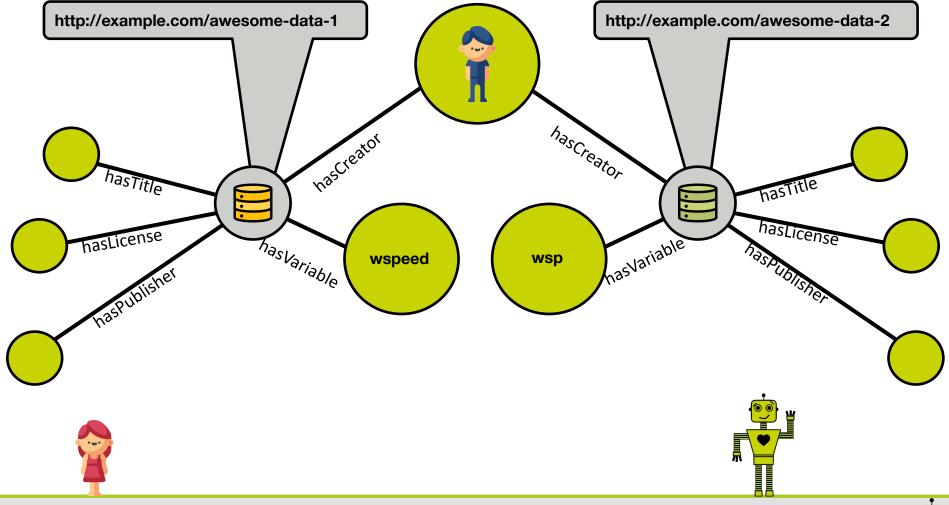
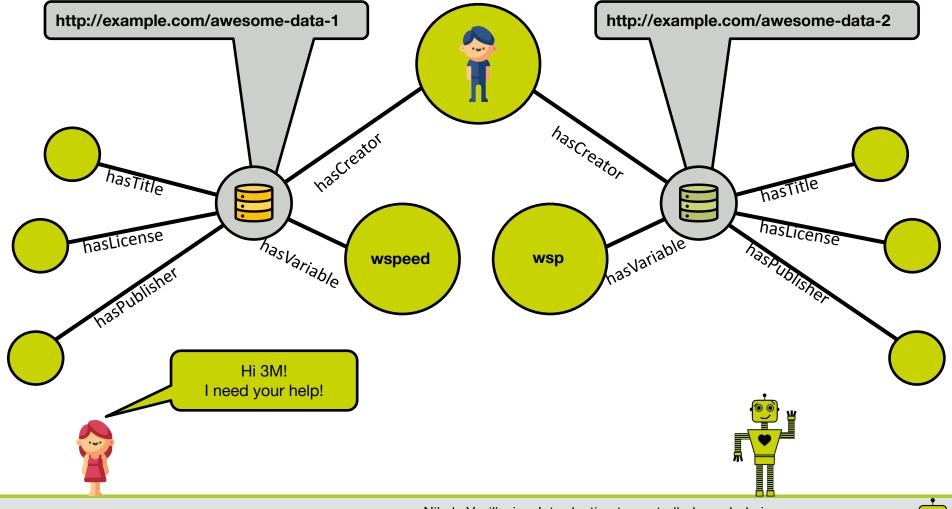
DOI:10.5281/zenodo.4705676

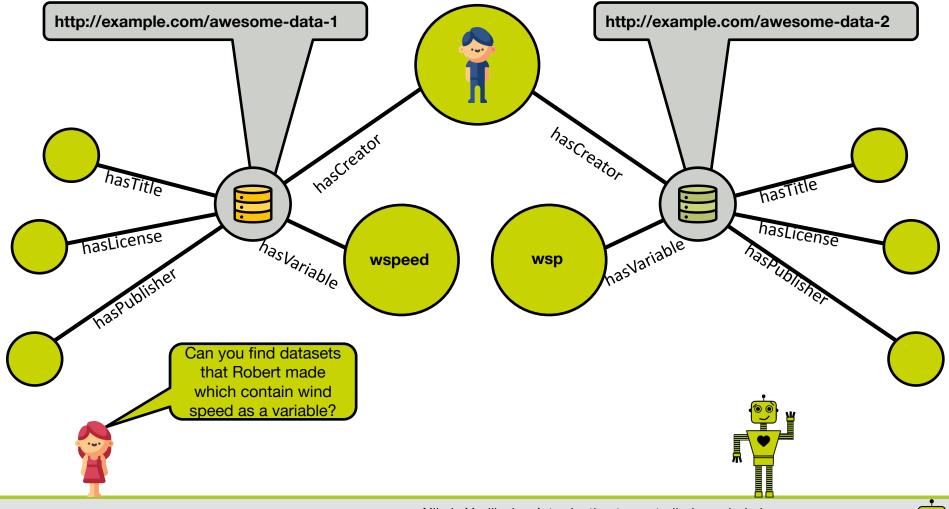
Introduction to controlled vocabularies and how to build them

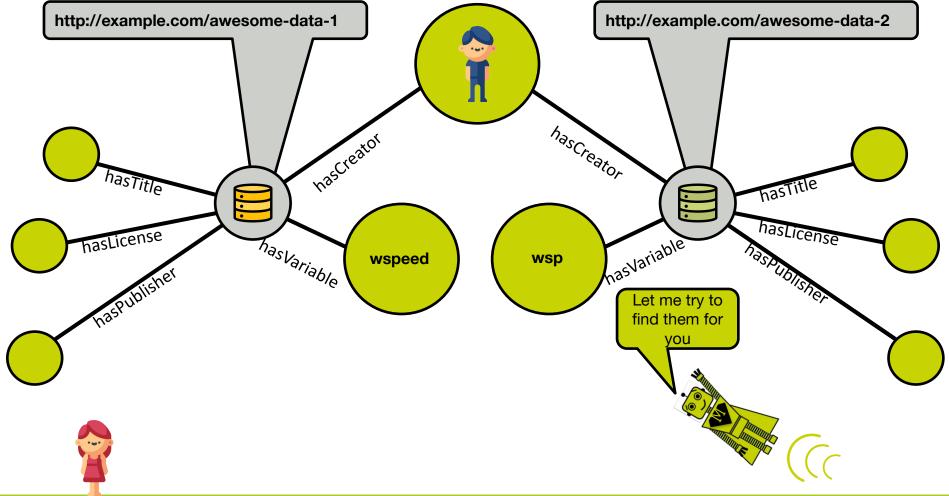
Nikola Vasiljević, DTU Wind Energy

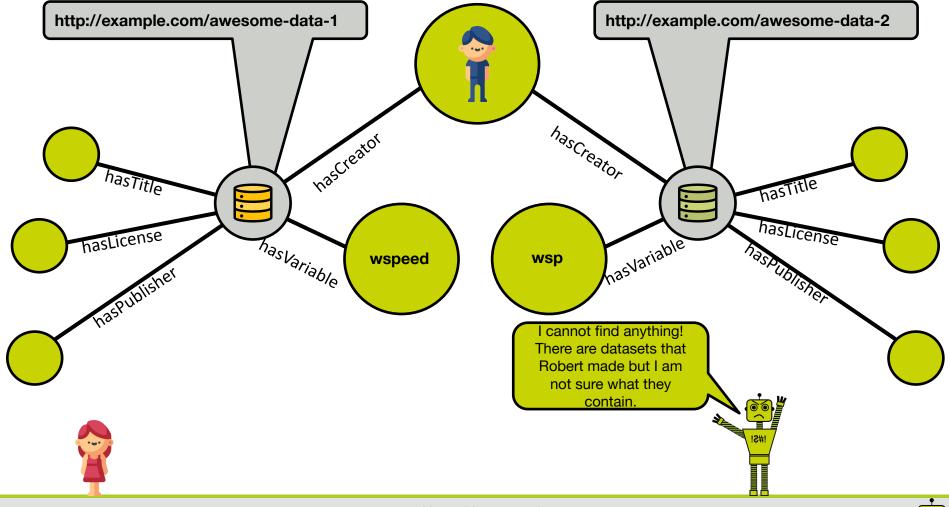




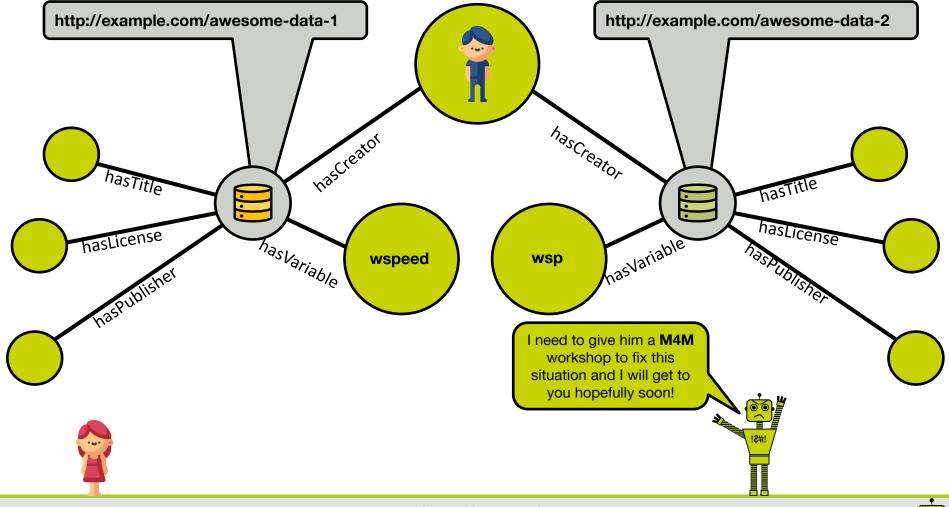


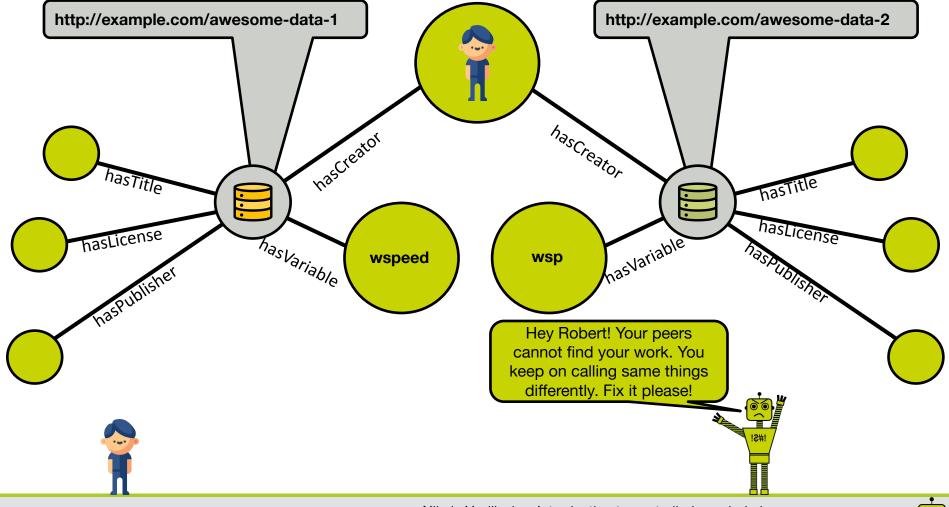


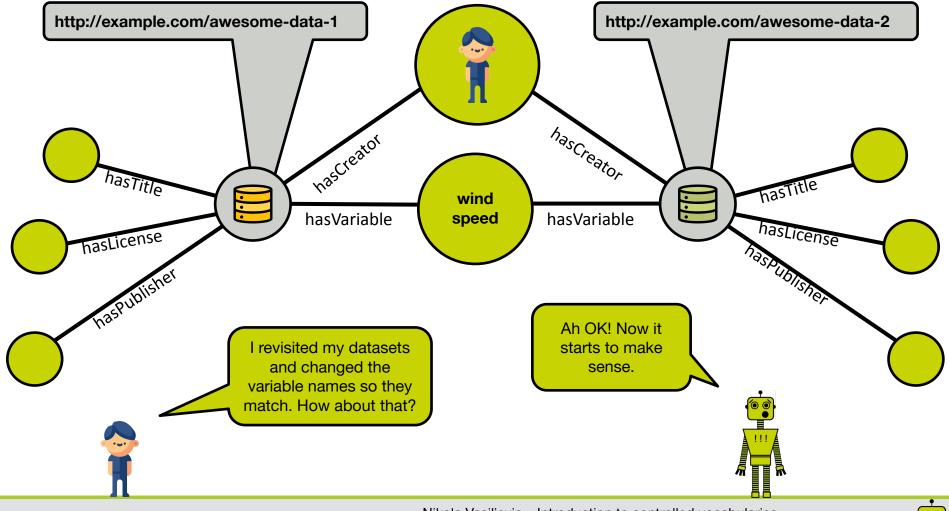


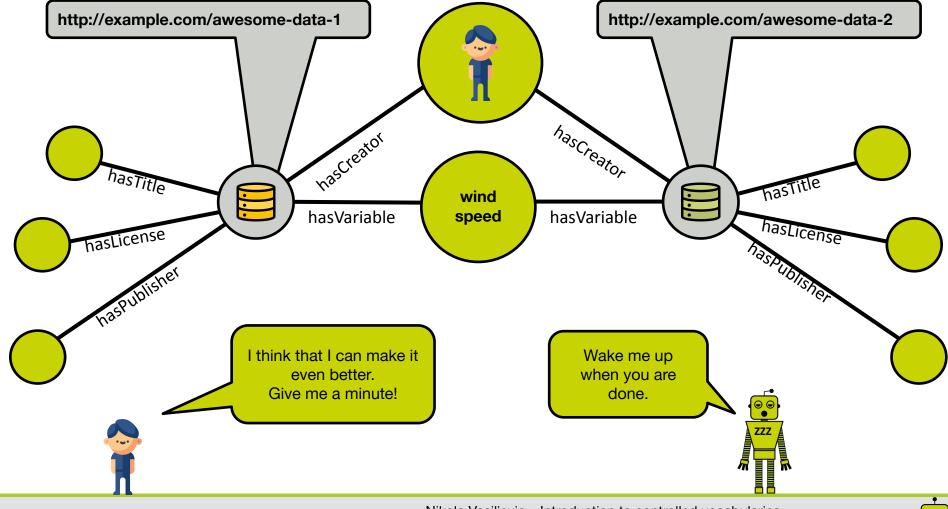


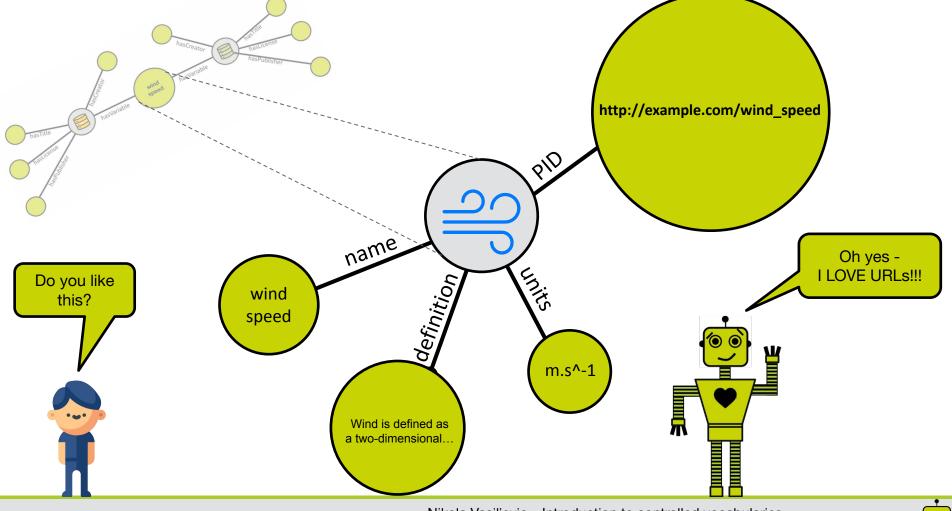


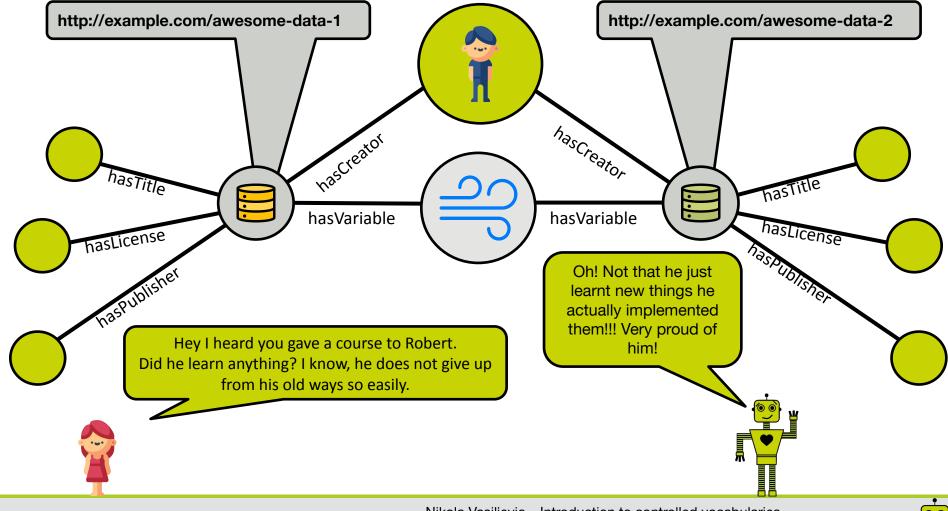


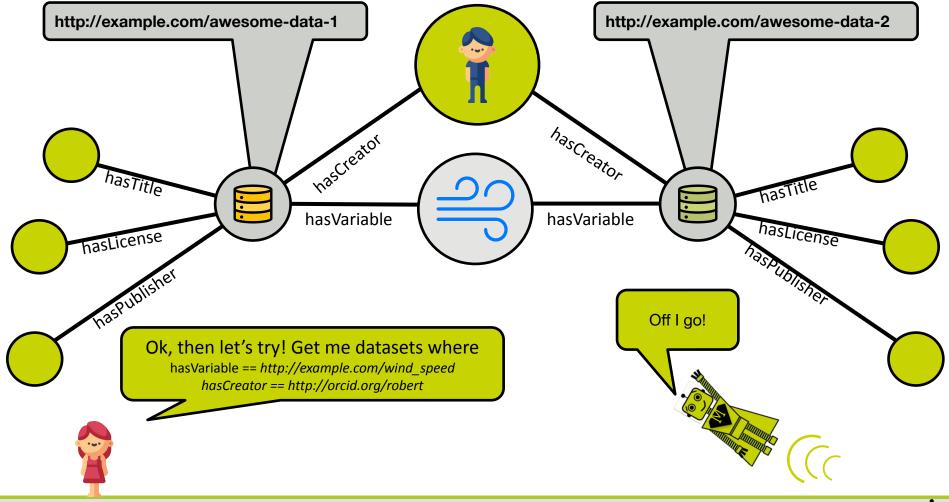


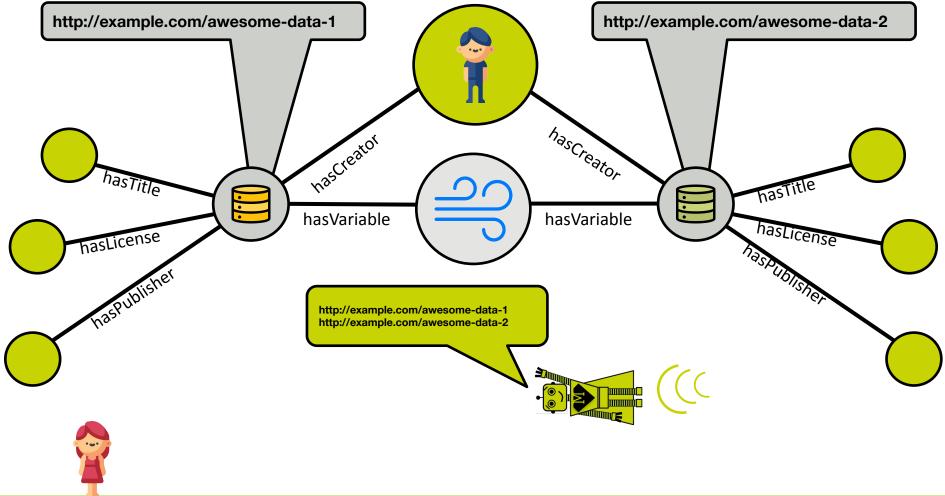






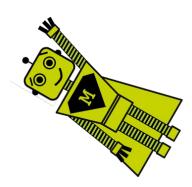






Why should vocabularies be FAIR?*

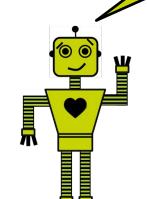
- > We need to know if concepts in different datasets mean the same thing
- > We want to annotate data using concepts from vocabularies
- > We want to use vocabularies that are standard-based and machine-actionable
- > We want to use vocabularies that are supported by the community



^{*} Ten Simple Rules for making a vocabulary FAIR. 2021, http://arxiv.org/abs/2012.02325

Why should vocabularies be FAIR?*

- 1. Governance arrangements for the legacy vocabulary
- 2. Licensing, legacy and new vocabulary
- 3. Check term definition completeness in legacy vocabulary
- 4. Establish a technical maintenance environment for the FAIR vocabulary
- 5. Assign a unique identifier to the vocabulary and to its each item
- 6. Create machine readable representations of the vocabulary terms
- Add rich metadata
- 8. Register the vocabulary in a semantic repository
- 9. Make the identifiers of vocabulary and terms resolvable
- 10. Implement a process for maintaining the FAIR vocabulary



* Ten Simple Rules for making a vocabulary FAIR. 2021, http://arxiv.org/abs/2012.02325

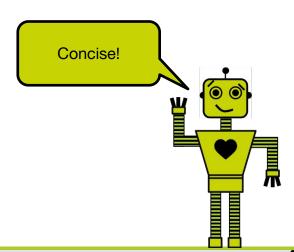


Instructions to

follow!!!

Characteristics of FAIR vocabularies

- ☐ FINDABLE: indexed and registered in a community or generic service
- □ ACCESSIBLE: retrievable on the web by both machine and humans
- ☐ INTEROPERABLE: encoded in standard representation, mapped to existing vocabularies
- □ REUSABLE: licensed under open license, maintained



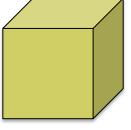
^{*} Ten Simple Rules for making a vocabulary FAIR. 2021, http://arxiv.org/abs/2012.02325

How to build controlled vocabularies?

RDF

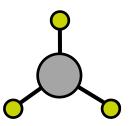


DATA MODEL TURTLE JSON-LD XML-RDF



FORMAT

SKOS OWL



REPRESENTATION LANGUAGE

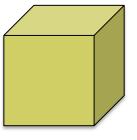
How to build controlled vocabularies?

RDF



DATA MODEL TURTLE

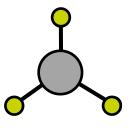
JSON-LD XML-RDF



FORMAT

SKOS

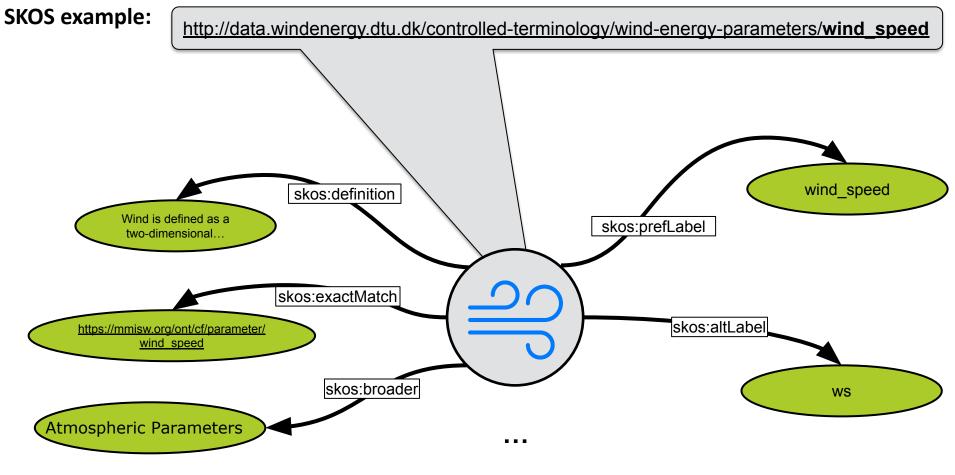
OWL



REPRESENTATION LANGUAGE

Why selecting RDF, Turtle and SKOS?

- > RDF (Resource Data Framework) is a standard model for data interchange on the Web
- > Turtle is a common, most human-readable and very compact data format for storing RDF data
- > **SKOS** (Simple Knowledge Organization System) is a W3C recommendation designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or **any other type of structured controlled vocabulary**.



This is a graphical representation of a **knowledge graph**.



Encoding graph in RDF Turtle format

@prefix : <http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/> .

@prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# .

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

:wind_speed a skos:Concept;

Yes I can!
Check tools in slides after the break.

skos:prefLabel "wind_speed";

skos:altLabel "ws", "horizontal wind speed";

skos:definition "Wind is ..."@en;

skos:exactMatch <https://mmisw.org/ont/cf/parameter/wind_speed>;

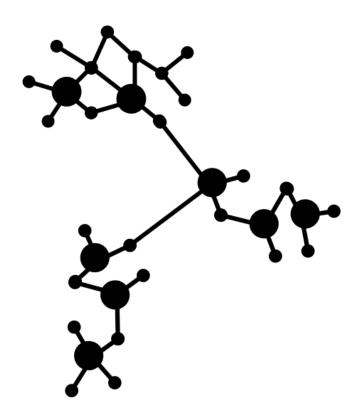
skos:broader:AtmosphericParameters.

*:wind_speed translates into http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/wind_speed

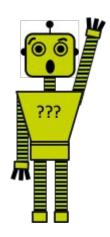
Hey Robert,
you said it is the 'most'
human readable
format! The course
participants will find
this very cryptic and
nerdy!!!

Can you make it simpler?!

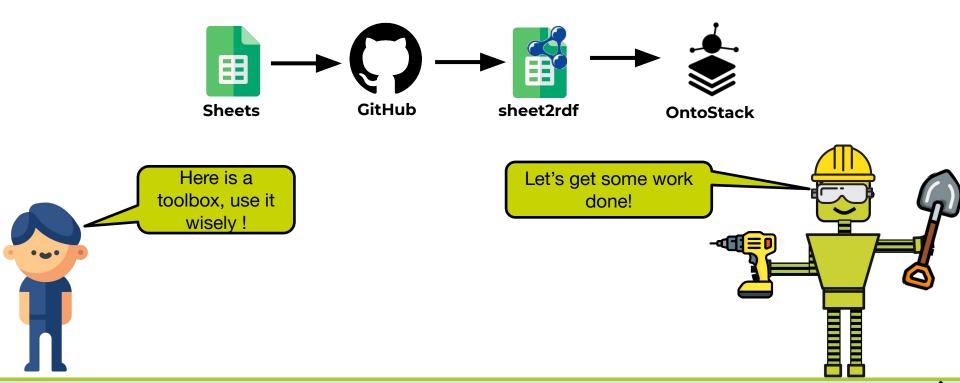




QUESTIONS???

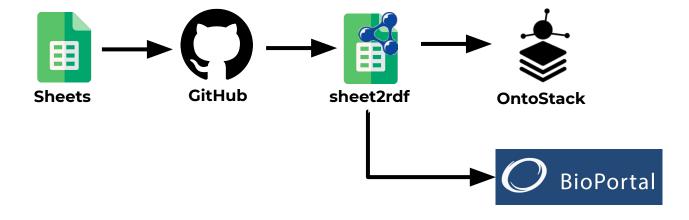


Effective workflow for generating and publishing controlled vocabularies



2/15/21 | **S 24**

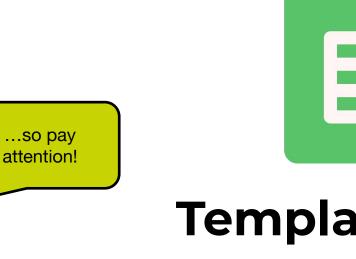
Effective workflow for generating and publishing controlled vocabularies





There will be quite some text now...

Template Sheet





Template Sheet



Template Sheet

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:	1			
vocab-name:	(1			
vocab-name:				



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/	General setup
PREFIX	owl	http://www.w3.org/2002/07/owl#	3 611 61 61 6 6 6
PREFIX	xsd	http://www.w3.org/2001/XMLSchema#	
PREFIX	skos	http://www.w3.org/2004/02/skos/core#	
dct:title			
dct:description			
dct:creator			Cautualladoradabolano
dct:rights			Controlled vocabulary
pav:version			on at a data
pav:createdOn			metadata
pav:lastUpdatedOn			
Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	
vocab-name:			
vocab-name:			
vocab-name:			Definition of terms and
vocab-name:			
vocab-name:			properties
vocab-name:			



General setup

Controlled vocabulary metadata

Definition of terms and properties

	<u>ntep.//ontology.deletolg/ov/vocab name/</u>
•	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 – <u>Simple Knowledge Organization System</u>
	dct – <u>Dublin Core Terms</u>
	pav – <u>Provenance Authoring and Versioning</u>

http://purl.org/pav/

ConceptScheme URI: This is a base URL for all your controlled terms (aka concepts in SKOS).

http://purl.org/dc/terms/

http://www.w3.org/2002/07/owl#

http://ontology.deic.org/cv/vocab-name/

http://www.w3.org/2001/XMLSchema#

http://www.w3.org/2004/02/skos/core#

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/

http://ontology.deic.org/cv/vocab-name/

vocab-name

http://ontology.deic.org/cy/yocab-name/

pay

dct

owl

xsd

owl - <u>Web Ontology Language</u> xsd - <u>XML Schema Definition</u>

skos

...

ConceptScheme URI

In this example:

PREFIX

PREFIX

PREFIX

PREFIX

PREFIX

PREFIX

dct:title dct:desc dct:crea dct:right pav:vers pav:crea pav:last	cription itor ts sion atedOn		Insert values here			
•	It is r			ogies to provide metadata a	bout a controlled	d vocabulary, e.g.:
		dct:title -	- controlled vocabulary title			
		dct:creato vocabular		ulary creator (full URL!), repe	at this row as ma	any time as there are
				abulary, preferably chose lice pdx.org/licenses/CC0-1.0 for		//spdx.org/licenses/
		pav:versi	on – version of the controlle	ed vocabulary, e.g. in a form o	of majorChange.	.minorChange.bugFix
		pav:creat	edOn – initial datetime of t	he vocabulary creation in ISC	D 8601 format (in	nclude time zone!)
		pav:lastU	pdatedOn – datetime of th	e last vocabulary update in I	SO 8601 format	(include time zone!)

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

Insert values here

Every time you update your control vocabulary update:

terms.

dct:title dct:description dct:creator dct:rights

pav:version pav:createdOn pav:lastUpdatedOn

Identifier	skos:prefLabel@en
vocab-name:	u e
vocab-name:	Insert values here
vocab-name:	
Column Identi	fier will be automatically g

Column Identifier will be automatically generated based on values in column skos:prefLabel, and it will be eventually turned into URL!

NB: The URL (and thus the skos:prefLabel) should be composed from a limited set of characters (US-ASCII), i.e. digits (0-9), letters(A-Z, a-z), special characters ("-", ".", "_", "~")

• If you really need to use characters in **skos:prefLabel** column which does not satisfy the above requirement, then you need to manually edit **Identifier** column such that the requirement is satisfied!!!



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

prefLabel = Århus, results in Identifier = vocab-name:Århus, which is then turned into http://ontology.deic.org/cv/vocab-name/Arhus and this is not a functional URL!

Therefore, if using 'Å' is important, you must manually replace **Identifier** with a value that can turn into a proper URL, for example: Identifier = vocab-name:AArhus



Identifier	skos:prefLabel@en
vocab-name:	
vocab-name:	
vocab-name:	
• @en part in sk	os:prefLabel@en indicates i

- To change preferred label language tag find appropriate two-letter tag in: https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry
- language tag!!!
 - Terms must have unique preferred labels, i.e. two different terms cannot have same preferred labels!

You can have preferred labels in multiple language, but you can only have one preferred label per



Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	
cab-name:			
cab-name:			
cab-name:	1		

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=",")

skos:altLabel is an alternative label for the term you are defining. Unlike skos:prefLabel you can have both

• If you prefer other ways of separating values in this column change this value accordingly.

2/15/21 | **S 37**

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

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

Change the current language tag (en) or add additional skos:definition columns associate with different

2/15/21 | **S 38**

language tags as you like/need.

definition per language tag!

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).						

skos:altLabel(separator=",")

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.
- in ease your term has several broader concepts, simply separate their corresponding race

If a term is a top concept (i.e., does not have broader concepts) leave this column empty.

skos:prefLabel@en



Identifier

vocab-name:

skos:definition@en skos:broader(separator=",")

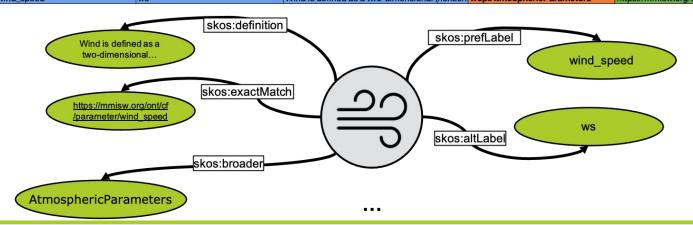
A few more columns you will encounter in the template...



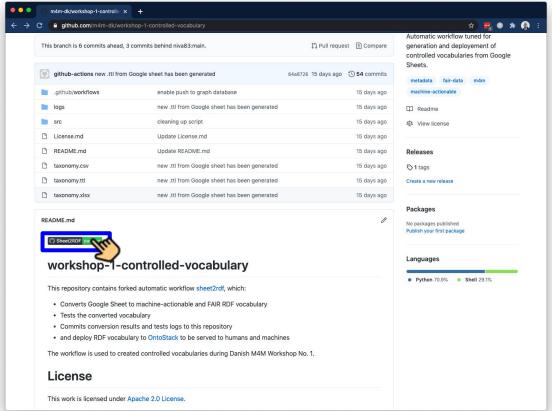
dct:source - place reference, such as URL, to the source material that you used to define a term. **skos:closeMatch** - place Identifier of term from this or another vocabulary to which the term you defined is closely related skos:exactMatch - place Identifier of term from this or another vocabulary to which the term you defined is identical to the term you are defining **owl:deprecated -** in case term is deprecated set value to **true** in this column dct:isReplacedBy - in case when deprecated term has been replaced by another terms, put URL of new term in this column dct:creator - put ORCID ID of term creator, for multiple contributors separate their IDs with comma dct:contributor - put ORCID ID of term contributor, for multiple contributors separate their IDs with comma

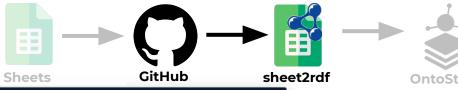
Example for wind speed

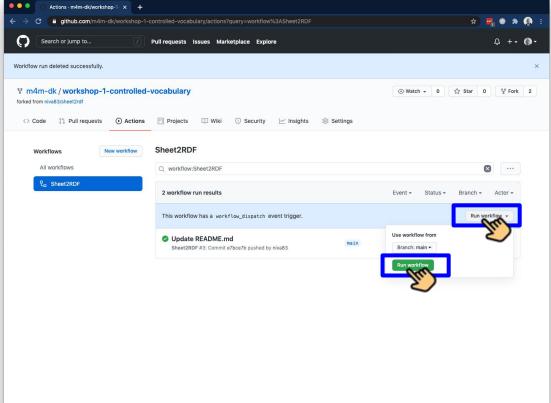
ConceptScheme URI	http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/				
PREFIX	wep	http://data.windenergy.dtu.dk/control	led-terminology/wind-energy-parameters/		
PREFIX	pav	http://purl.org/pav/			
PREFIX	dct	http://purl.org/dc/terms/			
PREFIX	owl	http://www.w3.org/2002/07/owl#			
PREFIX	xsd	http://www.w3.org/2001/XMLSchema	1#		
PREFIX	skos	http://www.w3.org/2004/02/skos/core	e#		
dct:title	Wind Energy Parameters				
dct:description	Controlled vocabulary of paramete	rs used in wind energy domain.			
dct:creator	https://ror.org/04qtj9h94				
dct:rights	https://spdx.org/licenses/CC0-1.0				
pav:version	0.1.0				
pav:createdOn	2020-10-22T22:00:00+01:00				
pav:lastUpdatedOn	2020-10-23T08:00:00+01:00				
Identifier	skos:prefLabel@en	skos:altLabel(separator=",")	skos:definition@en	skos:broader(separator=",")	skos:exactMatch(separator=",")
wep:AtmosphericParameters	Atmospheric Parameters		This is a category of parameters associated wi		
wep:wind_speed	wind speed	ws	Wind is defined as a two-dimensional (horizon	wep:AtmosphericParameters	https://mmisw.org/ont/cf/parameter/wind_speed











Sheet2RDF



- > Automatic workflow executed by means of GitHub actions
- > Contains underlying shell, python and java programs which:
 - (1) converts the previous Google Sheet to the machine-actionable controlled vocabulary
 - (2) tests the derived controlled vocabulary
 - (3) commits the conversion results and tests logs to a Git repository
 - (4) deploys the vocabulary to OntoStack to be served to humans and machines
- > **Sheet2RDF** is used by:
 - DTU Wind Energy
 - DeiC
 - International Energy Agency WIND Task 32
 - Dutch COVID program
 - _ ...

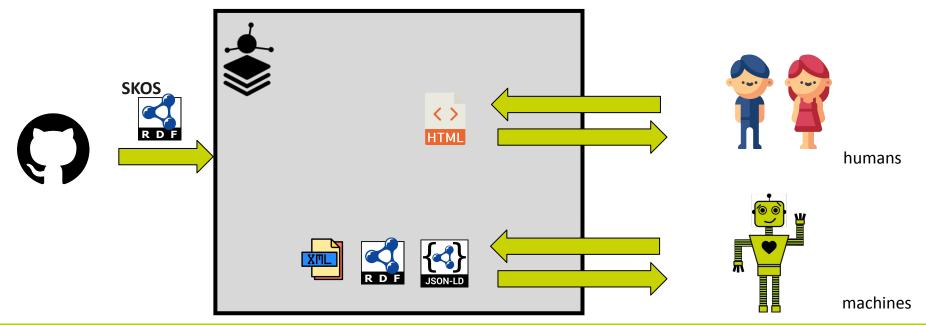
https://github.com/fair-data-collective/sheet2rdf



OntoStack



> A set of orchestrated micro-services configured and interfaced such that they can intake terminologies and serve them to humans or machines



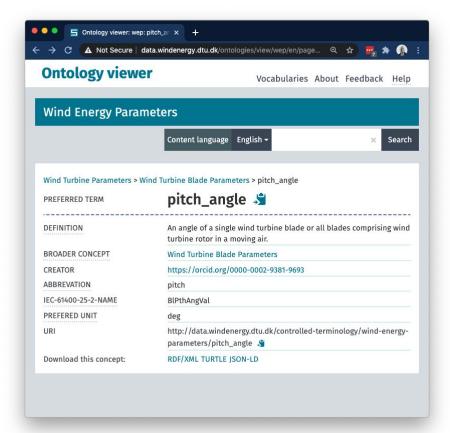
OntoStack in its core



- > A set of orchestrated micro-services:
 - Edge router (Traefik)
 - Graph database (Apache Jena Fuseki)
 - Web-based terminology browser/UI (SKOSMOS)
- > Four instances of OntoStack:
 - Departmental:http://data.windenergy.dtu.dk/ontologies/view
 - National: http://ontology.deic.dk
 - International:
 http://vocab.fairdatacollective.org/
 http://vocab.ieawindtask32.org/



OntoStack response to Resolvable Identifier



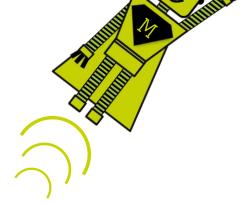




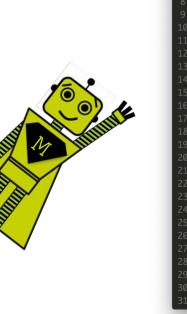
OntoStack response to Resolvable Identifier

```
data.windenergy.dtu.dk/ontolo x +
             A Not Secure | data.windenergy.dtu.dk/ontologies/view/rest/v1/wep/d... @
@prefix wep: <a href="http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/">energy-parameters/</a>>
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/terms/> .
wep:abbrevation
 skos:definition "A shortened form of a parameter label used as a short name of parameter
in different applications. "@en ;
 rdfs:label "abbrevation"@en .
wep:WindTurbineBladeParameters
  skos:prefLabel "Wind Turbine Blade Parameters"@en ;
 a skos:Concept;
  skos:narrower wep:pitch angle .
  skos:prefLabel "WindEnergyParameters"@en ;
  a skos:ConceptScheme .
wep:iec-61400-25-2-name
  skos:definition "IEC 61400-25-2 standardized name of a parameter"@en ;
  rdfs:label "iec-61400-25-2-name"@en .
wep:pitch angle
  skos:broader wep:WindTurbineBladeParameters;
 wep:long-name "Pitch Angle"@en ;
  skos:definition "An angle of a single wind turbine blade or all blades comprising wind
turbine rotor in a moving air. "@en ;
  wep:prefUnit "deg"@en ;
  dc:creator <https://orcid.org/0000-0002-9381-9693>;
  a skos:Concept;
  wep:iec-61400-25-2-name "BlPthAngVal"@en;
  wep:abbrevation "pitch"@en ;
  skos:inScheme wep: ;
  skos:prefLabel "pitch angle"@en .
wep:prefUnit
  skos:definition "A prefered unit for a given parameter"@en ;
  rdfs:label "prefered unit"@en .
```





OntoStack response to Resolvable Identifier

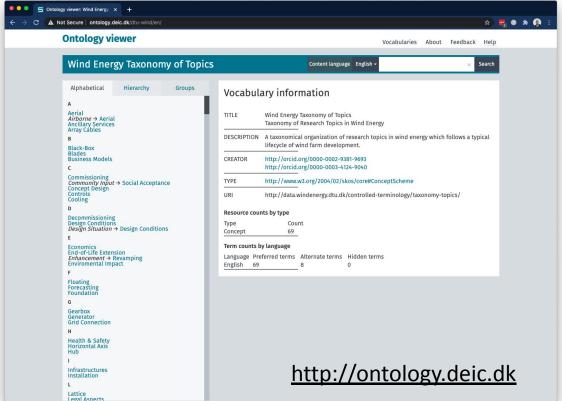


```
A data.windenergy.dtu.dk/ontolo × +
  "@context": {
  "skos": "http://www.w3.org/2004/02/skos/core#",
  "isothes": "http://purl.org/iso25964/skos-thes#",
  "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
  "owl": "http://www.w3.org/2002/07/owl#",
  "dct": "http://purl.org/dc/terms/",
  "dc11": "http://purl.org/dc/elements/1.1/",
  "uri": "@id",
  "type": "@type",
  "lang": "@language",
  "value": "@value",
  "graph": "@graph",
  "label": "rdfs:label",
  "prefLabel": "skos:prefLabel",
  "altLabel": "skos:altLabel",
  "hiddenLabel": "skos:hiddenLabel",
  "broader": "skos:broader",
  "narrower": "skos:narrower",
  "related": "skos:related",
  "inScheme": "skos:inScheme",
  "exactMatch": "skos:exactMatch",
  "closeMatch": "skos:closeMatch",
  "broadMatch": "skos:broadMatch",
  "narrowMatch": "skos:narrowMatch",
  "relatedMatch": "skos:relatedMatch"
"graph": [
```









BioPortal

BioPortal allows you to:

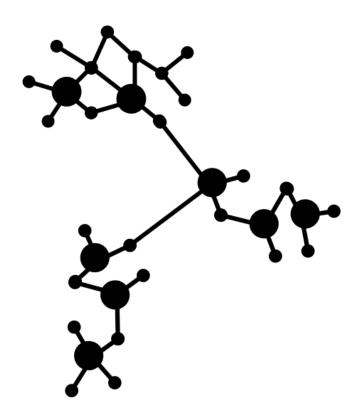
- Browse the library of ontologies
- Search for a term across multiple ontologies
- Browse mappings between terms in different ontologies
- Receive recommendations on which ontologies are most relevant for a corpus
- Annotate text with terms from ontologies
- ...
- Make use of controlled vocabularies in CEDAR Workbench

Configure BioPortal to automatically load sheet2rdf vocabularies

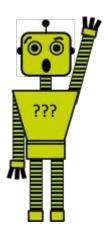
- 1. Go to the *Github* repository which is configured to run *sheet2rdf* workflow
- 2. Find *vocabulary.ttl* file in the repository
- 3. Click on vocabulary.ttl
- 4. Click on the Raw button
- 5. Copy the resulting link from the browser bar
- 6. Open another tab and go to BioPortal (make sure you have a BioPortal account that you are logged in)
- 7. Click on **Ontologies**
- 8. Click on Submit New Ontology
- 9. Fill in necessary fields
- 10. Click Create ontology
- 11. Fill in necessary fields
- 12. In Location select Load from URL
- 13. Past the link you copied before
- 14. Click Add submission
- 15. Click on the link that BioPortal generated for you which will be landing page for your vocabulary in BioPortal

Every night BioPortal will check if there is new version of the vocabulary and register it if there is.

A video walkthrough: http://bit.ly/m4m-bioportal-github



QUESTIONS???



CREDITS



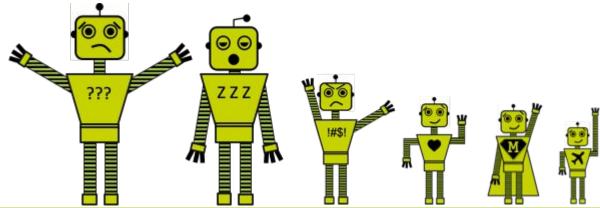
Source of graphical material for slides - MetaManMachine

Vasiljevic, Nikola. (2021).

MetaManMachine. Zenodo.

http://doi.org/10.5281/zenodo.4471098

Licensed under: <u>CC BY-SA 4.0</u>



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 CCO



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 CCO

