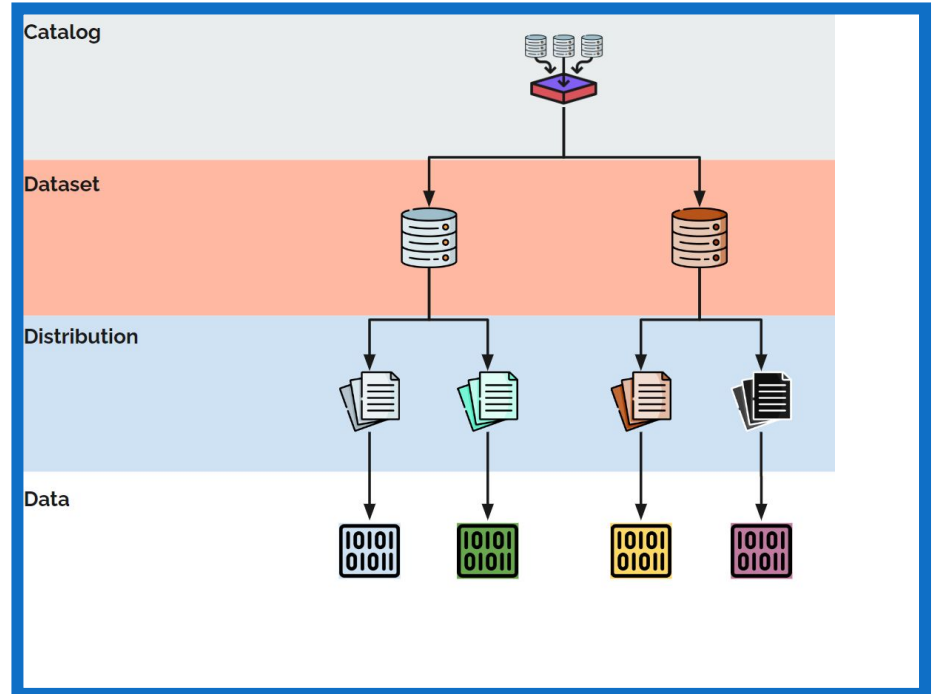
W3C Recommendation 04 February 2020



<https://www.w3.org/TR/vocab-dcat-2/>

DCAT Application Profile for data portals in Europe

DCAT-AP 2.1.0



The Trouble with Standards

So you've decided to adopt DCAT, and you made a form using the DCAT vocabularies. Great!

Now what if your users need to specify some additional metadata? Or some terms that aren't in the DCAT vocabulary, but are important subsets of it?

1 – Building machine-actionable MD templates using CEDAR



Image by [Dylan Leash](#) from [Pixabay](#)

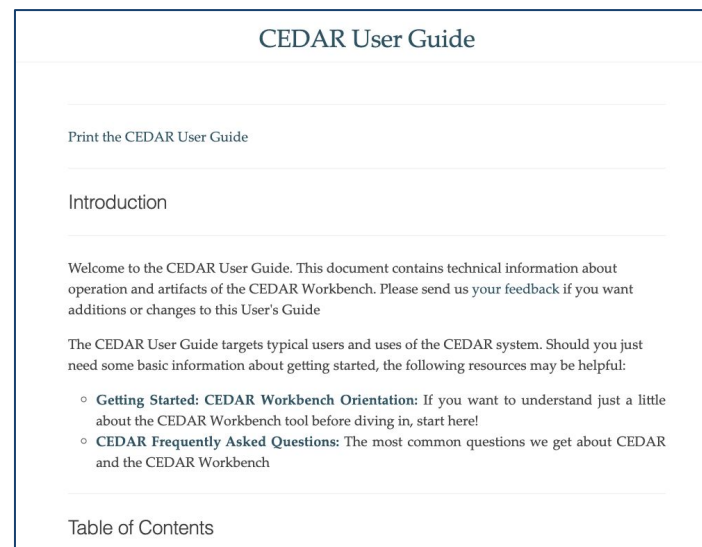
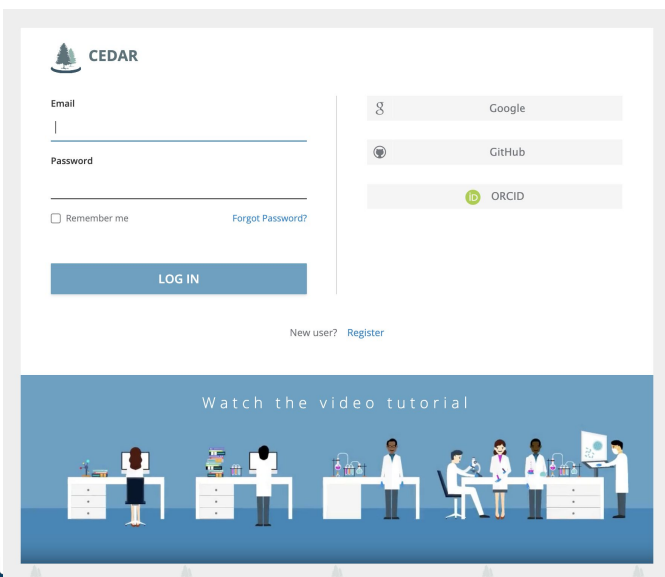
CEDAR

Courtesy of Barbara Magagna



CEDAR

- Web portal for developing templates and collecting metadata
- Human and machine readable output (RDF, JSON-LD)
- CEDAR account needed (cedar.metadatacenter.org — free!)
- Use the [CEDAR manual](http://metadatacenter.github.io/cedar-manual)
- Find many [CEDAR links](http://metadatacenter.org/references)



CEDAR metadata: the questions

Metadata schema = form to be filled in

Title*

Description

Publisher

Dataset Identifier

Catalog Identifier

CANCEL SAVE

JSON-LD

RDF

```
{
  "@context": {
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "pav": "http://purl.org/pav/",
    "schema": "http://schema.org/",
    "oslc": "http://open-services.net/ns/core#",
    "skos": "http://www.w3.org/2004/02/skos/core#",
    "rdfs:label": {
      "@type": "xsd:string"
    },
    ...
  },
  "datasetIdentifier": "http://www.w3.org/ns/dcat#dataset",
  "publisher": "http://purl.org/dc/terms/publisher",
  "catalogIdentifier": "http://www.w3.org/ns/dcat#catalog",
  "title": "http://purl.org/dc/terms/title",
  "description": "http://purl.org/dc/terms/description"
},
"datasetIdentifier": [
  {}
],
"publisher": {},
"catalogIdentifier": [
  {}
],
"title": {
  "@value": null
},
"description": {
  "@value": null
}
}
```

CEDAR form: human-readable

JSON-LD: machine-readable

CEDAR metadata: the answers

Drop-downs
Autocompletion
Recommendations

Controlled terms
Latest term updates



Scope

Focus Area

To which focus area of the ZonMw COVID-19 program does your project belong?*

Most projects have been classified by ZonMw in one of these focus areas. If you can not find your focus area, please contact us.

For all terms provided in the drop-down list it is useful to check the definitions in BioPortal <https://bioportal.bioontology.org/ontologies/ZONMW-CONTENT>.

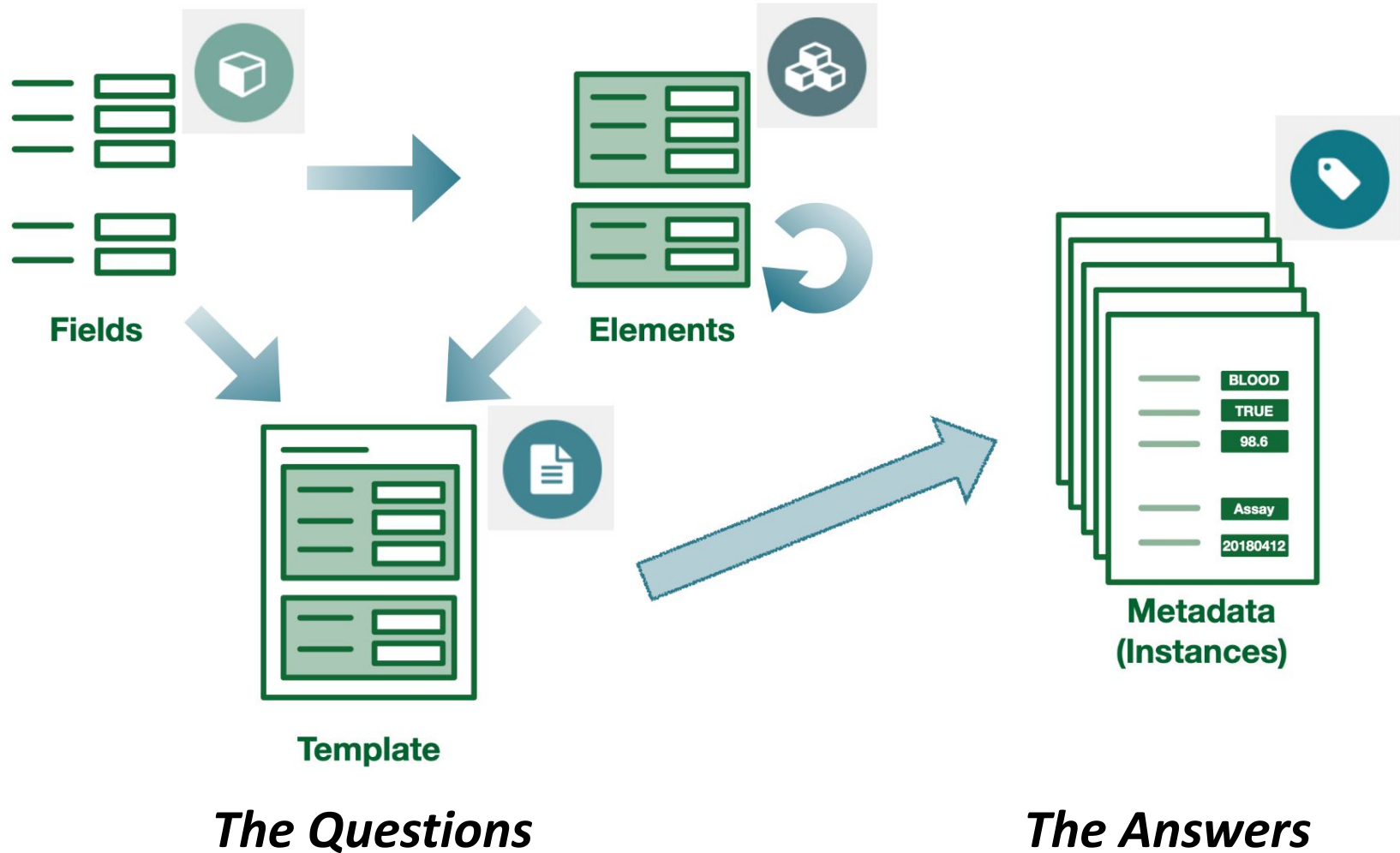
- care and prevention - care and prevention for vulnerable citizens
- care and prevention - organisation of care and prevention
- care and prevention - palliative care and bereavement
- care and prevention - transmission and epidemiology
- effects on society - impacts of measurements and strategies
- effects on society - impacts of measurements and strategies

Area Level(1..N)

Area Level - Multiple answers are allowed.

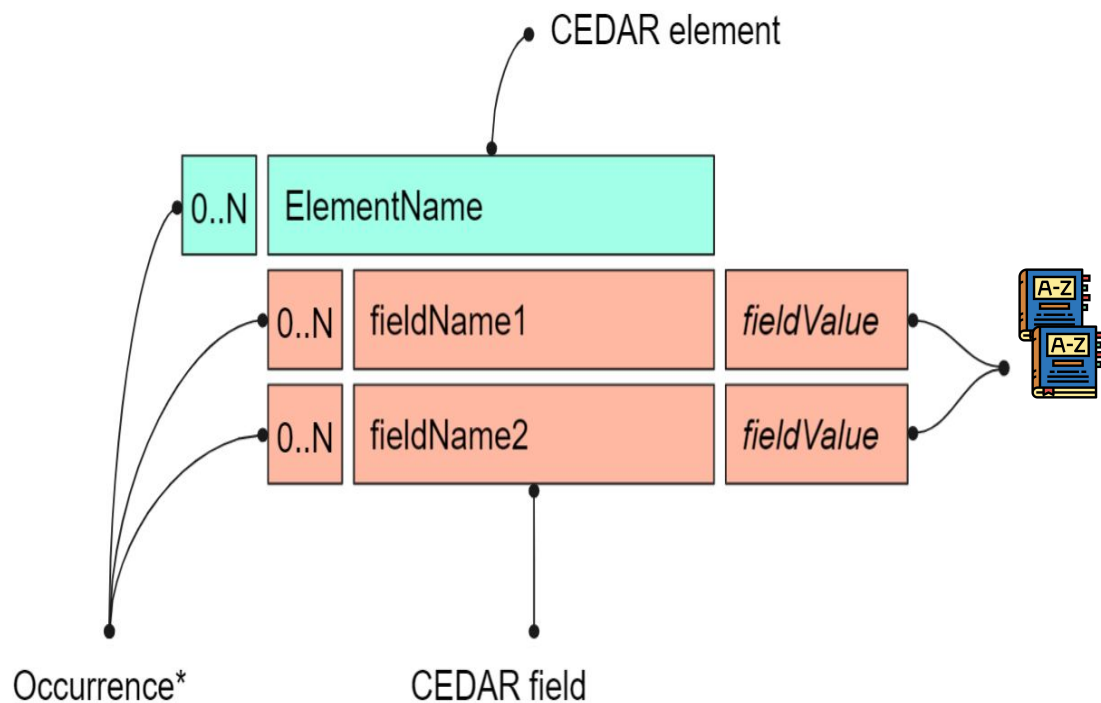
Disease (1..N)

CEDAR Artifact Composition: Templates, Elements, and Fields



CEDAR Templates, Elements and Fields

A CEDAR template is composed by elements and fields, which can be reused in different templates



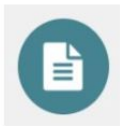
CEDAR artifacts



Metadata template **fields** (e.g., *title, description, start date*)



Metadata template **elements** (groups of fields and/or elements) (e.g., *Creator*)



Metadata **templates** (e.g., *GDMT*)

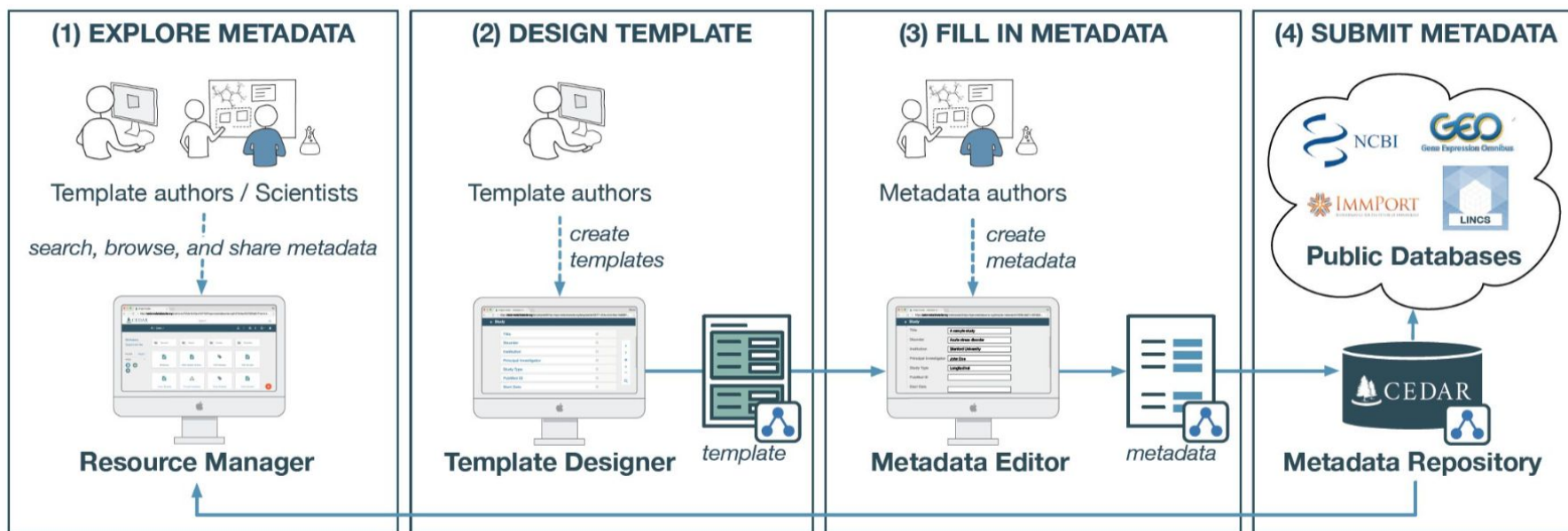


Metadata (aka template **instances**)

CEDAR system organization



CEDAR operations overview



cedar.metadatacenter.org

CEDAR metadata creation

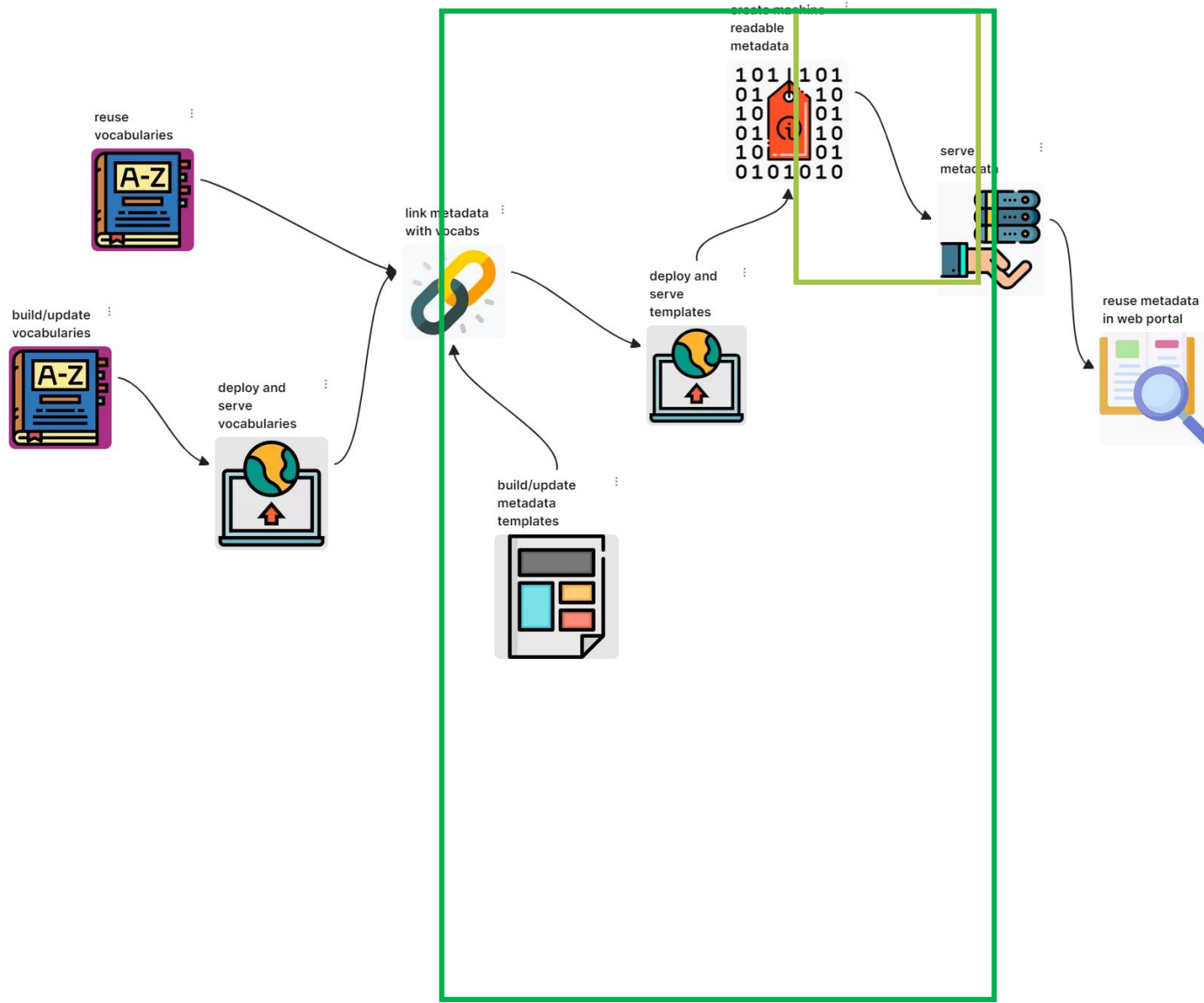
Barbara Magagna

see [guideline](#) and CEDAR [folder](#)



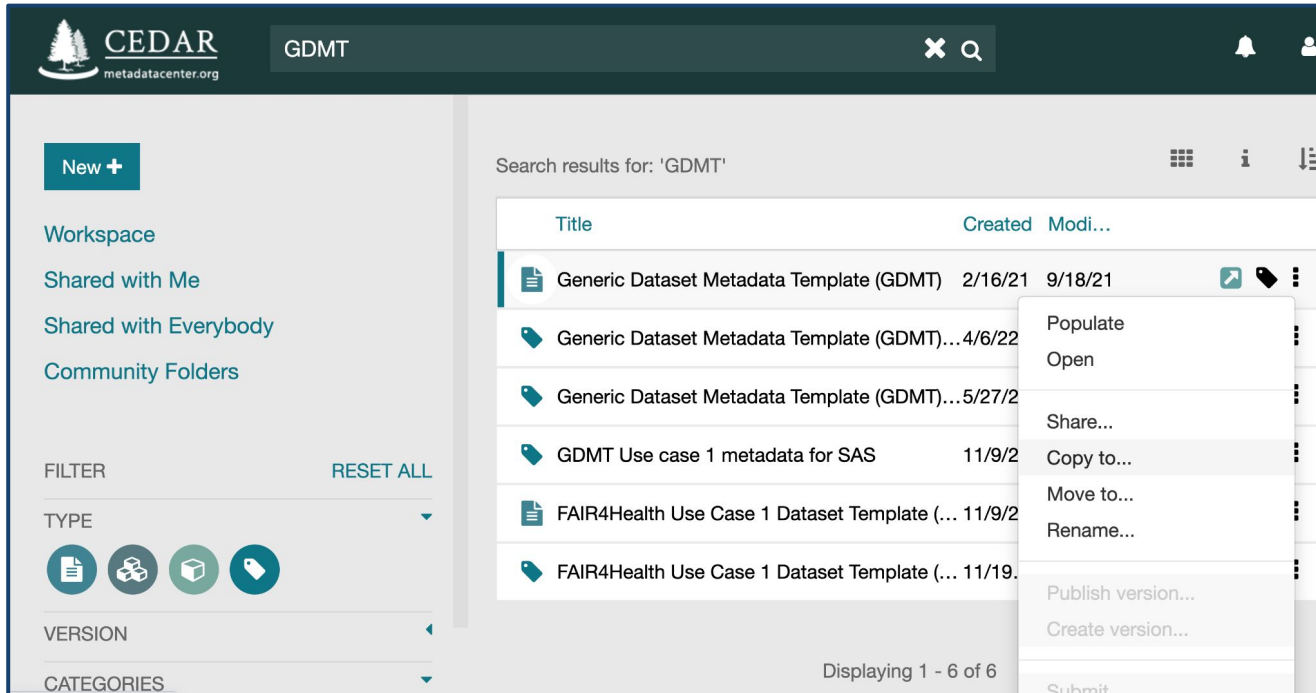
(Links are for M4M20 but interesting)

Create metadata



Step 1

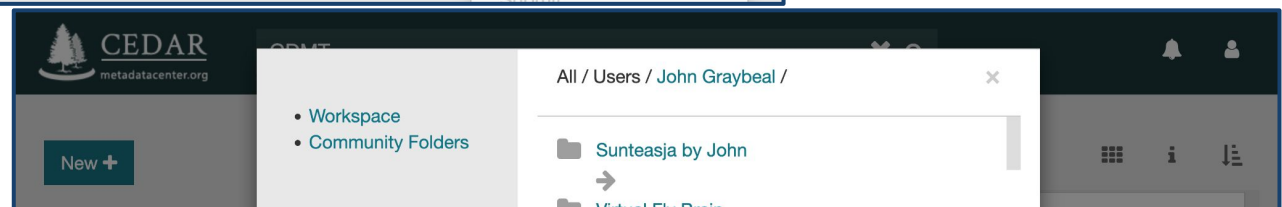
Create a copy of your favorite template (or GDMT)



Search results for: 'GDMT'

Title	Created	Modi...
Generic Dataset Metadata Template (GDMT)	2/16/21	9/18/21
Generic Dataset Metadata Template (GDMT)...4/6/22	4/6/22	
Generic Dataset Metadata Template (GDMT)...5/27/2	5/27/2	
GDMT Use case 1 metadata for SAS	11/9/2	
FAIR4Health Use Case 1 Dataset Template (... 11/9/2	11/9/2	
FAIR4Health Use Case 1 Dataset Template (... 11/19.	11/19.	

Displaying 1 - 6 of 6



All / Users / John Graybeal /

- Workspace
- Community Folders

Sunteasja by John

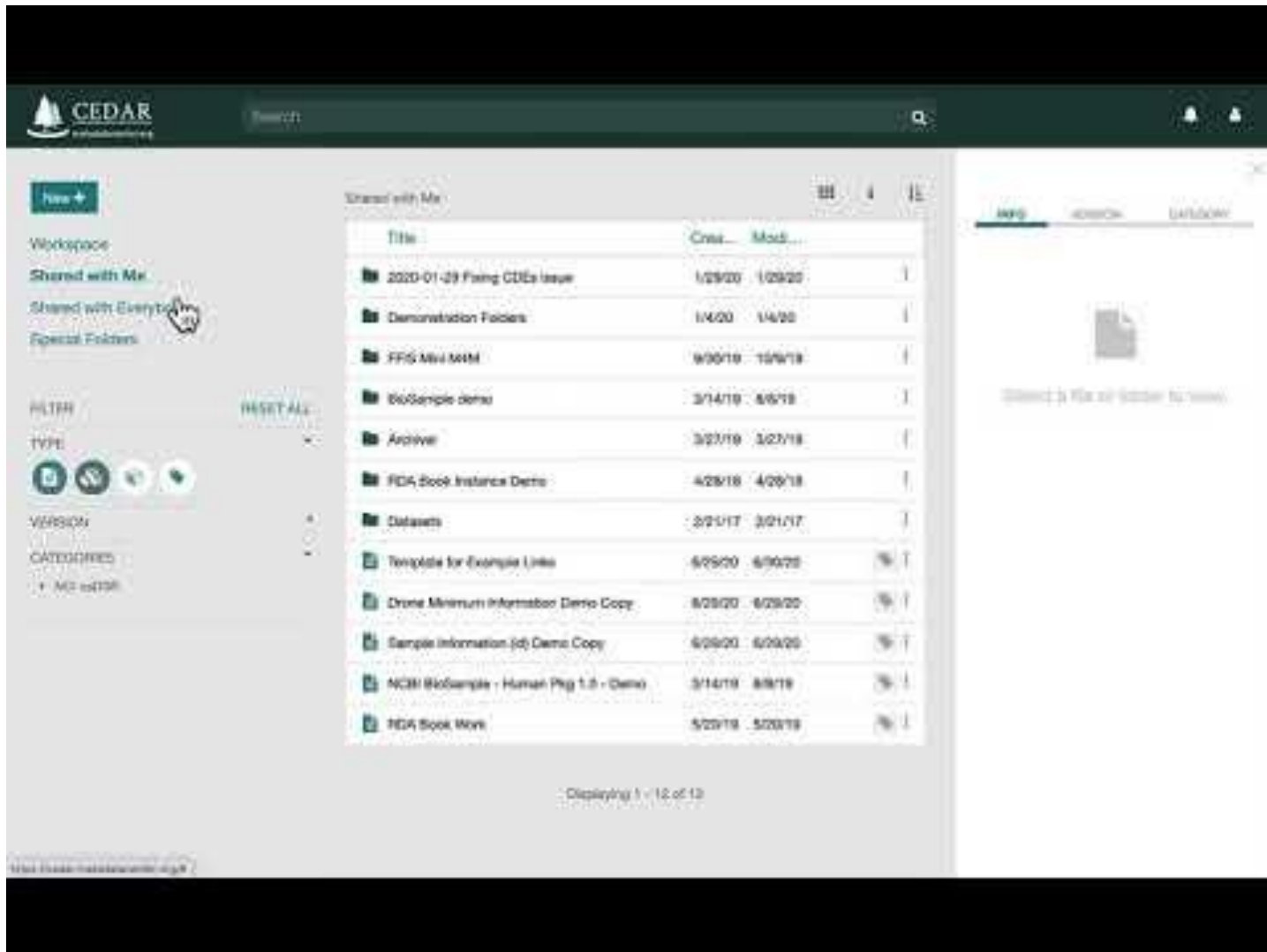
Step 1 Detailed Instructions

Create a copy of your favorite template (or GDMT)

- Log in to CEDAR
- Search for 'GDMT' in title bar
- Find the Generic Dataset Metadata Template in the list
- Use the drop-down 'kabob' menu on the right to select the Copy To option
- Choose to copy to your workspace
- Click on your Workspace icon to see your own files, and view the template to confirm it's real

Step 1 Basic Video 1: Find/Navigate

Create a copy of your favorite template (or GDMT)



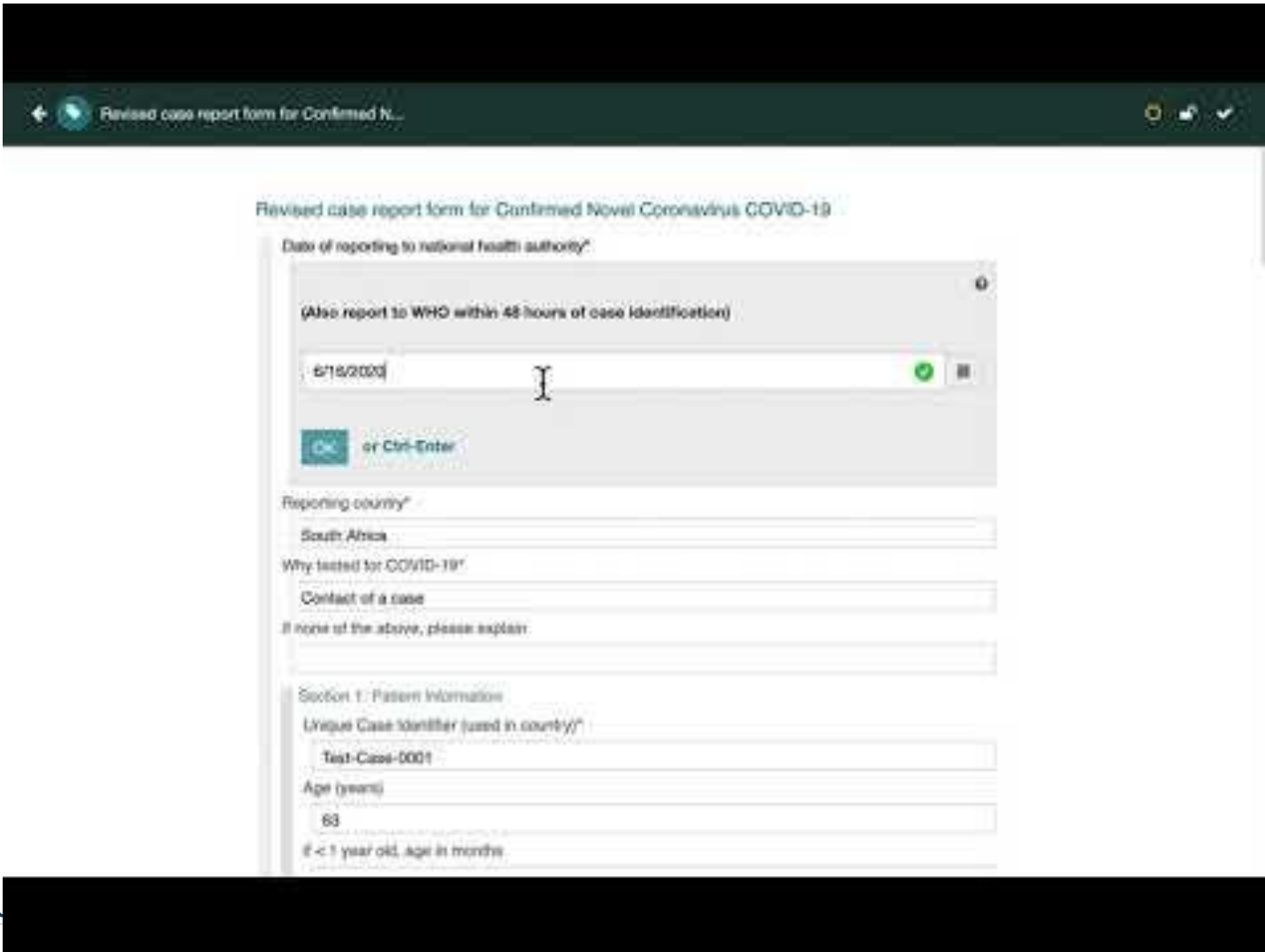
The screenshot shows the CEDAR web interface. The top navigation bar includes the CEDAR logo, a search bar, and notification icons. The main content area is titled "Shared with Me" and displays a table of files. The table has columns for "Title", "Created", "Modified", and "Actions". A sidebar on the left contains navigation options like "Workspace", "Shared with Me", and "Special Folders", along with filter sections for "TYPE" and "VERSION".

Title	Created	Modified	Actions
2020-01-29 Fxing CDEs Issue	1/29/20	1/29/20	
Demonstration Folders	1/4/20	1/4/20	
FFS Mini M4M	9/30/19	10/9/19	
EcSample demo	3/14/19	8/6/19	
Archive	3/21/19	3/21/19	
FDA Book Instance Demo	4/28/18	4/28/18	
Datasets	3/21/17	3/21/17	
Template for Example Links	6/29/20	6/29/20	%
Drone Minimum Information Demo Copy	6/29/20	6/29/20	%
Sample Information (d) Demo Copy	6/29/20	6/29/20	%
NCBI EcSample - Human Pig 1.2 - Demo	3/14/19	8/6/19	%
FDA Book Work	5/29/19	5/29/19	%

Displaying 1 - 12 of 13

Step 1 Basic Video 2: View

Create a copy of your favorite template (or GDMT)



Revised case report form for Confirmed Novel Coronavirus COVID-19

Date of reporting to national health authority*

(Also report to WHO within 48 hours of case identification)

6/16/2020

or Ctrl-Enter

Reporting country*

South Africa

Why tested for COVID-19*

Contact of a case

If none of the above, please explain:

Section 1: Patient Information

Unique Case Identifier (used in country)*

Test-Case-0001

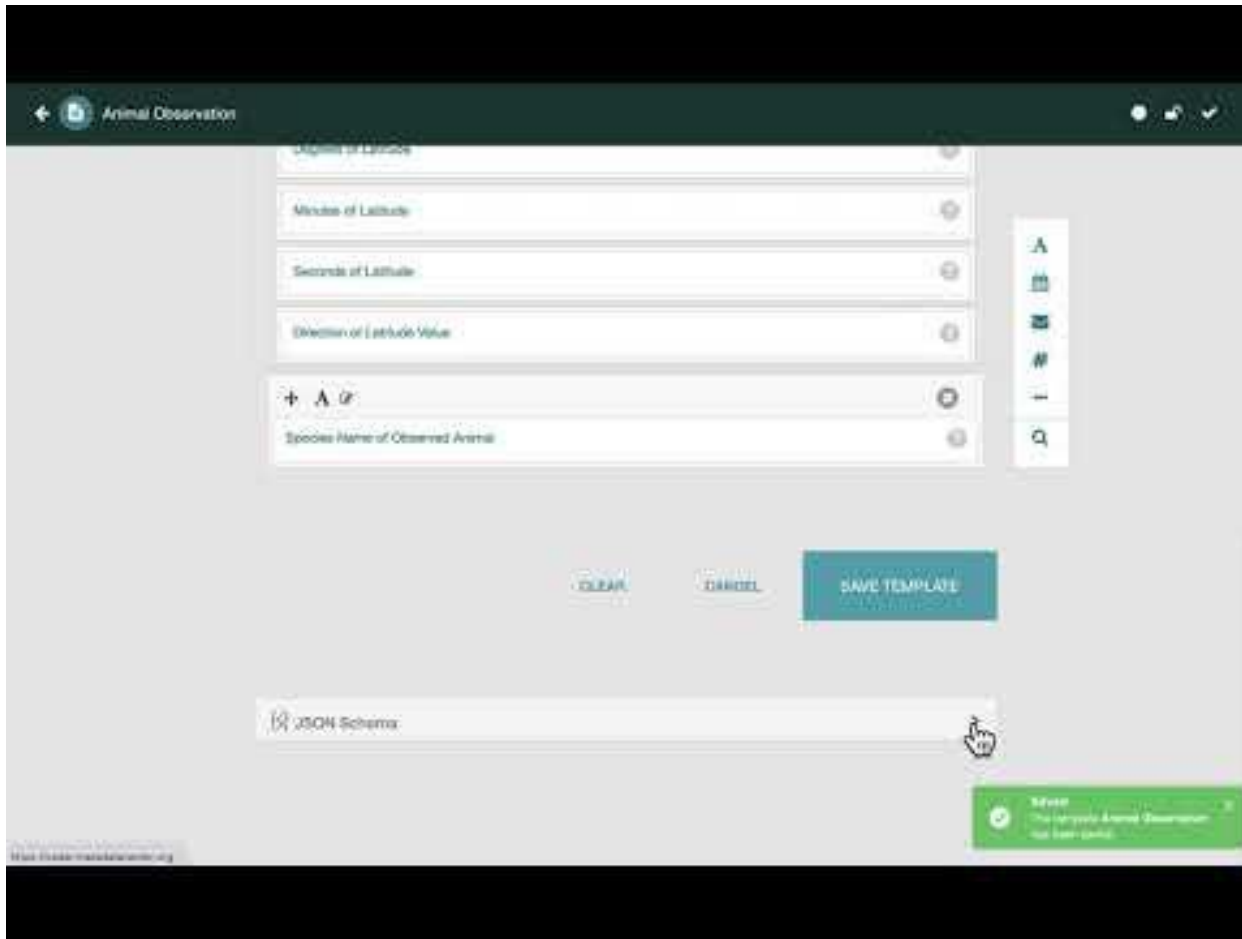
Age (years)

63

If < 1 year old, Age in months

Step 2

Create metadata fields and assign controlled vocabularies



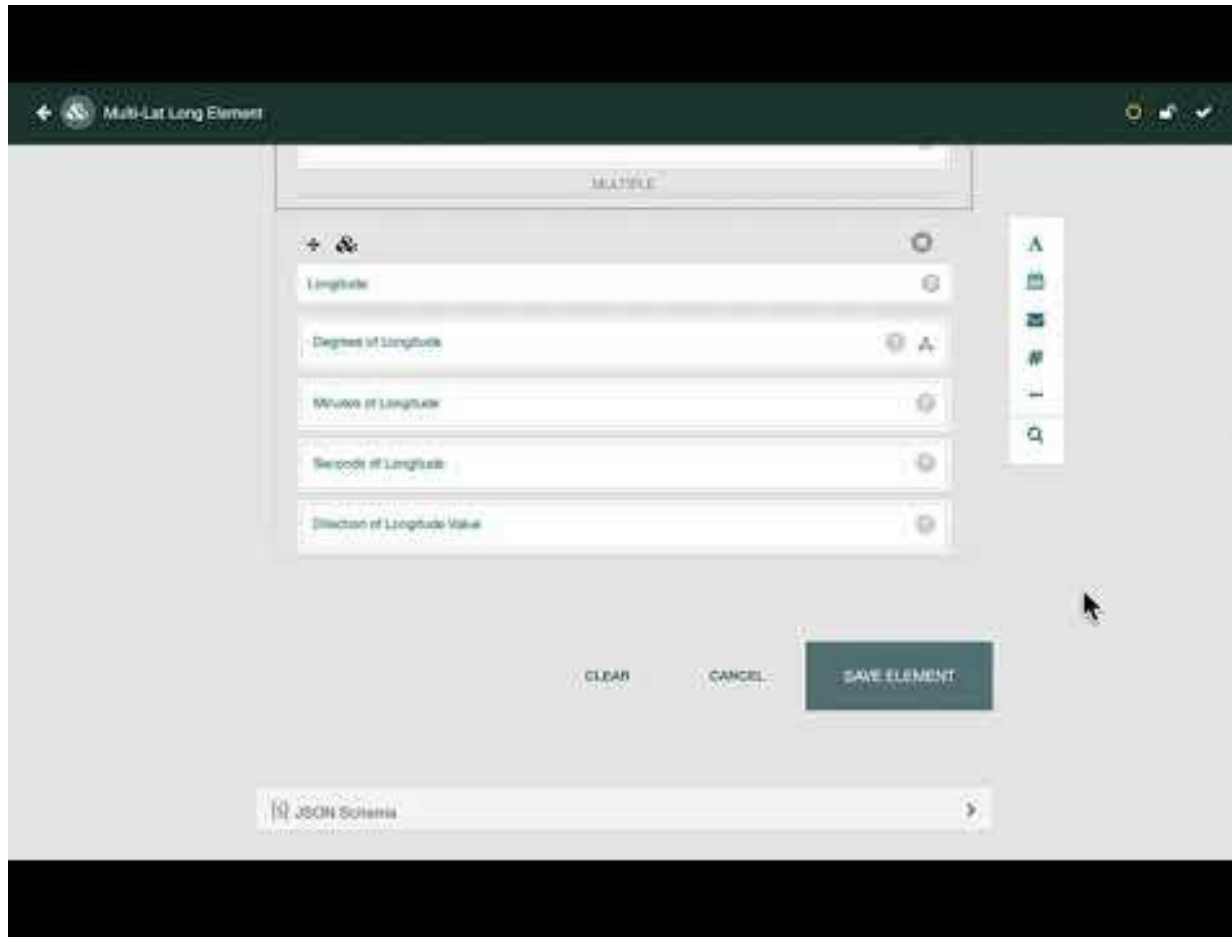
The screenshot shows the 'Animal Observation' metadata editor interface. The top bar includes a back arrow, a plus icon, and the title 'Animal Observation'. Below the title, there is a list of metadata fields: 'Minutes of Latitude', 'Seconds of Latitude', 'Direction of Latitude Value', and 'Species Name of Observed Animal'. A vertical toolbar on the right contains icons for adding, deleting, and searching fields. At the bottom, there are 'CLEAR', 'CANCEL', and 'SAVE TEMPLATE' buttons. A 'JSON Schema' field is visible at the bottom, and a green notification box in the bottom right corner states: 'Success The Animal Observation has been edited.'

Break



Step 3

Add your fields to metadata elements; add RDF properties



The screenshot shows the EarthCube metadata editor interface for a 'Multi-Lat Long Element'. The interface is titled 'Multi-Lat Long Element' and features a 'MULTIPLE' dropdown menu. Below this, there is a list of fields to be added to the metadata element:

- Longitude
- Degree of Longitude
- Mins of Longitude
- Seconds of Longitude
- Direction of Longitude Value

Each field has a plus icon and a gear icon. To the right of the list is a vertical toolbar with icons for adding, deleting, and searching. At the bottom of the interface, there are three buttons: 'CLEAR', 'CANCEL', and 'SAVE ELEMENT'. Below the buttons is a 'JSON Schema' dropdown menu.

Step 4

Make your final template by composing and replacing elements

Question: Is it better to build from the top down, or the bottom up? Why?

Step 5

Make machine-actionable metadata by filling out the form

The Trouble with Standards

Revisited

So you've decided to adopt DCAT, and you made a form using the DCAT vocabularies. Great!

Now what if your users need to specify some additional metadata? Or some terms that aren't in the DCAT vocabulary, but are important subsets of it?

Extending dataset metadata ... with additional metadata attributes

If you need to capture more metadata attributes, CEDAR gives you options!

- by modifying (a copy of) the template: you can add attributes into the form you created
- Or, if you included a repeating Attribute-Value field in your template, your users can add as many attributes as they want.

Extending dataset metadata ... with additional value selections

If you need to capture more metadata values:

- by modifying the template: you can extend the allowed terms with your own terms from your own vocabulary (use the VALUES tab)
- or, if they are attributes for all users of that vocabulary, you can just add them to the vocabulary, without changing the template
- You can add a placeholder empty vocabulary to preserve that option!



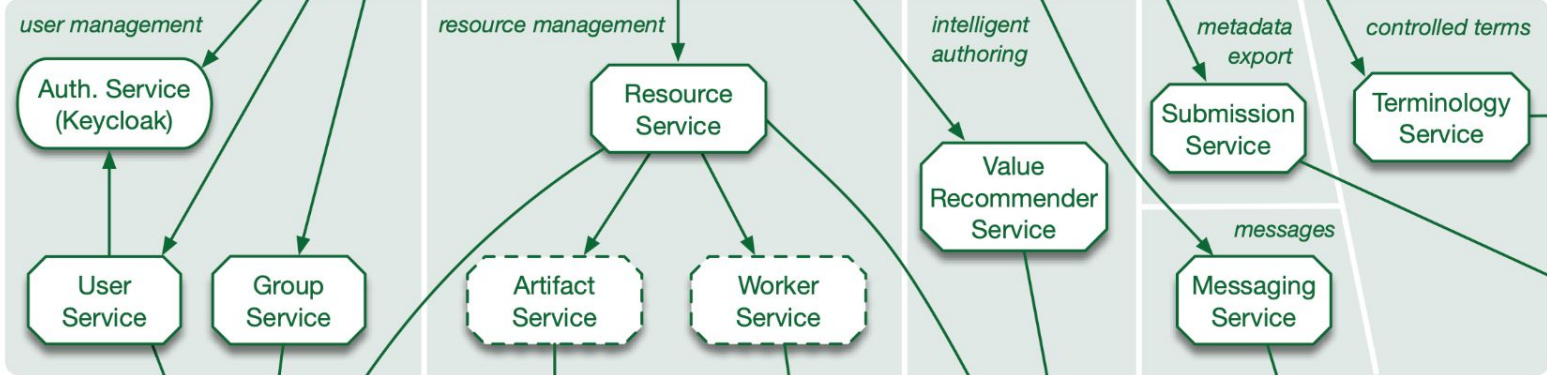
Integrations with CEDAR

What about scaling and integration?

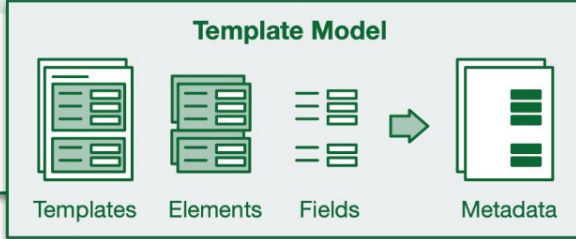
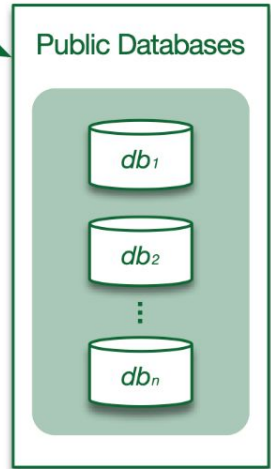
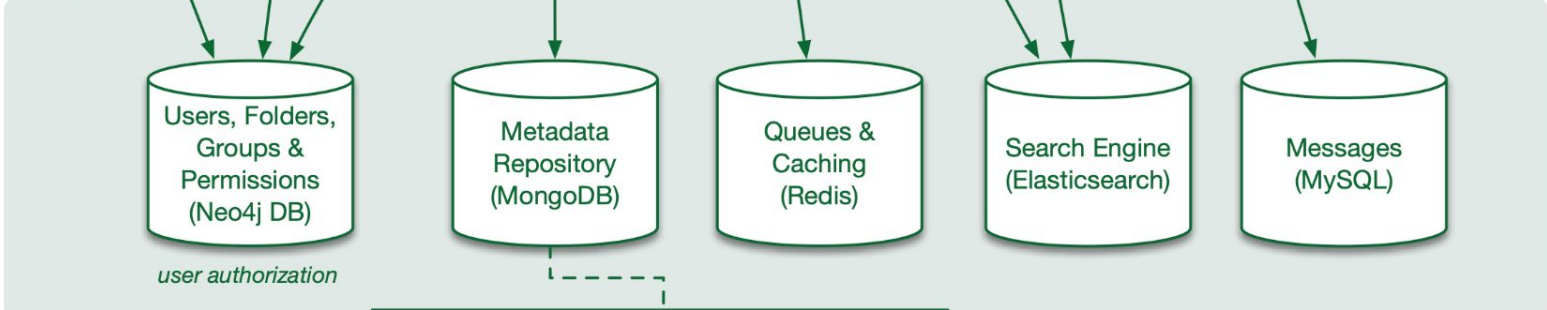
We're just coming to that, but first...



Open Services



Storage



- = CEDAR components
- = Third-party components
- = Only internal access



So many integration options

Here are some examples below. You can see the details at [this Google slides presentation](#)

- Swagger (API documentation on-line)
- Scripting with curl
- Viewing data in Google Sheets (Sheet script)
- Embedding in Applications (e.g., as front end)
- Exported as XML (we have conversion code)

4 - FAIR Orchestration and Integration with other systems

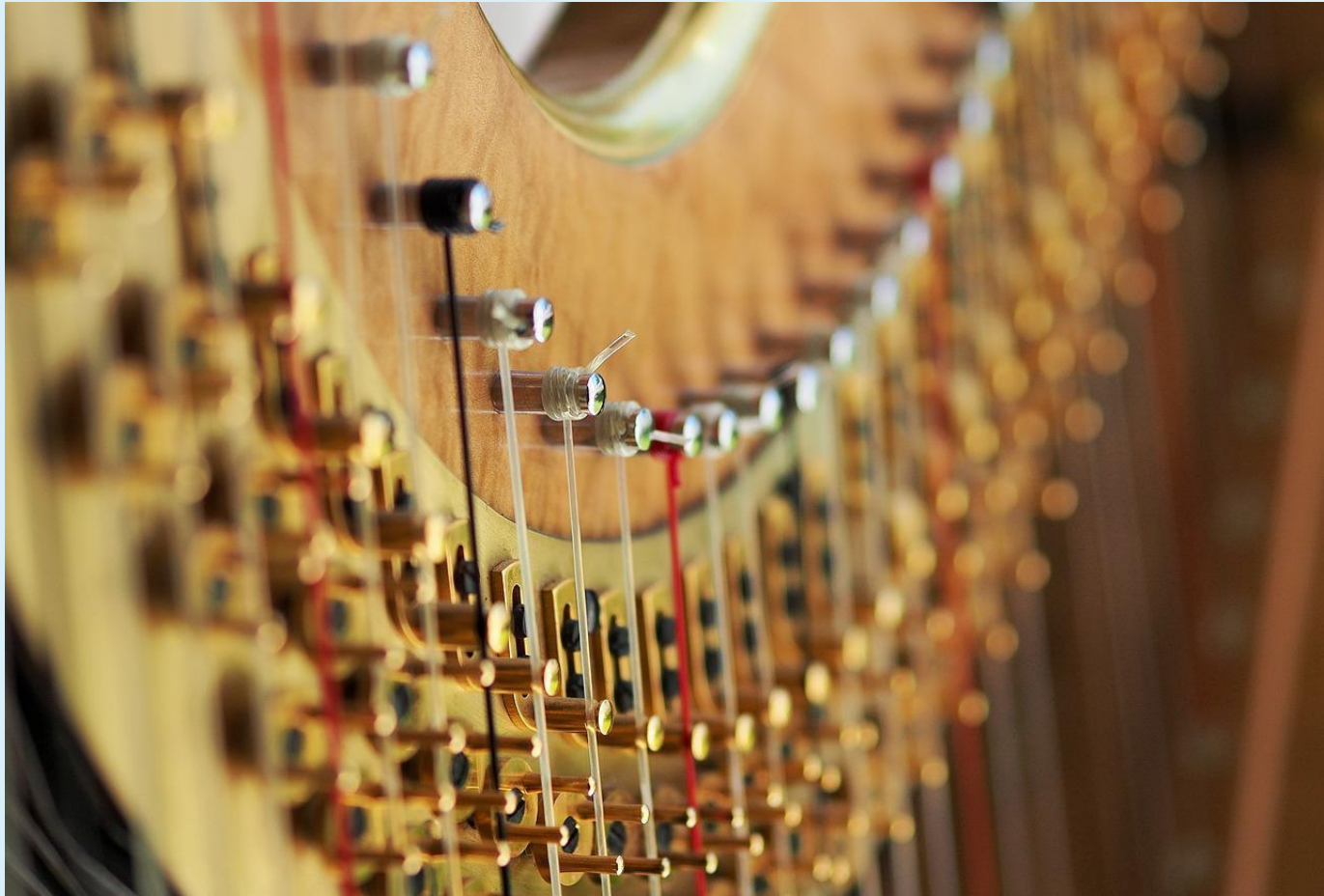
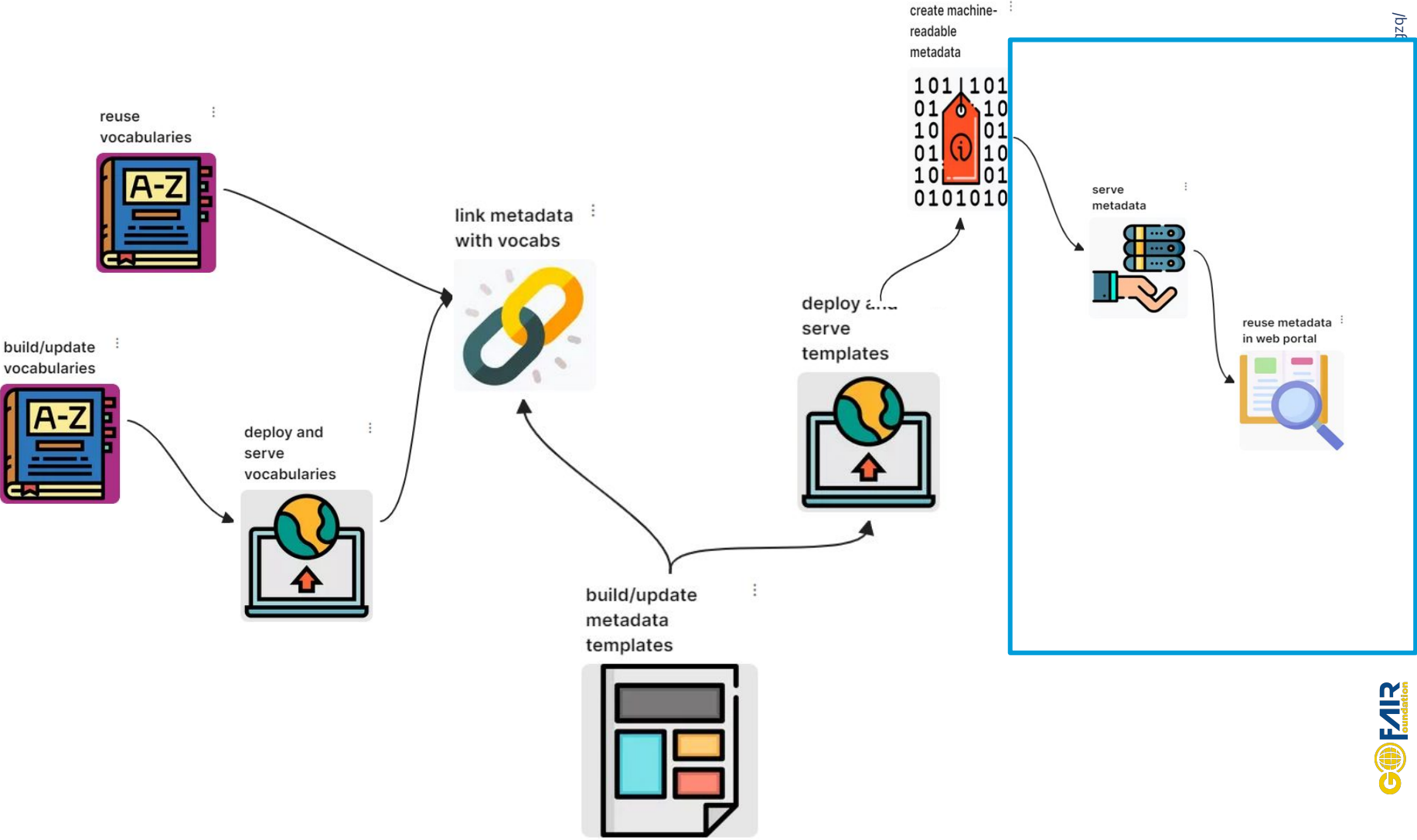


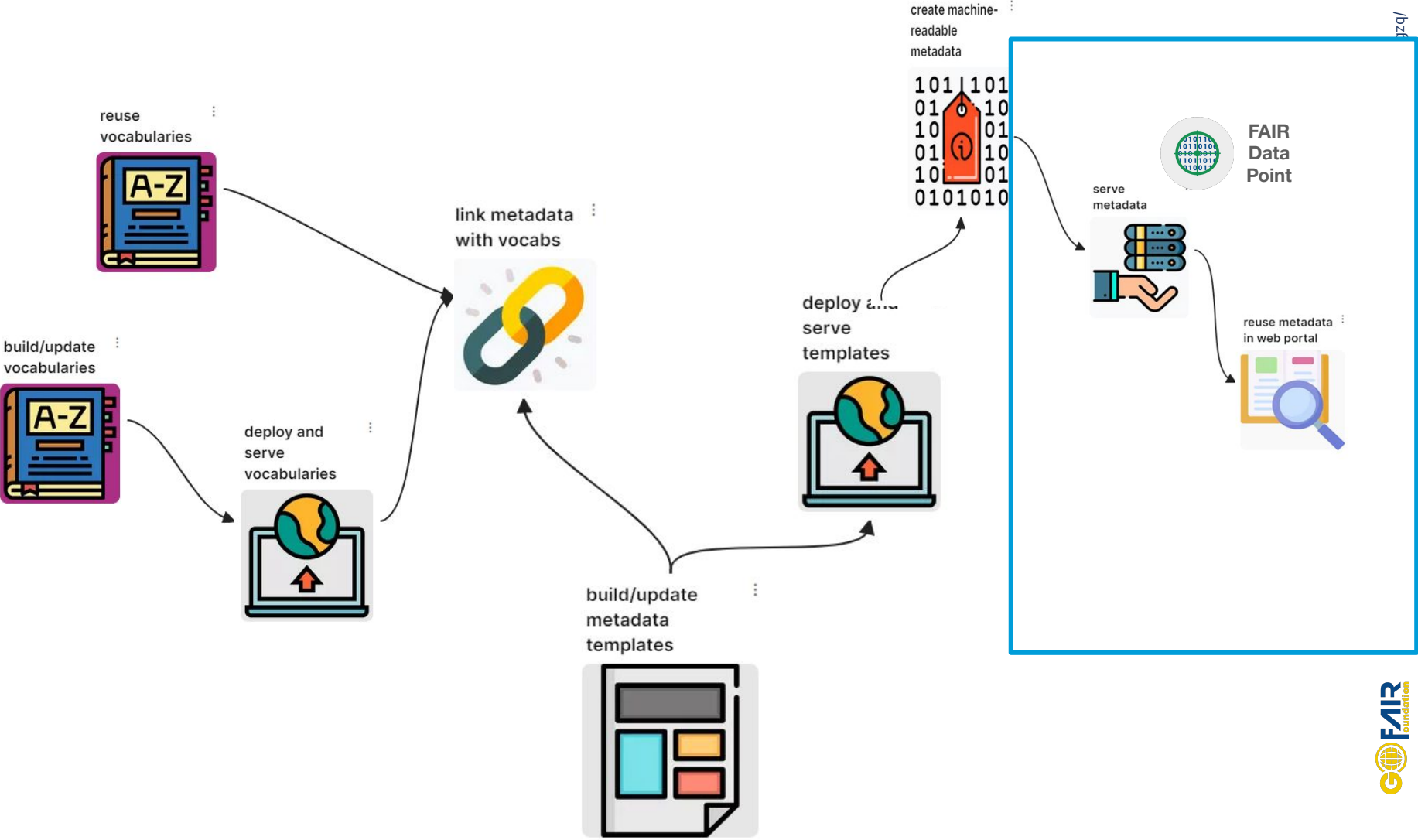
Image by [Andrew Martin](#) from [Pixabay](#)

FAIR Orchestration



g2q/

FAIR Orchestration



g2q/

The gateway to FAIR Orchestration

FAIR Data Point

<https://www.fairdatapoint.org>

FAIR Data Point About FAIR Data Point

FAIR Data Point

What is a FAIR Data Point

A FAIR Data Point (sometimes abbreviated to FDP) is the realisation of the vision of a group of authors of the [original paper on FAIR](#) on how (meta)data could be presented on the web using existing standards, and without the need of APIs.

A FAIR Data Point ultimately stores *information about data sets*, which is the definition of *metadata*. And just like the *webserver* in the WWW in the beginning of the 1990s brought the power of publishing text to anyone, a FAIR data point aims to give anyone the power of putting their own data on the web.

The system is called a **FAIR** data point because it takes care of a lot of the issues that need to be taken care of to make data FAIR; especially with the metadata needed for **Findability** and **Reusability**, and a uniform open way of **Accessing** the data. The FAIR data point also addresses the Interoperability of the metadata it stores, but it leaves the Interoperability aspects for the data itself to the data provider.

Components

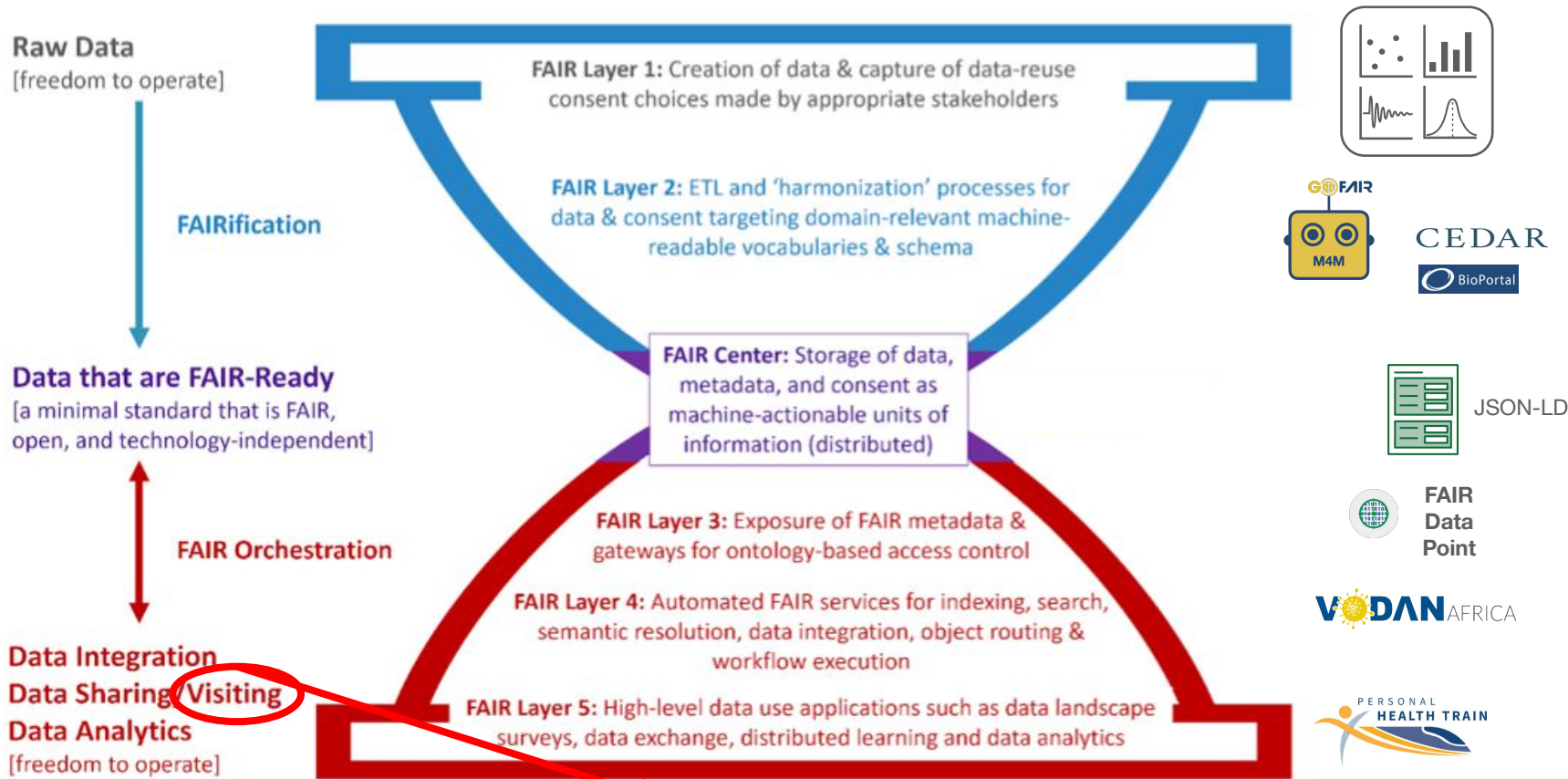
The FAIR Data Point as we have implemented it has three components.



Networks for FAIR data points: A web of data



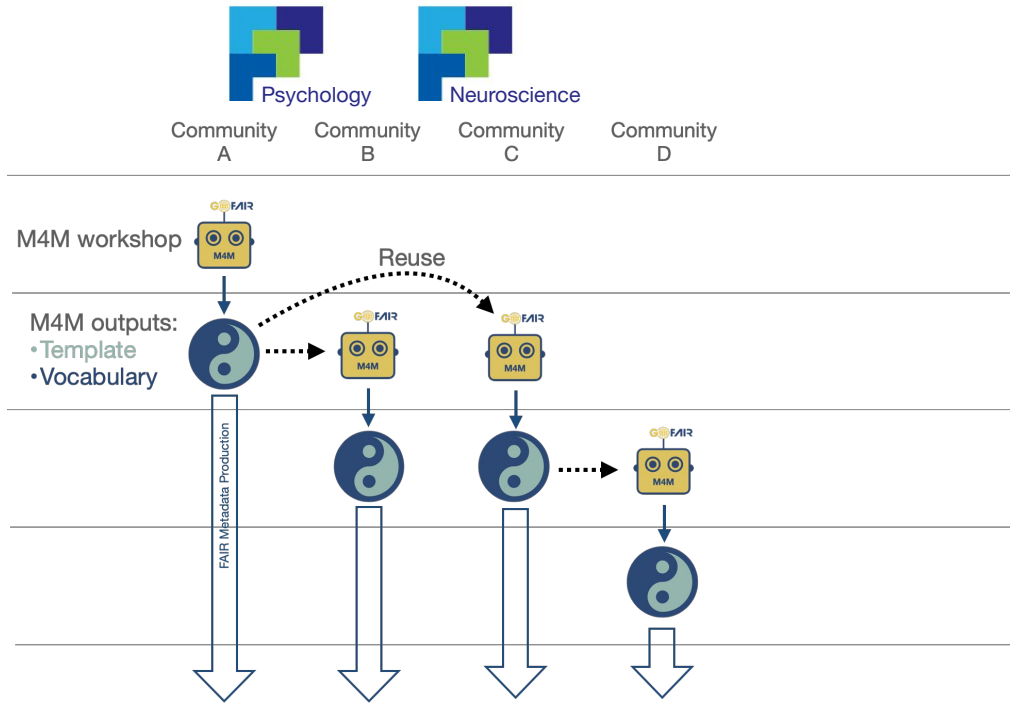
FAIR Hourglass



Your Post-M4M homework assignment
 SPHN Data Ecosystem for FAIR Data
<https://www.youtube.com/watch?v=pqV0qp4oisM>

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

Scaling machine-actionable metadata



Next steps - Scaling M4M

GO FAIR M4M workshop Progression for scaling machine-actionable metadata

Introduction M4M: FAIR vocab and metadata

Advanced M4M: vocabulary creation and governance

Advanced M4M: metadata creation and governance

Advanced M4M: variable descriptions using I-ADOPT

Capacity Building: would you like to lead the next M4M?



Summary - Day 2/ Next steps for you?

Acknowledgements

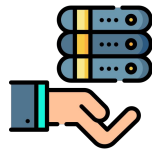


Icons made by:

<https://www.flaticon.com/authors/smalllikeart>



<https://www.flaticon.com/authors/mynampong>



<https://www.freepik.com>



<https://www.flaticon.com/authors/flat-icons>



Nikola Vasiljevic:

- 10.5281/zenodo.4705676
- 10.5281/zenodo.5624786



Thank you &
FAIR well

EarthCube Pilot Workshop evaluation form

<https://bit.ly/ECPilotM4MEvaluation>