DOI: 10.5281/zenodo.5031675

FarmConners

From Generic to Specific Metadata Template

Nikola Vasiljevic

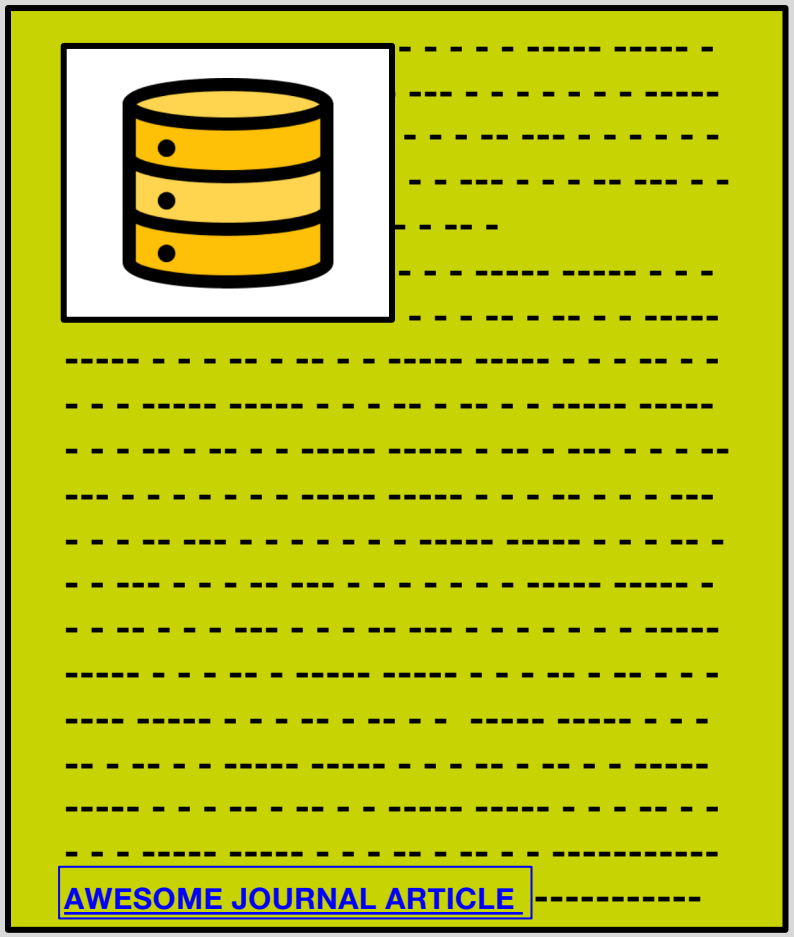
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857844.

FAIR Data Collective

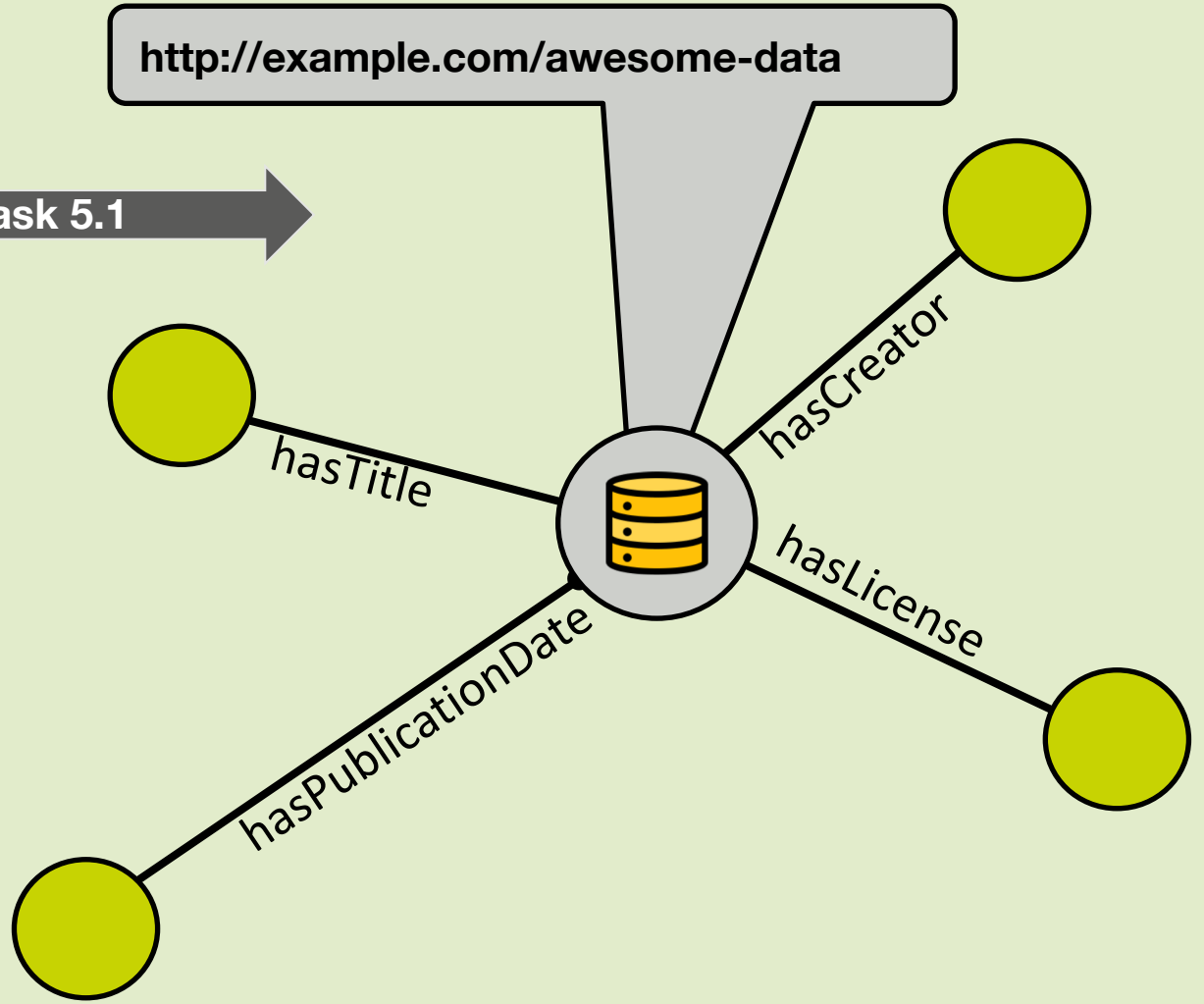
- A majority of slides curtesy of FAIR Data Collective with permission for adaptation

FarmConners

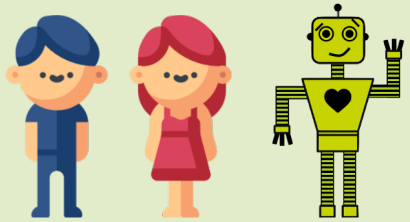
http://example.com/awesome-data



Transition in Task 5.1

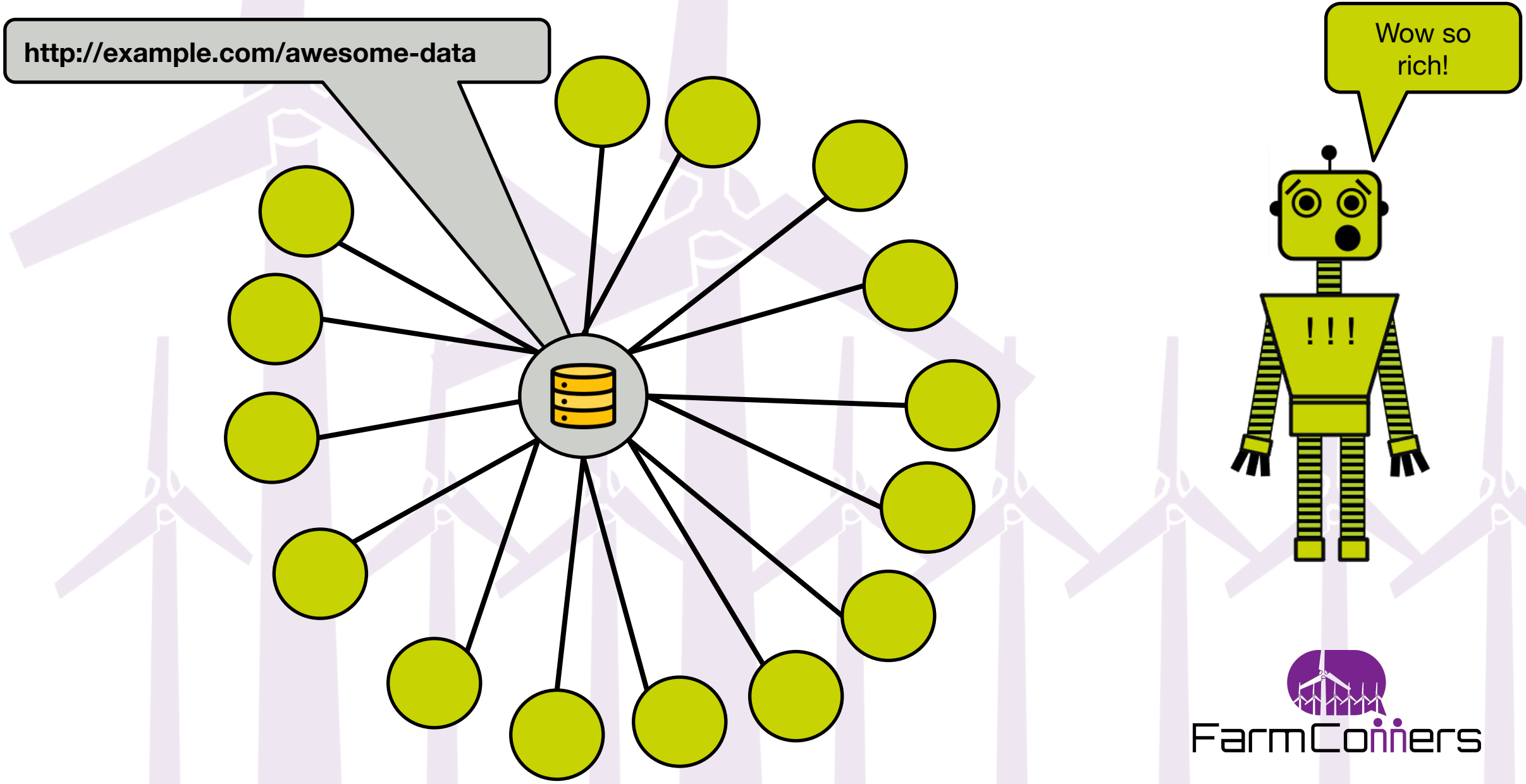


Human readable



Human and machine readable

*http://example.com/awesome-data is resolvable PID for metadata describing data



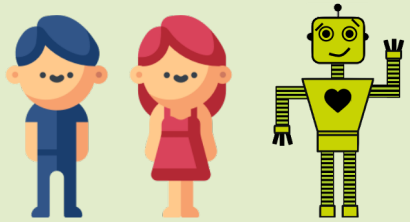
FarmConners

Enabling machine actionability

Field	Value
Title	*insert_title*
Creator	*insert_creator*
Publication date	*insert_publication_date*
License	*insert_license*
Subject	*insert_subject*
Variable	*insert_variable*
...	...

Human readable metadata template

Field	Value
http://purl.org/dc/elements/1.1/title	Free text
http://purl.org/dc/elements/1.1/creator	URL representing ORCID ID
http://purl.org/gdmt/hasDatasetDate	datetime string
http://purl.org/dc/elements/1.1/rights	**https://spdx.org/licenses/**
http://purl.org/dc/elements/1.1/subject	**http://purl.org/neat**
http://purl.org/gdmt/hasVariableInfo	**http://purl.org/aspect**
...	...



Human and machine readable metadata template

Enabling machine actionability

Field	Value
Title	*insert_title*
Creator	*insert_creator*
Publication date	*insert_publication_date*
License	*insert_license*
Subject	*insert_subject*
Variable	*insert_variable*
...	...

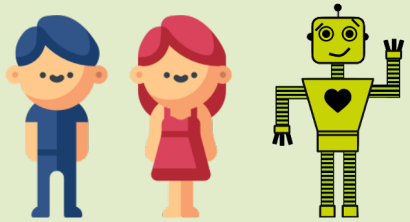
Human readable metadata template

Field	Value
http://purl.org/dc/elements/1.1/title	Free text
[illegible]	URL representing ORCID ID
[illegible]	datetime string
[illegible]	https://spdx.org/licenses/
[illegible]	http://purl.org/neat
[illegible]	http://purl.org/aspect
...	...

SPDX License List

NEAT: wi**N**d **E**nergy t**A**xonomy of **T**opics

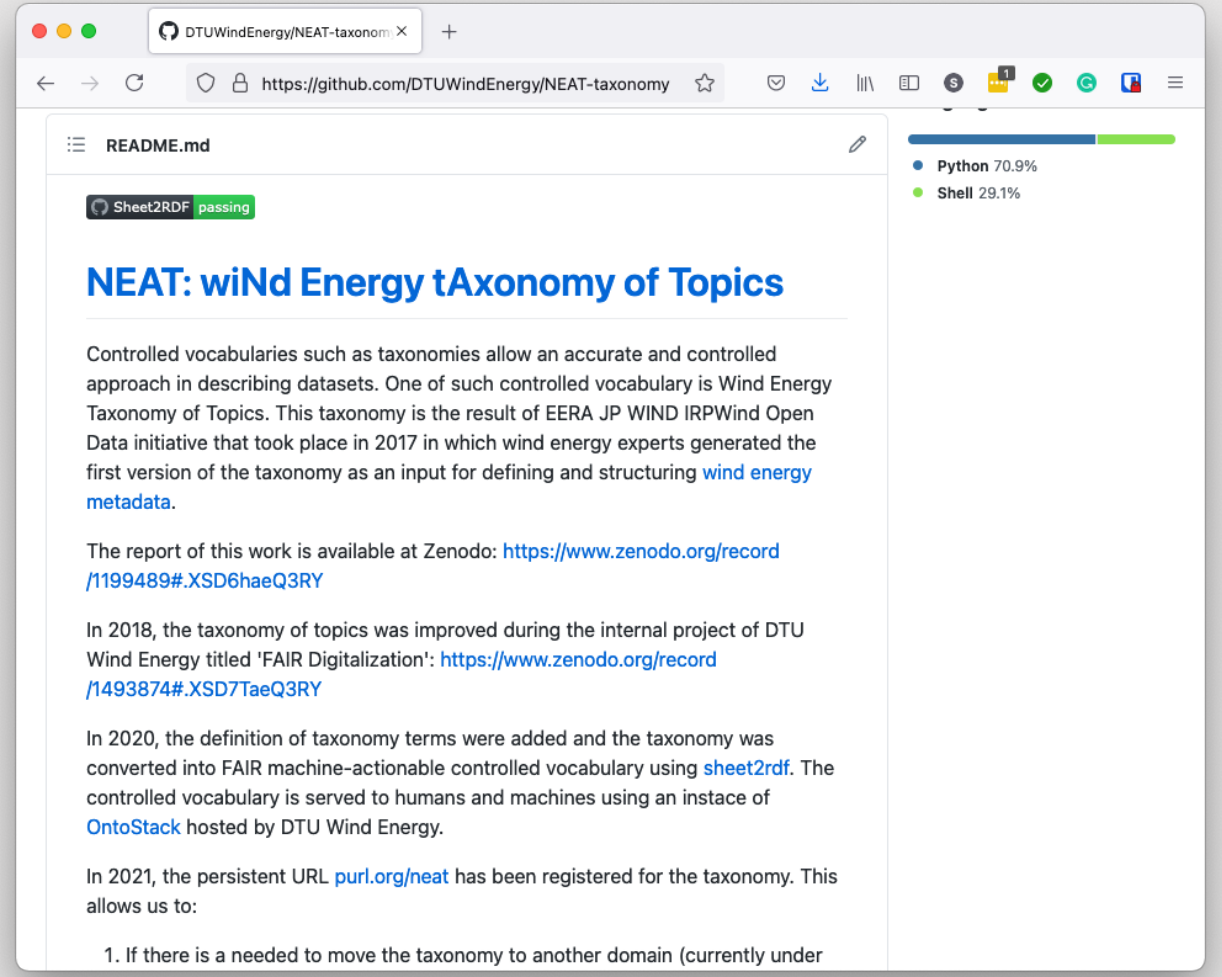
ASPECT: wind energy v**A**riable**S** **P**aramet**E**rs **C**ons**T**ants



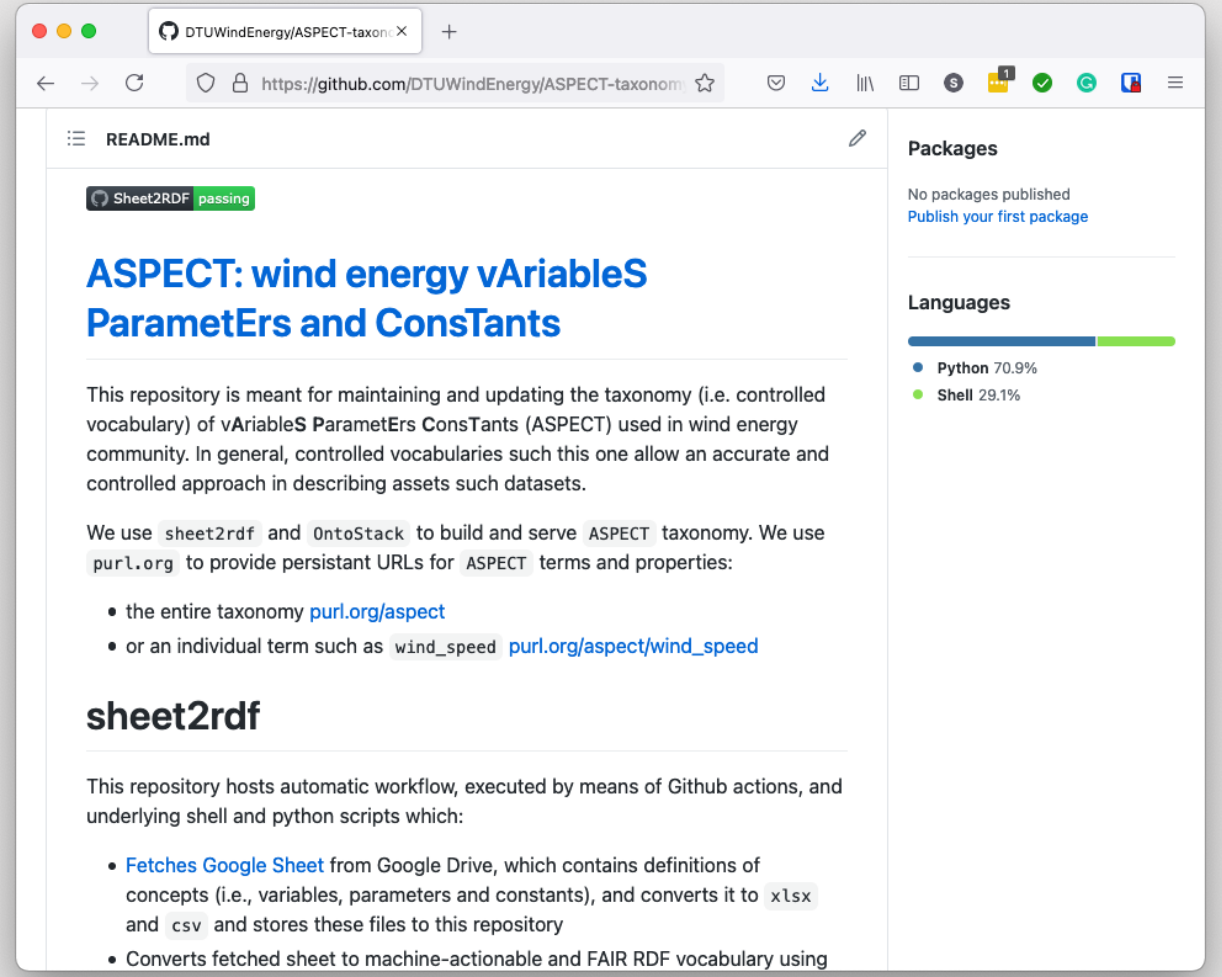
Human and machine readable metadata template

Vocabularies built and maintained on GitHub using **sheet2rdf** by

FAIR Data Collective



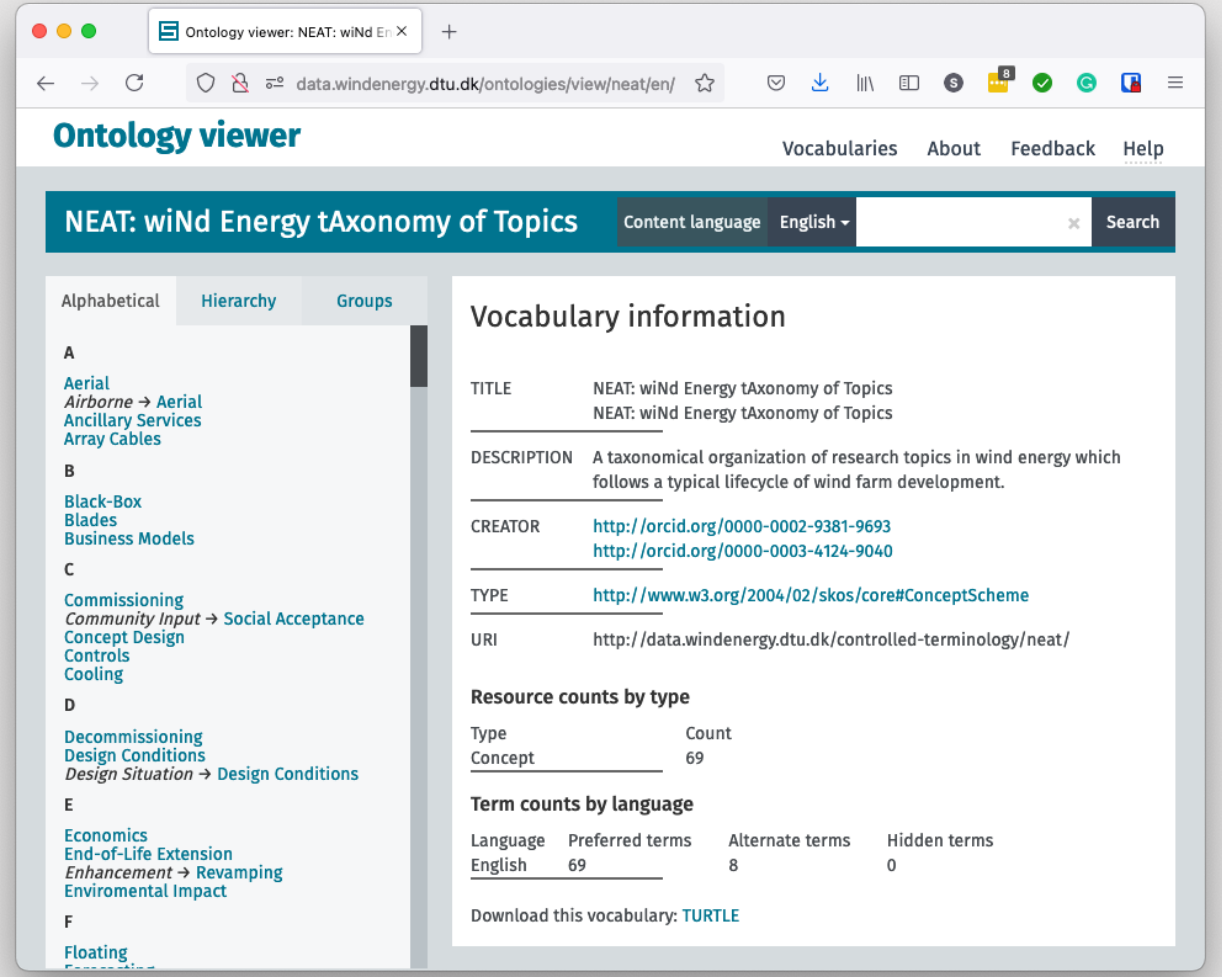
https://github.com/DTUWindEnergy/**NEAT**-taxonomy



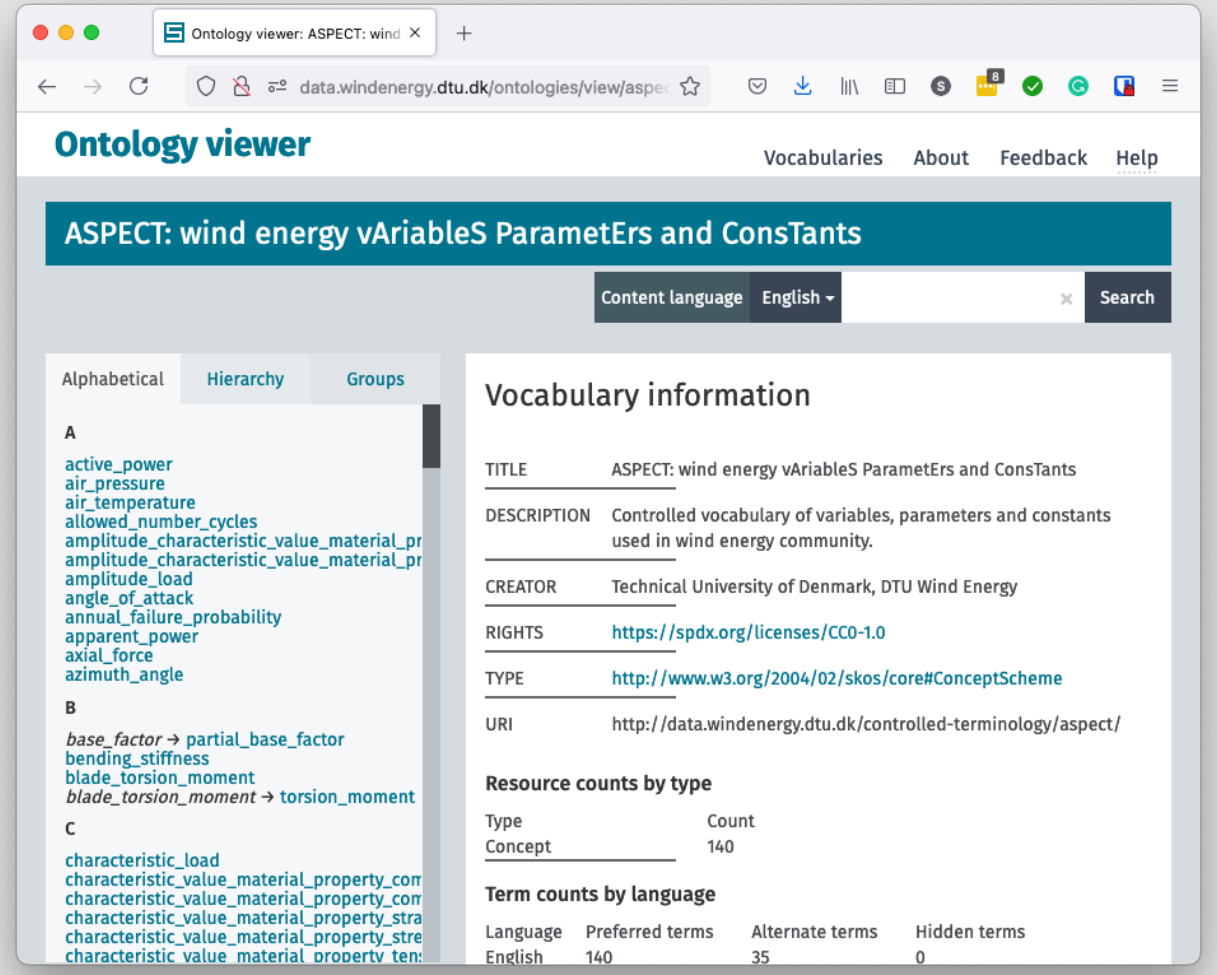
https://github.com/DTUWindEnergy/**ASPECT**-taxonomy

Vocabularies resolved via **OntoStack** by

FAIR Data Collective

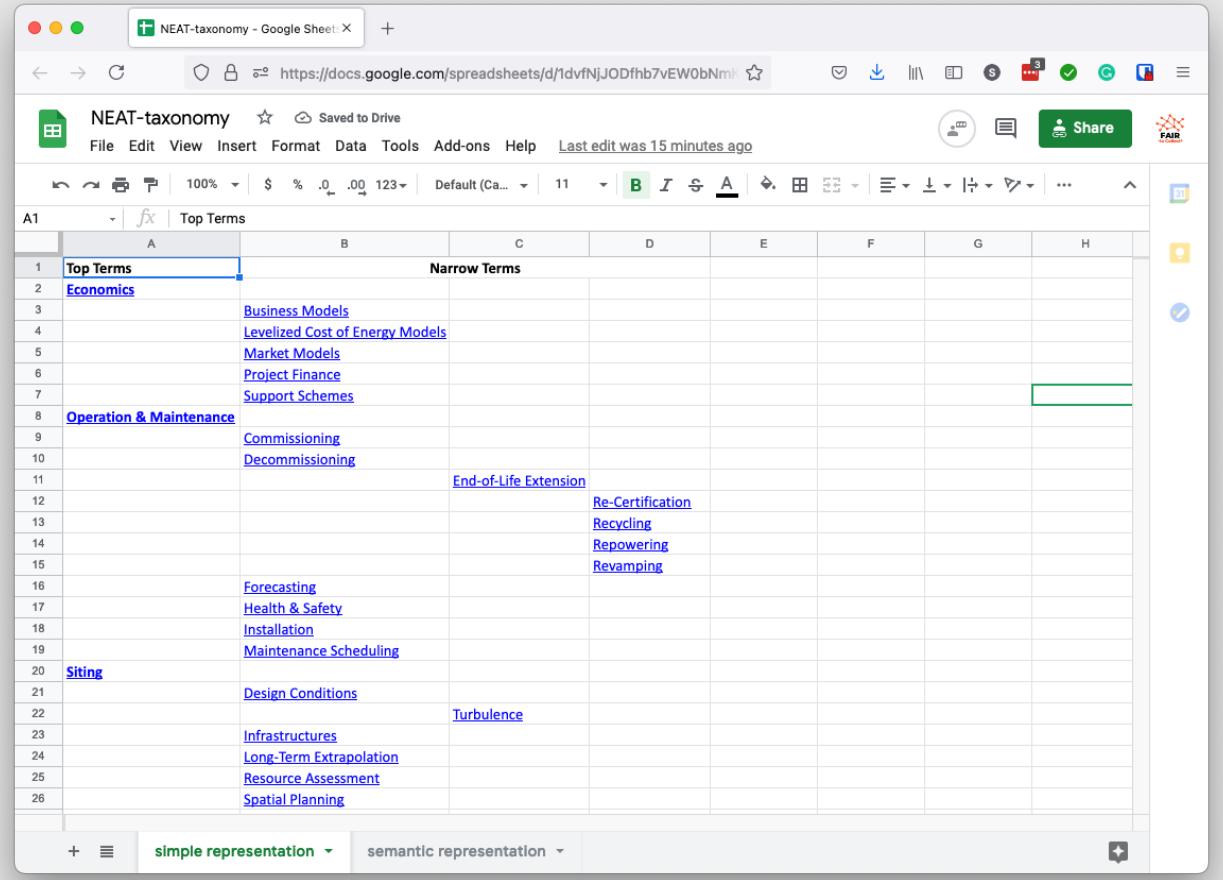


http://purl.org/neat

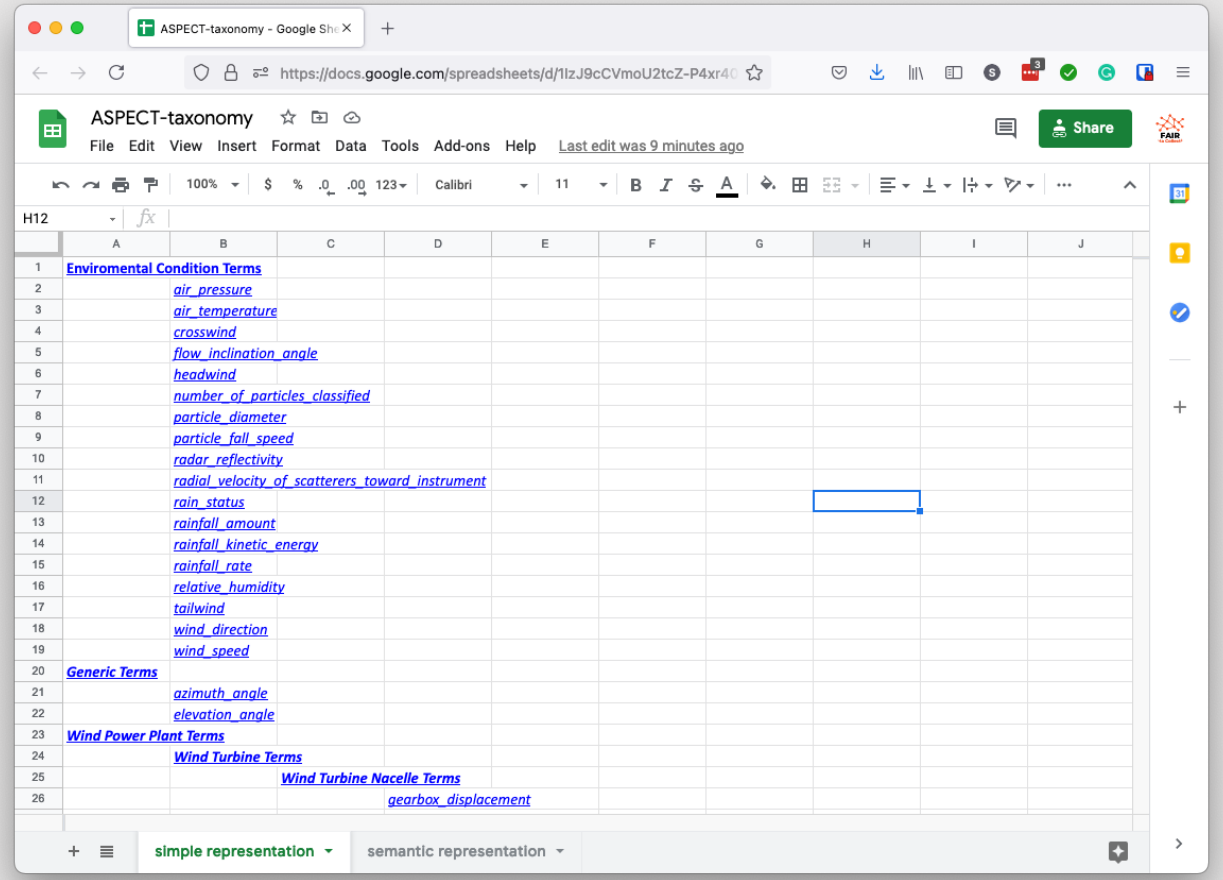


http://purl.org/aspect

Vocabularies for review on Google Drive / **simple representation**

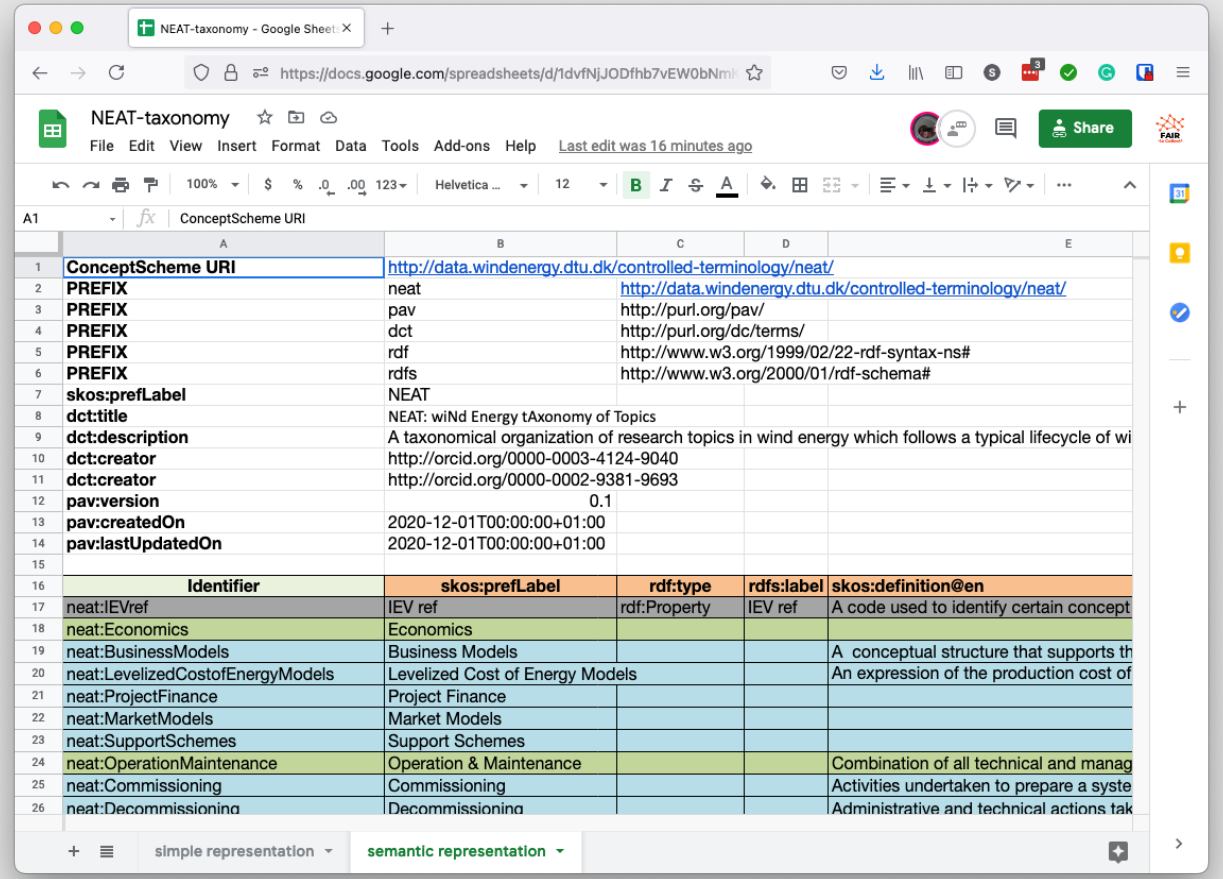


NEAT taxonomy



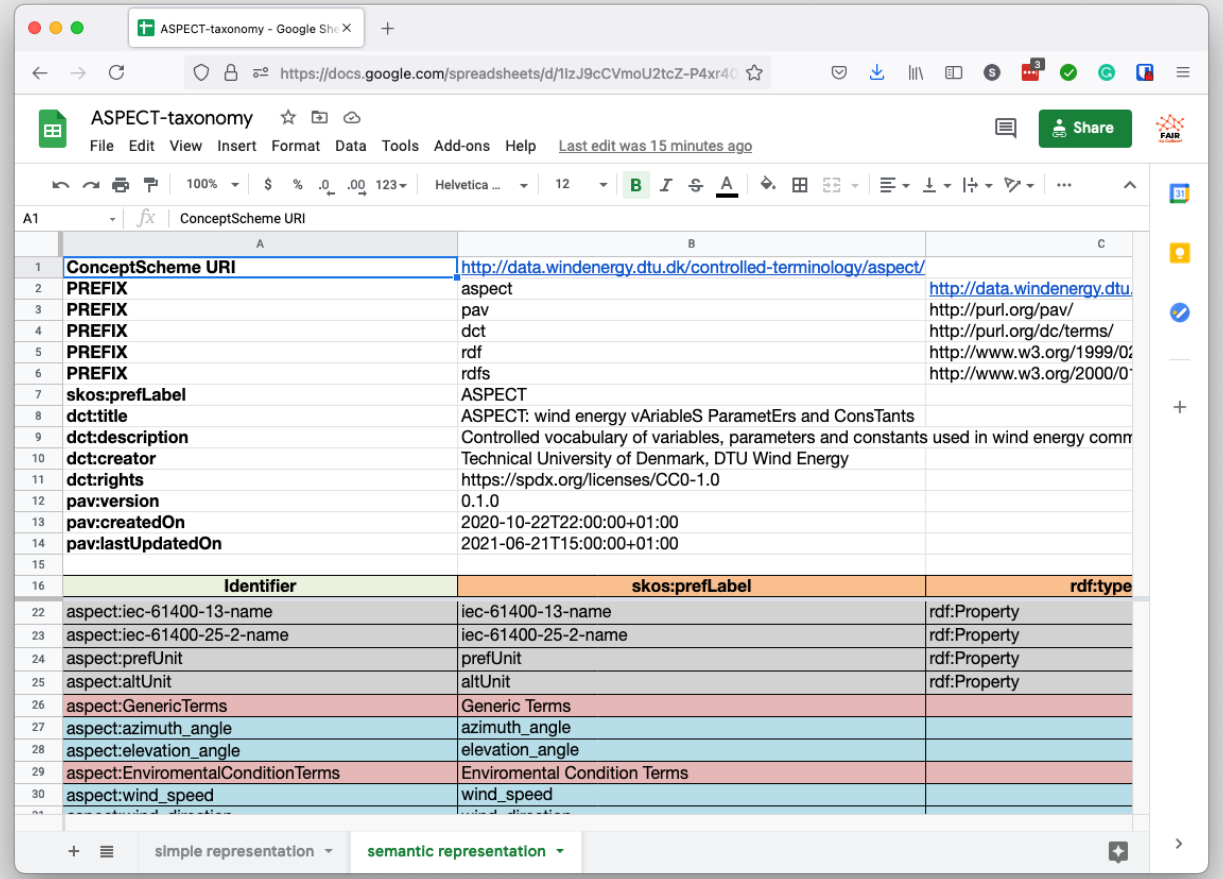
ASPECT taxonomy

Vocabularies for review on Google Drive / **semantic representation**



	A	B	C	D	E
1	**ConceptScheme URI**	http://data.windenergy.dtu.dk/controlled-terminology/neat/			
2	**PREFIX**	neat	http://data.windenergy.dtu.dk/controlled-terminology/neat/		
3	**PREFIX**	pav	http://purl.org/pav/		
4	**PREFIX**	dct	http://purl.org/dc/terms/		
5	**PREFIX**	rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#		
6	**PREFIX**	rdfs	http://www.w3.org/2000/01/rdf-schema#		
7	**skos:prefLabel**	NEAT			
8	**dct:title**	NEAT: wiNd Energy tAxonomy of Topics			
9	**dct:description**	A taxonomical organization of research topics in wind energy which follows a typical lifecycle of wi			
10	**dct:creator**	http://orcid.org/0000-0003-4124-9040			
11	**dct:creator**	http://orcid.org/0000-0002-9381-9693			
12	**pav:version**	0.1			
13	**pav:createdOn**	2020-12-01T00:00:00+01:00			
14	**pav:lastUpdatedOn**	2020-12-01T00:00:00+01:00			
15					
16	**Identifier**	**skos:prefLabel**	**rdf:type**	**rdfs:label**	**skos:definition@en**
17	neat:IEVref	IEV ref	rdf:Property	IEV ref	A code used to identify certain concept
18	neat:Economics	Economics			
19	neat:BusinessModels	Business Models			A conceptual structure that supports th
20	neat:LevelizedCostofEnergyModels	Levelized Cost of Energy Models			An expression of the production cost of
21	neat:ProjectFinance	Project Finance			
22	neat:MarketModels	Market Models			
23	neat:SupportSchemes	Support Schemes			
24	neat:OperationMaintenance	Operation & Maintenance			Combination of all technical and manag
25	neat:Commissioning	Commissioning			Activities undertaken to prepare a syste
26	neat:Decommissioning	Decommissioning			Administrative and technical actions tak

NEAT taxonomy



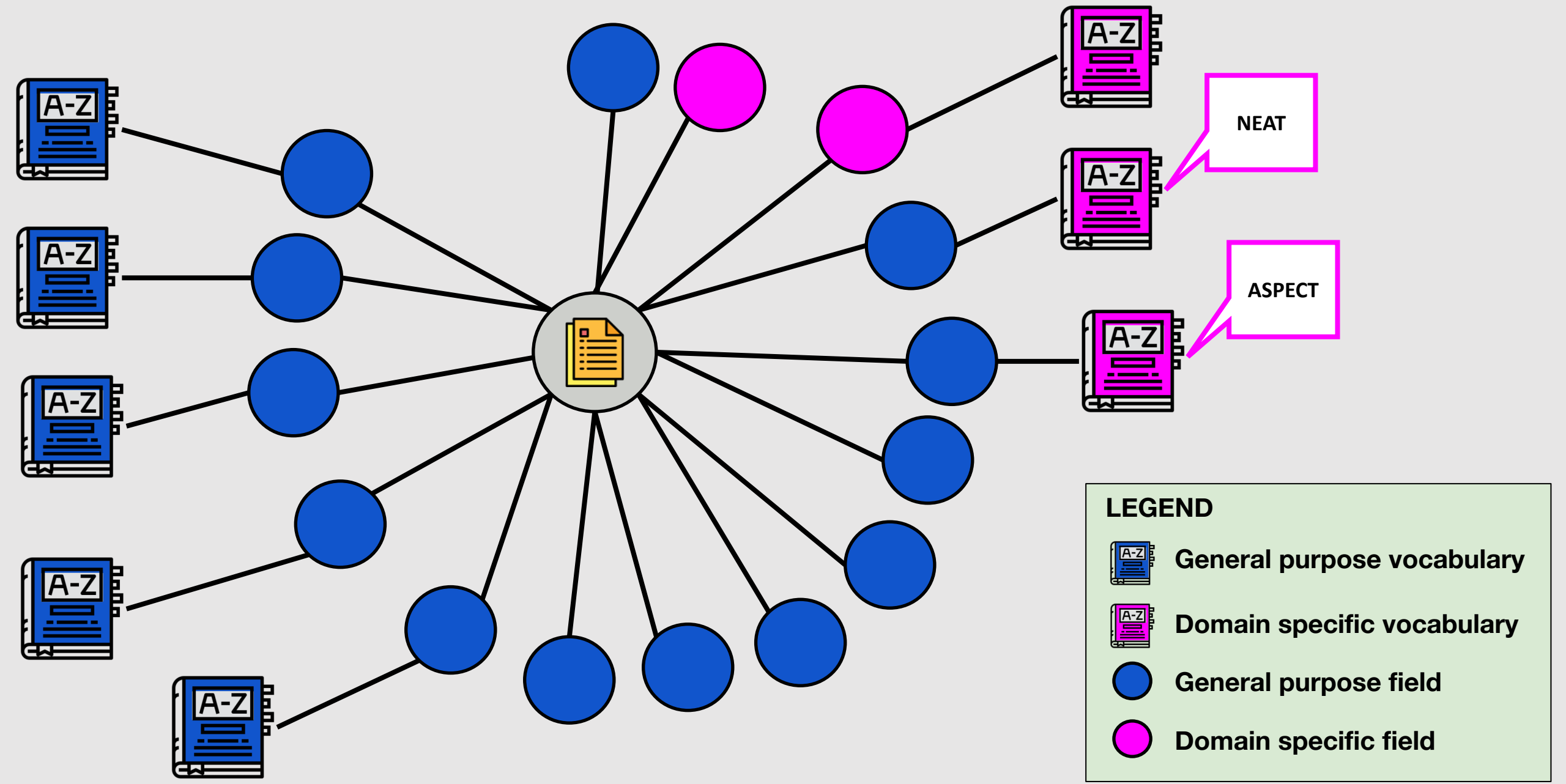
	A	B	C
1	**ConceptScheme URI**	http://data.windenergy.dtu.dk/controlled-terminology/aspect/	
2	**PREFIX**	aspect	http://data.windenergy.dtu.
3	**PREFIX**	pav	http://purl.org/pav/
4	**PREFIX**	dct	http://purl.org/dc/terms/
5	**PREFIX**	rdf	http://www.w3.org/1999/02
6	**PREFIX**	rdfs	http://www.w3.org/2000/01
7	**skos:prefLabel**	ASPECT	
8	**dct:title**	ASPECT: wind energy vAriableS ParametErs and ConsTants	
9	**dct:description**	Controlled vocabulary of variables, parameters and constants used in wind energy comm	
10	**dct:creator**	Technical University of Denmark, DTU Wind Energy	
11	**dct:rights**	https://spdx.org/licenses/CC0-1.0	
12	**pav:version**	0.1.0	
13	**pav:createdOn**	2020-10-22T22:00:00+01:00	
14	**pav:lastUpdatedOn**	2021-06-21T15:00:00+01:00	
15			
16	**Identifier**	**skos:prefLabel**	**rdf:type**
22	aspect:iec-61400-13-name	iec-61400-13-name	rdf:Property
23	aspect:iec-61400-25-2-name	iec-61400-25-2-name	rdf:Property
24	aspect:prefUnit	prefUnit	rdf:Property
25	aspect:altUnit	altUnit	rdf:Property
26	aspect:GenericTerms	Generic Terms	
27	aspect:azimuth_angle	azimuth_angle	
28	aspect:elevation_angle	elevation_angle	
29	aspect:EnviromentalConditionTerms	Enviromental Condition Terms	
30	aspect:wind_speed	wind_speed	

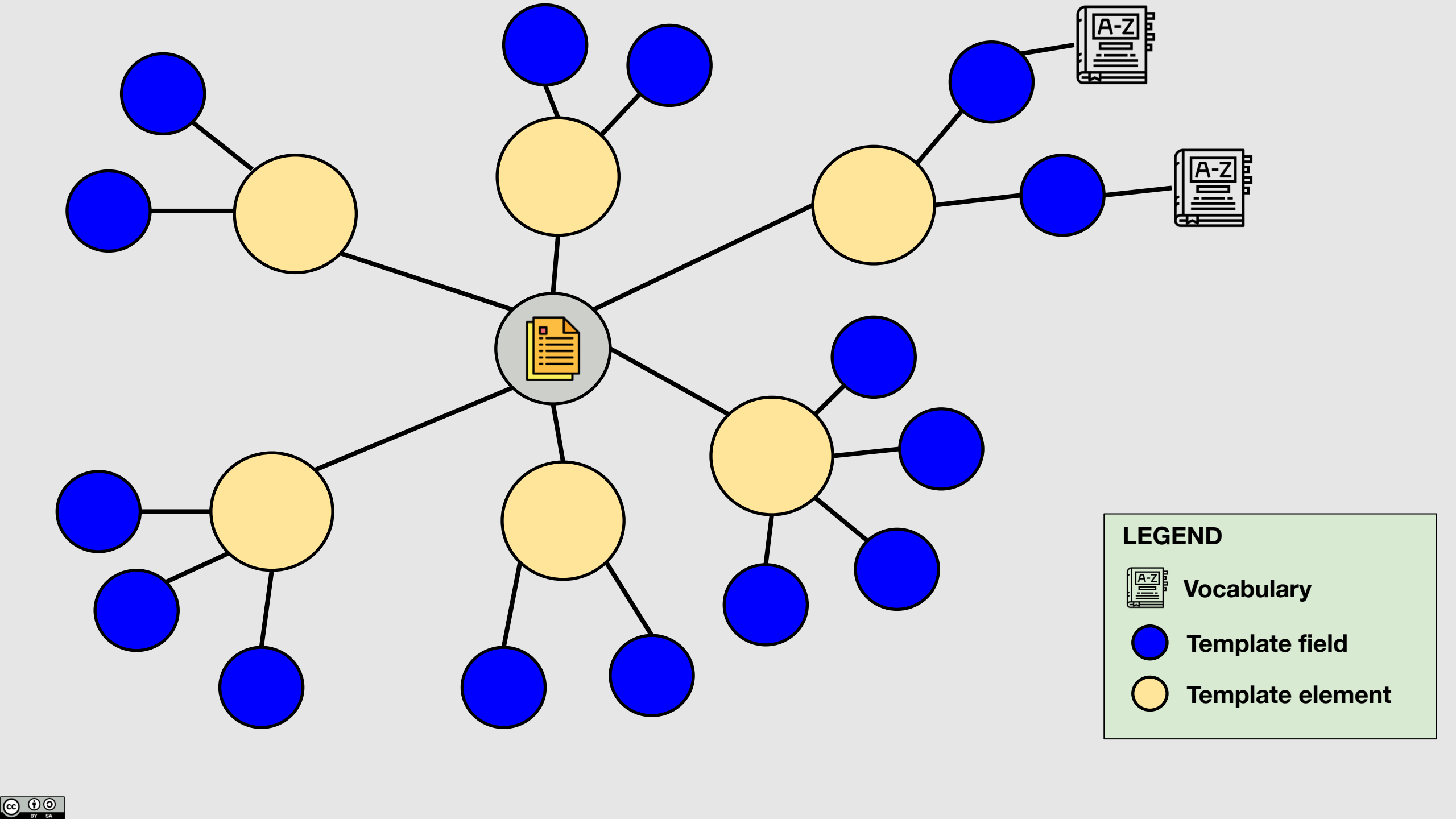
ASPECT taxonomy

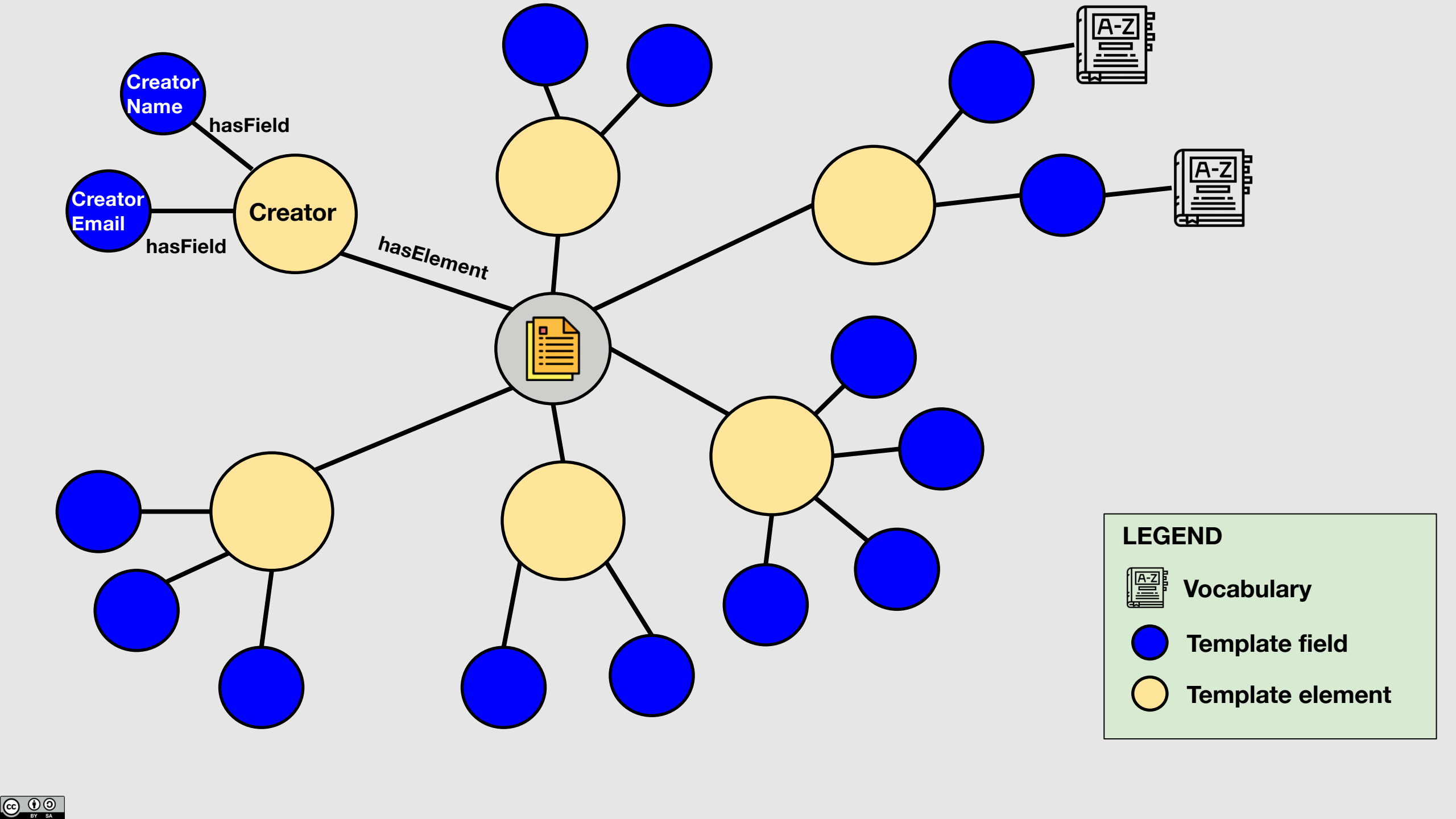
WE NEED YOUR INPUTS ON EXTENDING NEAT AND ASPECT !!!

PLEASE COMMENT THE PREVIOUS GOOGLE SHEET DOCS!!!

START WITH SIMPLE REPRESENTATIONS AND OPTIONALY LOOK AT SEMANTIC REPRESENTATION


A-Z
A-Z
A-Z
A-Z
A-Z
A-Z
A-Z
NEAT
ASPECT
LEGEND
General purpose vocabulary
Domain specific vocabulary
General purpose field
Domain specific field


A-Z
A-Z
LEGEND
Vocabulary
Template field
Template element

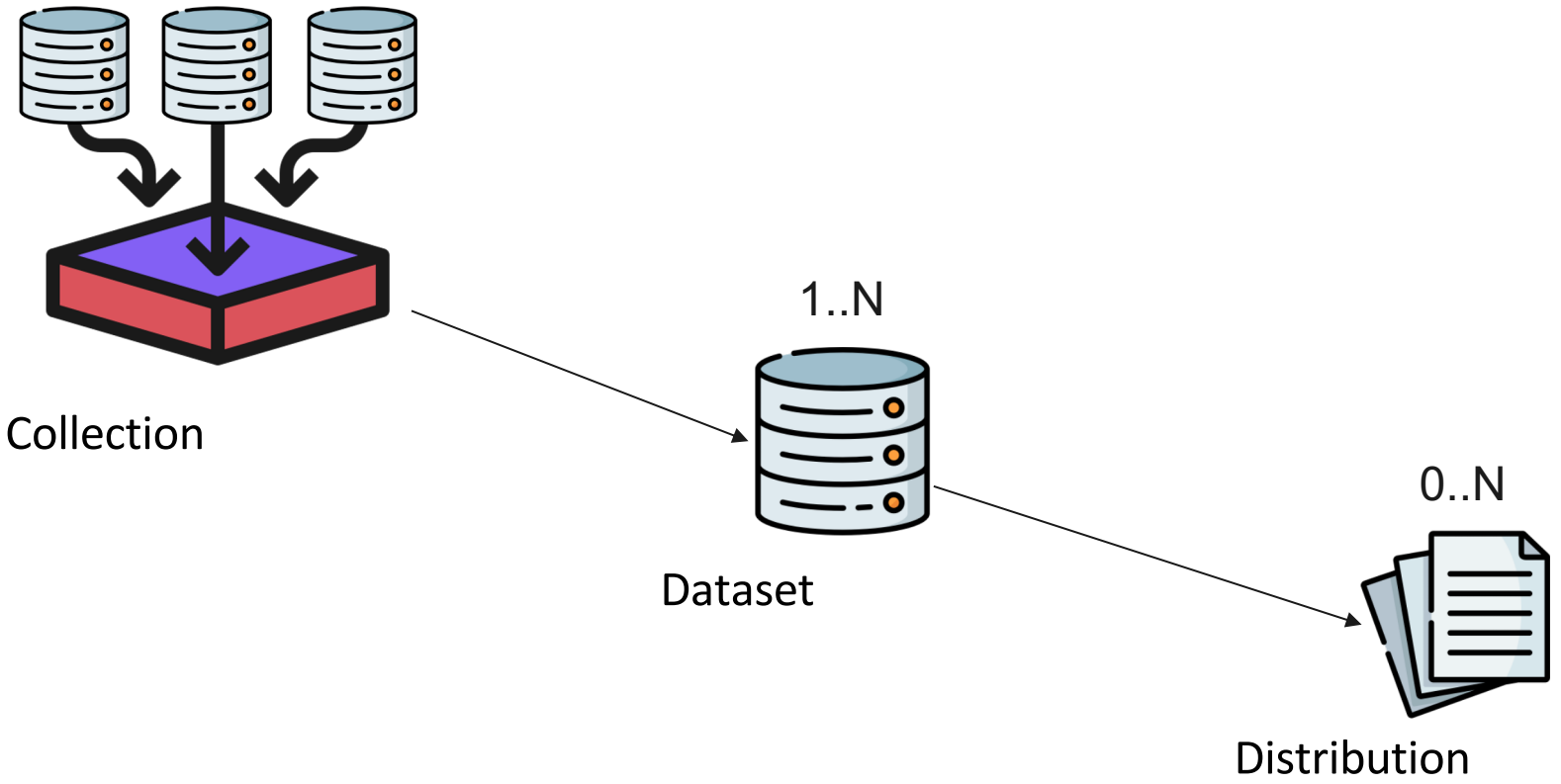

Creator Name
hasField
Creator Email
hasField
Creator
hasElement
A-Z
A-Z
LEGEND
Vocabulary
Template field
Template element

Generic Dataset Metadata Template (GDMT)

- Developed by FAIR Data Collective (FDC)
- Made by fusing and improving **DataCite** and **DCAT** scheme
- GDMT contains 100 fields ('only' 13 mandatory) grouped in ~20 elements
- Unlike the DataCite template, GDMT is **MACHINE-ACTIONABLE**, details at:
  - **CEDAR**
  - **GitHub**
- GDMT contains a **'back-end' ontology** that enables machine-actionability and has:
  - ~**120** RDF properties (i.e., defined fields and elements with resolvable PIDs)
  - ~**1000** controlled terms (with resolvable PIDs)

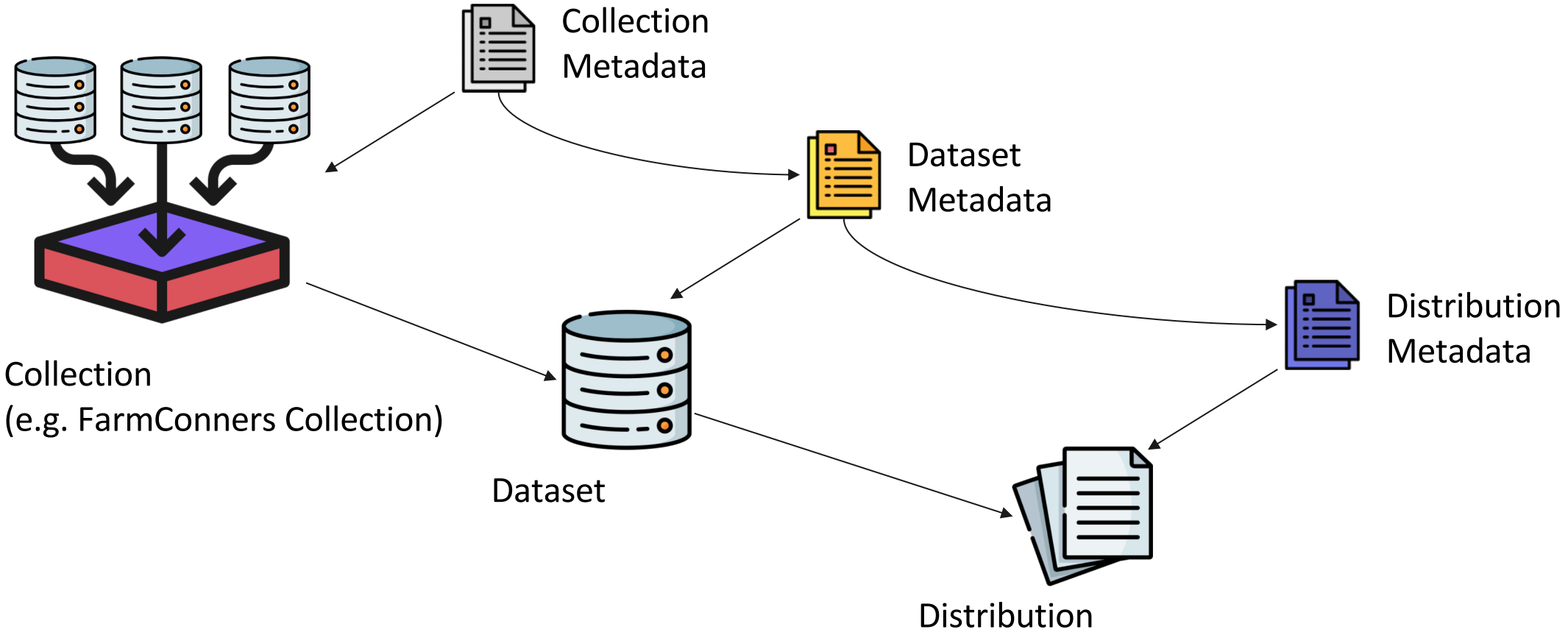
**PID = Persistent Identifier, URL = resolvable PID (e.g., your ORCiD ID)*

DCAT (Data Catalog Vocabulary) organizations



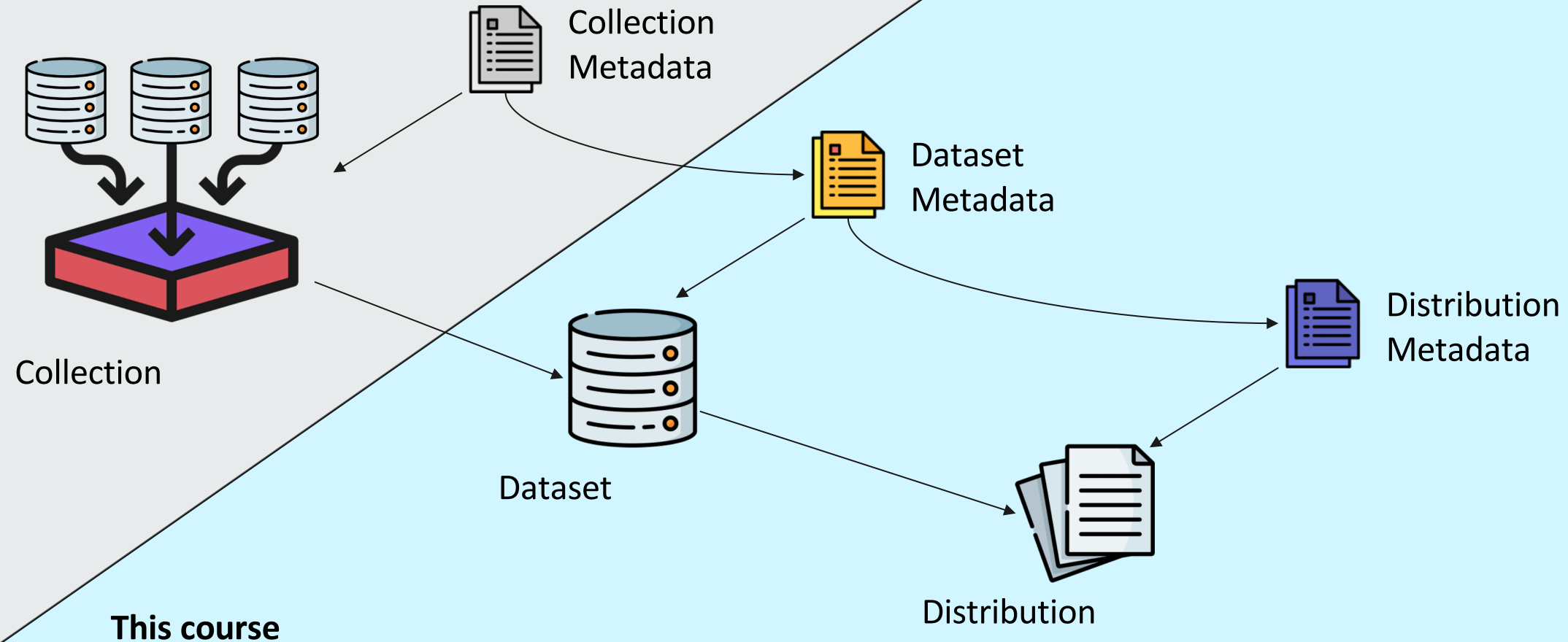
https://www.w3.org/TR/vocab-dcat-3/

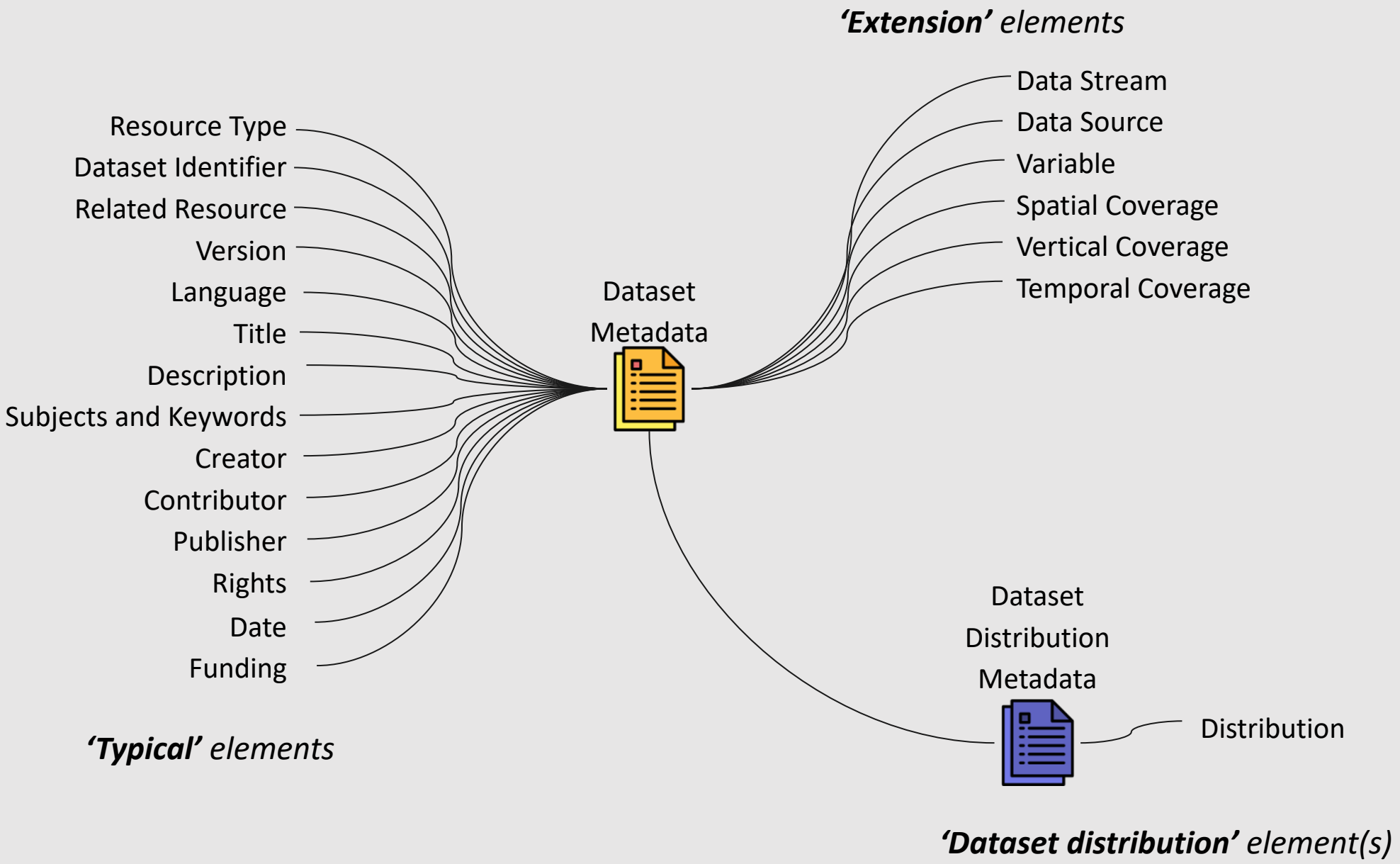
DCAT organizations



https://www.w3.org/TR/vocab-dcat-3/

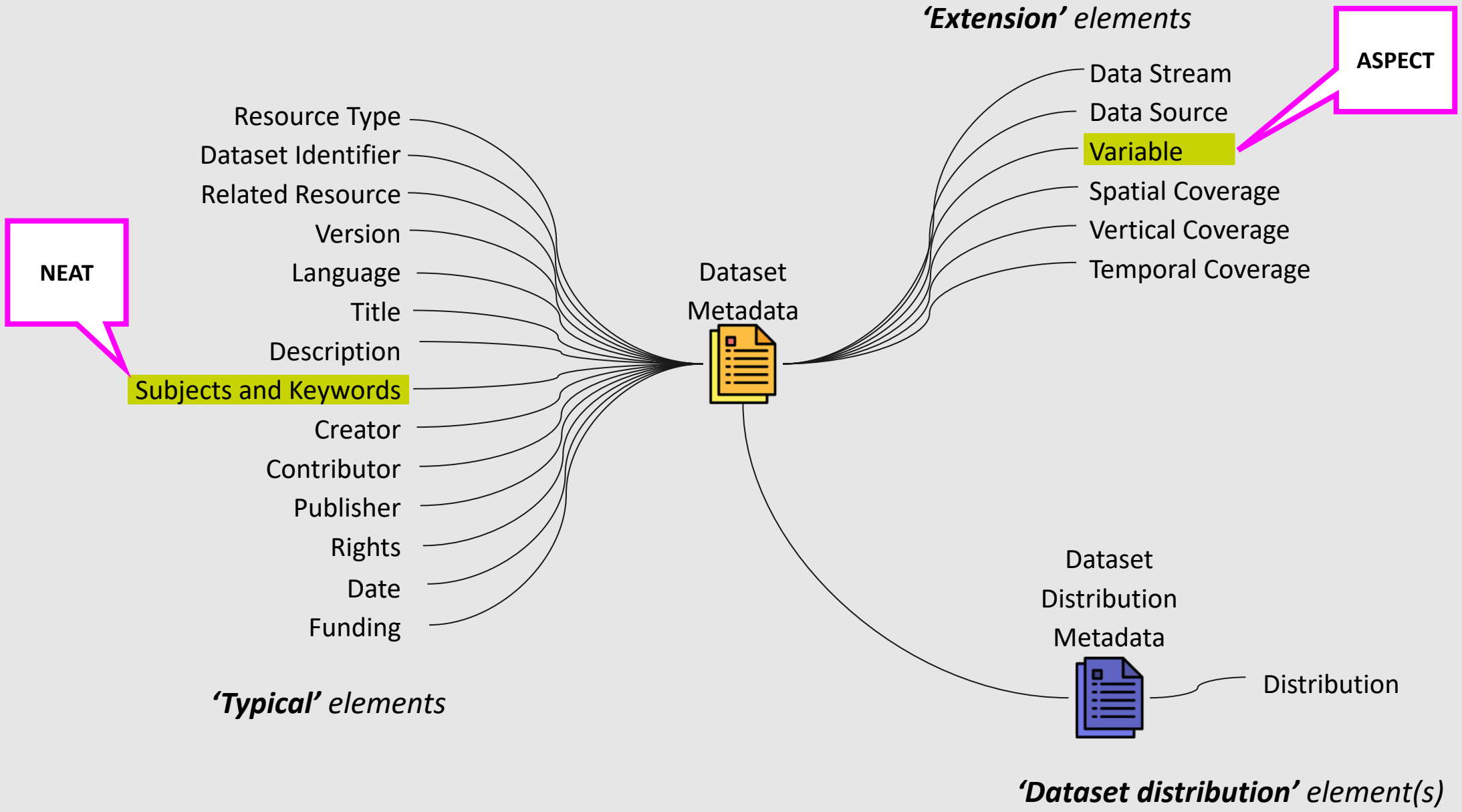
DCAT organizations



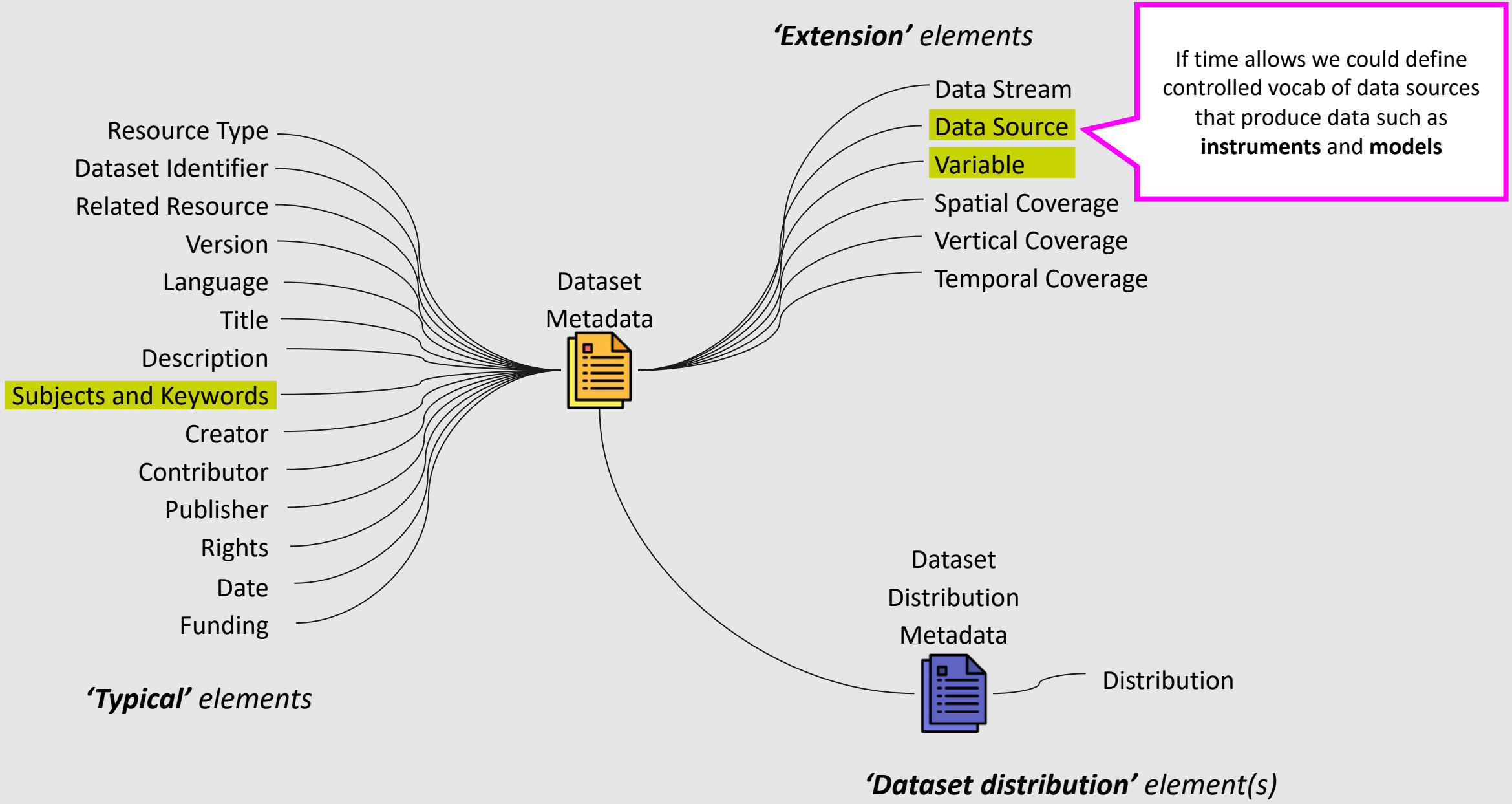


You can find definitions of elements and fields on CEDAR and GitHub.

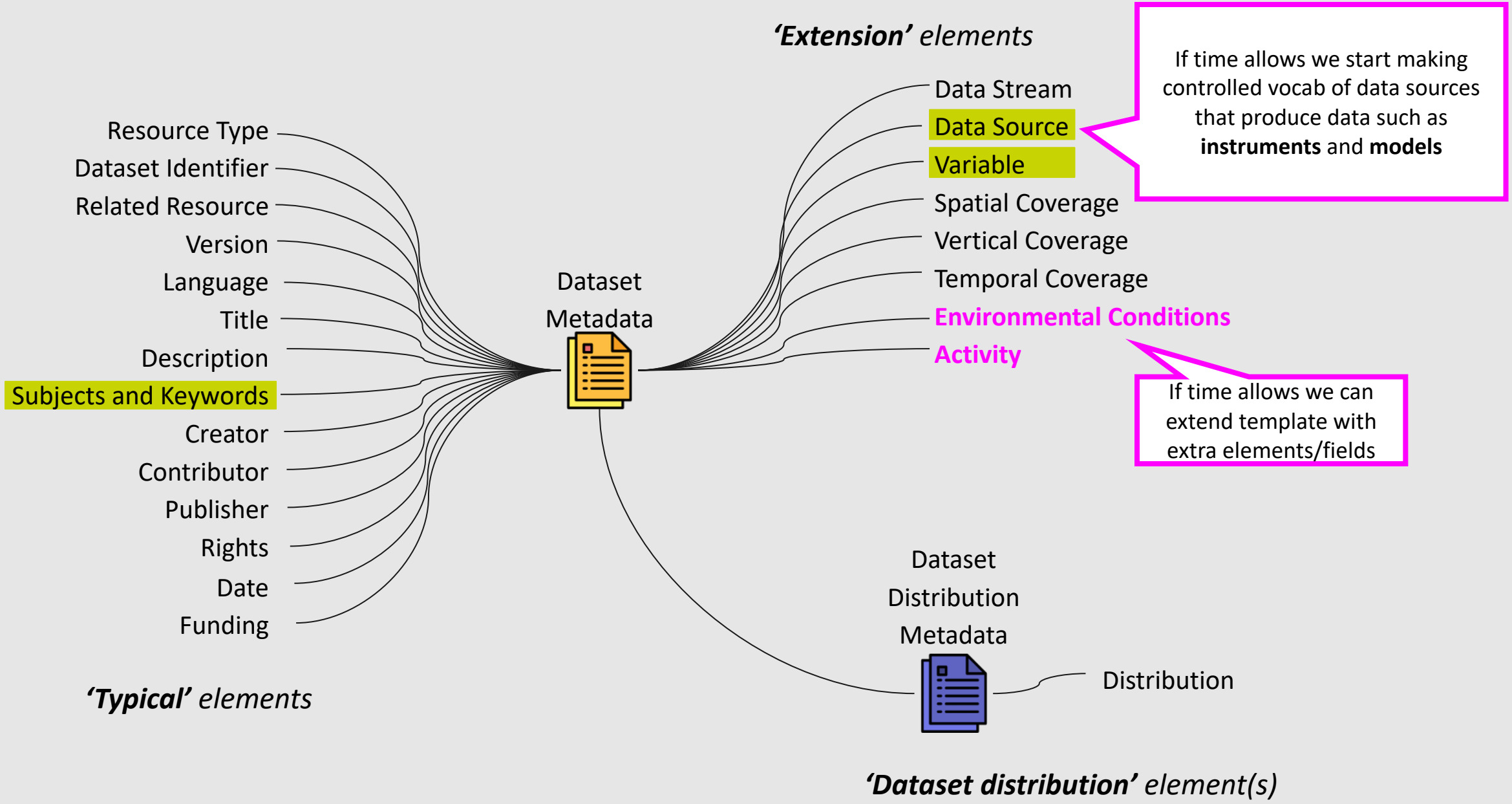
OntoStack serves the GDMT ontology, which contains a number of controlled terms and RDF properties that enable machine-actionability .



*By creating domain specific controlled vocabularies and updating **GDMT** to use them, we turn GDMT to Specific Dataset Metadata Template (**SDMT**)*



*By creating domain specific controlled vocabularies and updating **GDMT** to use them, we turn GDMT to Specific Dataset Metadata Template (**SDMT**)*



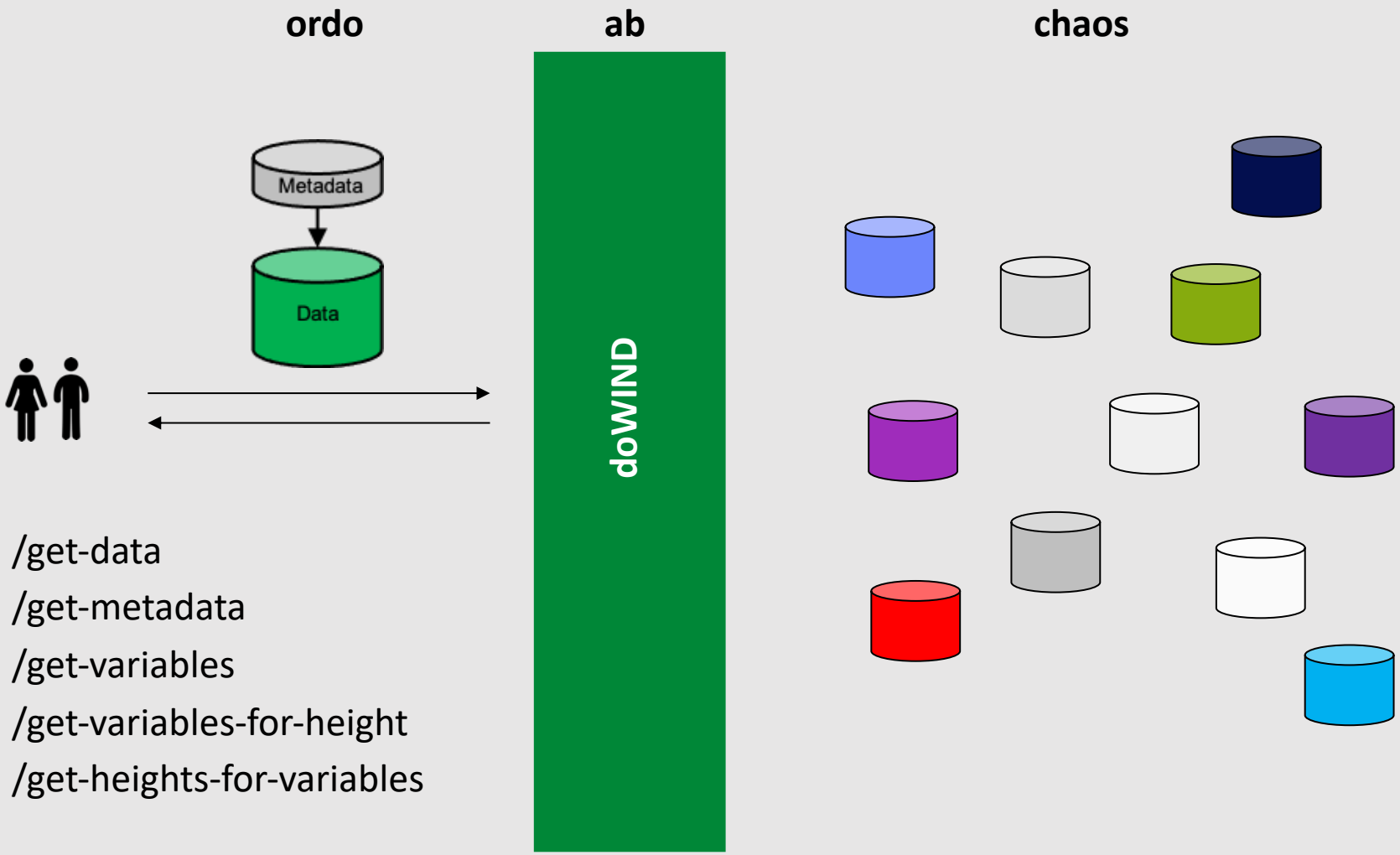
*By creating domain specific controlled vocabularies and updating **GDMT** to use them, we turn GDMT to Specific Dataset Metadata Template (**SDMT**)*

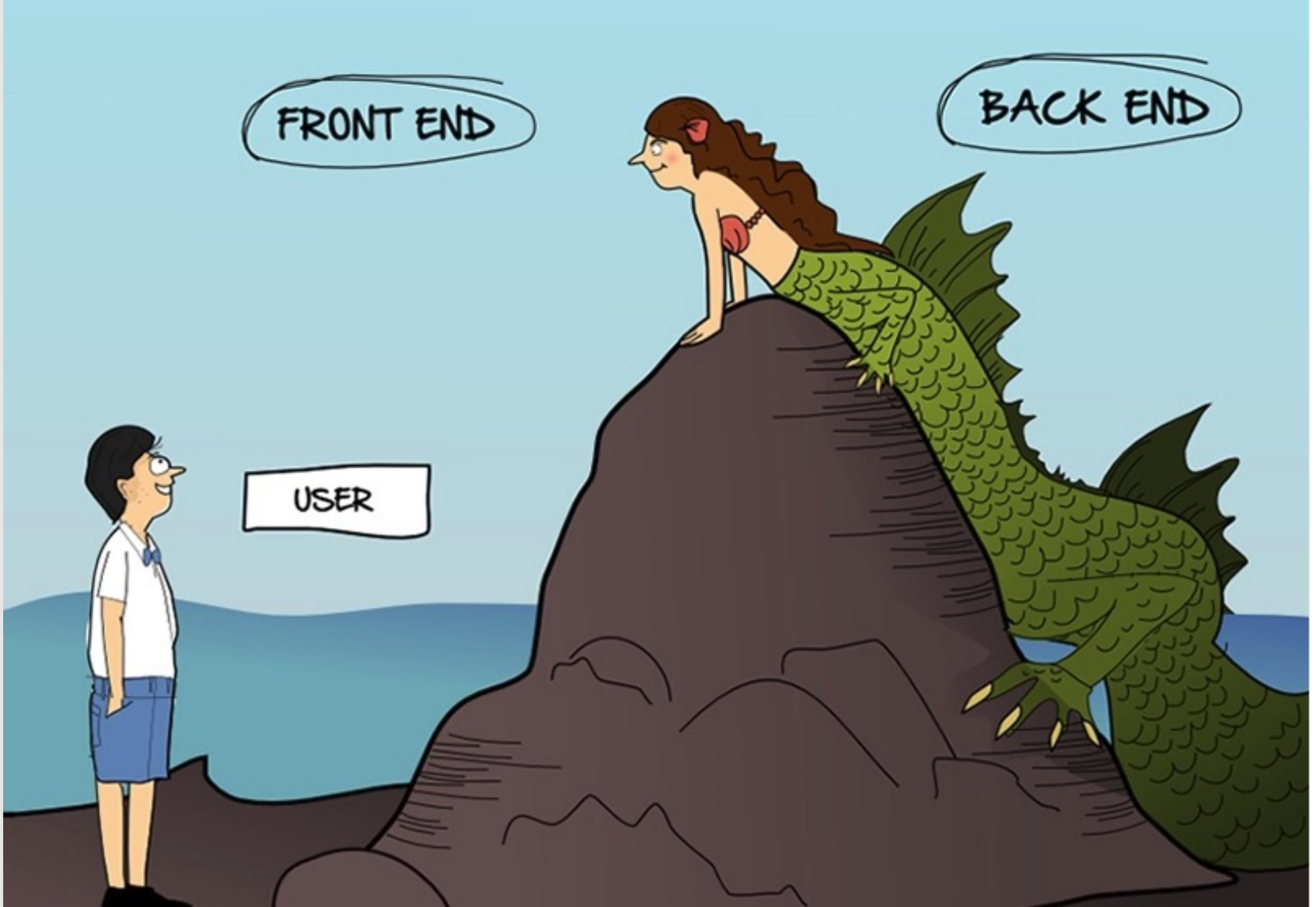
Questions/Comments?

Why should I do this? What is the point of wasting time on this? Research cannot be standardized, that is for industry! This is only for academia and not for industry! Why do you think 'your way' is better than 'my way'? I am busy! We have our own standard! There is a header in a CSV file and it is readable by my computer! Our data are not open! This is difficult and I don't want to spend time learning it! Scientific papers is what matters! This is a utopian idea!

Implementation of semantics

doWIND: data aggregation and subsetting API





doWIND: key features

Data owners

- API provided as a docker container / python package
- Configuration of the API through a simple YAML file
- Ability to control maximum file size served to user
- Ability to control data access using ***auth*** package
- Back-end data can be in various formats:
  GeoTiff, NetCDF, Zarr, …
- **Metadata make use of controlled vocabularies**

Data users

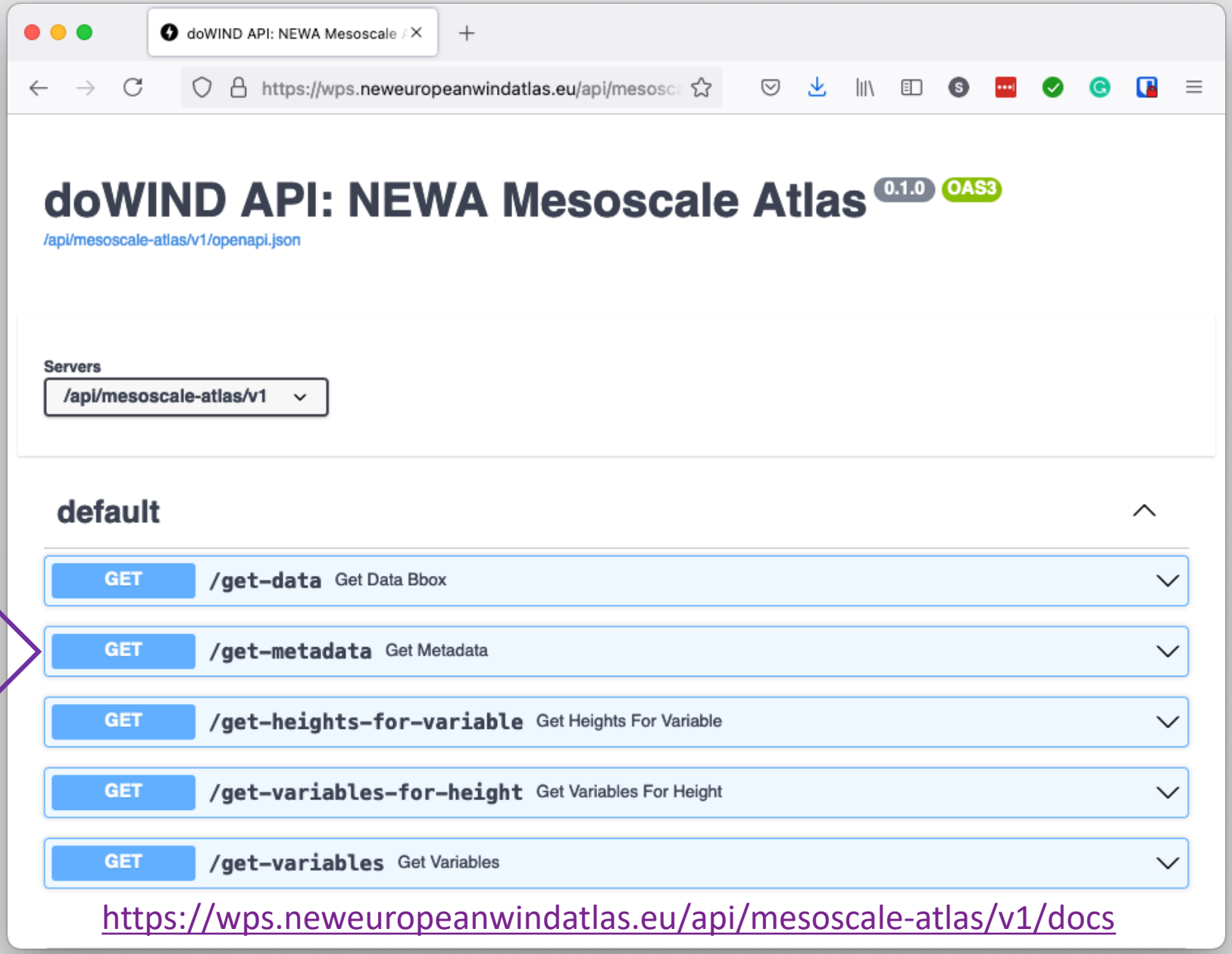
- Ability to aggregate and subset data based on (currently):
  – Variables
  – Geospatial coverage (Bounding box, Height)
- Access to data through HTTP calls (basically links)
- **Receive NetCDF file with rich metadata**
- **NetCDF conforms to metadata schema**
- **NetCDF variable attributes linked to controlled vocabularies**

doWIND instances

- New European Wind Atlas (NEWA):
  - NEWA microscale atlas
  - NEWA mescoscale atlas

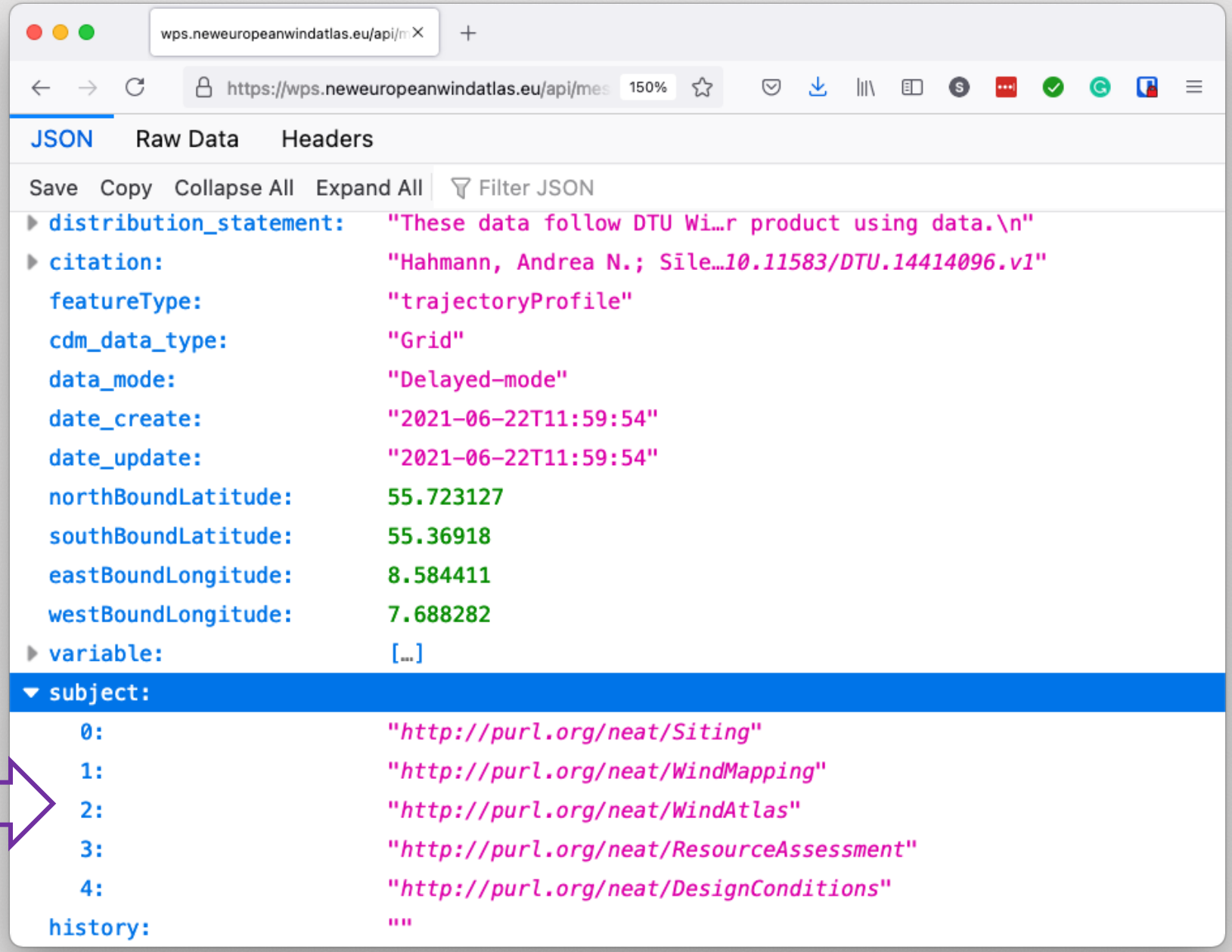
Implementation of semantics in **doWIND**

https://.../get-metadata



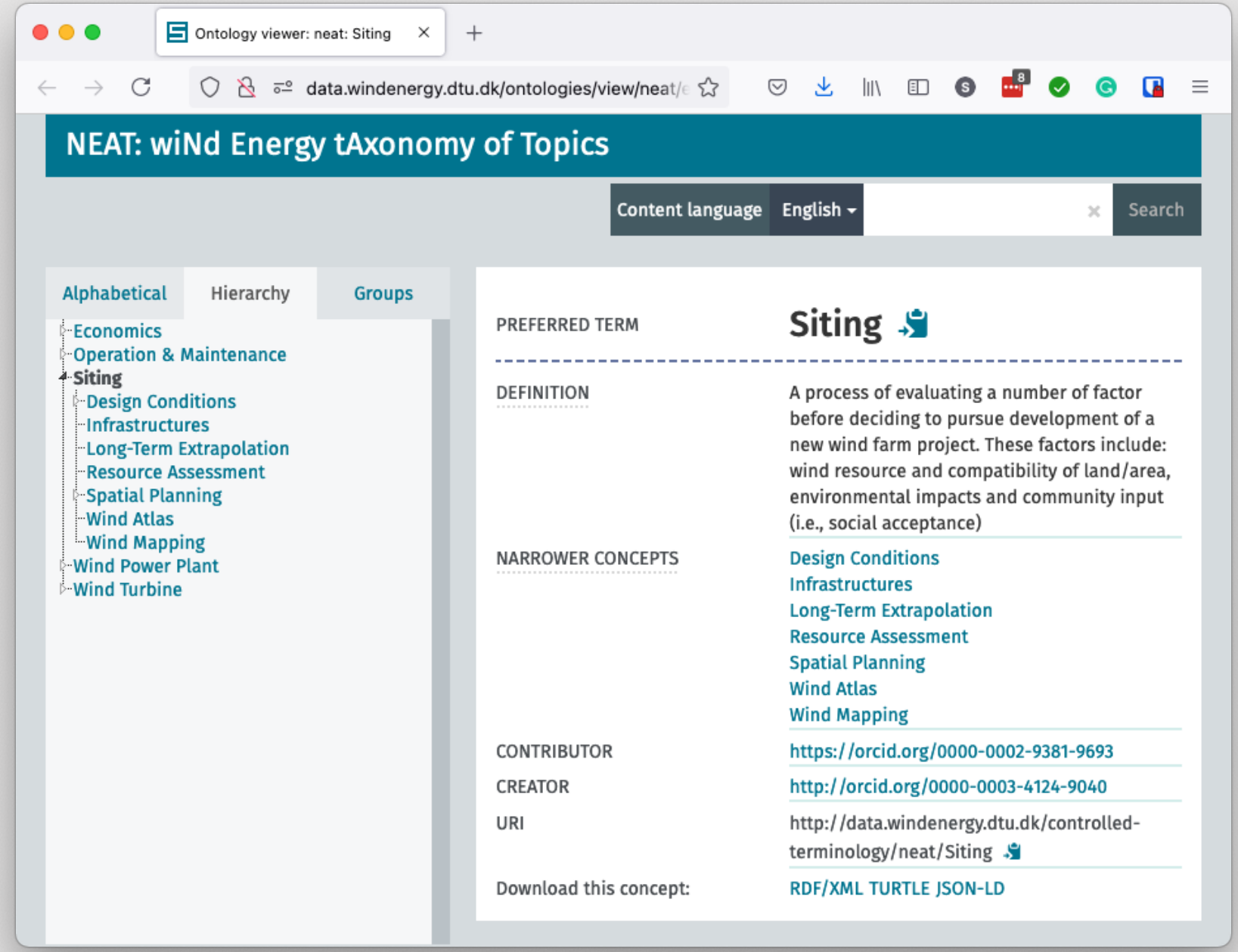
https://wps.neweuropeanwindatlas.eu/api/mesoscale-atlas/v1/docs

Subject resolvable PIDs

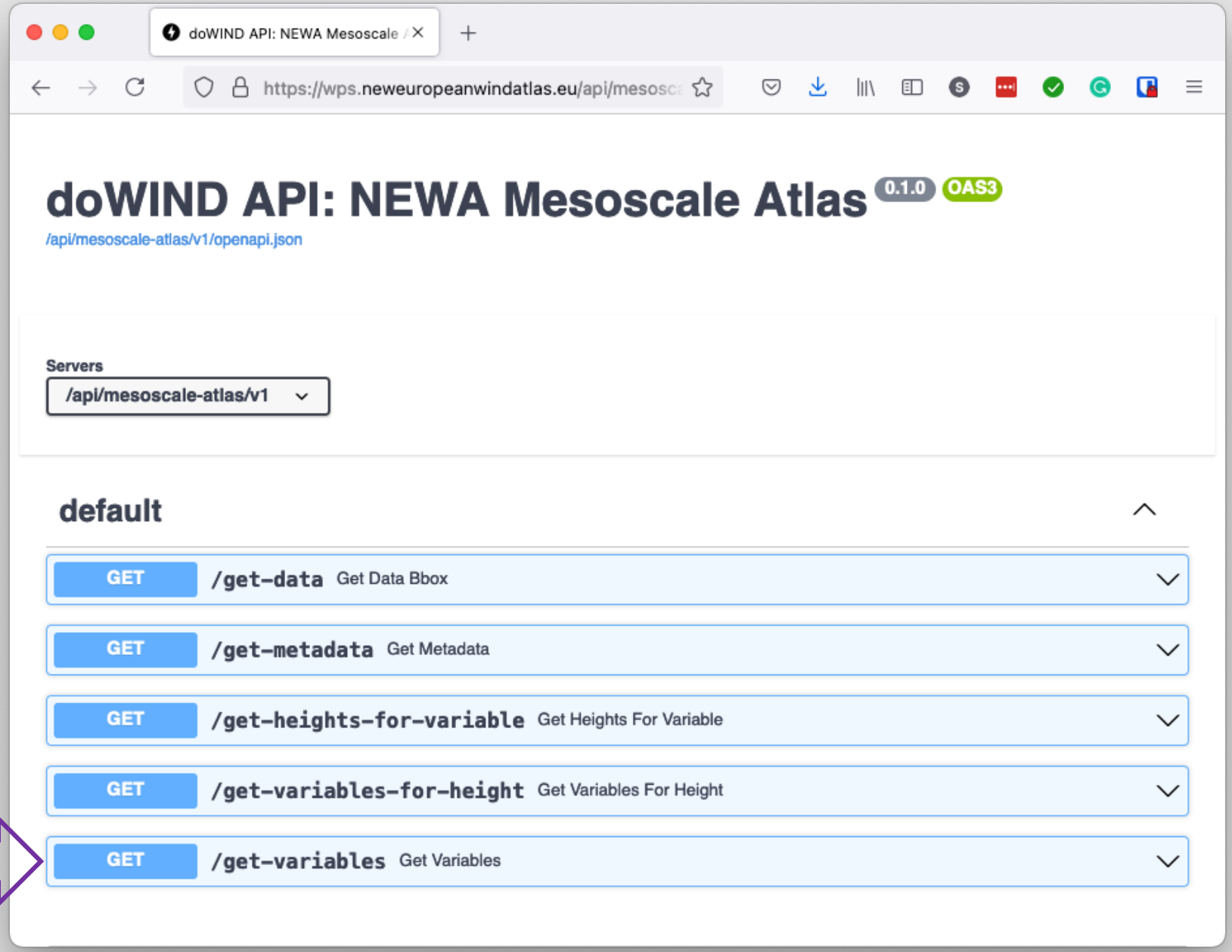


Subject PIDs

Resolvable PIDs



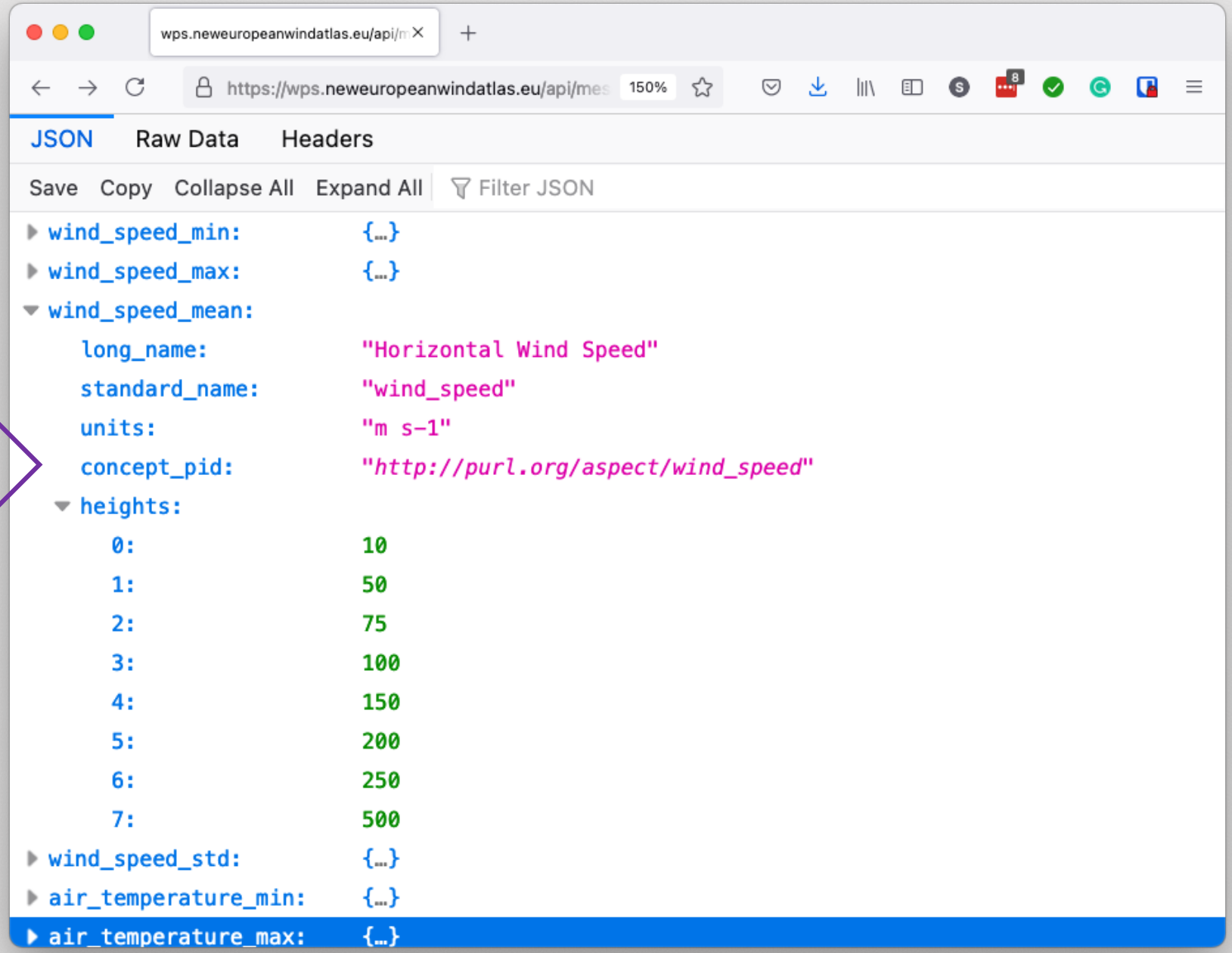
Variables metadata



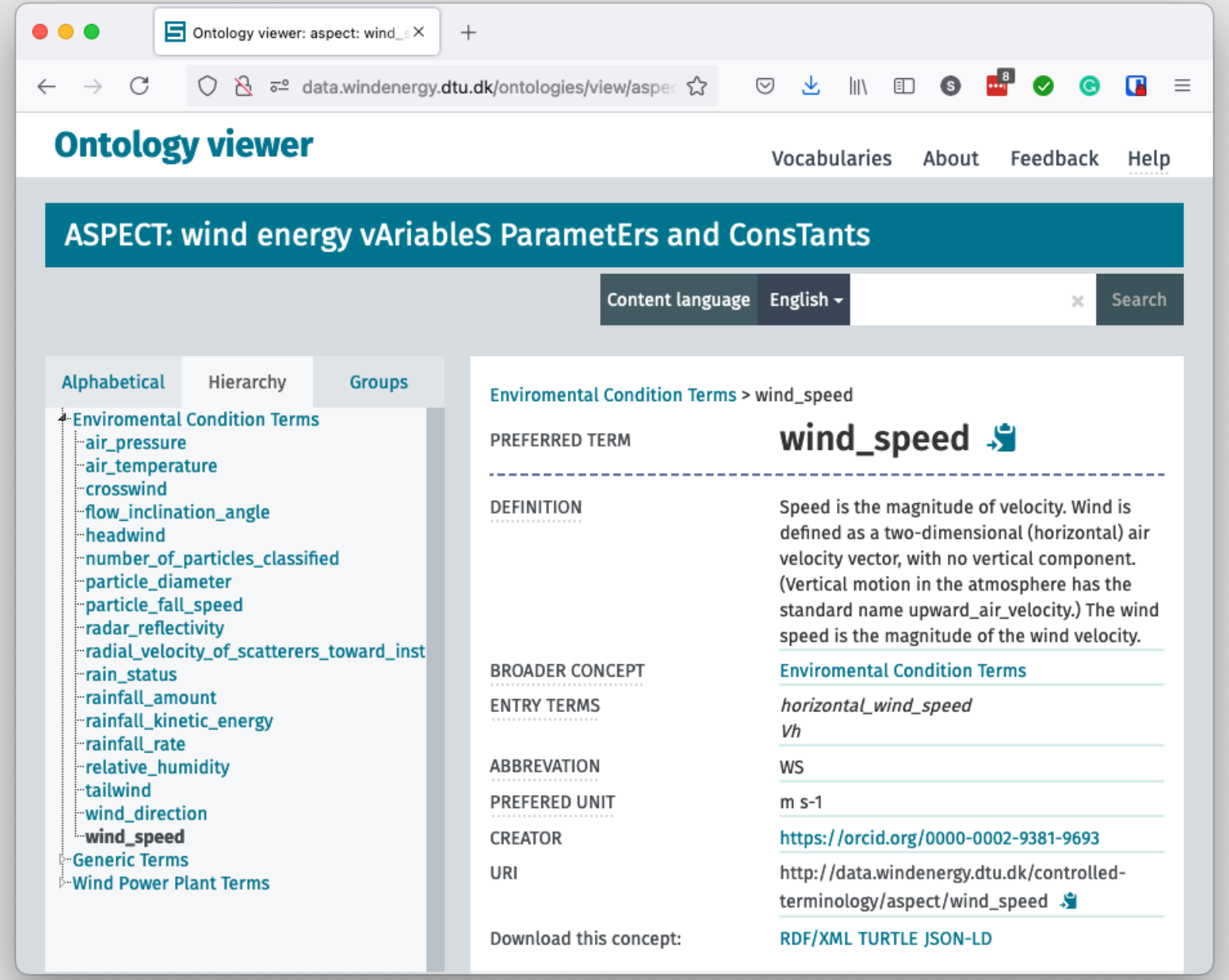
https://.../get-variables

Variables resolvable PIDs

Variable PID

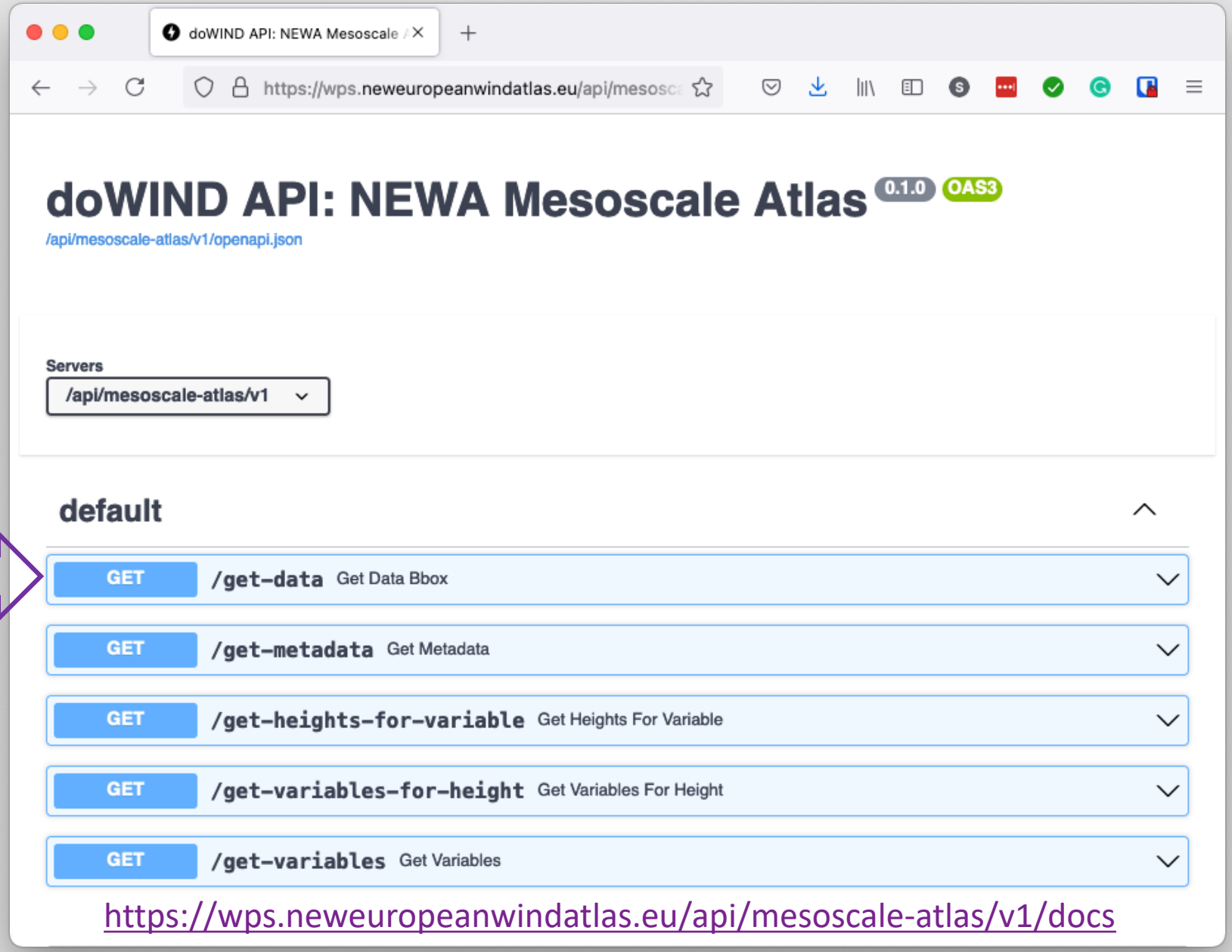


Variables resolvable PIDs



Embedded metadata in data

https://.../data



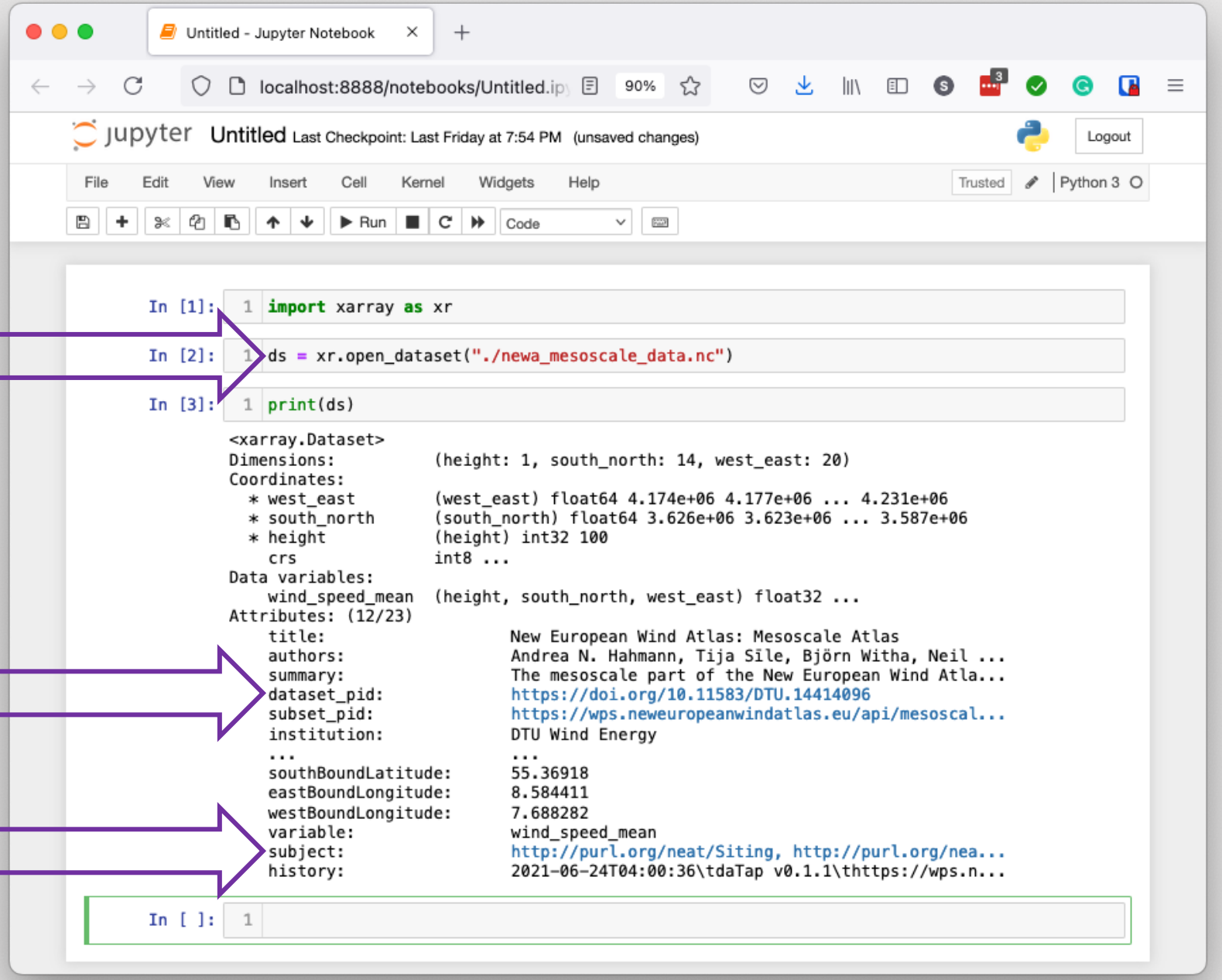
https://wps.neweuropeanwindatlas.eu/api/mesoscale-atlas/v1/docs

NetCDF global and variable attributes

Resulting NetCDF

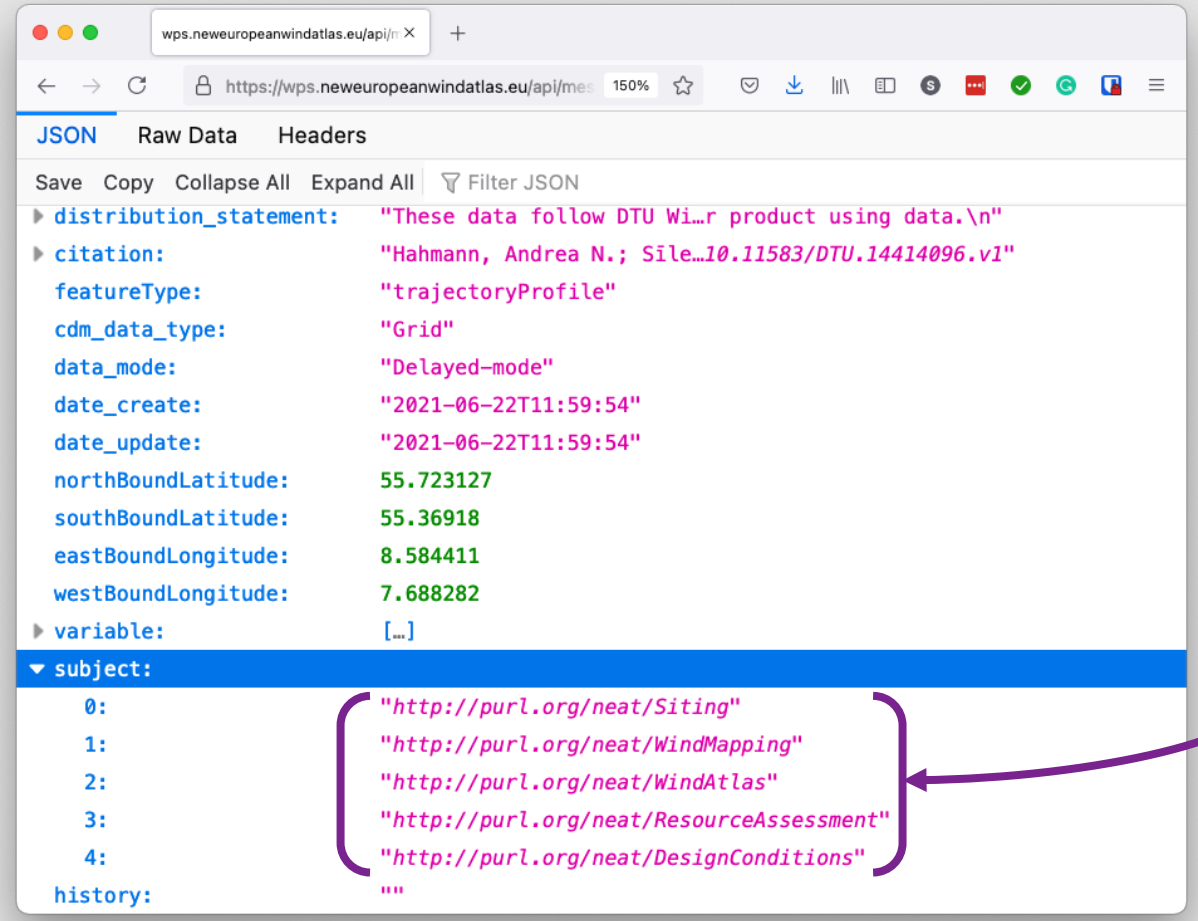
Metadata embedded as attributes

subject attribute

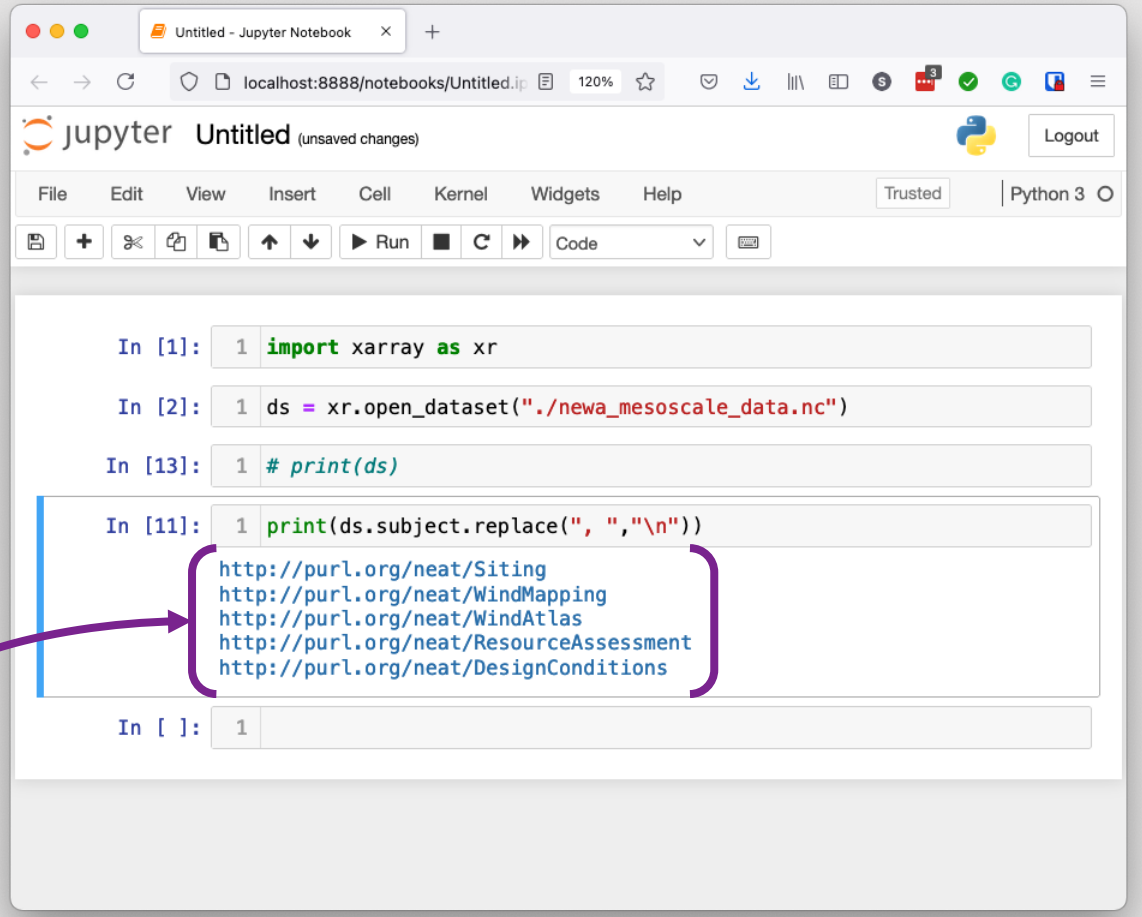


Embedded and stand-off metadata linking

DTU Wind Energy
Department of Wind Energy



=



JSON

NetCDF (variable attribute)

Technical University of Denmark

FarmConners

Thank you!