

DTU





IG RDM in Engineering - Online seminar series - "Exploring annotation and metadata initiatives for engineering data"  
Seminar 3 "Building metadata standards within Engineering disciplines and communities"

Nikola Vasiljevic

# DTU Wind Energy metadata and terminology

# Disclaimer

*This presentation contains personal opinion which not necessarily reflect the DTU Wind Energy or Research Data Alliance politics.*

# DTU Wind Energy – who are we

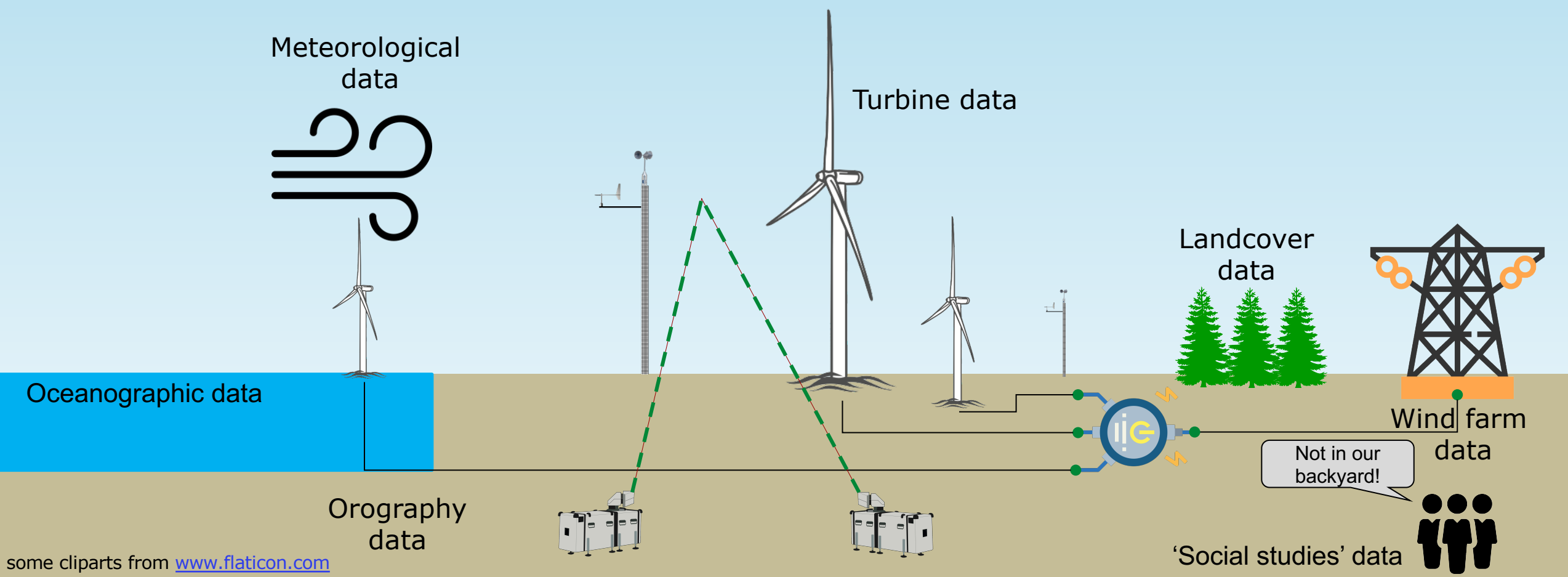
- DTU Wind Energy is **one of the largest and most well-known** university department for wind energy in the world with **250 employees**.
- DTU Wind Energy consists of three divisions (and many sections):
  - Wind Energy Systems
  - Wind Energy Materials and Components
  - Wind Turbine Design
- **We 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 wind power plant lifecycle

# 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)

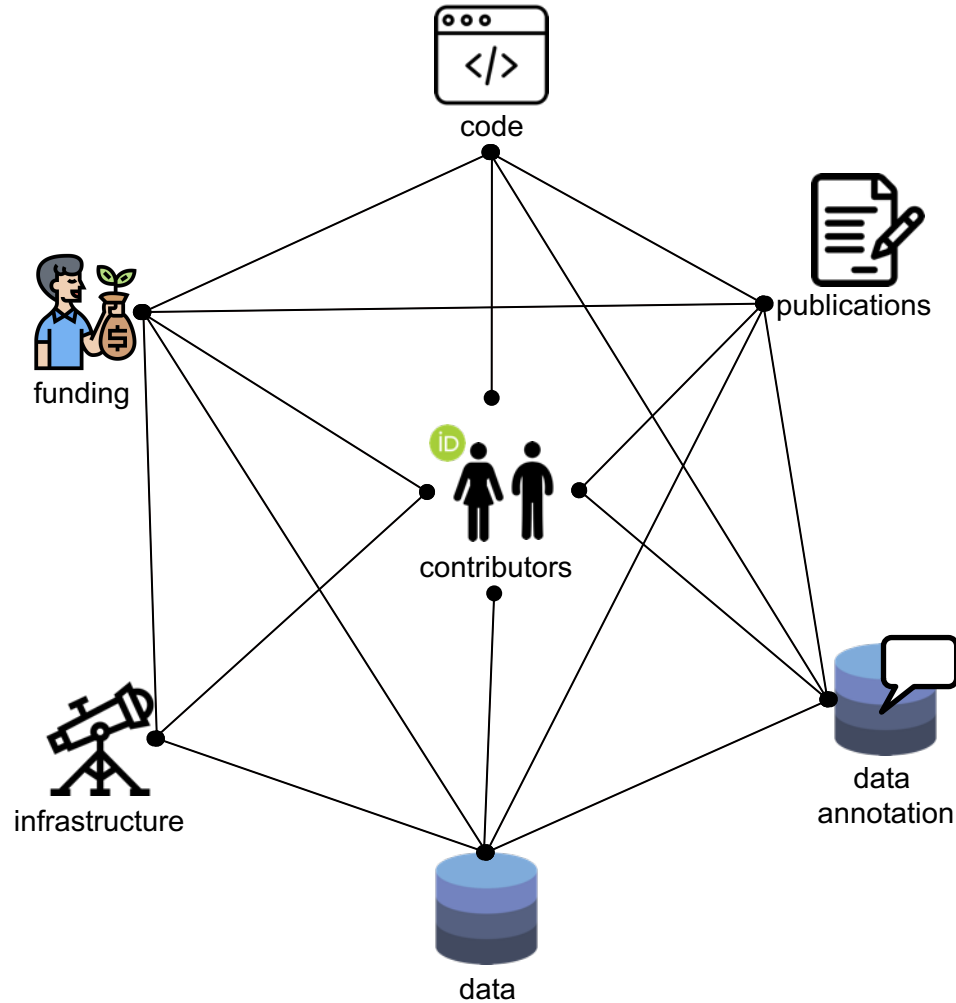
# Wind Energy domain

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



# Where we would like to go...

# Interconnected and richly described resources

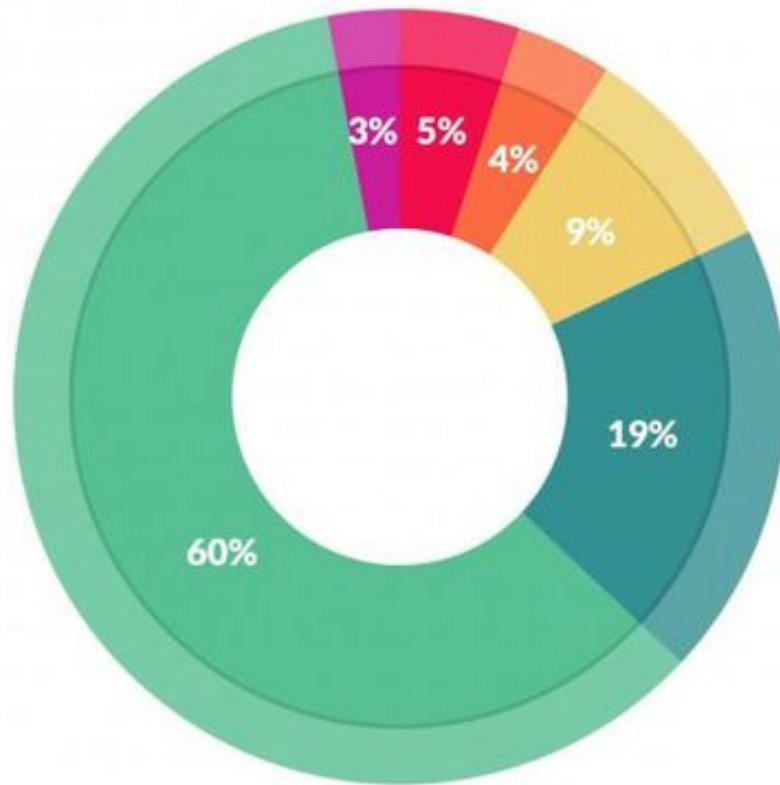


Full chain of custody  
 Researcher at center



# Why?

# How much time we spend analysing data?

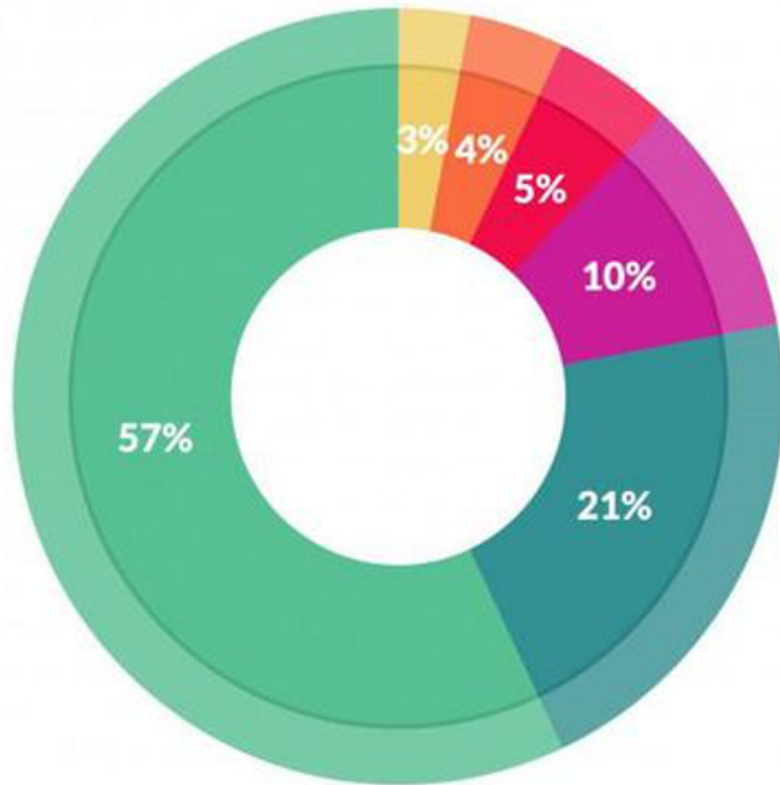


## What data scientists spend the most time doing

- 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 is not perceived as 'cool' activity compering to data analytics, however it has much more lasting impact then trendy data analytics methods.**

# How?

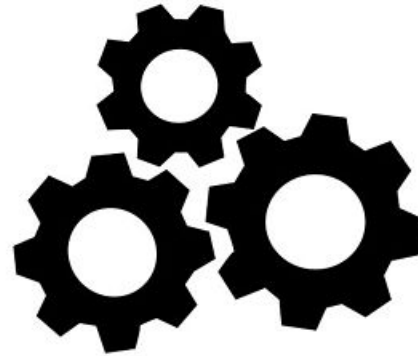
F  
indable



A  
ccessible



I  
nteroperable



R  
eusable



Source: [wikimedia](https://www.wikimedia.org/)

# Our interaction with the FAIR principles

- 2017 EERA JP WIND IRP Wind Open Data (European level)
- 2018 Internal project and FAIR ambassadors (University/Department level)
- 2019 RDA Ambassadorship ('Individual' level)
- 2020 Running M4M workshop at the moment, interaction with the RDA (National/International level)

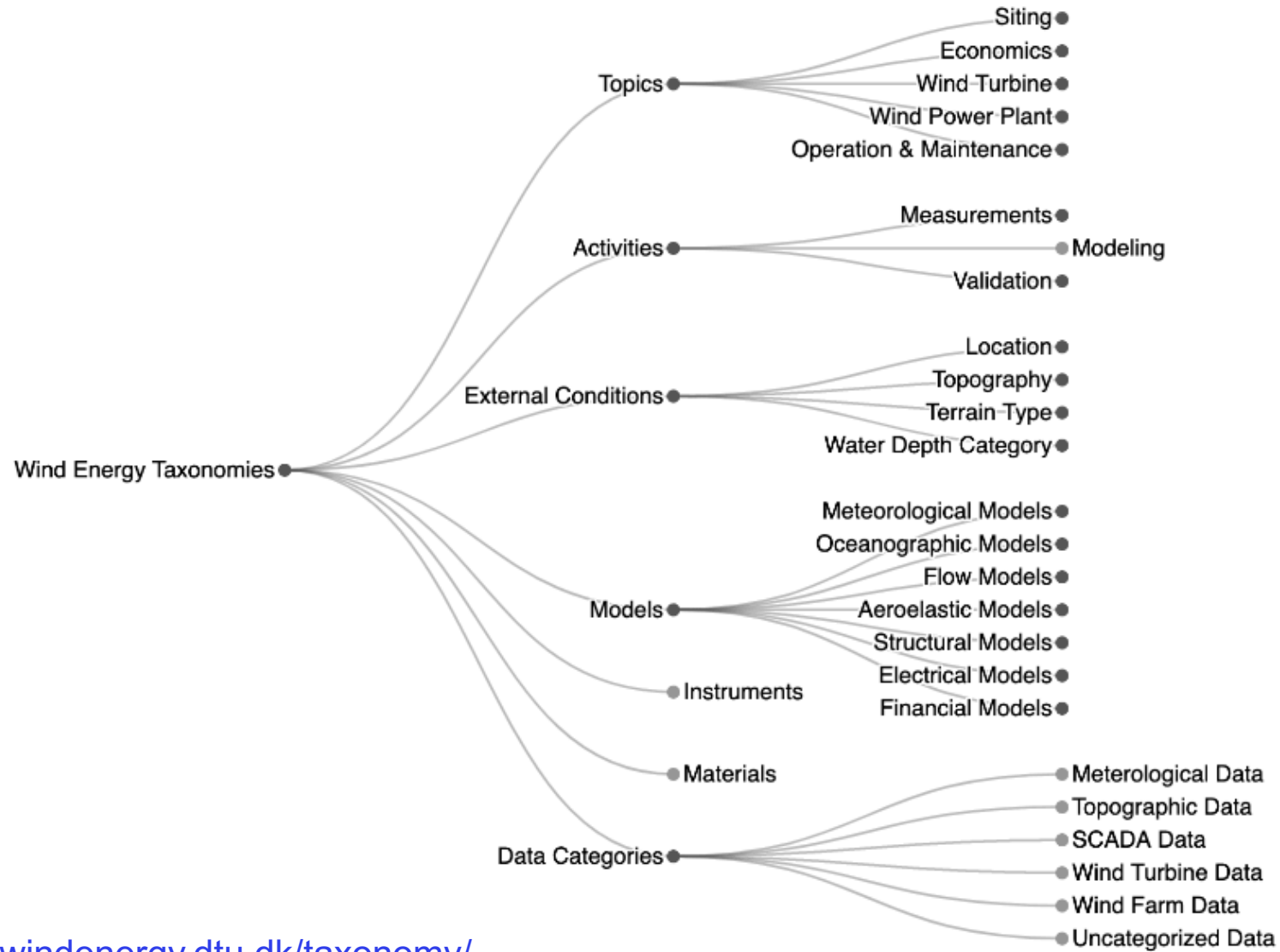
*Disclaimer: I cherry-picked some activities, there were other activities related to data.*

## 2017 – tackling 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 facet-search / term-search
- 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/>

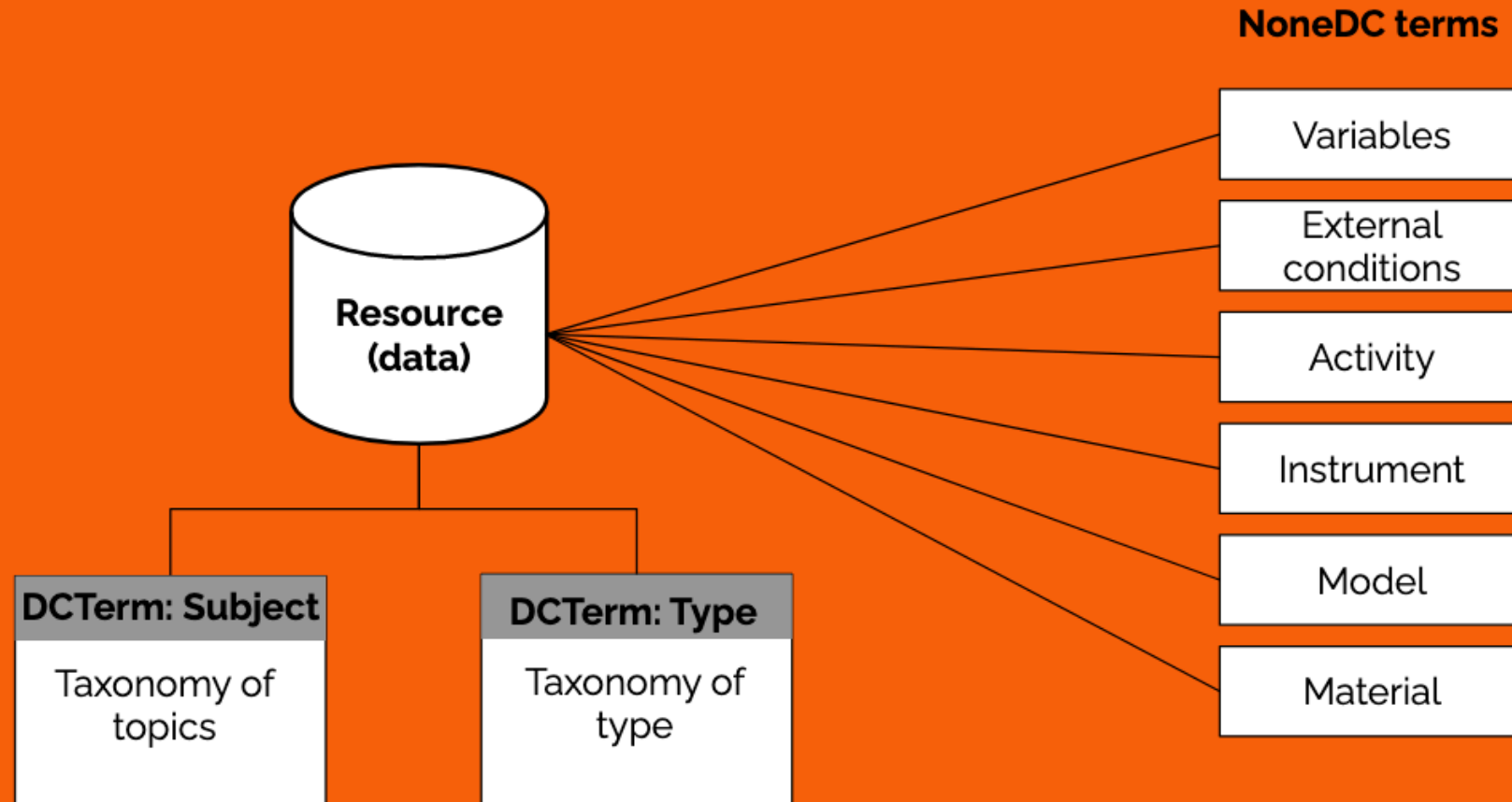
# Dublin Core Application Profile

		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 Application Profile

GOAL: describe datasets with metadata cards




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

## 2018 – implementing FA of FAIR

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

Technical University of Denmark | [https://data.dtu.dk/DTU\\_Wind\\_Energy](https://data.dtu.dk/DTU_Wind_Energy)

DTU Browse  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 Gözçü ▾ 20/04/2020

**DATASET**

**LES of wake flow behind 2.3MW wind turbine**  
Søren Juhl Andersen 26/03/2020

**DATASET**

**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 Ahsbøhn ▾ 20/01/2020

**Animation of lidar and simulation data of complex flow over the...**  
Robert Menke ▾ 15/11/2019

**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](https://data.dtu.dk/articles/dataset/The_DeRisk_Database/10322033)

Animation of lidar and simulation data of complex flow over the Perdigão measurement site

1716 views | 261 downloads | 1 citations

**DTU Data**  
- a service from DTU Library

**Version 2** ▾ Media posted on 15.11.2019, 14:30 by Robert Menke, Jose Laginha Palma, Jakob Mann, Nikola Vasiljevic, Alexandre Silva Lopes, Carlos Veiga Rodrigues, V.M. Costa Gomes

Animation of lidar and simulation data of complex flow over the Perdigão measurement site for a 24h period on May 14-15, 2017. Details on the lidar measurements and the simulation can be found in the attached reference.

A detailed description of the measurement setup and the computer model, such as a description of flow process observed during the period can be found under the following link:  
<https://doi.org/10.1088/1742-6596/1222/1/012006>

**FUNDING**  
The Perdigão-2017 field campaign was primarily funded by the US National Science Foundation, European Commission (ENER/FP7/618122/NEWA), Danish Energy Agency, German Federal Ministry of Economy and Energy, FCT-Portuguese Foundation for Science and Technology, (NEWA/1/2014), and US Army Research Laboratory.

**TIMELINE**

Date	Event
19.03.2019	Submission date
22.03.2019	First online date
15.11.2019	Posted date

**REFERENCES**

- <https://doi.org/10.1088/1742-6596/1222/1/012006>

**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

**CATEGORIES**

- Atmospheric Dynamics
- Atmospheric Sciences
- Meteorology
- Renewable Power and Energy Systems Engineering (excl. Solar Cells)

**KEYWORD(S)**

- complex terrain
- complex flow
- Perdigão
- scanning lidar
- WindScanner
- simulation results
- VENTOS

**LICENCE**  
CC BY 4.0


**EXPORT**

- RefWorks
- BibTeX
- Ref. manager
- Endnote
- DataCite
- NLM
- DC



Technical University of Denmark | [https://data.dtu.dk/DTU\\_Wind\\_Energy](https://data.dtu.dk/DTU_Wind_Energy)

DTU Browse  Log in

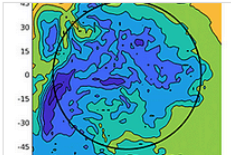


Discover research from **DTU Wind Energy** + Follow

ALL CATEGORIES SEARCH sort Posted date ↓

DATASET

**HAWC2 Input data and statistics of time series for slightly modified...**  
Ozan Gözçü ▾ 20/04/2020



**LES of wake flow behind 2.3MW wind turbine**  
Søren Juhl Andersen 26/03/2020

DATASET

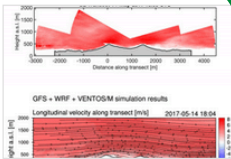
**The DeRisk Database**  
Fabio Pierella ▾ 29/01/2020

DATASET

**Drone-based wind lidar proof-of-concept measurements**  
Nikola Vasiljevic ▾ 29/01/2020

DATASET

**SAR wind atlas US East Coast**  
Tobias Torben Ahsbøhn ▾ 20/01/2020



**Animation of lidar and simulation data of complex flow over the...**  
Robert Menke ▾ 15/11/2019

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](https://data.dtu.dk/articles/dataset/The_DeRisk_Database/10322033)

Animation of lidar and simulation | [https://data.dtu.dk/articles/media/Animation\\_of\\_lidar\\_and\\_sim](https://data.dtu.dk/articles/media/Animation_of_lidar_and_sim)

**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 mechanic I of FAIR

- Switching from JSON to RDF representation
- SKOS Lexicalization to describe our controlled terminologies
- Selected and running in staging env a stack of ontology technologies to
  - Build and maintain controlled terminologies
  - Serve terminologies to humans
  - Serve terminologies to machines
- Will be available for public at [data.windenergy.dtu.dk](https://data.windenergy.dtu.dk) domain

DTU

DTU Wind Energy

DTU Wind Energy metadata and terminology

15 September 2020

Web app for building and maintaining controlled terminologies

VocBench

Projects Data Metadata SPARQL History Tools

Class Concept Scheme Collection Property Datatype

Operation & Maintenance (en)

Economics (en)

Wind Power Plant (en)

Wind Turbine (en)

Siting (en)

Spatial Planning (en)

Enviromental Impact (en)

Social Acceptance (en)

Nature Impacts (en)

Noise Perception (en)

Legal Aspects (en)

Infrastructures (en)

Design Conditions (en)

Resource Assessment (en)

Long-Term Extrapolation (en)

Wind Atlas (en)

Wind Mapping (en)

Enviromental Impact (en) http://data.windenergy.dtu.dk/taxonomy/topics/EnviromentalImpact

Types:

rdf:type

skos:Concept

Top Concept of:

Schemes:

skos:inScheme

TaxonomyTopics (en)

Broaders:

skos:broader

Spatial Planning (en)

Lexicalizations:

skos:prefLabel

Enviromental Impact

Notes:

Other properties:

ResView TermView Code

Status:



## Wind Energy Taxonomy of Topics

Content language

English



Search

Alphabetical

Hierarchy

Groups

- ↳ Economics
- ↳ Operation & Maintenance
- ↳ Siting
  - ↳ Design Conditions
  - ↳ Infrastructures
  - ↳ Long-Term Extrapolation
    - ↳ Wind Indices
  - ↳ Resource Assessment
  - ↳ Spatial Planning
    - ↳ Environmental Impact
    - ↳ Legal Aspects
  - ↳ Wind Atlas
  - ↳ Wind Mapping
- ↳ Wind Power Plant
- ↳ Wind Turbine
  - ↳ Concept Design
  - ↳ Controls
  - ↳ Nacelle
    - ↳ Cooling
    - ↳ Gearbox
    - ↳ Generator
    - ↳ Main Shaft
    - ↳ Power Electronics
    - ↳ Turbine Control
    - ↳ Yaw
  - ↳ Rotor
  - ↳ Support Structure
  - ↳ Tower

### Vocabulary information

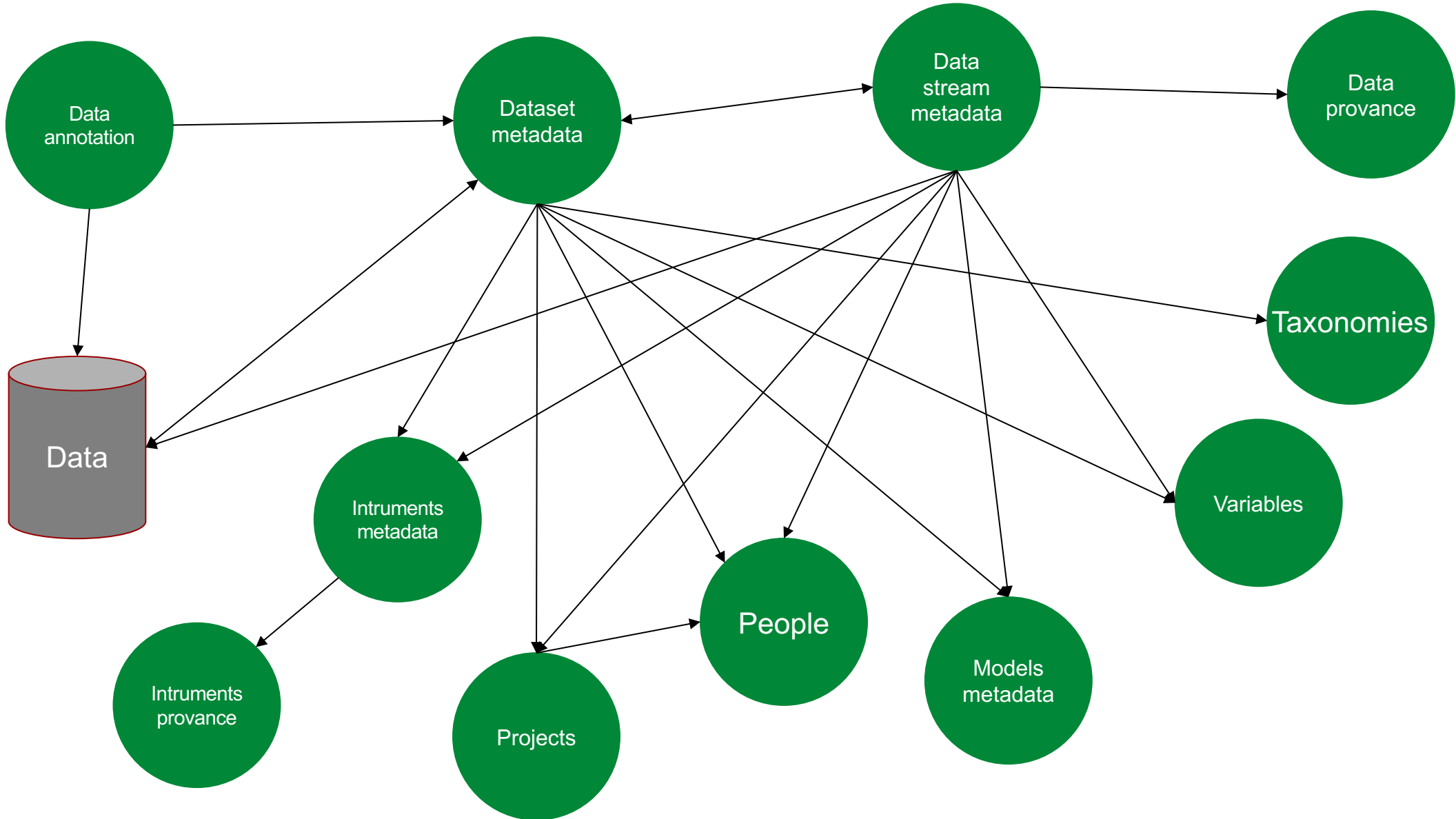
TITLE [Wind Energy Taxonomy of Topics](#)TYPE <http://www.w3.org/2004/02/skos/core#ConceptScheme>URI <http://data.windenergy.dtu.dk/taxonomy/topics/WindEnergyTaxonomyOfTopics>Download this vocabulary: [TURTLE](#)

Web app for serving controlled terminologies to humans

# 2020 – tackling machinic I of FAIR / metadata templates

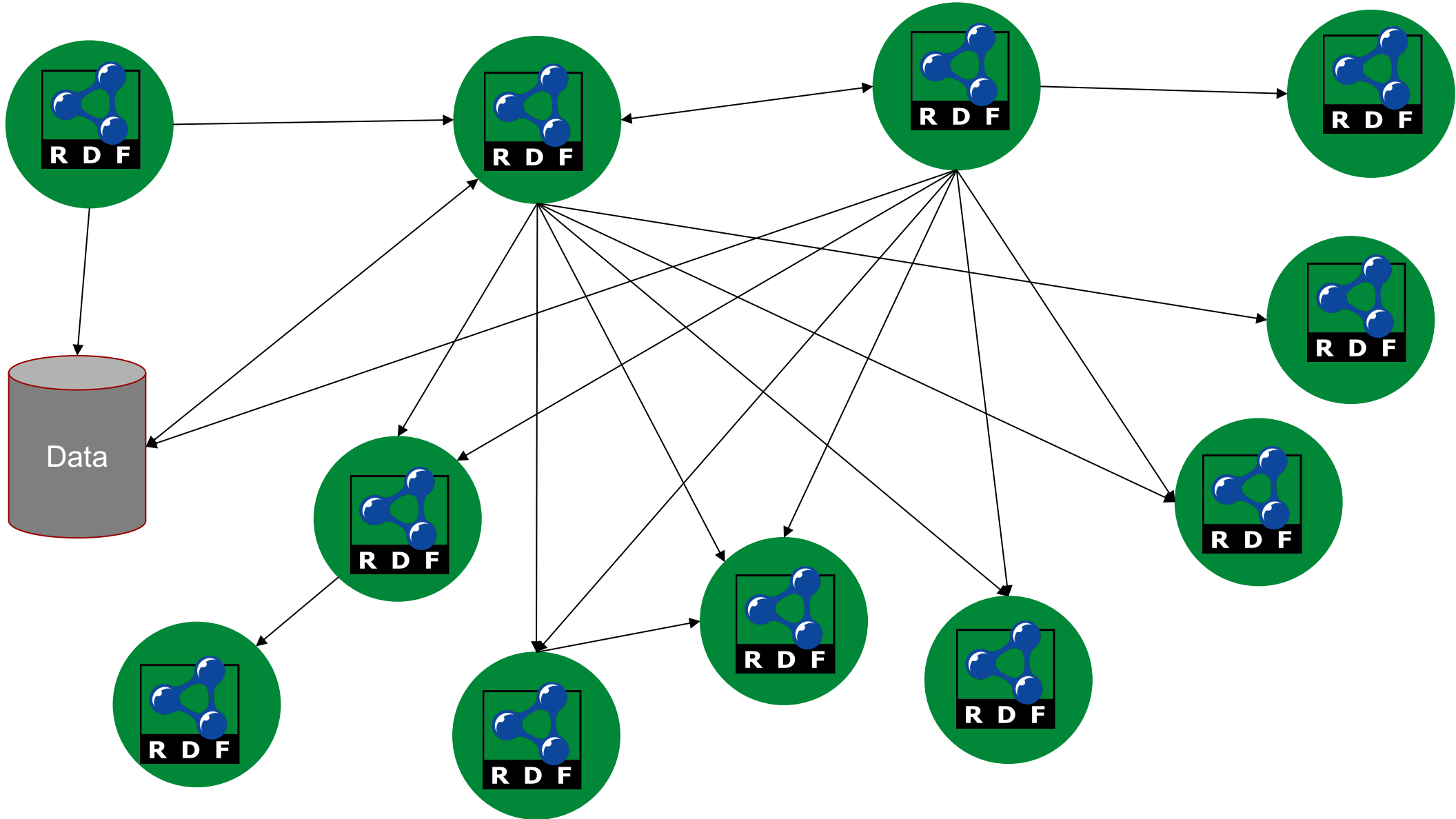
- Selected NetCDF data format for sharing & publishing data
- Extending our Dublin Core Application Profile making it machine actionable
- Using DataCite metadata schema (4.3) as a base template for dataset metadata
- Extend template with missing fields that will provide information about:
  - (geo)spatio-temporal 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)
- Plenty of room for further automation
- Provided use-case for [the RDA WG PIDINST](#) (mapping of internal instruments db to PIDINST scheme)

# Where are we going with all of this?



Arrows are semi-random

# Where are we going with all of this?



Arrows are semi-random

# Summary

- Everything that has been done so far it has been done:
  - with ‘bottom-up’, ‘divide & concur’ and ‘fail-fast’ approach
  - by knowing things good enough to be ‘dangerous’
  - largely by a few persons
  - and mainly in our ‘free time’ (we had very limited funding)
- If you want to be ‘FAIR’ you:
  - must roll up your sleeves and start reusing, adapting, building and testing solutions
  - not avoid being technical
  - 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
- Alternatively you can hope that someone else will do the hard work for you (good luck!)

# Summary

- If you are up for hard but rewarding work:
  - decide what you want to do
  - decide how you would like to do it
  - assess what you can build/design considering constraints
  - think on how you are going to use it
  - think on how you will maintain it and update it
  - think on how you will eventually replace it

# Thank you for your attention!

## Questions?

[niva@dtu.dk](mailto:niva@dtu.dk)

