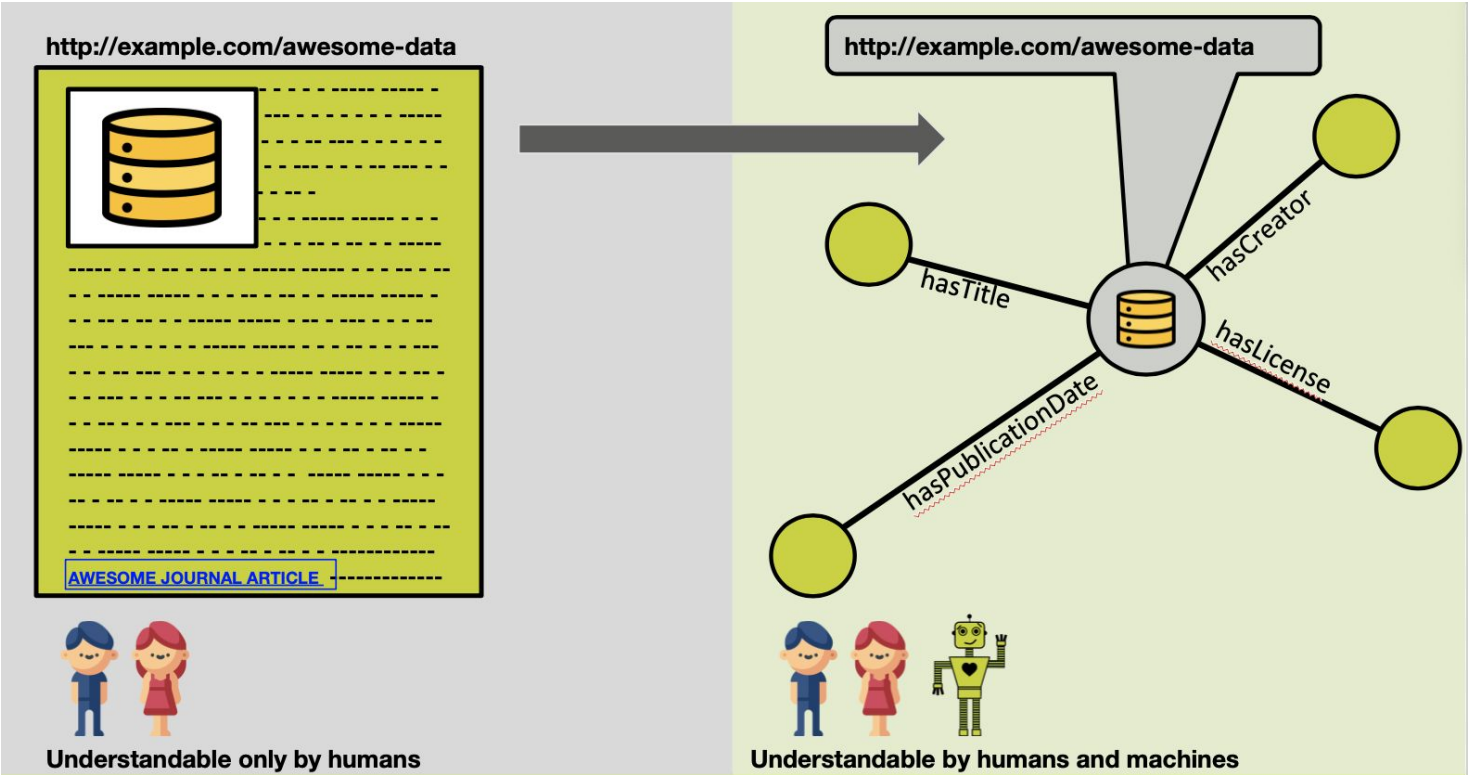
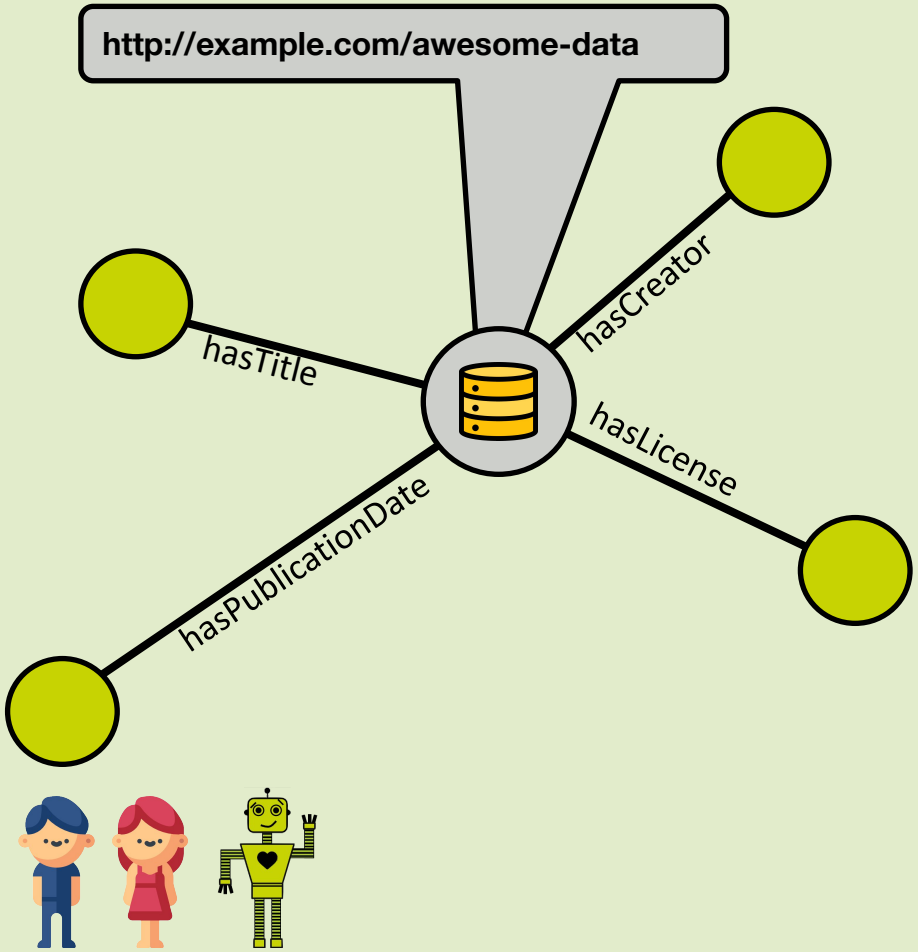
DOI:10.5281/zenodo.5533930

> Introduction to machine actionable metadata

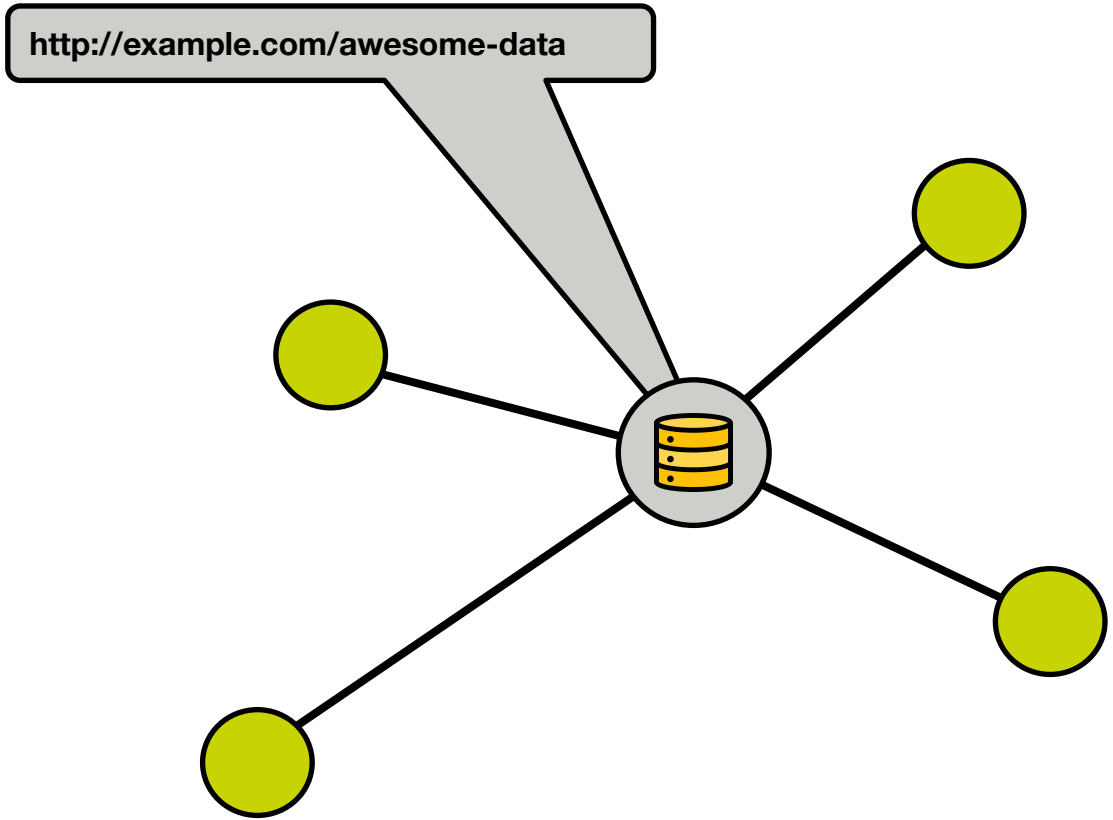
> Nikola Vasiljevic, DTU Wind Energy/FAIR Data Collective

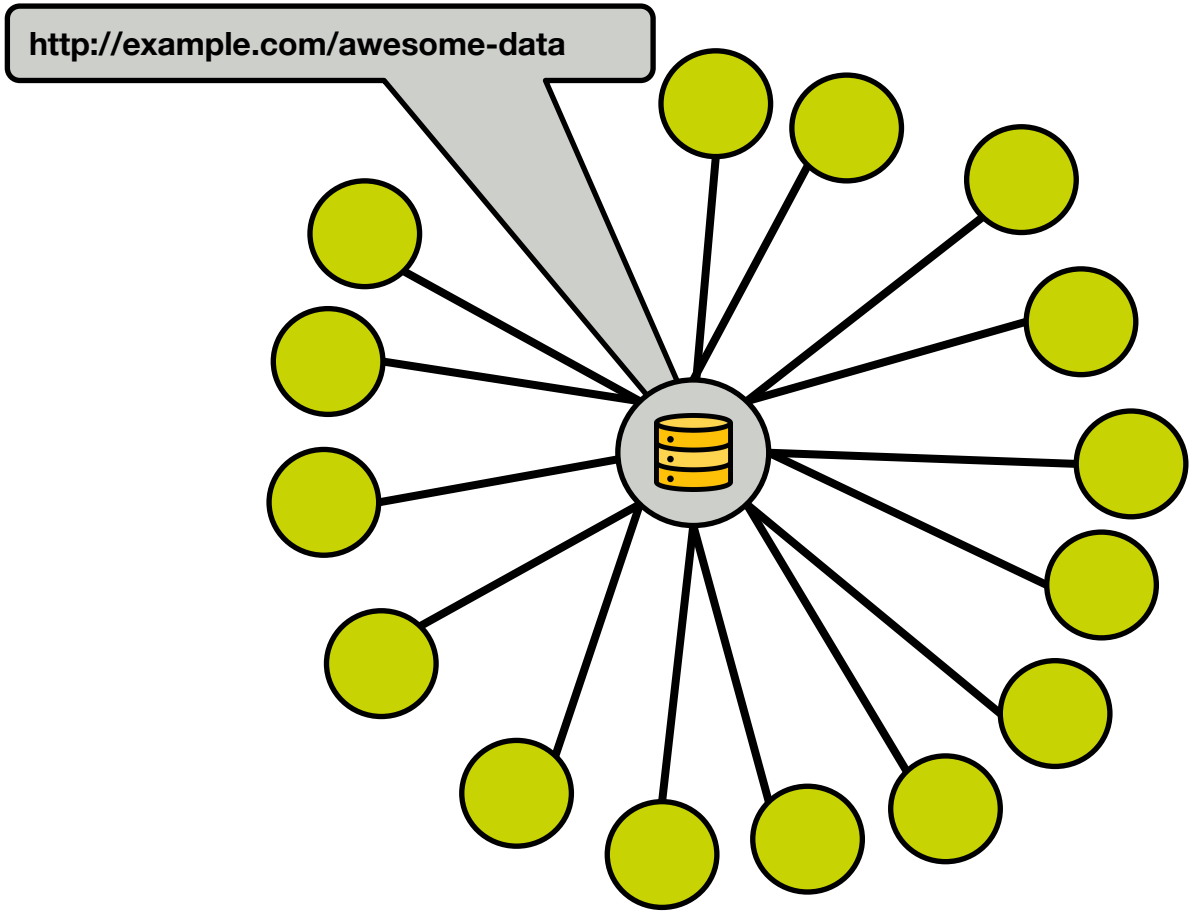
The M4M mission is changing practices

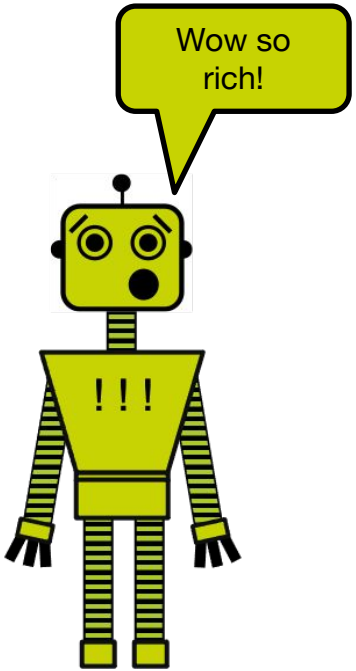




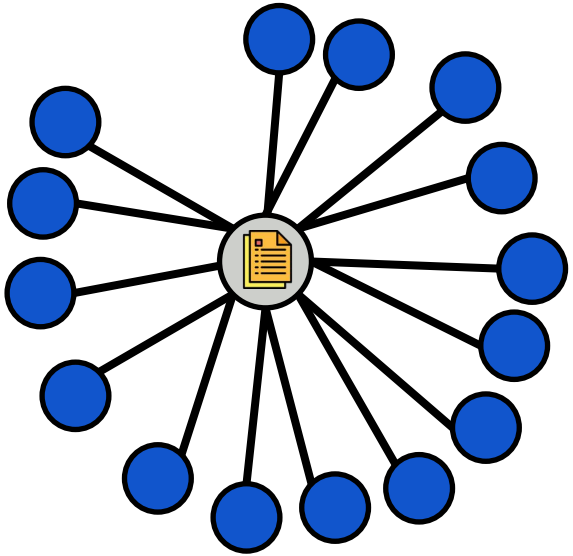
Understandable by humans and machines




http://example.com/awesome-data


Wow so rich!
! ! !

Metadata Template



Graph representation

Field*	Value
Identifier	*insert identifier (e.g., DOI)*
Title	*insert title*
Creator	*insert creator*
Publication date	*insert publication_date*
License	*insert license*
...	...

(Web) Form representation

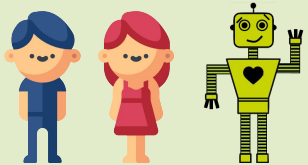
***Field = Property**

Transition to machine actionable template

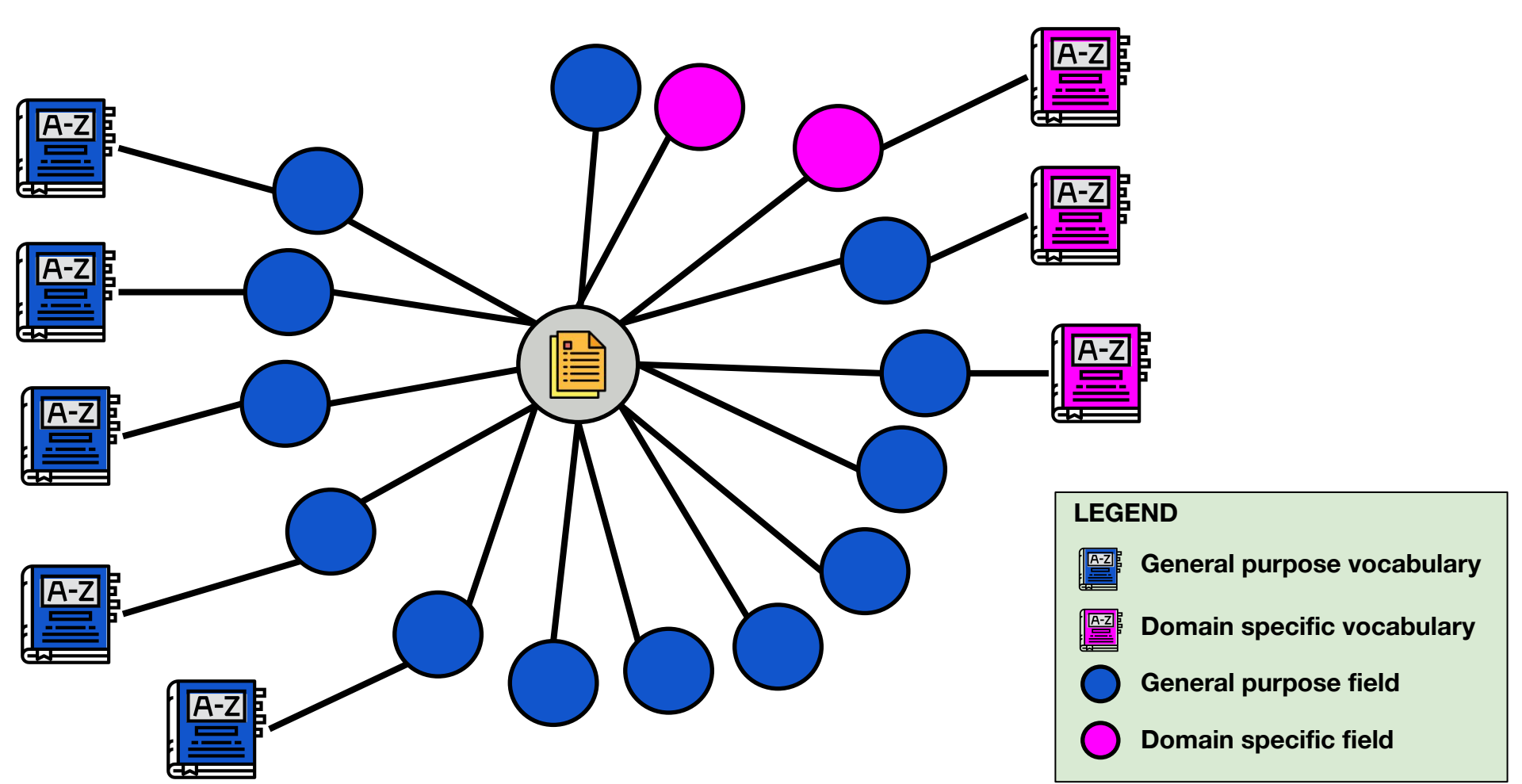
Field	Value
Title	*insert_title*
Creator	*insert_creator*
Publication date	*insert_publication_date*
License	*insert_license*
Subject	*insert_subject*
Variable	*insert_variable*
...	...

Human readable template

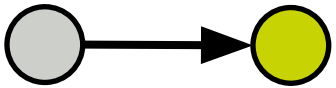
Field	Value
http://purl.org/dc/terms/title	Free text
http://purl.org/dc/terms/creator	URL representing ORCID ID
http://purl.org/dc/terms/date	datetime string (ISO 8601)
http://purl.org/dc/terms/license	**https://spdx.org/licenses/**
http://purl.org/dc/terms/subject	**http://purl.org/neat/**
http://purl.org/gdmt/hasVariable	**http://purl.org/aspect/**
...	...



Human and machine actionable template

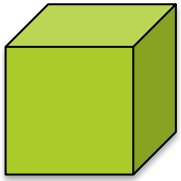

A-Z
A-Z
A-Z
A-Z
A-Z
A-Z
A-Z
A-Z
LEGEND
General purpose vocabulary
Domain specific vocabulary
General purpose field
Domain specific field

LINKED DATA



APPROACH

JSON-LD

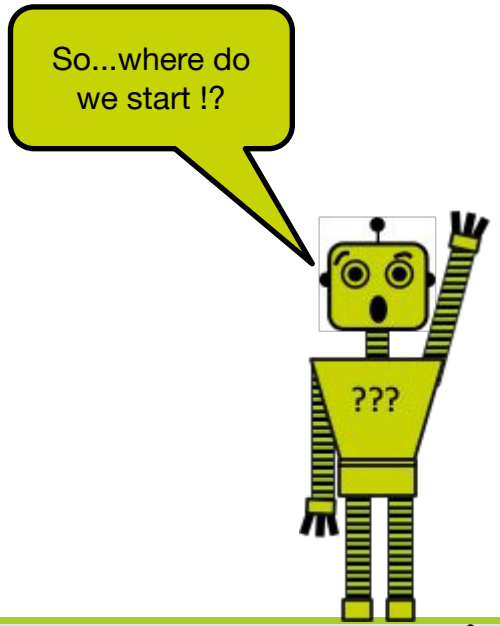


FORMAT

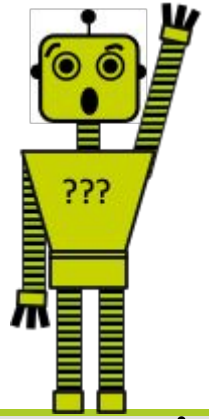
Why LINKED DATA and JSON-LD?

- **LINKED DATA** builds upon standard Web technologies such as HTTP and URIs/IRIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by machines. This enables data from different sources to be connected and queried.

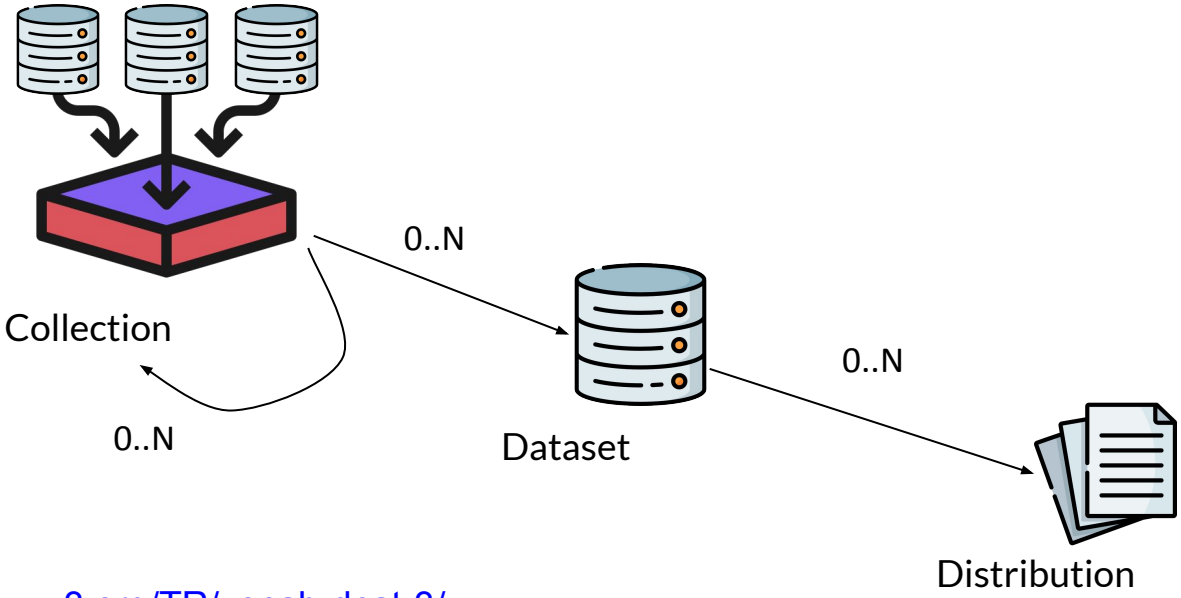
- **JSON-LD** is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as Apache CouchDB and MongoDB.



We start with a short intro to GMT (Generic Metadata Templates) templates based on DCAT organization

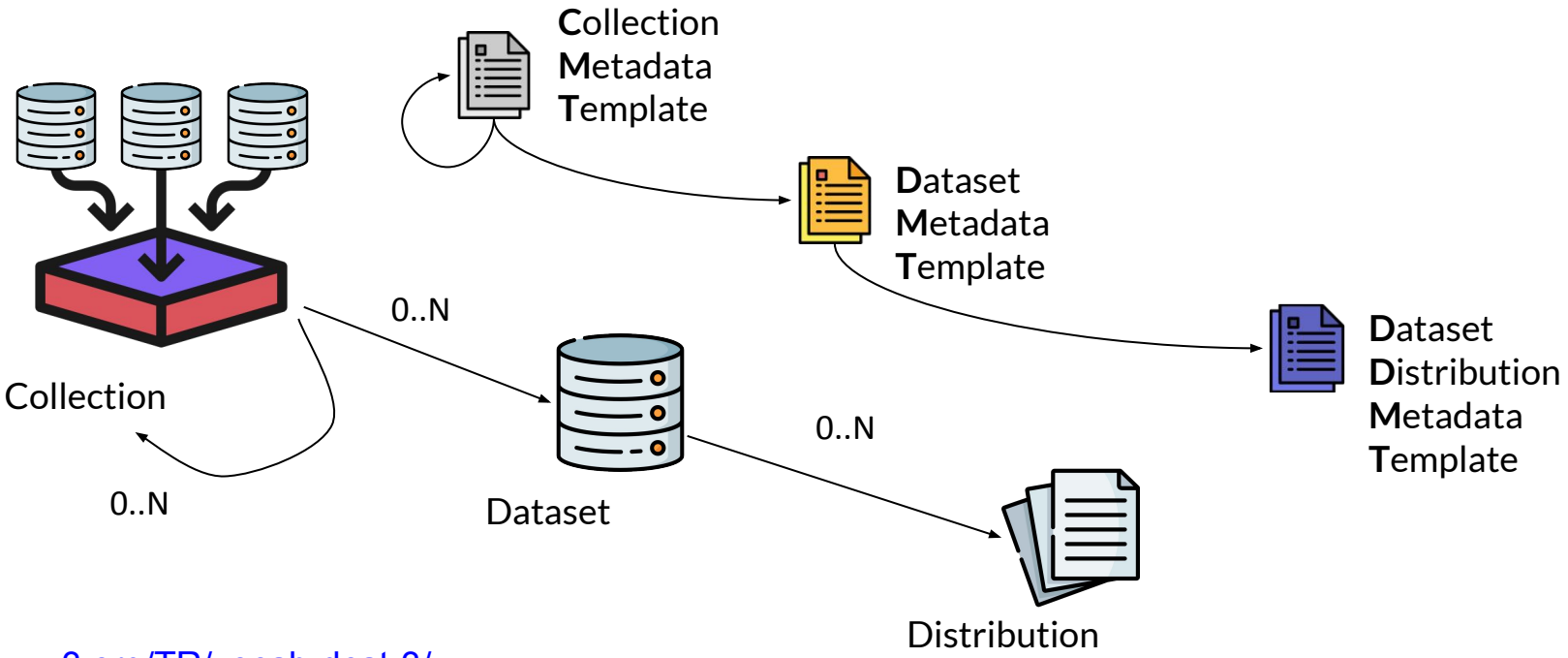


DCAT (Data Catalog Vocabulary) organizations

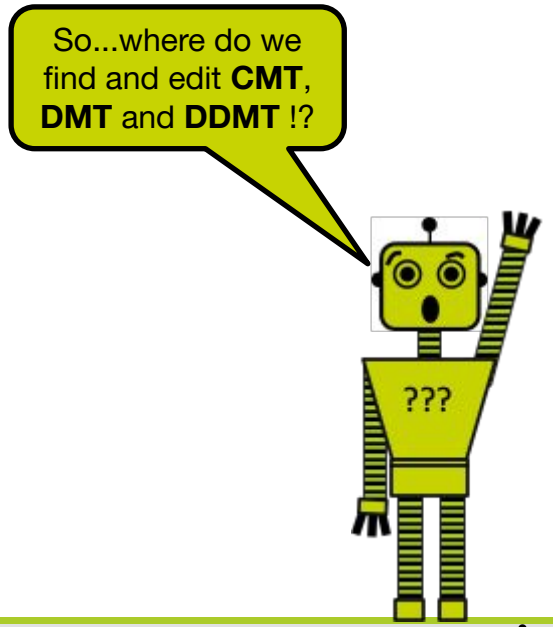


https://www.w3.org/TR/vocab-dcat-3/

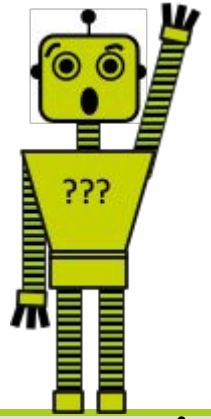
DCAT (Data Catalog Vocabulary) organizations



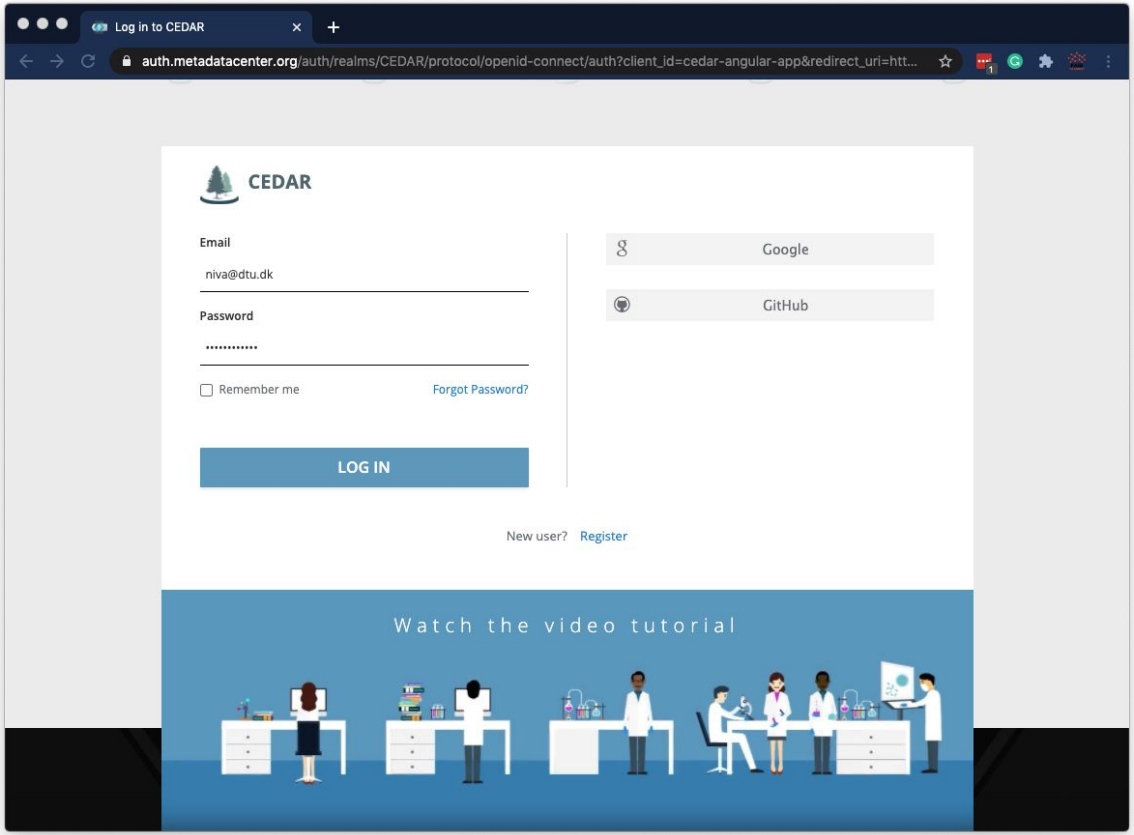
https://www.w3.org/TR/vocab-dcat-3/



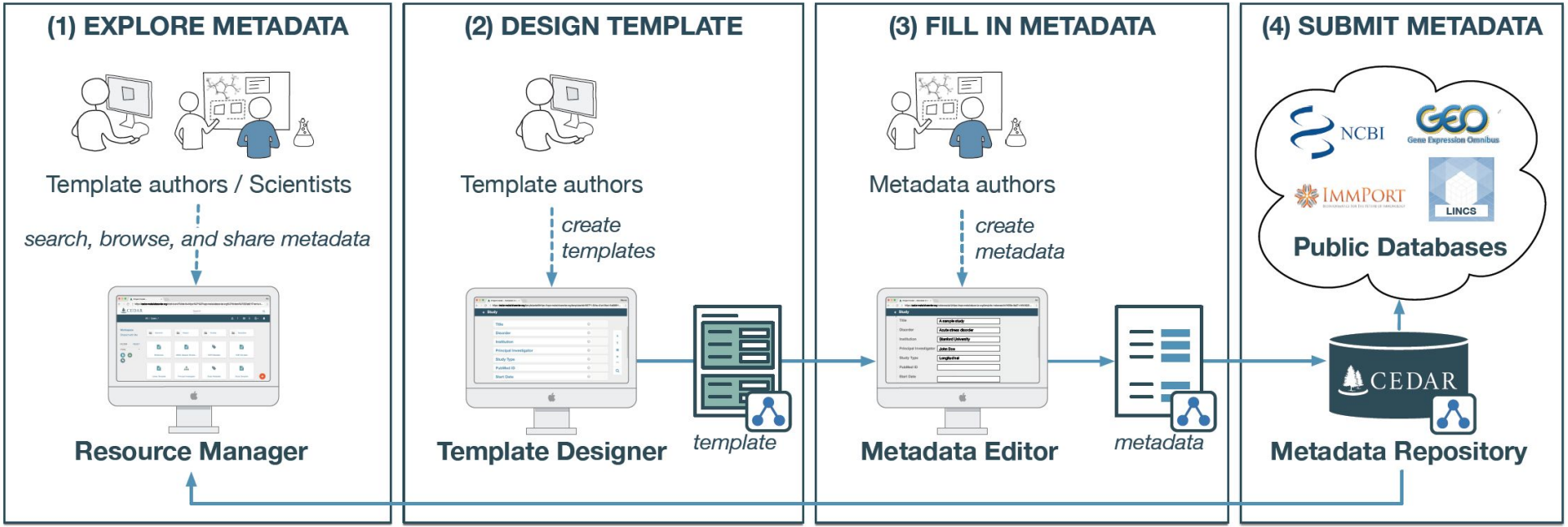
You find and work with them in the CEDAR WorkBench.



CEDAR


Log in to CEDAR
auth.metadatacenter.org/auth/realms/CEDAR/protocol/openid-connect/auth?client_id=cedar-angular-app&redirect_uri=htt...
CEDAR
Email
niva@dtu.dk
Password
Remember me
Forgot Password?
LOG IN
Google
GitHub
New user? Register
Watch the video tutorial

Workflow for metadata authoring and submission



https://cedar.metadatacenter.org

CEDAR artifacts

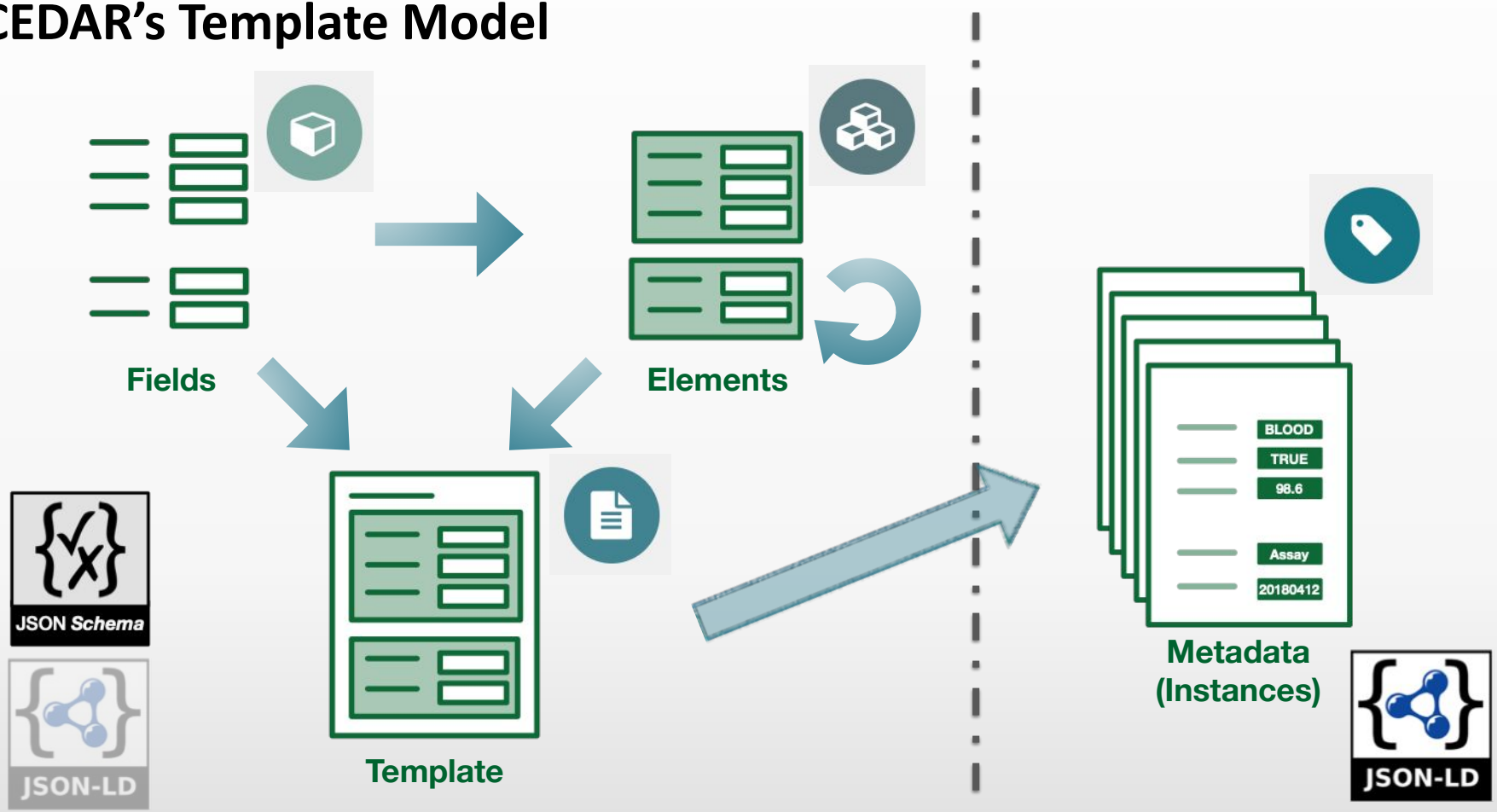
Metadata template **fields** (e.g., *title*, *description*, *start date*)

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

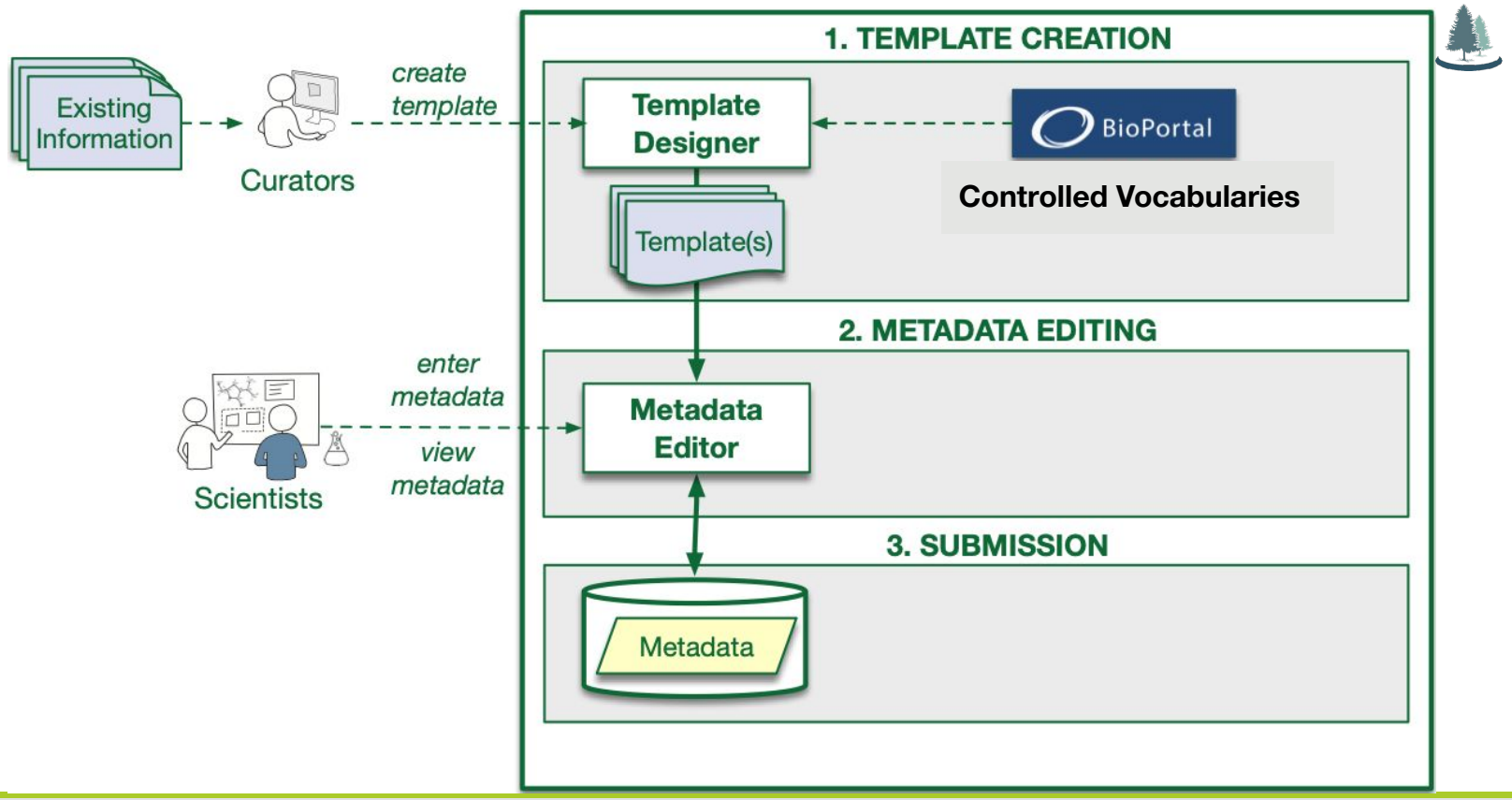
Metadata **templates** (e.g., *GDMT*)

Metadata (aka template **instances**)

CEDAR's Template Model



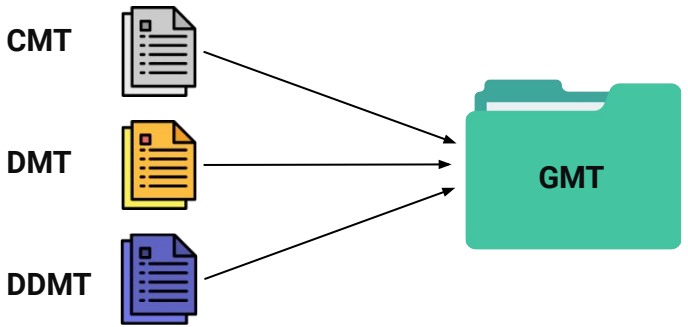
CEDAR-Only Workflow



Exercise 1: **Make your copy of the GMT templates**

0. Login to **CEDAR**
1. Click New
2. Select Folder
3. Write **GMT** and click save
4. In the search bar write CMT and hit enter
5. Select "Catalog Template (**CMT**)"
6. Press on three dots, and click on Copy To
7. Select GDMT folder and press copy
8. Repeat this process for the remaining two templates, which are **DMT** and **DDMT**
9. Upon completion of the previous step click on Workspace
10. Go to the GDMT folder where you should find your own copy of **CMT**, **DMT** and **DDMT**

[A video walkthrough]



Exercise 2: **Create** subject and variable **fields**

1. Click New
2. Select Field
3. In “Enter Field Name” write *subject*
4. In “Enter Preferred Label” write *Subject*
5. In "Enter Field Help Text" write "*Concepts (terms, classification, themes, topics, etc.) that define the dataset or purpose (i.e., subjects which can be addressed) using the dataset.*"
6. Click on Values
7. Click on settings icon (looks like a small cog) which is at the end of the search bar
8. In Advanced Search Option menu select "Search for an ontology in BioPortal..."
9. In the search bar write "neat"
10. From results select "*NEAT: wiNd Energy tAxonomy of Topics*", scroll to the bottom of the page and click Add
11. Click Save Field
12. Go back to the GDMT folder
13. Repeat the previous steps for the field variable using following information:
    - Field Name: *variable*
    - Preferred Label: *Variable*
    - Help text: *A single entity being measured, computed or described.*
    - Ontology: *ASPECT: wind energy vAriableS, ParametErs and ConsTants*
14. Once you create and save field variable, go back to the GDMT folder

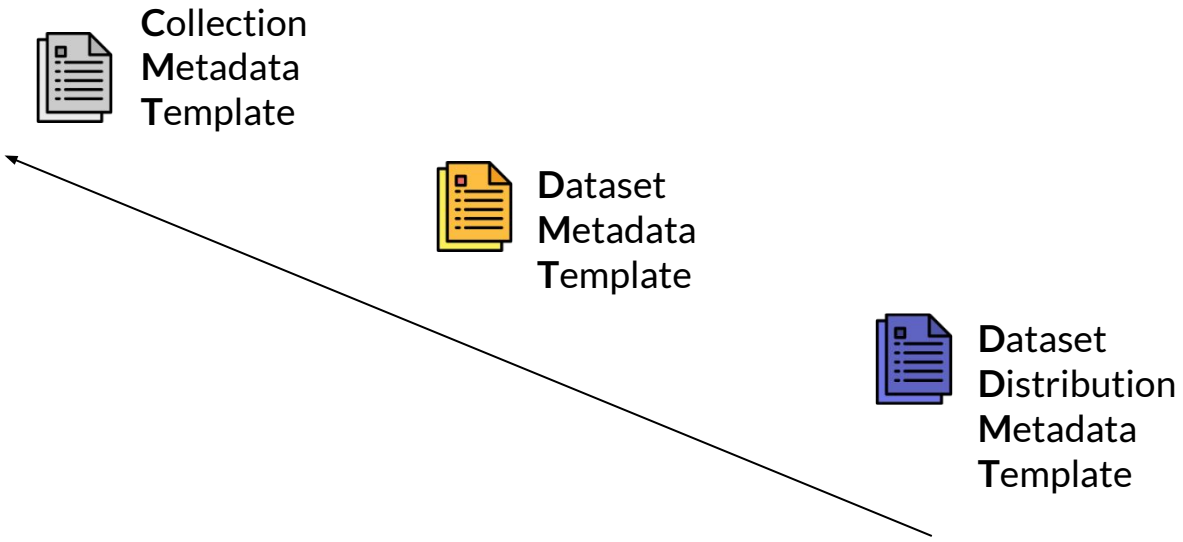
A video walkthrough

Exercise 3: **Update DMT template with domain-specific** fields

1. Double click on the "Dataset Template (DMT)"
2. In the template find current variable and subject fields and delete them
3. Click on magnifying glass icon and find the subject and variable fields that you've just created
4. Add one by one to the template
5. Double click on the new subject field and set Multiple option to Yes
6. Optionally you can set the subject field to be required
7. Afterward click on the icon represented with a triangle, as you hover over it a text "Add property" will be shown
8. In the search bar write "subject"
9. In Advance Search Options menu write and select "DCT" in "Narrow your search to specific ontologies" bar
10. In the resulting list select property Subject which contains definition and press OK
11. Repeat the previous steps for the variable field, but when adding property find and select property "hasVariable" which is a part of FDC-GDMT ontology
12. Drag the new fields to the position where the old fields were located
13. Save the edited template and go back to the GDMT folder

A video walkthrough

Now we instantiate templates and make metadata
We go from bottom up, starting with the distribution level

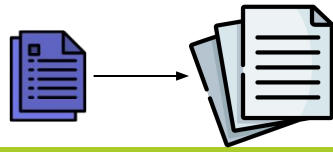


Exercise 4: **Instantiate your local** Distribution Template (**DDMT**)

1. Click on the 'label' icone which will bring you to the field entry screen
2. Enter requested metadata
3. Click Save and go back to the GDMT folder
4. Enable OpenView for the instance (i.e. metadata) of DDMT template
5. Try visiting OpenView to see values you entered
6. In case you did not enable OpenView for DDMT template, you will not be able to see corresponding instance in OpenView, therefore enable OpenView for all templates in the GDMT folder
7. Upon enabling OpenView for templates, visit again DDMT instance in OpenView
8. From the navigation bar copy the resulting **link***, which will be GUPRI of your distribution metadata
9. Go back to the GDMT folder

A video walkthrough

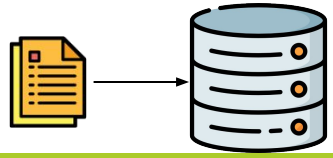
**CEDAR has globally unique and persistent identifiers for all artefacts, however they are not resolvable. We use OpenView URLs instead to make them resolvable, a bit of hack but gets work done! Nevertheless, please 'nudge' CEDAR team to make their identifier resolvable by upvoting the following CEDAR ticket on Github: https://github.com/metadatacenter/cedar-project/issues/1151*



Exercise 5: **Instantiate your local** Dataset Template (**DMT**)

1. Click on the 'label' icon which will bring you to the field entry screen
2. Scroll all the way down the bottom of the page and for field Distribution Identifier paste the previously copied URL for distribution metadata from OpenView
3. Enter values for the remaining fields
4. Click Save and go back to the GDMT folder
5. Enable OpenView for the instance (i.e. metadata) of **DMT** template
6. Try visiting OpenView to see values you entered
7. From the navigation bar copy the resulting link (CEDAR has globally unique and persistent identifiers for all artefacts, however they are not resolvable. By using corresponding OpenView URL instead you getting a resolvable version of your instance identifier)
8. Go back to the GDMT folder

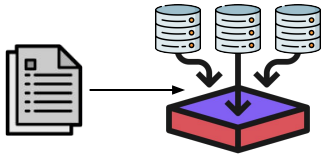
A video walkthrough



Exercise 6: **Instantiate your local** Catalog Template (CMT)

1. Click on the 'label' icon which will bring you to the field entry screen
2. Paste the previously copied URL for dataset metadata from OpenView to the Dataset Identifier field
3. Enter values for the remaining fields
4. Click Save and go back to the GDMT folder
5. Enable OpenView for the instance (i.e. metadata) of CMT template
6. Try visiting OpenView to see values you entered, and navigate back and forth between three layers of metadata which represents DCAT organization
7. **Congrats you have completed rapid M4M workshop! Well Done!**

[A video walkthrough]

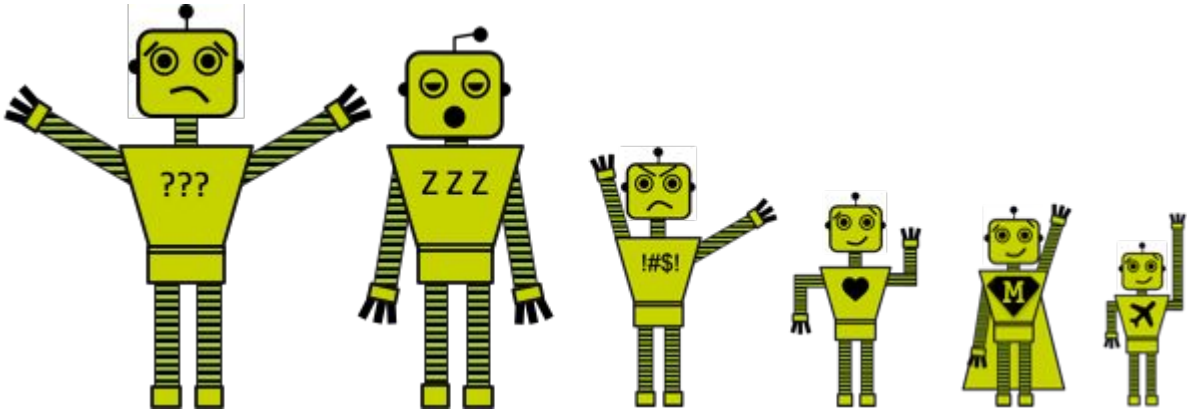


CREDITS

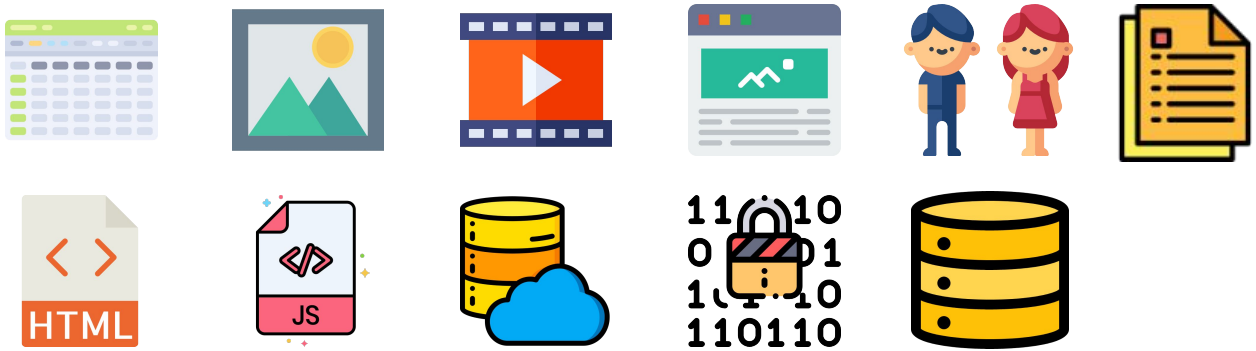
John Graybeal for the slides on CEDAR Workbench

Source of graphical material for slides - MetaManMachine

Vasiljevic, Nikola. (2021).
MetaManMachine. Zenodo.
http://doi.org/10.5281/zenodo.4471098
Licensed under: CC BY-SA 4.0



Source of graphical material for slides - Icons



Icons made by https://www.freepik.com

Logo made by Bill Schwappacher <bill@tracermedia.com> provided by W3C for public use

Released by https://json-ld.org/ under CC0