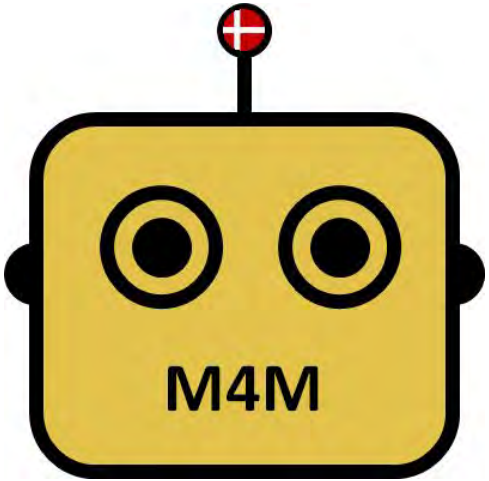
Metadata for Machines

> DeiC-webinar 21 January 2021



Program

- Introducing the DeiC rational: Why M4M?
- M4M in the flowchart of a FAIRification process
- GoFAIR & DeiC M4M workshop – Summer 2020
- Three research case studies
- The need for local, national and international support functions
- Future directions and plans

Introducing the DeiC rational: Why M4M?

- Requests from researchers
- Continued collaboration with GO FAIR after Data Stewardship course
- FAIR strategy:
  - Support for academic and professional groups in defining their own implementation of FAIR principles
- Funding opportunity

M4M in the flowchart of a FAIRification process

1. **Define the FAIRification rational**
   a. Clarify why FAIRification and/or open access to the research data is pursued.
   b. Assess FAIRness of data to decide which metrics need improvement.
   c. Prioritise and choose datasets for FAIRification.
   d. Define who should have access to your data (domain-specific vs. broader audience).

2. **Define all data elements and their relations**
   a. Analyse the content of the data in terms of structure and concepts represented.
   b. Check, if there is an already existing vocabulary available, preferably from your research domain.
   c. If necessary, add vocabulary to an existing one to fit your needs, or build a new vocabulary.

M4M in the flowchart of a FAIRification process

3. **Prepare your metadata**
   a. Use an existing metadata template or, if not available, create one preferably in agreement with your research domain.
   b. Use the chosen vocabulary to describe the meaning of data elements and relations – accurately, unambiguously, preferably in a computer-actionable way
   c. Decide on licensing (who can access data how it can be used).
   d. Link metadata to datasets.

4. **Make decisions about software and hardware**
   a. Decide in which database or data repository your data/metadata should be stored.
   b. Secure operations (service level agreements, costing etc.), preferably also after the original research grant runs out.

M4M in the flowchart of a FAIRification process

5. **Implementation: Hosting the FAIR data and metadata**
   a. Implement and test operational databases, including external access, queries etc.
   b. Or, export data to a repository that is well suited for hosting FAIR data.

6. **Assess FAIRness of data, considering the objective**
   a. Re-assess FAIRness of your data.
   b. If there is still room for improvement, restart from the top.

GoFAIR & DeiC M4M workshop – Summer 2020

- Rene

Three research case studies - AnaEE

- Klaus Steenberg Larsen, Associate Professor
- Coordinator of AnaEE Denmark
- Department of Geosciences and Natural Resource Management
- Section for Forest Nature and Biomass
- University of Copenhagen

UNIVERSITY OF COPENHAGEN

AnaEE
Analysis and Experimentation on Ecosystems
Denmark

Example of starting a FAIRification process from scratch – with limited resources

Klaus Steenberg Larsen
Associate Professor, AnaEE Denmark Coordinator
Department of Geosciences and Natural Resource Management
University of Copenhagen

UNIVERSITY OF COPENHAGEN
AARHUS UNIVERSITY
RUC
DTU

What is AnaEE Denmark?

AnaEE (Analysis and Experimentation on Ecosystems):

- A pan-European research infrastructure for experimental field-scale research facilities (anaee.com)
- ERIC (European Research Infrastructure consortium) in preparation with planned start in 2021
- AnaEE Denmark is both a national Danish research infrastructure and the Danish contribution to AnaEE international
- Partners are UCPH, AU, DTU and RUC
- Official start of AnaEE Denmark 1 Jan 2018

Financing: 45 mill. DKK
Jan 2018 – Dec 2022
20 mill. DKK from Ministry of Higher Education and Science
25 mill. DKK from participating universities

AnaEE Analysis and Experimentation on Ecosystems (International)

AnaEE Analysis and Experimentation on Ecosystems Denmark

What is AnaEE Denmark?

AnaEE Denmark Funding will be used for:

- Updating platforms with new instrumentation for climate manipulations, sensors and greenhouse gas measurement technologies
- Opening up experimental platforms to outside users – develop common access policy and procedure for users
- International membership fees
- Hosting the AnaEE Technology Centre
- Consortium coordination
- Common protocols for usage of instruments, technologies and data analysis
- …and in-kind resources only for creating FAIR data!

Platforms

AnaEE Denmark FAIR data strategy

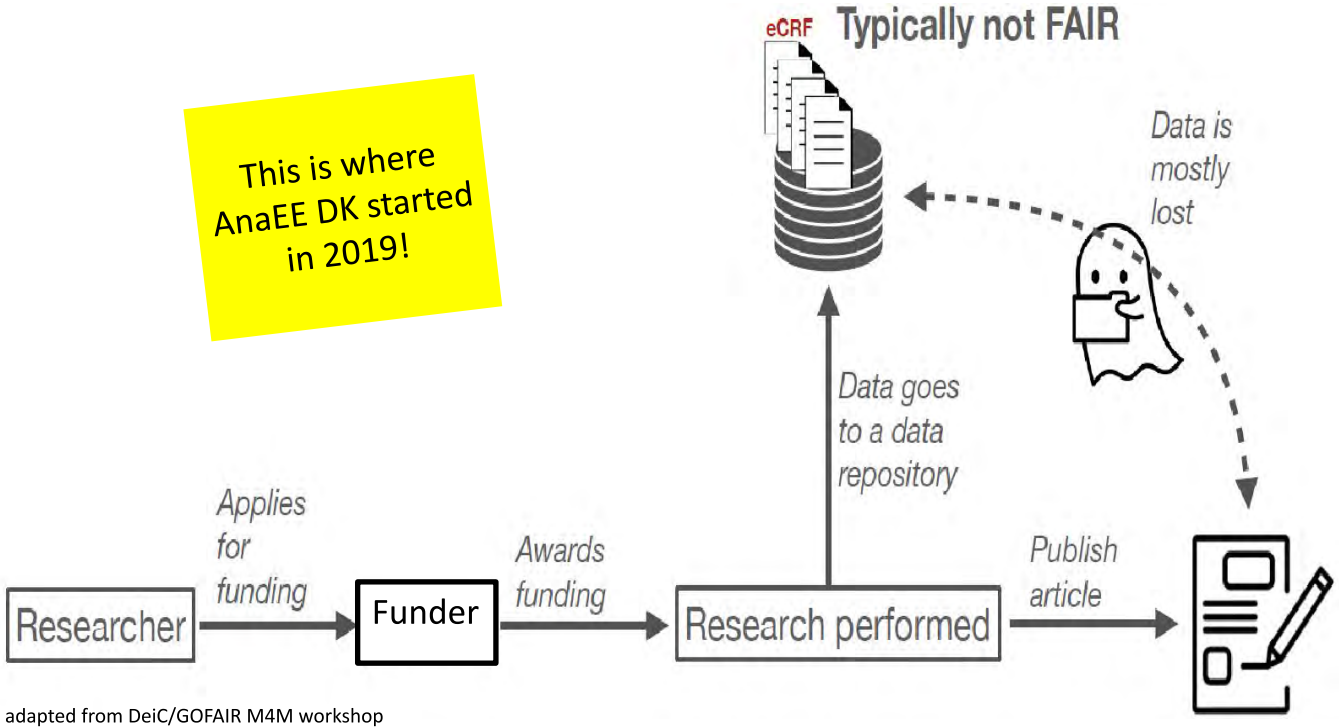
AnaEE Denmark FAIRification started from scratch

- Decentralised data storage – no common database (same at international level)
- Go for DOIs – resources and repositories are now becoming available at the Universities, e.g. ERDA at UCPH, LOAR (Royal Library, AU), and DTU Data at DTU
- Develop standardized metadata descriptions across platforms
- Data in several categories:
  - ”Easy”: Standard meteorological data from experiments
  - ”Difficult”: Measured response data (highly variable)

Start here!

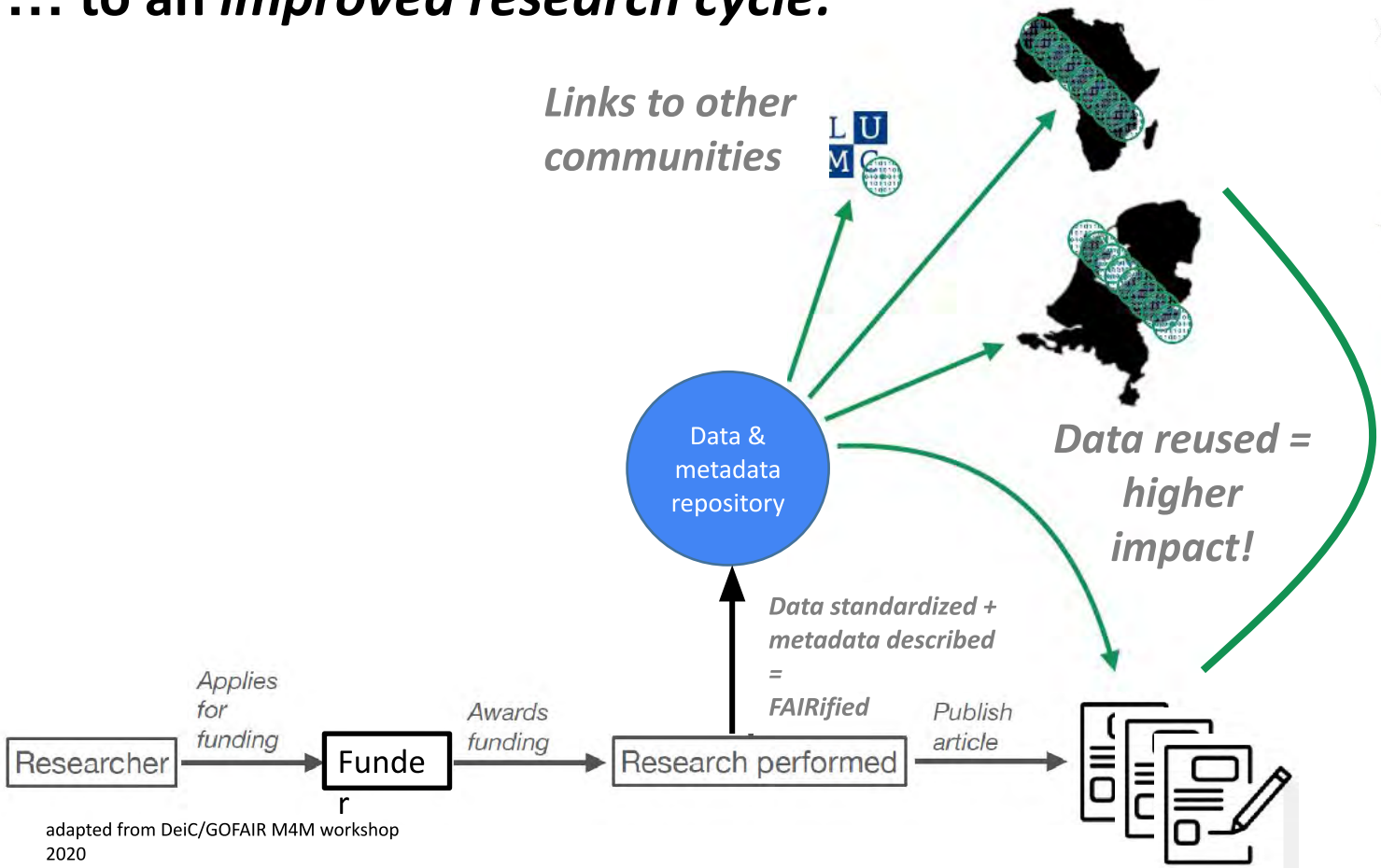
Better and more efficient science – moving from a *typical research cycle:*

This is where AnaEE DK started in 2019!



adapted from DeiC/GOFAIR M4M workshop 2020

… to an *improved research cycle:*

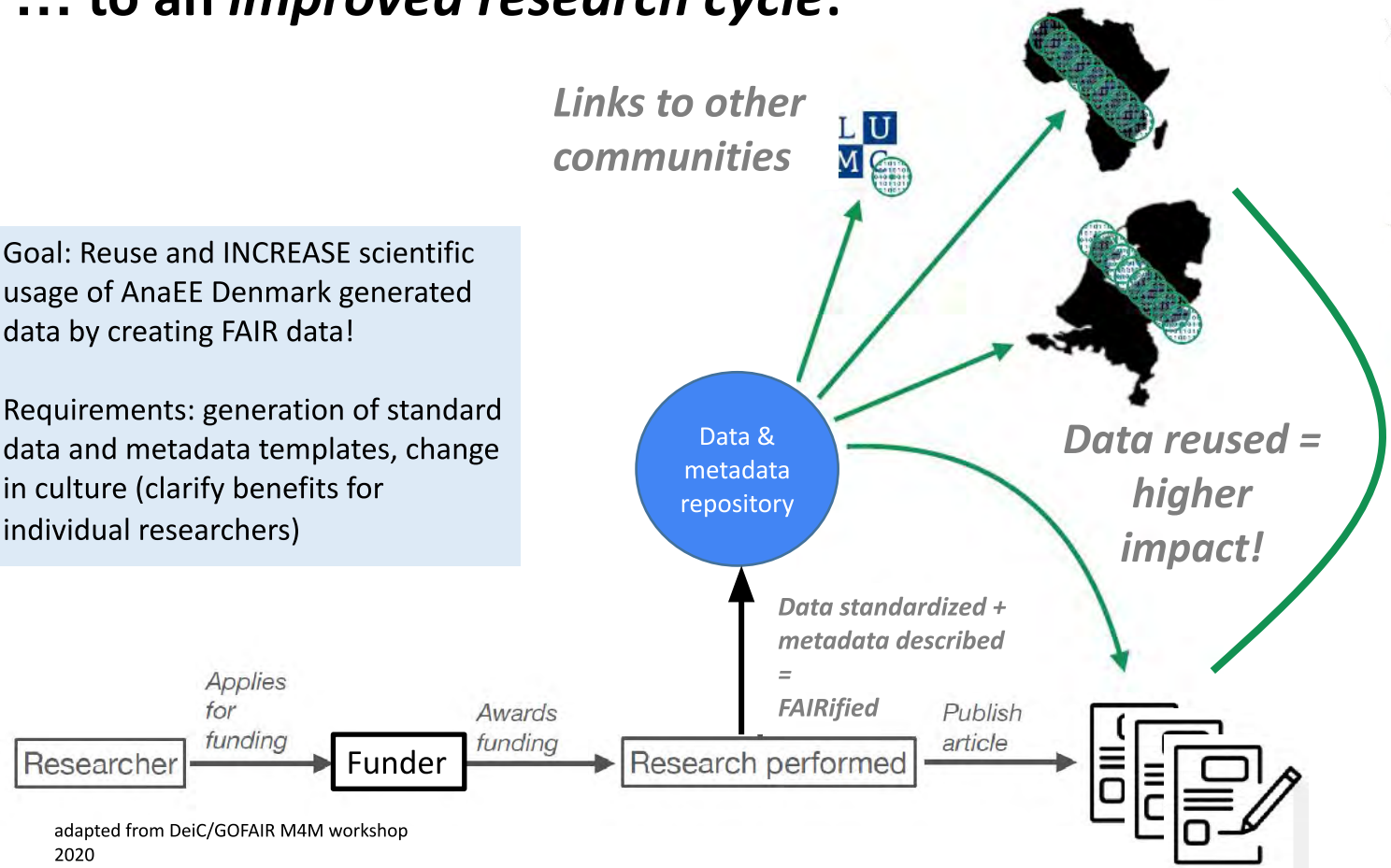


adapted from DeiC/GOFAIR M4M workshop 2020

… to an *improved research cycle*:

Goal: Reuse and INCREASE scientific usage of AnaEE Denmark generated data by creating FAIR data!

Requirements: generation of standard data and metadata templates, change in culture (clarify benefits for individual researchers)



adapted from DeiC/GOFAIR M4M workshop 2020

AnaEE Denmark FAIR data projects – co-funded by DeiC

- Project 1: 10 MM in total in 2019
- Project 2: 5 x 2-3 days workshops in June-October 2020 with DeiC + GOFAIR. No MM, but free workshops

Strategy: Lowest hanging fruits first:
Do what you can – skip what you can't

Main outcomes

- A lot of learning!
- 3 "FAIRified" meteorological datasets
- A FAIRification roadmap for AnaEE Denmark
- First machine-readable metadata (work in progress)
- FAIRified greenhouse gas exchange data (work in progress)

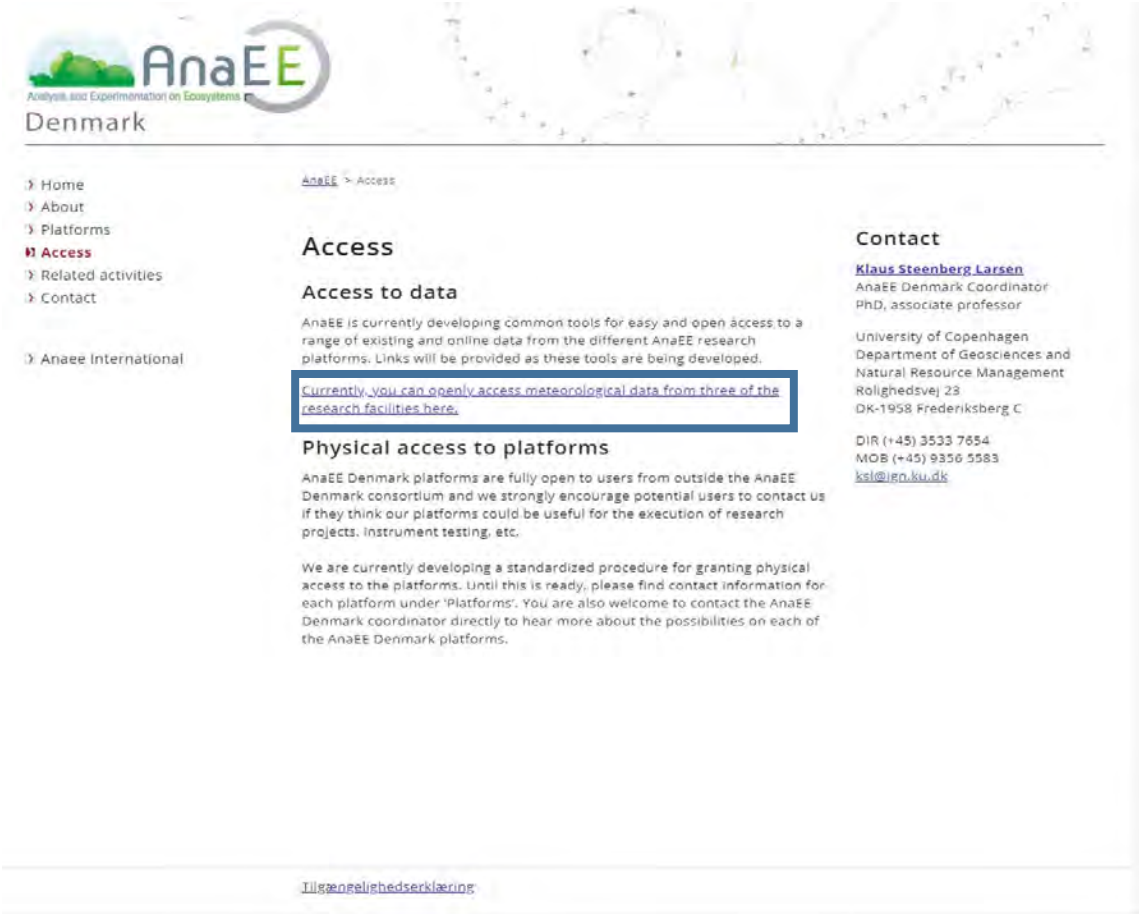
The steps we went through (in short):

- ☑ Find a persistent data repository, i.e. ERDA (UCPH) and LOAR (Royal Library for AU)
- ☑ Create templates for data and meta data
- ☑ Get a PID (in our case DOI – ERDA and Royal Library can provide)
- ☐ Create machine-readable metadata (M4M)

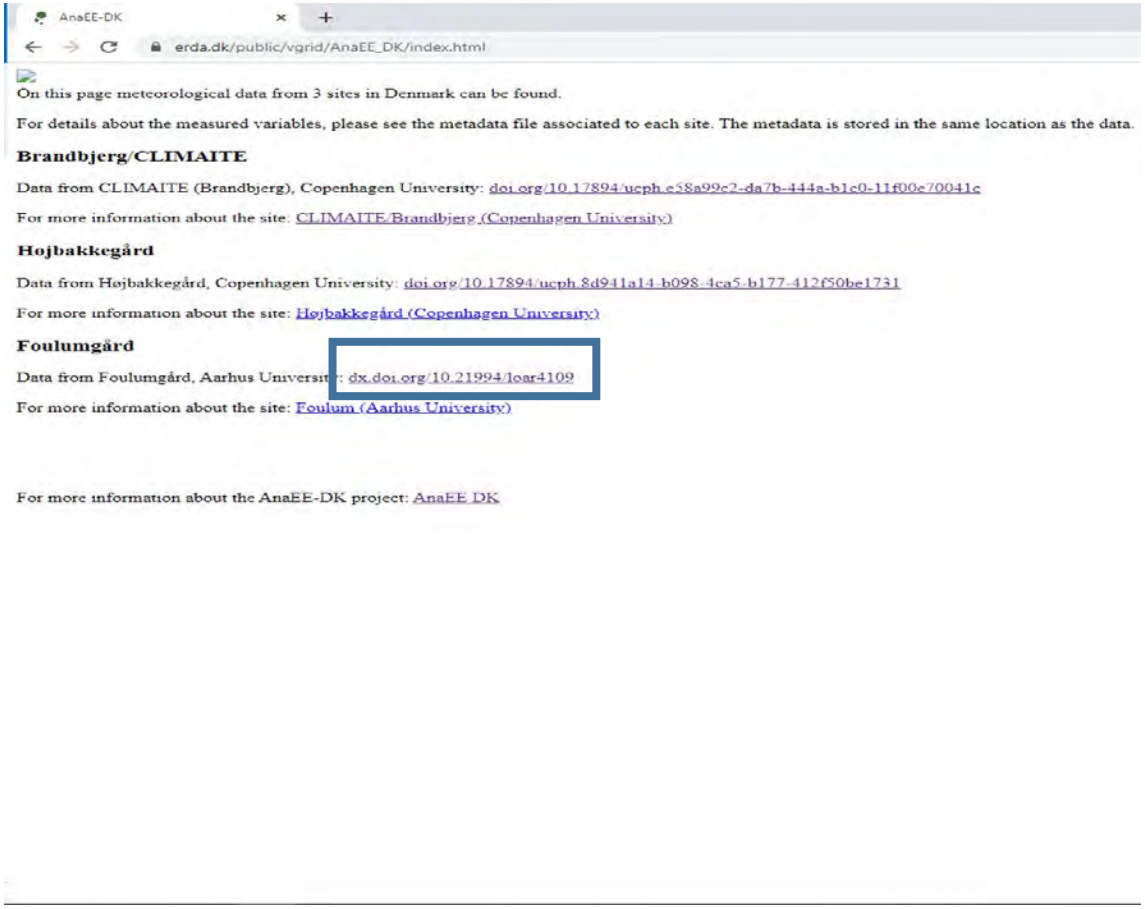
Link to data: anaee.dk/access/

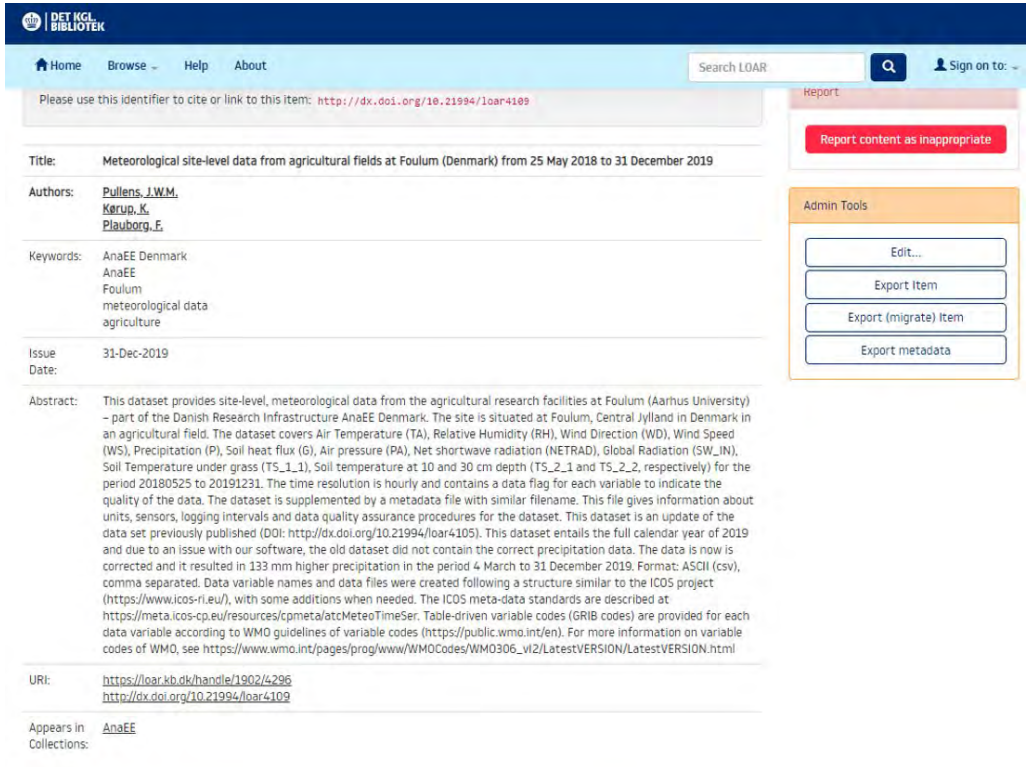
Data repositories and PIDs (access at anaee.dk):

Data repositories and PIDs (access at anaee.dk):



Data repositories and PIDs (access at anaee.dk):





metadata

2 files for download

DET KGL. BIBLIOTEK

Home Browse Help About Search LOAR

period 20180525 to 20191231. The time resolution is hourly and contains a data flag for each variable to indicate the quality of the data. The dataset is supplemented by a metadata file with similar filename. This file gives information about units, sensors, logging intervals and data quality assurance procedures for the dataset. This dataset is an update of the data set previously published (DOI: http://dx.doi.org/10.21994/loar4105). This dataset entails the full calendar year of 2019 and due to an issue with our software, the old dataset did not contain the correct precipitation data. The data is now is corrected and it resulted in 133 mm higher precipitation in the period 4 March to 31 December 2019. Format: ASCII (csv), comma separated. Data variable names and data files were created following a structure similar to the ICOS project (https://www.icos-ri.eu/), with some additions when needed. The ICOS meta-data standards are described at https://meta.icos-cp.eu/resources/cpmeta/atcMeteoTimeSer. Table-driven variable codes (GRIB codes) are provided for each data variable according to WMO guidelines of variable codes (https://public.wmo.int/en). For more information on variable codes of WMO, see https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/LatestVERSION/LatestVERSION.html

URI: https://loar.kb.dk/handle/1902/4296
http://dx.doi.org/10.21994/loar4109

Appears in Collections: AnaEE

Files in This Item:

File	Description	Size	Format	
FOULUM_20180525_20191231_HOURLY.csv	Meteorological site-level data from agricultural fields at Foulum (Denmark) from 25 May 2018 to 31 December 2019	1.42 MB	CSV	View/Open
FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv	Metadata for the meteorological site-level data from agricultural fields at Foulum (Denmark) from 25 May 2018 to 31 December 2019	3.69 kB	CSV	View/Open

Show full item record

DET KGL. BIBLIOTEK

Home Browse Help About Search LOAR

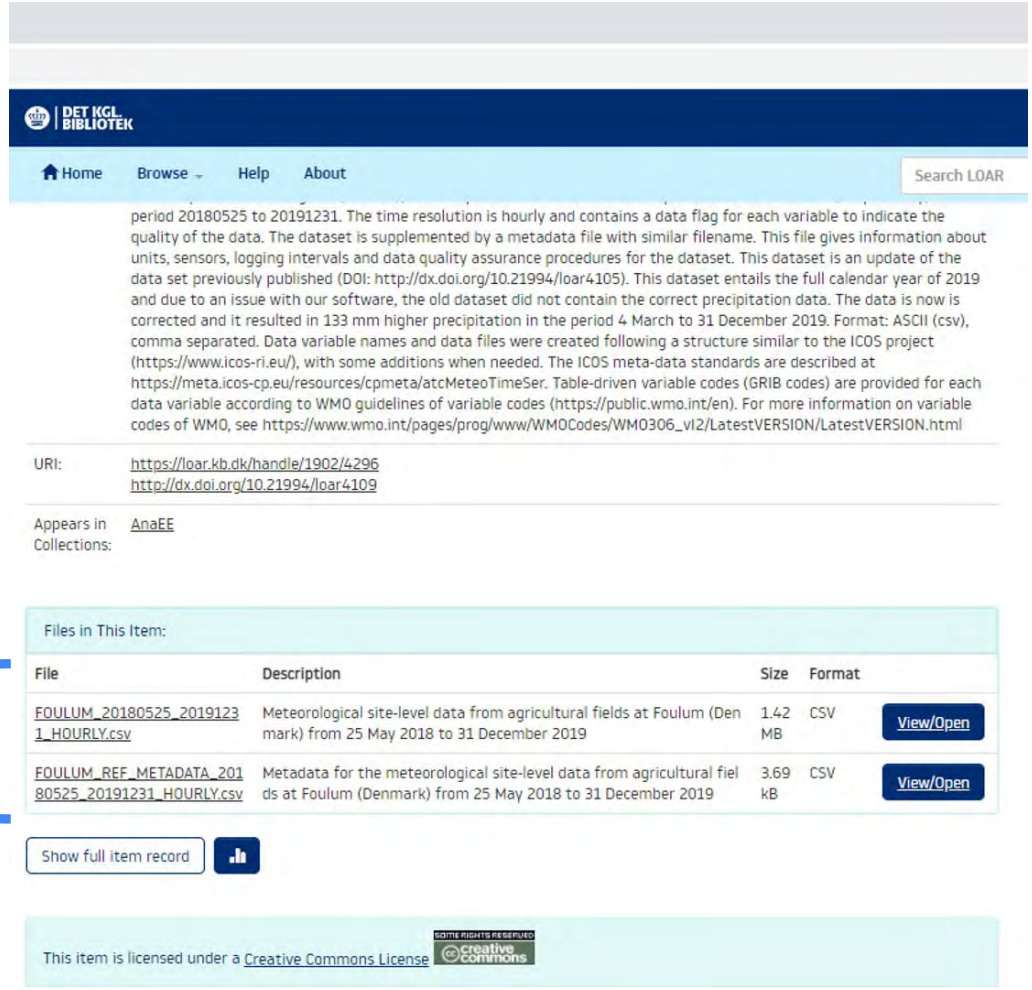
situated at Foulum, Central Jylland in Denmark in an agricultural field. The dataset covers Air Temperature (TA), Relative Humidity (RH), Wind Direction (WD), Wind Speed (WS), Precipitation (P), Soil heat flux (G), Air pressure (PA), Net shortwave radiation (NETRAD), Global Radiation (SW_IN), Soil Temperature under grass (TS_1_1), Soil temperature at 10 and 30 cm depth (TS_2_1 and TS_2_2, respectively) for the period 20180525 to 20191231. The time resolution is hourly and contains a data flag for each variable to indicate the quality of the data. The dataset is supplemented by a metadata file with similar filename. This file gives information about units, sensors, logging intervals and data quality assurance procedures for the dataset. This dataset is an update of the data set previously published (DOI: http://dx.doi.org/10.21994/loar4105). This dataset entails the full calendar year of 2019 and due to an issue with our software, the old dataset did not contain the correct precipitation data. The data is now is corrected and it resulted in 133 mm higher precipitation in the period 4 March to 31 December 2019. Format: ASCII (csv), comma separated. Data variable names and data files were created following a structure similar to the ICOS project (https://www.icos-ri.eu/), with some additions when needed. The ICOS meta-data standards are described at https://meta.icos-cp.eu/resources/cpmeta/atcMeteoTimeSer. Table-driven variable codes (GRIB codes) are provided for each data variable according to WMO guidelines of variable codes (https://public.wmo.int/en). For more information on variable codes of WMO, see https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/LatestVERSION/LatestVERSION.html

dc.language.iso	en	en_US
dc.relation	Derived from https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/LatestVERSION/LatestVERSION.html	en_US
dc.relation	Derived from https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/LatestVERSION/LatestVERSION.html	en_US
dc.relation	Update of http://dx.doi.org/10.21994/loar4105	en_US
dc.rights	Attribution 2.5 Denmark	*
dc.rights.uri	http://creativecommons.org/licenses/by/2.5/dk/	*
dc.subject	AnaEE Denmark	en_US
dc.subject	AnaEE	en_US
dc.subject	Foulum	en_US
dc.subject	meteorological data	en_US
dc.subject	agriculture	en_US
dc.title	Meteorological site-level data from agricultural fields at Foulum (Denmark) from 25 May 2018 to 31 December 2019	en_US
dc.type	Dataset	en_US
Appears in Collections:	AnaEE	

Files in This Item:

File	Description	Size	Format

More metadata

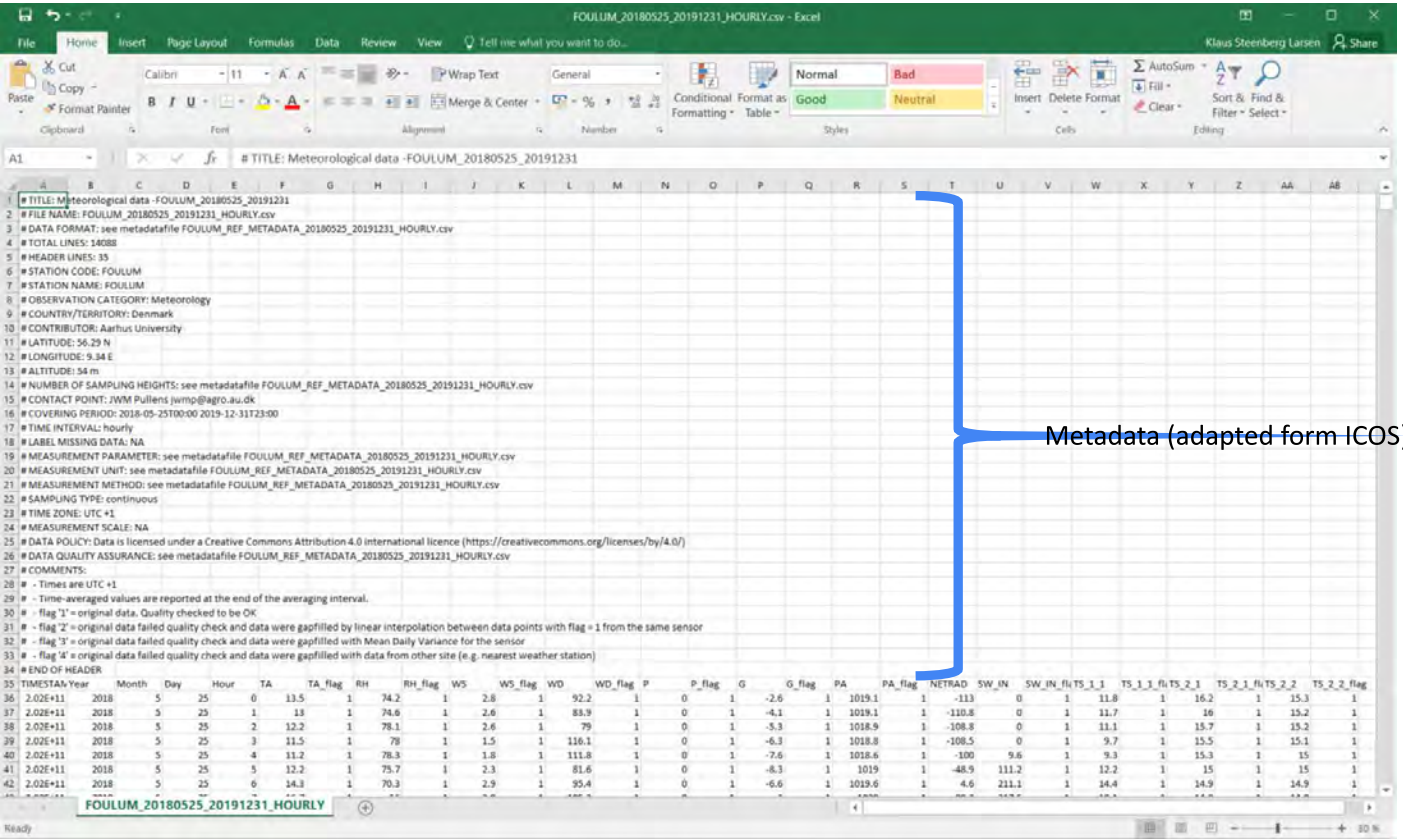


2 files for download

data file

metadata file

Data file


FOULUM_20180525_20191231_HOURLY.csv - Excel
TITLE: Meteorological data -FOULUM_20180525_20191231
FILE NAME: FOULUM_20180525_20191231_HOURLY.csv
DATA FORMAT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
TOTAL LINES: 14088
HEADER LINES: 35
STATION CODE: FOULUM
STATION NAME: FOULUM
OBSERVATION CATEGORY: Meteorology
COUNTRY/TERRITORY: Denmark
CONTRIBUTOR: Aarhus University
LATITUDE: 56.29 N
LONGITUDE: 9.34 E
ALTITUDE: 54 m
NUMBER OF SAMPLING HEIGHTS: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
CONTACT POINT: JWM Pullens jwmp@agro.au.dk
COVERING PERIOD: 2018-05-25T00:00 2019-12-31T23:00
TIME INTERVAL: hourly
LABEL MISSING DATA: NA
MEASUREMENT PARAMETER: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
MEASUREMENT UNIT: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
MEASUREMENT METHOD: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
SAMPLING TYPE: continuous
TIME ZONE: UTC +1
MEASUREMENT SCALE: NA
DATA POLICY: Data is licensed under a Creative Commons Attribution 4.0 international licence (https://creativecommons.org/licenses/by/4.0/)
DATA QUALITY ASSURANCE: see metadatafile FOULUM_REF_METADATA_20180525_20191231_HOURLY.csv
COMMENTS:
- Times are UTC +1
- Time-averaged values are reported at the end of the averaging interval.
- flag '1' = original data. Quality checked to be OK
- flag '2' = original data failed quality check and data were gapfilled by linear interpolation between data points with flag = 1 from the same sensor
- flag '3' = original data failed quality check and data were gapfilled with Mean Daily Variance for the sensor
- flag '4' = original data failed quality check and data were gapfilled with data from other site (e.g. nearest weather station)
END OF HEADER
TIMESTAMP Year Month Day Hour TA TA_flag RH RH_flag WS WS_flag WD WD_flag P P_flag G G_flag PA PA_flag NETRAD SW_IN SW_IN_flag TS_1_1 TS_1_1_flag TS_2_1 TS_2_1_flag TS_2_2 TS_2_2_flag
2.02E+11 2018 5 25 0 13.5 1 74.2 1 2.8 1 92.2 1 0 1 -2.6 1 1019.1 1 -113 0 1 11.8 1 16.2 1 15.3 1
2.02E+11 2018 5 25 1 13 1 74.6 1 2.6 1 83.9 1 0 1 -4.1 1 1019.1 1 -110.8 0 1 11.7 1 16 1 15.2 1
2.02E+11 2018 5 25 2 12.2 1 78.1 1 2.6 1 79 1 0 1 -5.3 1 1018.9 1 -108.8 0 1 11.1 1 15.7 1 15.2 1
2.02E+11 2018 5 25 3 11.5 1 78 1 1.5 1 116.1 1 0 1 -6.3 1 1018.8 1 -108.5 0 1 9.7 1 15.5 1 15.1 1
2.02E+11 2018 5 25 4 11.2 1 78.3 1 1.8 1 111.8 1 0 1 -7.6 1 1018.6 1 -100 9.6 1 9.3 1 15.3 1 15 1
2.02E+11 2018 5 25 5 12.2 1 75.7 1 2.3 1 81.6 1 0 1 -8.3 1 1019 1 -48.9 111.2 1 12.2 1 15 1 15 1
2.02E+11 2018 5 25 6 14.3 1 70.3 1 2.9 1 95.4 1 0 1 -6.6 1 1019.6 1 4.6 211.1 1 14.4 1 14.9 1 14.9 1
FOULUM_20180525_20191231_HOURLY
Metadata (adapted form ICOS)

Metadata file (metadata on variables – adapted from ICOS)

FOULUM_REF_METADATA_20180525_20191231_HOURLY (3).csv - Excel

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do... Klaus Steenberg Larsen Share

Paste Cut Copy Format Painter | Calibri 11 | Wrap Text Merge & Center | General | Conditional Formatting Format as Table | Normal Bad Good Neutral Calculation Check Cell Explanatory T... Input | Insert Delete Format | AutoSum Fill Clear | Sort & Filter Find & Select

Clipboard Font Alignment Number Styles Cells Editing

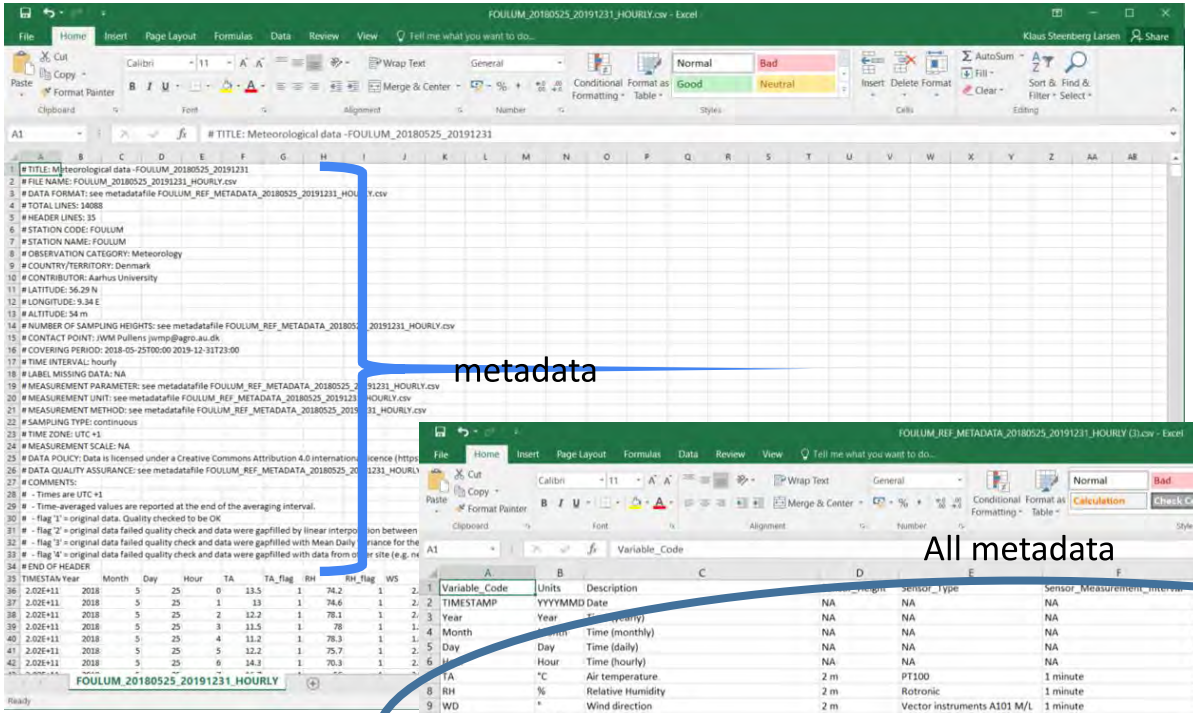
A1 Variable_Code

	A	B	C	D	E	F	G	H	I
1	Variable_Code	Units	Description	Sensor_Height	Sensor_Type	Sensor_Measurement_Interval	Data_Analysis	Data_Quality_Assurance	Grib_code
2	TIMESTAMP	YYYYMMD	Date	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	NA
3	Year	Year	Time (yearly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2004
4	Month	Month	Time (monthly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2003
5	Day	Day	Time (daily)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2002
6	Hour	Hour	Time (hourly)	NA	NA	NA	NA	All times are UTC +1 (always wintertime only)	04.04.2001
7	TA	°C	Air temperature	2 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.0.0
8	RH	%	Relative Humidity	2 m	Rotronic	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.1.1
9	WD	°	Wind direction	2 m	Vector instruments A101 M/L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.2.0
10	WS	m s-1	Wind speed	2 m	Vector Insturments W200P	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.2.1
11	P	mm	Precipitation	1.5 m	OTT Pluvio2	0.1 mm	1 minute values, 10 minute sum	Data were flagged for gapfilling by applying vith vis	4.2.0.1.8
12	G	W m-2	Soil heat flux	5 cm depth	Campbell scientific: HFT3-L	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.2.3.26
13	PA	hPa	Air pressure	2 m	Vaisala PTB101B	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.3.0
14	NETRAD	W m-2	Net shortwave radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were calculated and not flagged	4.2.0.4.0
15	SW_IN	W m-2	Global radiation	1.5 m	Kipp & Zonen CNR1	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.0.4.7
16	TS_1_1	°C	Soil temperature under grass	- 0.05 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.2.0.2
17	TS_2_1	°C	Soil temperature at 10 cm depth below bare soil	- 0.10 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.2.0.2
18	TS_2_2	°C	Soil temperature at 30 cm depth below bare soil	- 0.30 m	PT100	1 minute	1 minute values, 10 minute mean	Data were flagged for gapfilling by applying vith vis	4.2.2.0.2
19									
20	For WMO code tables. Grib code table FM_92 (Ver.23.0.0) see https://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/LatestVERSION/LatestVERSION.html								
21									

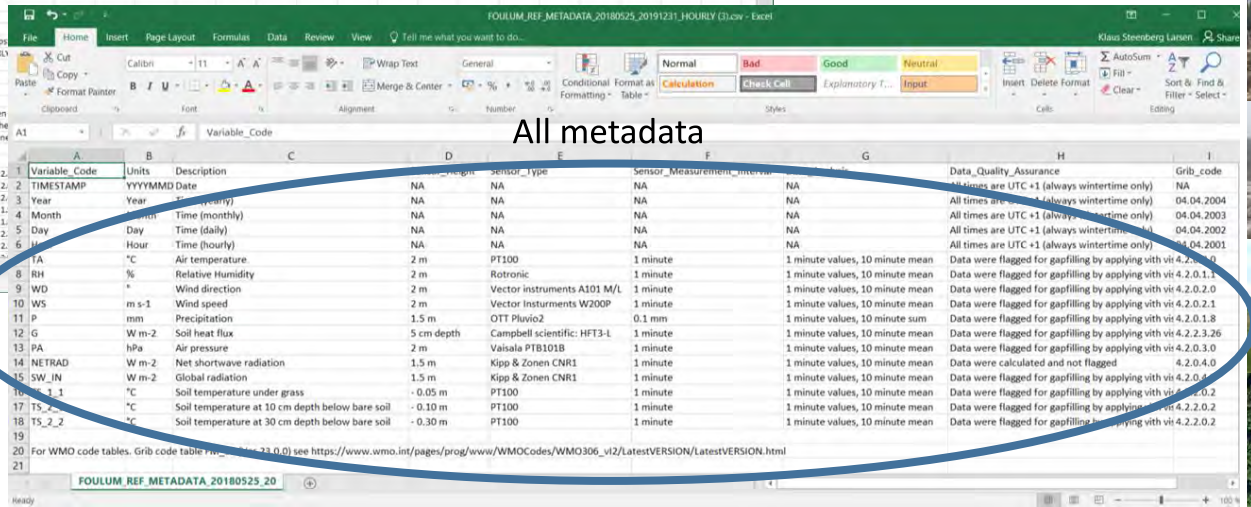
FOULUM_REF_METADATA_20180525_20

Ready

Metadata in data file and metadata file are not M4M!



metadata



Machine-readable metadata: Openly available tools:

BioPortal – a catalogue of ontologies from 1000 different communities

https://bioportal.bioontology.org/

Training during 2020 DeiC/GOFAIR M4M workshops

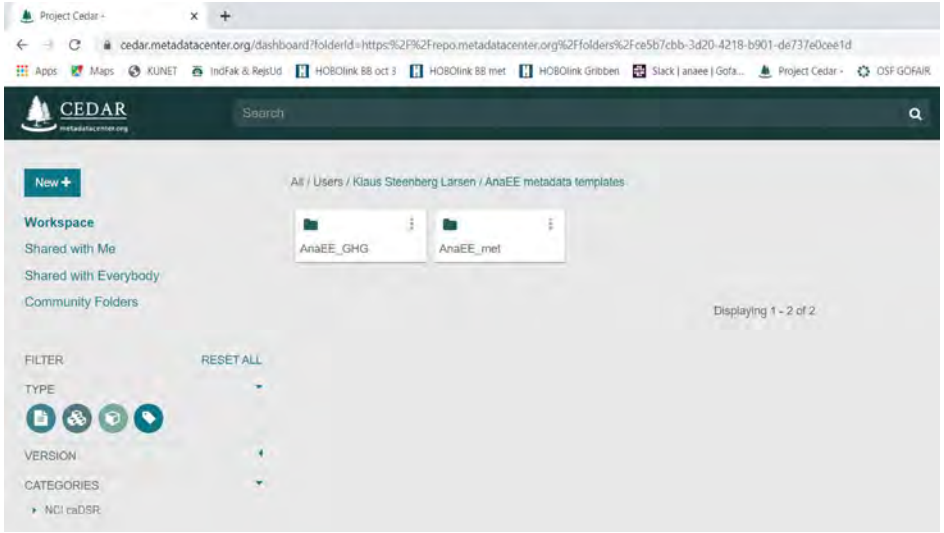
CEDAR:

cedar.metadatacenter.org

CEDAR can help you

map your data to ontologies

= create machine-readable metadata code – without doing code!

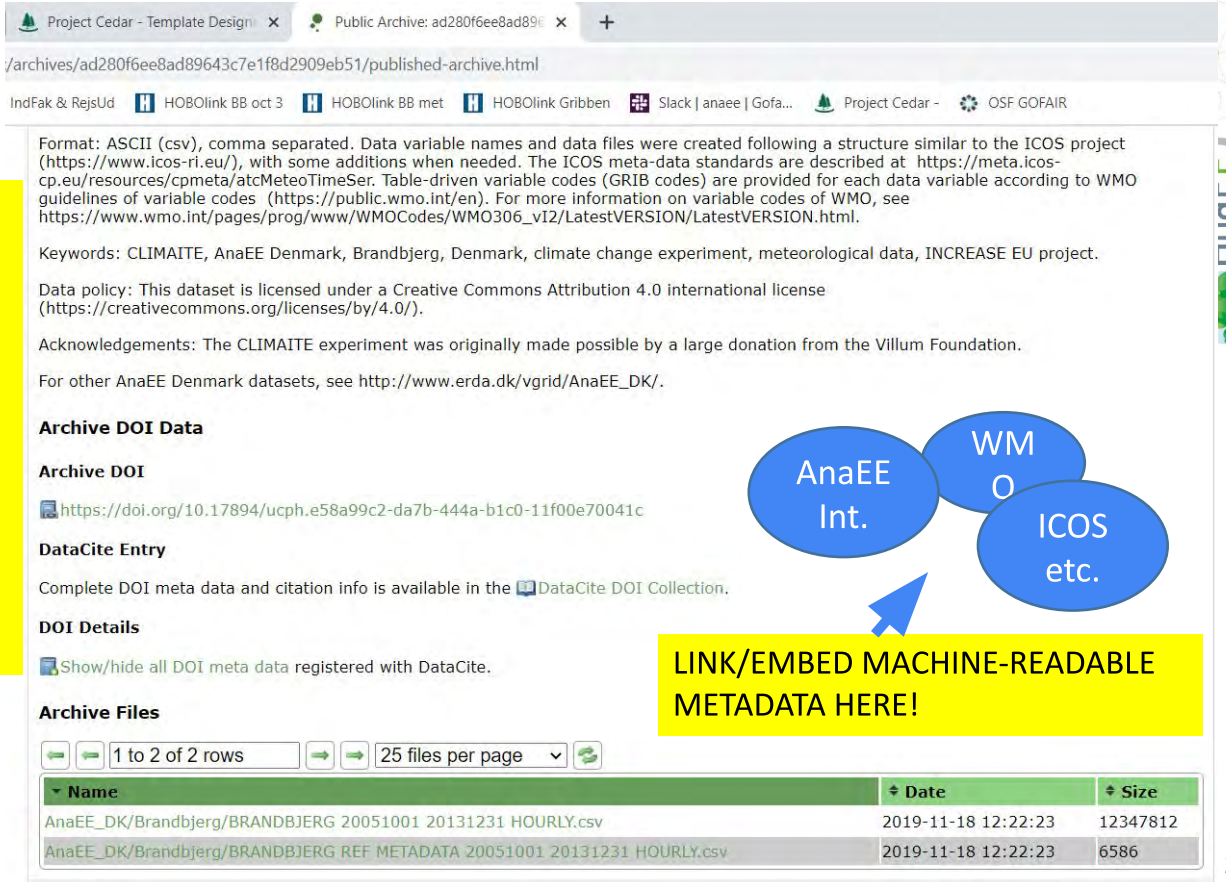


FAIR data with machine-readable metadata

Major challenge because ERDA not optimized for exposing /sharing metadata

UCPH now has a pilot to implement Dataverse

Time will show if we can use the CEDAR templates in Dataverse…



Final product from DeiC/GOFAIR M4M workshops 2020: A *FAIRification* roadmap for AnaEE Denmark

Content:

AnaEE
Analysis and Experimentation on Ecosystems
Denmark

Thank you!

UNIVERSITY OF COPENHAGEN
AARHUS UNIVERSITY
RUC
DTU

Three research case studies – DTU Wind Energy

- Nikola Vasiljevic, Special Consultant for Digitalization
- Department of Wind Energy
- Technical University of Denmark

Technical University of Denmark

Disclaimer

This presentation contains opinions which not necessarily reflect the DTU Wind Energy

DTU Wind Energy – who are we

- DTU Wind Energy is **one of the largest and most well-known** university department for wind energy in the world with **250 employees**.

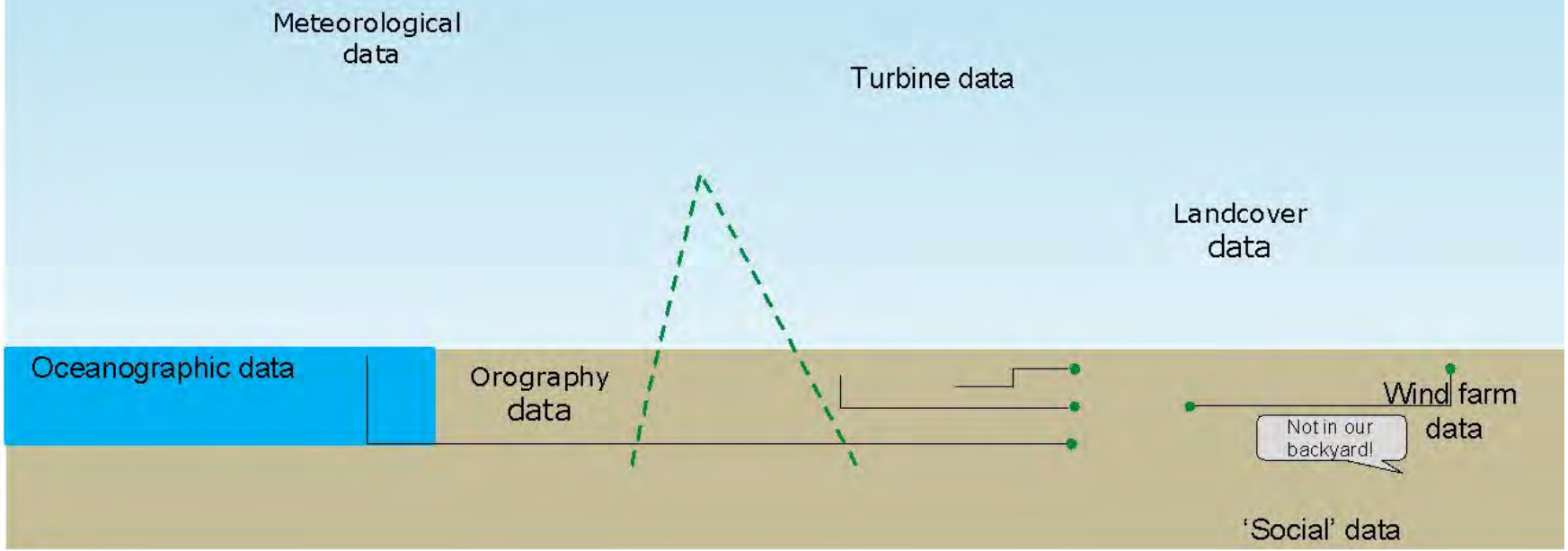
- DTU Wind Energy consists of three divisions (and many sections):
  - Wind Energy Systems
  - Wind Energy Materials and Components
  - Wind Turbine Design

- **We deal with:** Aerodynamic Design • Composite Materials • Composite Mechanics and Structures • Fluid Mechanics • Wind Turbine Structures • Component Design • Wind Turbine Loads and Control • Meteorology • Remote Sensing • Resource Assessment Modelling • Test and Measurements • Integration and Planning • Social impact • …

- We cover entire lifecycle of wind power plant

Wind Energy domain

- Large majority of our data has a spatio-temporal structure
- All sort of spatio-temporal scales
- Open, confidential, personal data

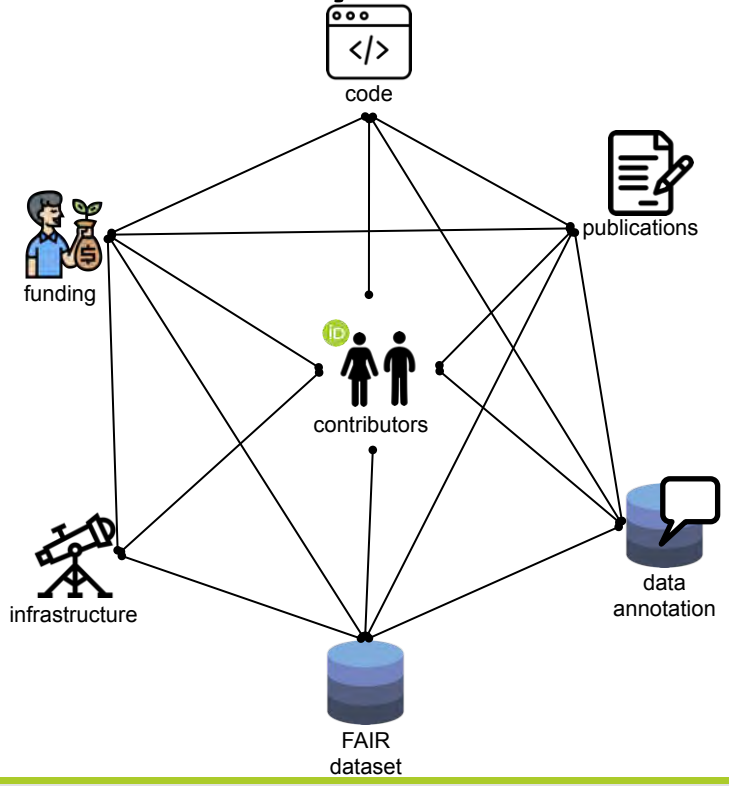


Physical infrastructures we operate

- Test Centre for large wind turbines Høvsøre
- Test Centre for large wind turbines Østerild
- Research Facilities:
  - Composite Laboratories
  - Drivetrain
  - Large Scale Facility
  - Poul la Cour Wind Tunnel
  - Research Wind Turbine V52 (Risø test station)
  - Windscanner

…
(+ a number of virtual infrastructures)

Interconnected and richly described resources



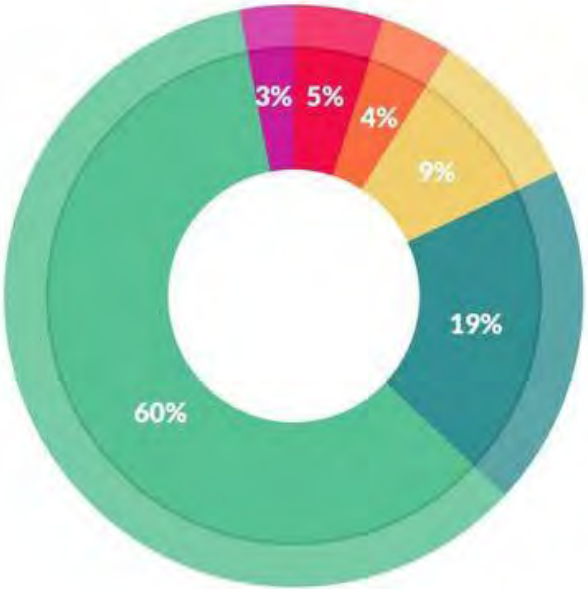
Full chain of custody

Researcher at center

clipart from www.flaticon.com

>Why?

How much time we spend analysing data?

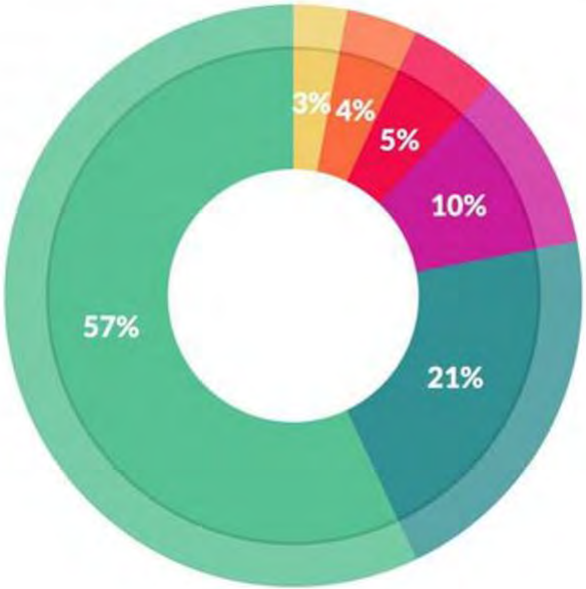


What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Source: Forbes

What we don't like to do?



What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

Source: Forbes

- If we don’t spend time handling data at the moment of their creation, we will ‘waste’ 80% of resources anytime we or anyone else need to use them (again).

Data engineering/stewardship is not perceived as ‘cool’ activity compering to data analytics, however it has much more lasting impact then trendy data analytics methods.

Our interaction with the FAIR principles

- 2017 EERA JP WIND IRPWind Open Data Initiative (European level)
- 2018 Internal project and FAIR ambassadors (University/Deprtment level)
- 2019 RDA Ambassadorship (‘Individual’ level)
- 2020 Running/participating in M4M workshop, interaction with the RDA (National/International level)

Our interaction with the FAIR principles

- 2017 EERA JP WIND IRPWind Open Data Initiative (European level)
- 2018 Internal project and FAIR ambassadors (University/Deprtment level)
- 2019 RDA Ambassadorship (‘Individual’ level)
  - Participation in FAIR Data Stewardship training organized by NeiC
- 2020 Running/participating in M4M workshops, interaction with the RDA (National/International level)
  - Developing and demonstrating **sheet2rdf** and **OntoStack**
  - Generated controlled vocabulary of 100+ wind energy parameters
  - Training Dutch Covid Program data stewards

2017 – dealing with FA of FAIR

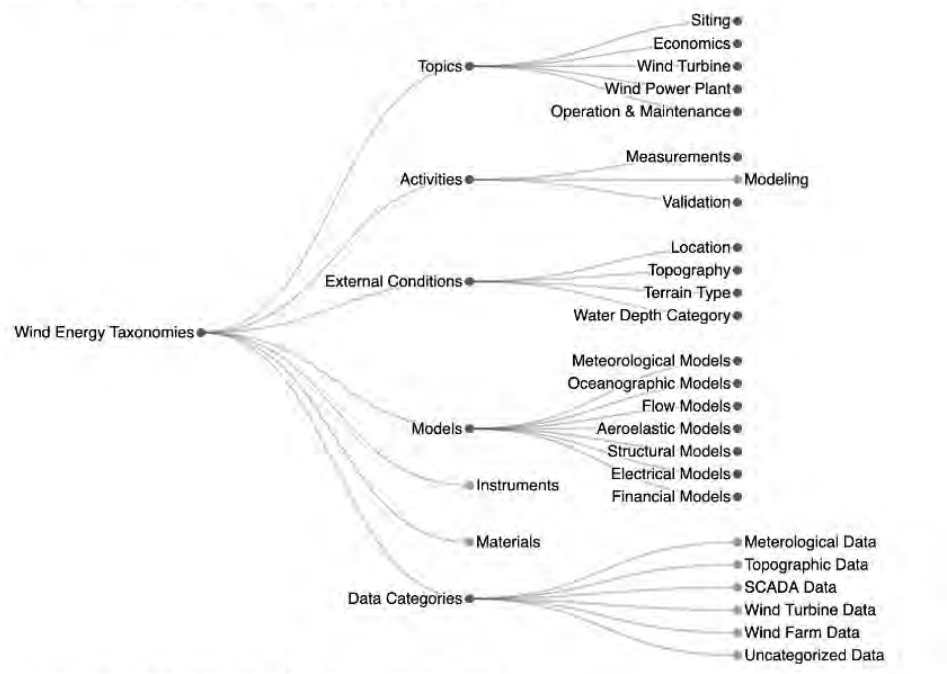
- As part of the IRPWind Open Data initiative Drafted Wind Energy Taxonomies of:
  - Topics
  - Activities
  - External Conditions
  - Models

- The taxonomies were drafted by employing the expert elicitation (16 international domain experts)

- The taxonomies were drafted with the ambition to use them as controlled terminologies to ‘tag’ data enabling search by means of controlled terms

- For the purpose of ‘tagging’ data, thus using controlled terminologies we have drafted Dublin Core Wind Energy Application Profile, in other words a **metadata template for datasets**

Taxonomies



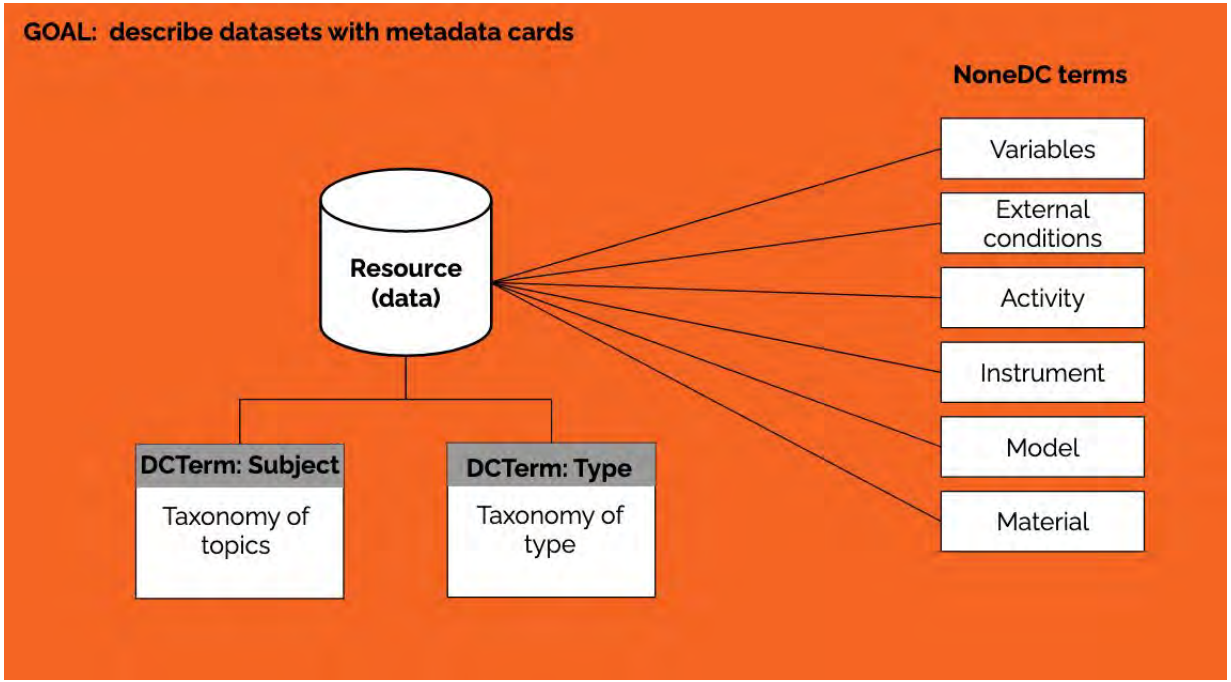
Source: http://data.windenergy.dtu.dk/taxonomy/

Dublin core profile

		Mandatory	Highly Likely	Maybe	Highly Unlikely	Impossible
1	Title					
2	Creator			ORCID?		
3	Subject	WE Taxonomy				
4	Description					
5	Publisher			Custom?		
6	Contributor			ORCID?		
7	Date		W3C DTF			
8	Type	WE Taxonomy				
9	Format	MIME/IMT				
10	Identifier			URI/DOI?		
11	Source					
12	Language	RFC 3066 (?)				
13	Relation					
14	Coverage		ISO 3166?			
15	Rights					

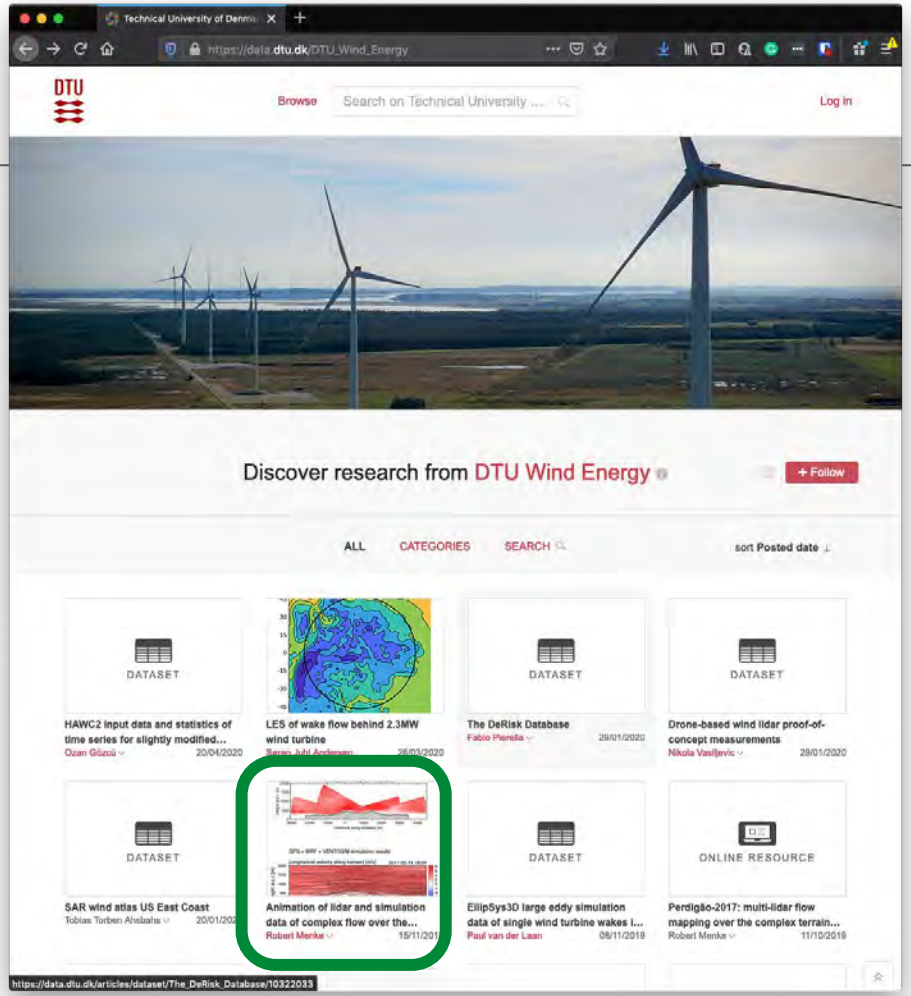
Source: https://zenodo.org/record/4013191

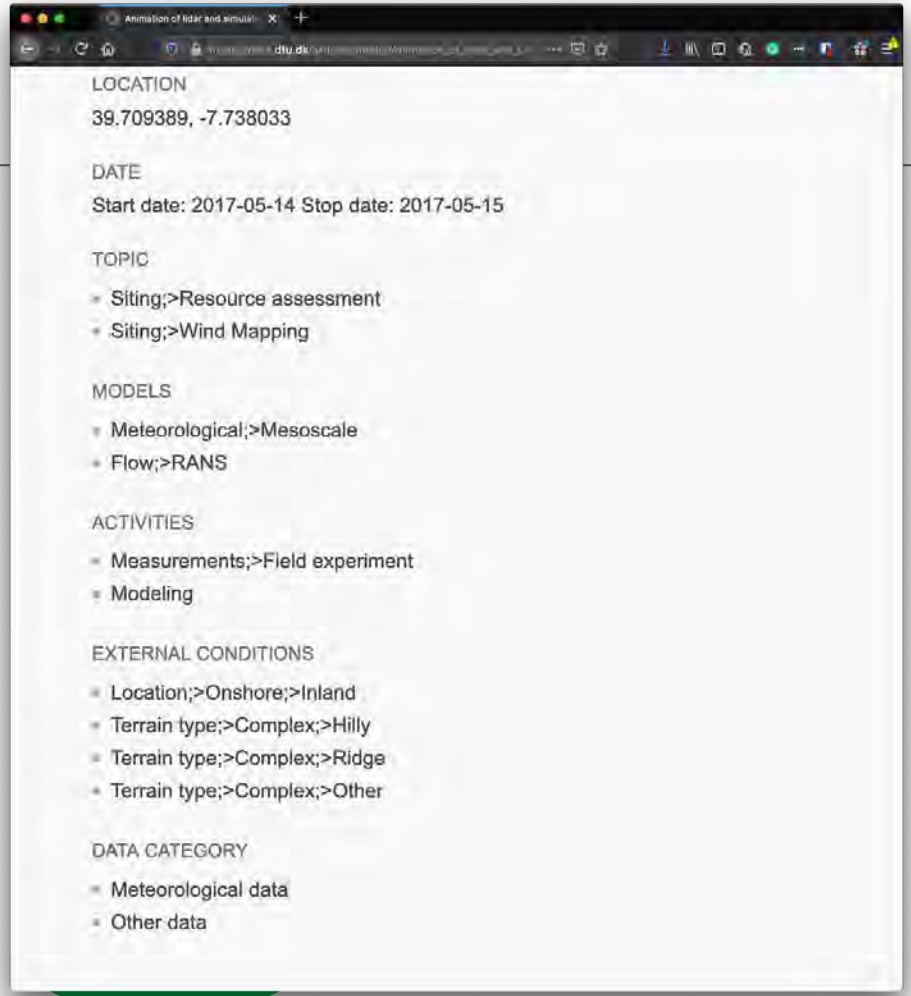
Dublin core profile



2018 – implementing FA of FAIR

- Updated taxonomies with a pull of depertamental FAIR ambassadors
- A part of DTU Library pilot project for implementation of DTU-data , university data publishing platform
- DTU-data is an instance of figshare
- Extended FigShare metadata template to take in account our taxonomies


Technical University of Denma
https://data.dtu.dk/DTU_Wind_Energy
DTU
Browse
Search on Technical University ...
Log in
Discover research from DTU Wind Energy
+ Follow
ALL
CATEGORIES
SEARCH
sort Posted date
DATASET
HAWC2 input data and statistics of time series for slightly modified...
Ozan Gözcü
20/04/2020
LES of wake flow behind 2.3MW wind turbine
The DeRisk Database
Fabio Pierella
29/01/2020
DATASET
Drone-based wind lidar proof-of-concept measurements
Nikola Vasiljevic
29/01/2020
DATASET
SAR wind atlas US East Coast
Tobias Torben Ahsbahs
Animation of lidar and simulation data of complex flow over the...
Robert Menke
15/11/201
DATASET
EllipSys3D large eddy simulation data of single wind turbine wakes i...
Paul van der Laan
08/11/2019
ONLINE RESOURCE
Perdigão-2017: multi-lidar flow mapping over the complex terrain...
Robert Menke
11/10/2019
https://data.dtu.dk/articles/dataset/The_DeRisk_Database/10322033


Animation of lidar and simulati
dtu.dk
LOCATION
39.709389, -7.738033
DATE
Start date: 2017-05-14 Stop date: 2017-05-15
TOPIC
Siting;>Resource assessment
Siting;>Wind Mapping
MODELS
Meteorological;>Mesoscale
Flow;>RANS
ACTIVITIES
Measurements;>Field experiment
Modeling
EXTERNAL CONDITIONS
Location;>Onshore;>Inland
Terrain type;>Complex;>Hilly
Terrain type;>Complex;>Ridge
Terrain type;>Complex;>Other
DATA CATEGORY
Meteorological data
Other data

> 2020 – tackling machinic I of FAIR / controlled terminologies

- Switched from JSON to RDF

- Configured and deployed OntoStack to:
  - Build and maintain controlled terminologies:
    GitLab CI/CD or Github actions executing sheet2rdf
    sheet2rdf is also used in Dutch Covid Program

  - Serve terminologies to humans and machines:
    Edge routing with traefik in front of Jena Fuseki and *SKOSMOS*

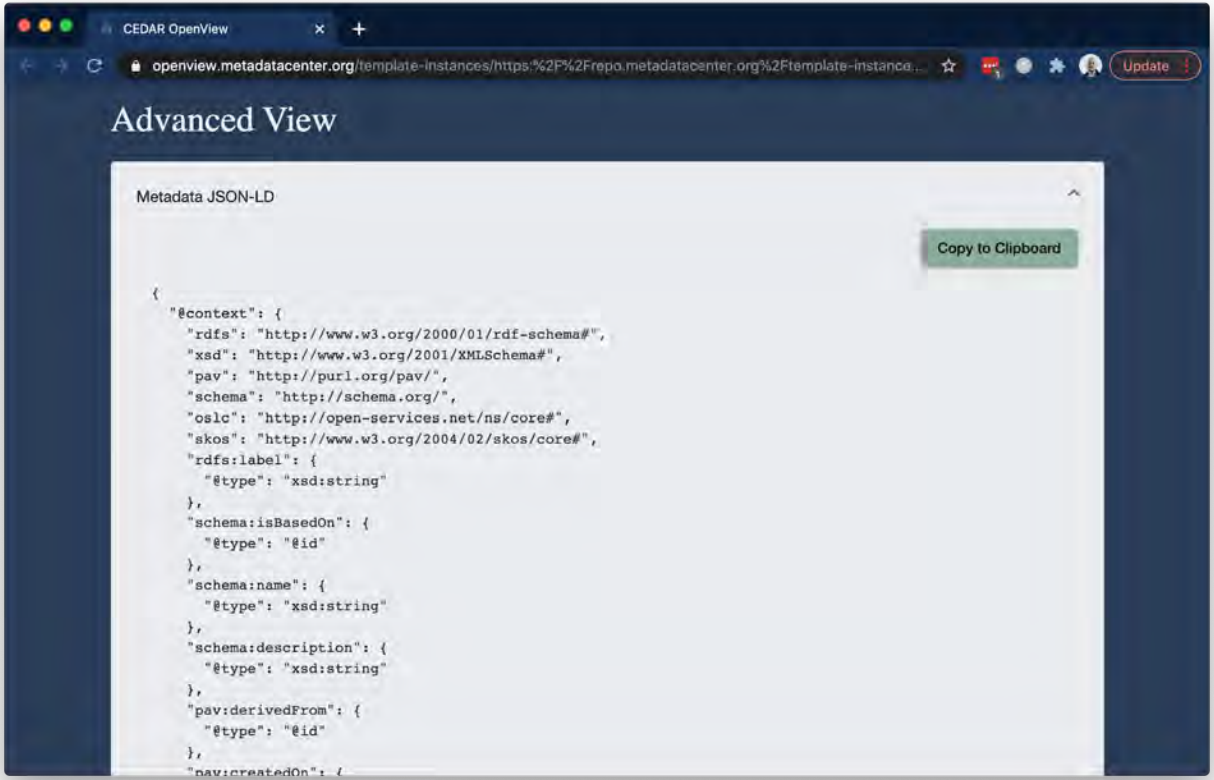
- OntoStack runs on:
  - DTU Web Server: http://data.windenergy.dtu.dk/ontologies/
    Servers several wind energy related ontologies

  - DeiC VM: http://ontology.deic.dk/
    To start it will be used for Danish M4Ms

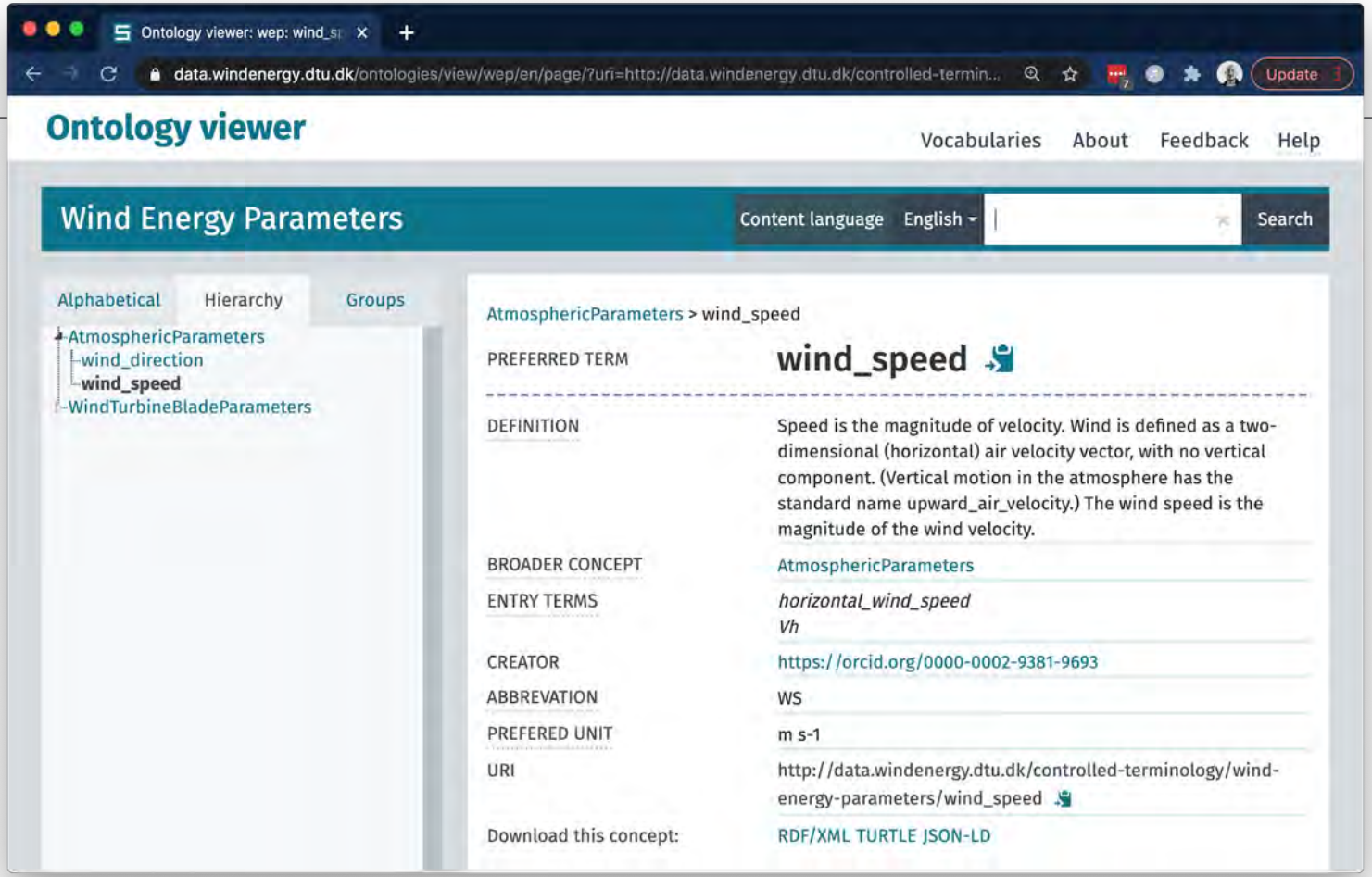
> 2020 – tackling machinic I of FAIR / metadata templates

- Selected NetCDF data format for sharing / publishing data
- Extending our Dublin Core Application Profie making it machine actionable
- Using DataCite metadata schema (4.3) as a base template for dataset metadata
- Extend template with missing fields that will provide information about:
  - (geo)spatio-temporal structure of data
  - data quality
  - what (and how) produced data
- 90 fields belonging to 20 groups of which, 23 mandatory (5 manual entries), 24 recommended (4 manual entries) and 43 optional (15 manual entries)
- Created **Generic Dataset Metadata Template** accessible in CEDAR Workbench

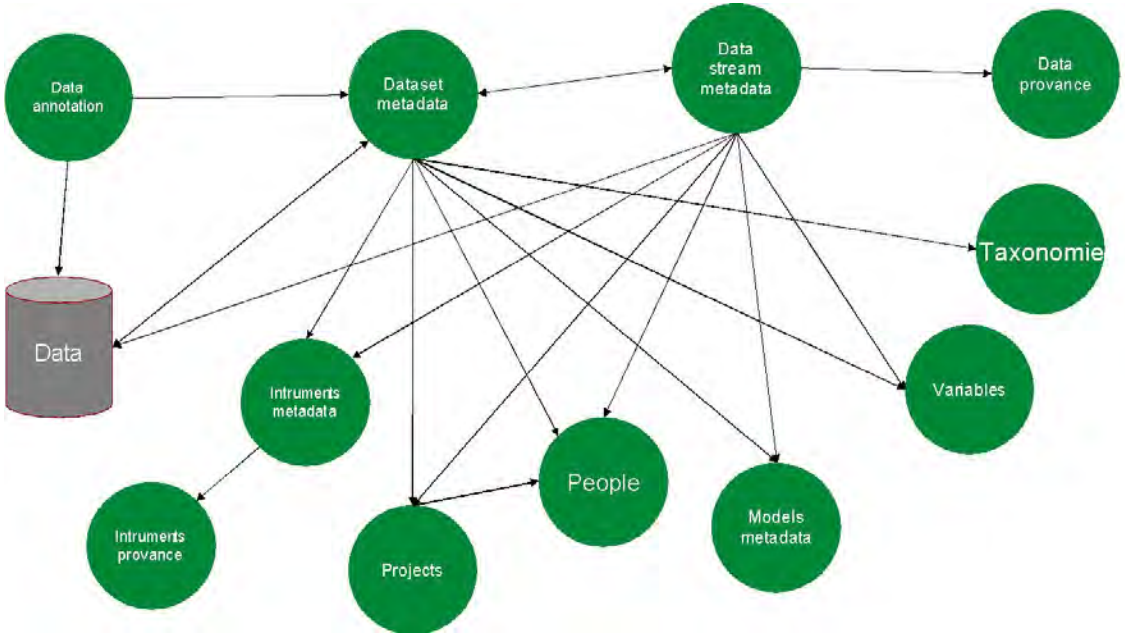
>Machinic metadata



```yaml
wind_speed:
  database_column: ['theta']
  dimensions: ['time','y','z']
  type: 'float64'
  attributes:
    standard_name: 'wind_speed'
    concept_id: 'http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/wind_speed'
    units: 'm.s-1'
```


Ontology viewer: wep: wind_sp
data.windenergy.dtu.dk/ontologies/view/wep/en/page/?uri=http://data.windenergy.dtu.dk/controlled-termin...
Update
Ontology viewer
Vocabularies
About
Feedback
Help
Wind Energy Parameters
Content language
English
Search
Alphabetical
Hierarchy
Groups
AtmosphericParameters
wind_direction
wind_speed
WindTurbineBladeParameters
AtmosphericParameters > wind_speed
PREFERRED TERM
wind_speed
DEFINITION
Speed is the magnitude of velocity. Wind is defined as a two-dimensional (horizontal) air velocity vector, with no vertical component. (Vertical motion in the atmosphere has the standard name upward_air_velocity.) The wind speed is the magnitude of the wind velocity.
BROADER CONCEPT
AtmosphericParameters
ENTRY TERMS
horizontal_wind_speed
Vh
CREATOR
https://orcid.org/0000-0002-9381-9693
ABBREVATION
WS
PREFERED UNIT
m s-1
URI
http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/wind_speed
Download this concept:
RDF/XML TURTLE JSON-LD

Where are we going with all of this?



Arrows are semi-random, did not want to overthink.

Summary

- Everything that has been done so far it has been done with:
  - ‘bottom-up’ **guerilla** approach (there was no ’top-down’ initiative)
  - largely by 1-2 persons
  - mainly done in ‘leisure time’ (no firm funding)

- Solutions which started in wind energy are now used in other domains
  *…and they were also produced by **re-using** someone else work*

- So there is abs no point in “silosing” and building ‘your own’ solutions since ‘your domain’ so special
  *…grab solutions from front-runners, **adapt**!, that’s what I did*

- If you want to be ‘FAIR’ you can:
  - Start reusing, adapting, building and testing solutions
  - Don’t count on (massive) funding to do this work (expect to work for ‘free’)
  - Don’t expect that your first solution is going to be perfect

Three research case studies – Danish BioImaging

- Clara Prats, Associate Professor

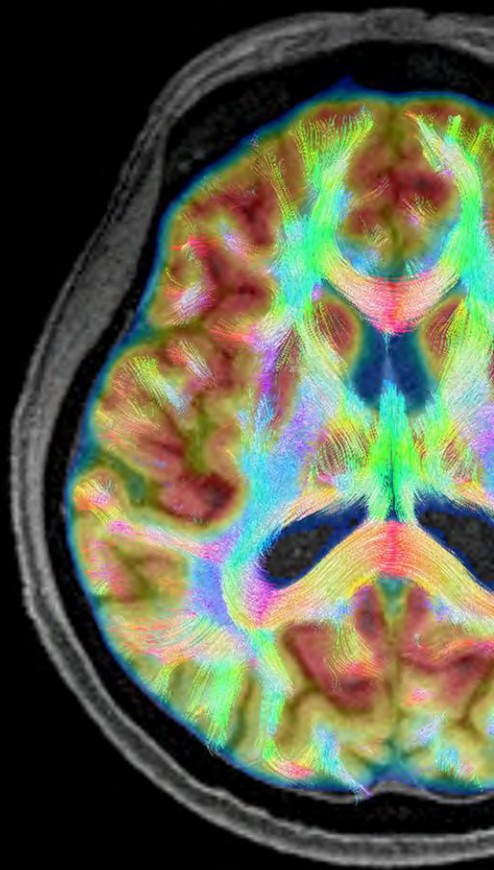
- Core Facility for Integrated Microscopy
- Faculty of Health and Medical Sciences
- University of Copenhagen

UNIVERSITY OF COPENHAGEN

BioImaging Research in a FAIR Data perspective

Clara Prats, Associate Professor
Core Facility for Integrated Microscopy
Faculty of Health and Medical Sciences
University of Copenhagen

www.danishbioimaging.dk
www.eurobioimaging.eu



Danish BioImaging

EuroBioImaging

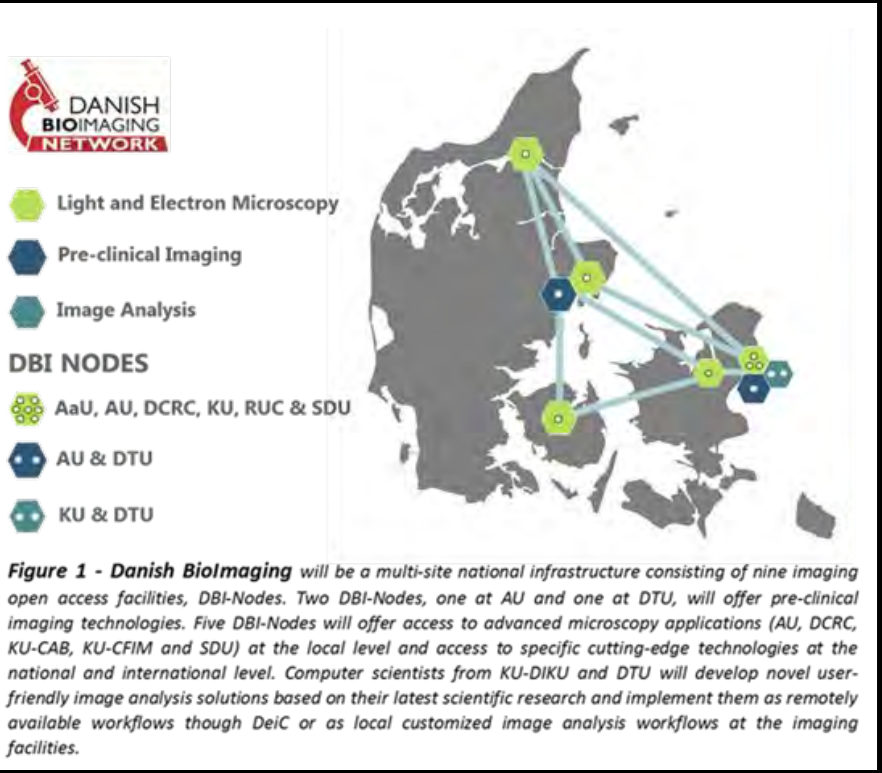


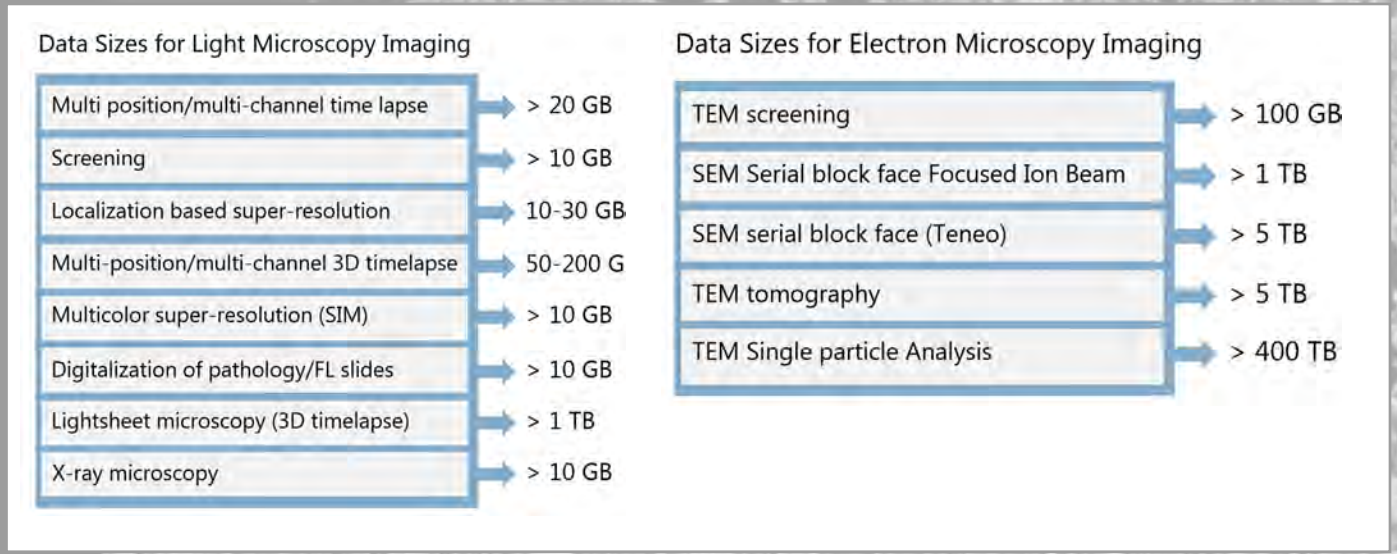
Figure 1 - Danish BioImaging *will be a multi-site national infrastructure consisting of nine imaging open access facilities, DBI-Nodes. Two DBI-Nodes, one at AU and one at DTU, will offer pre-clinical imaging technologies. Five DBI-Nodes will offer access to advanced microscopy applications (AU, DCRC, KU-CAB, KU-CFIM and SDU) at the local level and access to specific cutting-edge technologies at the national and international level. Computer scientists from KU-DIKU and DTU will develop novel user-friendly image analysis solutions based on their latest scientific research and implement them as remotely available workflows though DeiC or as local customized image analysis workflows at the imaging facilities.*

EURO BIOIMAGING

DANISH BIOIMAGING NETWORK

BioImaging

New imaging technologies are accelerating the life sciences as never before. They also generate a rapidly increasing amount of data (images), which need to be stored, documented and shared with the scientific community to comply with the FAIR principles



DANISH BIOIMAGING NETWORK

Expected DBI data production in TB:

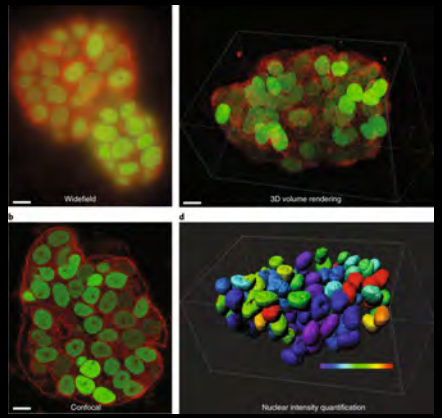
year	DBI
2021	7.400
2022	8.140
2023	8.954
2024	9.849
2025	10.834

~45 Petabytes

We need to define guidelines

- For how long does data need to be kept?
- What data needs to be kept to ensure Interoperable & Reusable?

RAW
Metadata



EURO BIOIMAGING

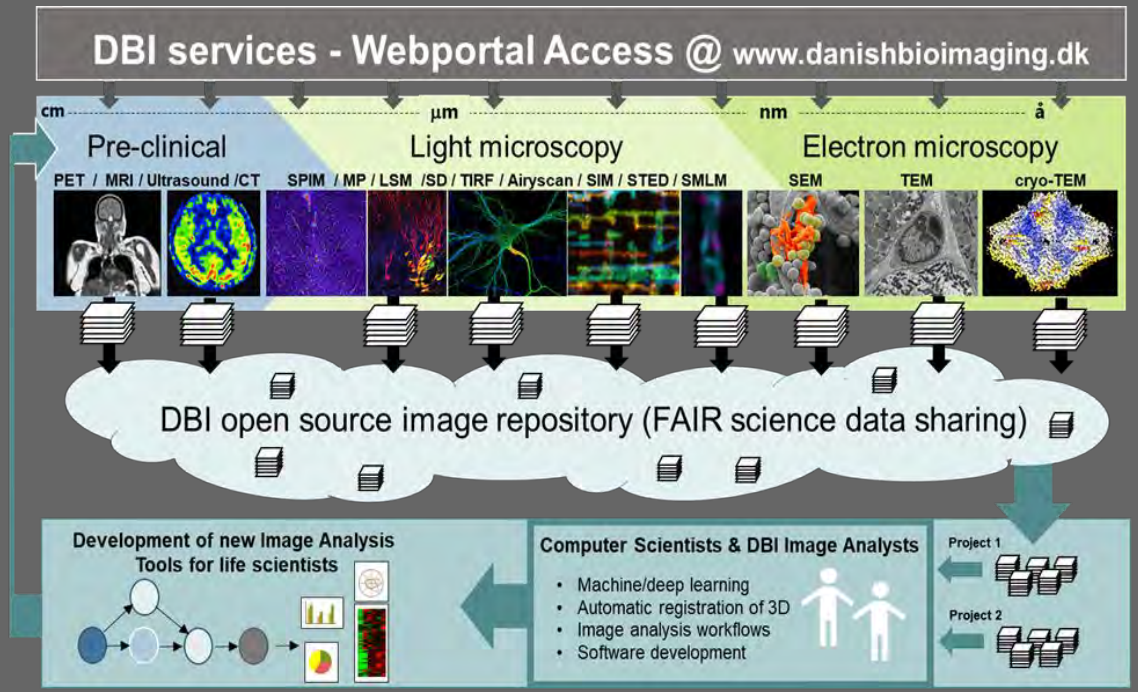
Danish BioImaging- image repository

Findable

Accessible

Interoperable

Reusable

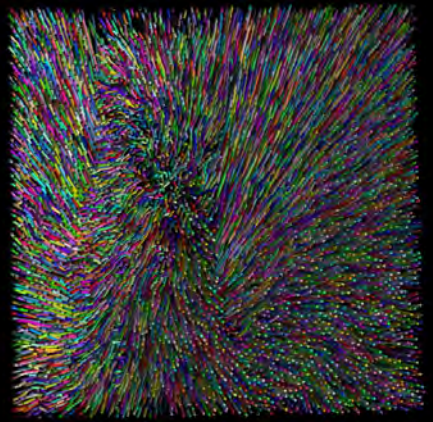


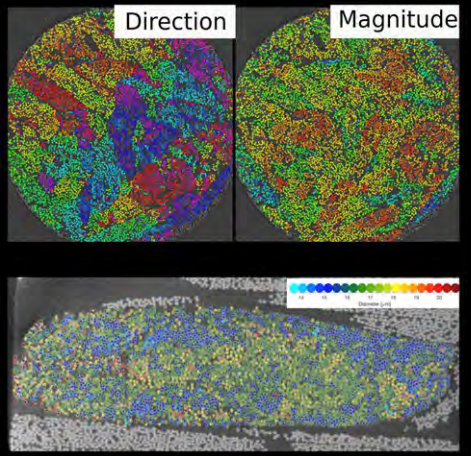
Center for Quantification of Imaging Data from Max IV (QIM)

DTU and DIKU are already building platforms to support 3D imaging analysis focused on MAX IV generated data, as part of the 3D Imaging Center at DTU and the ERDA system at Niels Bohr Institute, KU.

Professor at DTU
Anders Bjorholm Dahl

Professor at KU
Jon Sporring





EURO BIOIMAGING

DANISH BIOIMAGING NETWORK

FAIR on the bioimaging perspective – thoughts:

- Data ownership
- Data quality check
- Raw *versus* processed data – annotated data
- Big -to- Huge data

  Centralised *versus* decentralised

  - minimise duplication = storage space
  - minimise data transfer = time consuming

DANISH BIOIMAGING NETWORK

National & International repository of images

- accessible and safe
- Intelligent image data archival and retrieval
- Quality / integrity of data

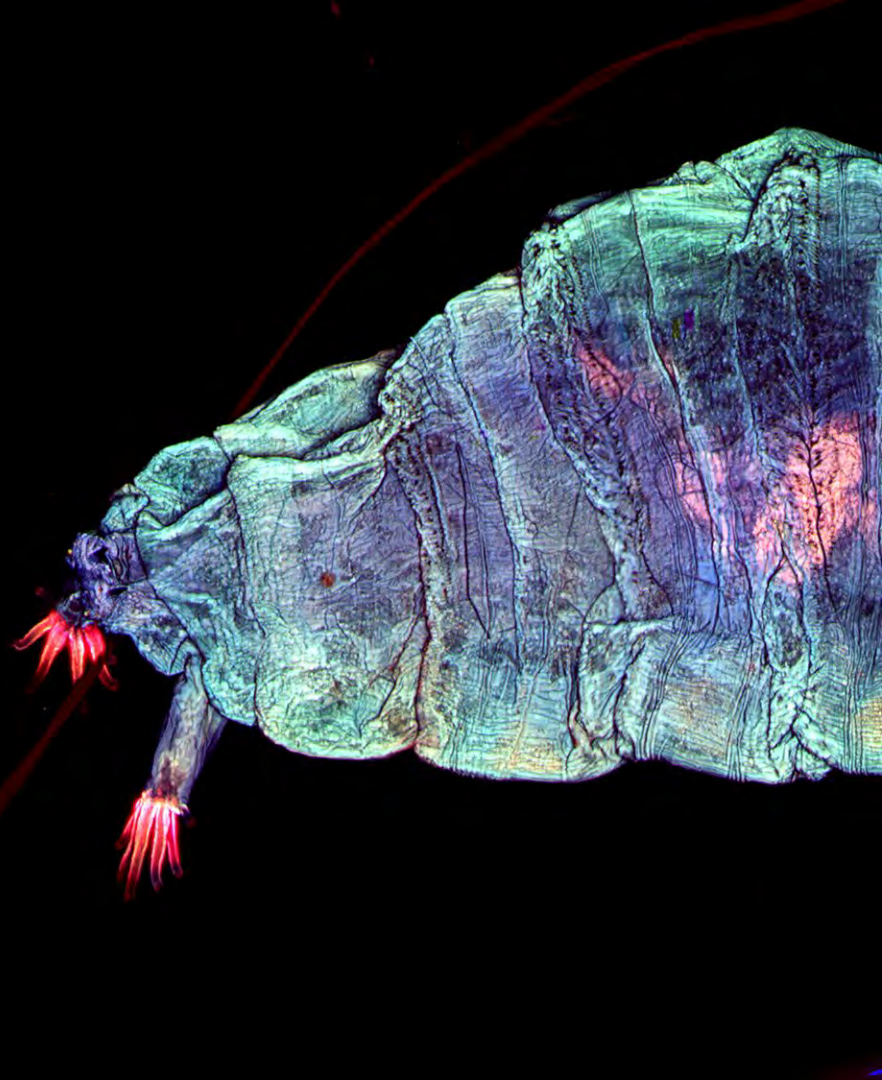
Sharing computing solutions

- **Image processing and analysis solutions for bioimaging data quantification and modeling**
- High performance infrastructure dedicated to massive computational demands including interactive and non interactive remote computing
- GPU based computing for deep learning and AI-based image analysis (XAI)

EURO BIOIMAGING

Questions?

Clara Prats –
Cprats@sund.ku.dk



DANISH BIOIMAGING NETWORK

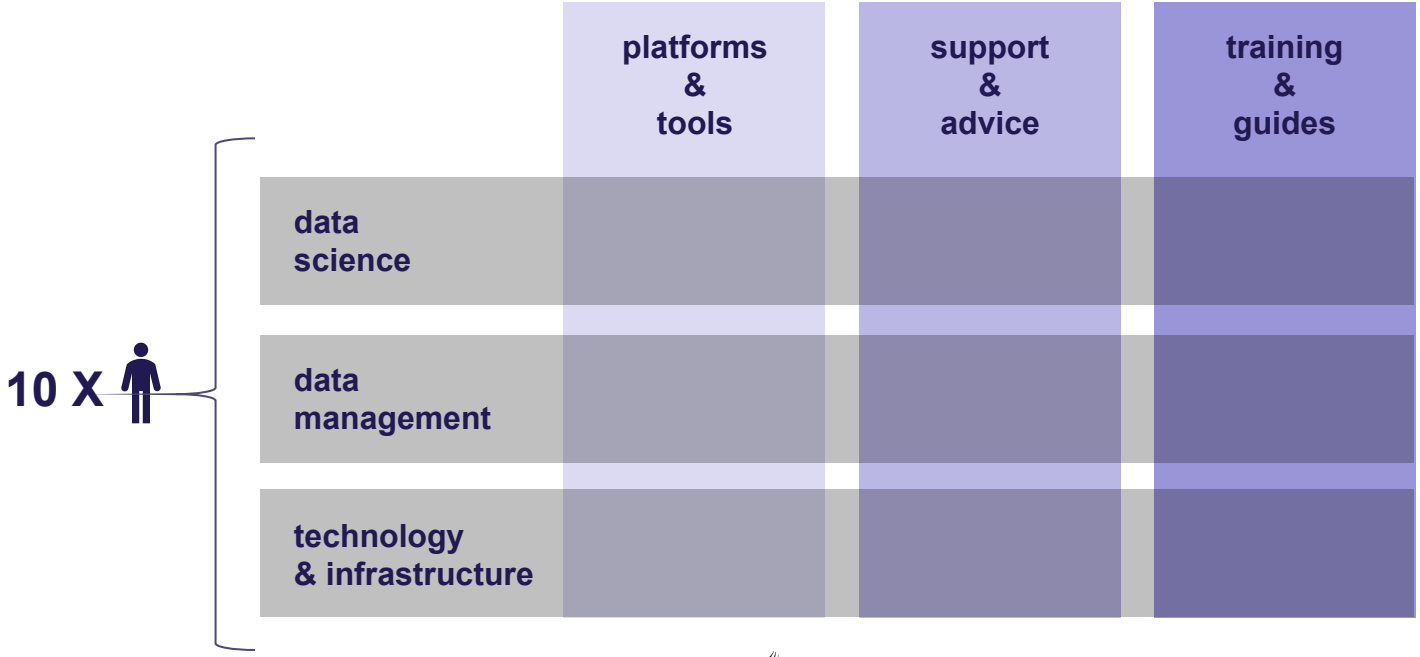
The need for local, national and international support functions

Diba Terese Markus, RDM consultant
CLAAUDIA Research Data Services

AALBORG
UNIVERSITY

CLAAUDIA
- a cross-functional team

CLAAUDIA
RESEARCH DATA SERVICES



AALBORG UNIVERSITET

The need for local, national & international support functions

CLAAUDIA
RESEARCH DATA SERVICES

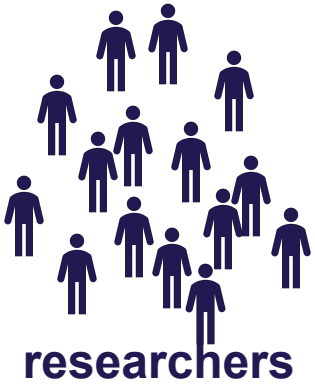
SUPPORT FOR WHOM?

SUPPORTING WHAT?

AALBORG UNIVERSITET

M4M support needs

CLAAUDIA
RESEARCH DATA SERVICES



	training & guides	platforms & tools	support & advice
international		●	
national		○	
local	●		●

AALBORG UNIVERSITET

CLAAUDIA
RESEARCH DATA SERVICES

An easily accessible portal to controlled vocabularies used in research

A controlled vocabulary reflects agreement on terminology used to label concepts. When research communities agree to use common language for the concepts in datasets, then the discovery, linking, understanding and reuse of research data are improved.

Research Vocabularies Australia (RVA) makes it easy to find and use controlled vocabularies used in research. It also makes it possible for Australian research organisations to publish, re-purpose, create, and manage their own controlled vocabularies. Vocabularies change over time, so the service enables management of new versions while retaining superseded versions.

Over time the RVA portal aims to describe any controlled vocabularies commonly used by or relevant to Australian researchers. Some vocabularies are also accessible directly from this portal (ie downloadable and queryable); some are to be accessed elsewhere and are simply described here. The ARDC can work with research organisations to enable software interaction with hosted vocabularies, and to develop the service in response to need.

Ways of using the service

Use vocabularies

- Find and learn about controlled vocabularies relevant to research
- **Access** those vocabularies and **reuse** them in your community
- Integrate vocabularies into your local information systems at a technical level

Provide vocabularies

- Upload and describe a vocabulary to **share** with others
- Make a vocabulary **machine readable** (more easily integrated into other's systems).
- Create new or import existing vocabularies and **manage** them with your community's input.

Give feedback on vocabularies

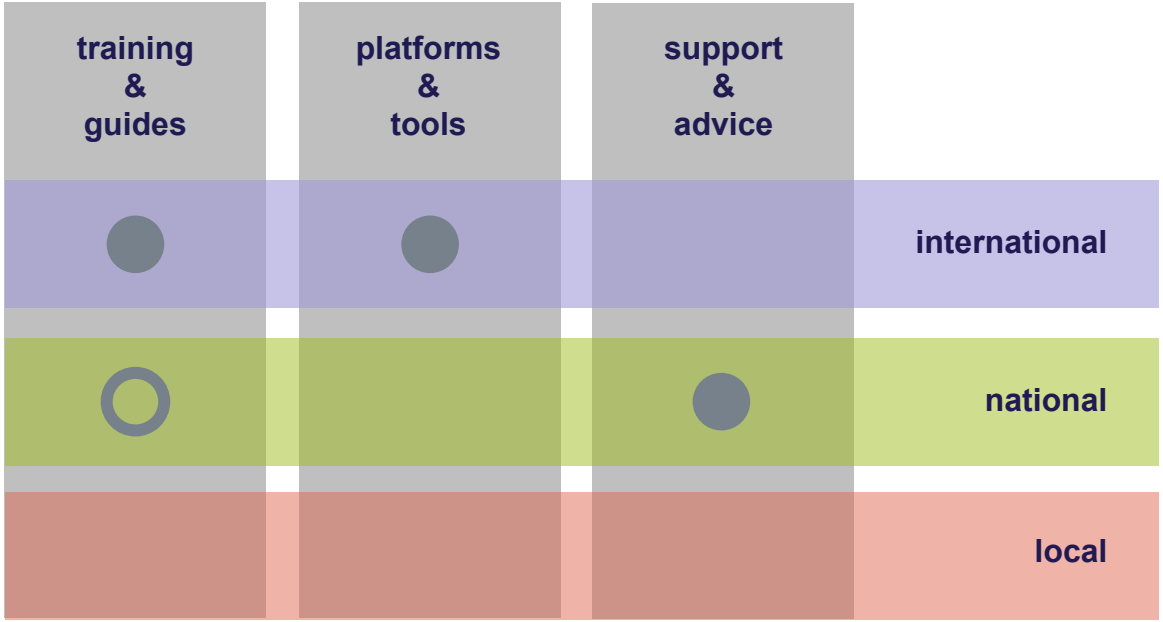
You can also Request the addition of a vocabulary or suggest changes to a vocabulary.

Anyone can search, browse and access the vocabularies described in the RVA portal. You can also self-register to describe or upload a vocabulary. However access to the editor is restricted to the ARDC partner institutions (Australian research organisations, including universities, research institutes, collecting organisations and government agencies). To use the RVA editor to create a new machine-readable vocabulary you will need to register here. Contact us to sign up for a free account.

https://ardc.edu.au/services/research-vocabularies-australia/

Support needs when ramping up M4M workshops

CLAAUDIA
RESEARCH DATA SERVICES



data stewards
RDM consultants

AALBORG UNIVERSITET

Going forward with M4M workshops - challenges @ local level

CLAAUDIA
RESEARCH DATA SERVICES

GETTING ORGANISED

- Gathering the team
- Acquiring the competencies needed

MARKETING

- Timing
- Getting the message right
- Avoiding the ‘yet another thing we researchers have to do’ pitfall

FACILITATING

- Mastering the WS format
- Handling differences in participant skill set levels
- Making it work with BYOD (Bring Your Own Data)

Future directions and plans

- Goal: Make FAIRification tools and methods easier available for everyone to use
- Create general tools to use for discipline specific metadata definitions
- Scale up process to include more researchers and (potential) data stewards
- Possible for non-techies to join, lower barriers
- Practical approach, both in terms of time and content of workshops
- Easy onboarding to FAIR for machines
- Collaborative effort

Plans for future

- Condensed M4M to be held for BioImaging community April 2021
  - still experimental, but already including lots of experience
- If successful, complete and publish course material and toolbox
- If there is interest: Establish training program for workshop facilitators
- Continued cooperation with GO FAIR
  - co-designing the M4M format, integrate with GO FAIR M4M Handbook
  - learn from GO FAIR's facilitator training program
  - cooperation with other active parties, e.g. GO FAIR US
  - consider future options for e.g. certification, GO FAIR DK office
- Support continued FAIR strategy and implementation plan development