

# **Find the data: Metadata and taxonomies for **FAIR** data reuse in Wind Energy.**

**Anna Maria Sempreviva & Many More**

Technical University of Denmark

DTU Wind Energy

Coordinator of metadata and taxonomy creation

# Content

- Context:
  - The digital transformation of the Wind Energy sector,
  - The stakeholders: data owner, data user, the funding body
  - From Open to FAIR, the culture of sharing
- Metadata and Taxonomies for the Wind Energy what and why?
- Data Portal and data registry
- The FAIRification process
- Conclusive remarks

**PO.094**

**OPEN SCIENCE: SHARING DATA, TOOLS AND WORKFLOWS  
A STRATEGY TO INSPIRE EFFICIENT COLLABORATION**

Anna Maria Sempreviva<sup>2</sup>, Nikolay Dimitrov<sup>1,2</sup>, Nikola Vasiljevic<sup>2</sup>, Neil Davis<sup>2</sup>, Falco Hüser<sup>4</sup>, Paula Lavanchy<sup>1</sup>  
Technical University of Denmark, DTU, <sup>1</sup>Presenting author, <sup>2</sup>Department of Wind Energy, <sup>3</sup>Office for Innovation and Sector Services

**Abstract**

The H2020 Work Programme is a milestone for the transition to Science 2.0: the era of **Open Science (OS)** will increase the drive towards **Open Science**. In particular, the **Open Data (OD)** policy aims at making digital assets, i.e. data, tools and workflows **FAIR (Findable, Accessible, Interoperable and Re-usable)** [1]: available to everyone in Europe. FAIR assets allow to reuse data in multiple applications which multiply the data value thus optimizing the impact of projects funded by public money. **NO PANIC**: to respect industry foregrounds **Research data must be as open as possible and as closed as necessary!** The wind energy community generally agrees that sharing assets would shorten the time from new ideas to innovation, making the work more efficient; BUT assets provide a competitive advantage, leading to a reluctance to sharing important data. **WHAT TO DO THEN?** We suggest a simple strategy to energize the free flow of information amongst the European wind energy stakeholders, increasing collaboration by sharing research data including data, software and workflows.

**Objectives: Share digital objects to multiply their value.**

The main goal: increase collaboration by sharing the three main type of digital objects e.g. data sets, software, and workflows. These three digital objects have a lifecycle: design, management, publication and discovery for reuse. This cycle happens at different levels e.g. at group, at section, at departmental, at Organizational level and Global level.

**Method: the DTU Wind Energy FAIR strategy**

**Results: Taxonomy and metadata, connecting stakeholders. Meet the data owner and the data user**

**DTUdata, a research data repository** with a metadata catalog has the potentiality to connect users to data owners, starting or reinforcing cooperation.

**Data owner /creator**

- Can tag and make visible data via metadata, choosing suitable terms from standard vocabularies; and
- Can maintain control on data access without necessarily uploading any data

**Data Market Place?**  
€ € \$ ?  
Services?  
Co-creation?

**Data user**

- Can find data by searching the same terms as used by the data owner
- Can retrieve information on available data
- Can work efficiently

**Conclusions: FAIR data makes work more efficient, connected, and visible**

- Research Data are assets giving competitive advantage but can also insure visibility
- It is possible to **secure information sharing** in collaborative environments increasing the work **efficiency** and broadening the **Impact** of research
- Findable and Accessible in FAIR do not mean **directly** accessible. On the other hand, you and your company will be **directly visible** and **marketable**.

**References**

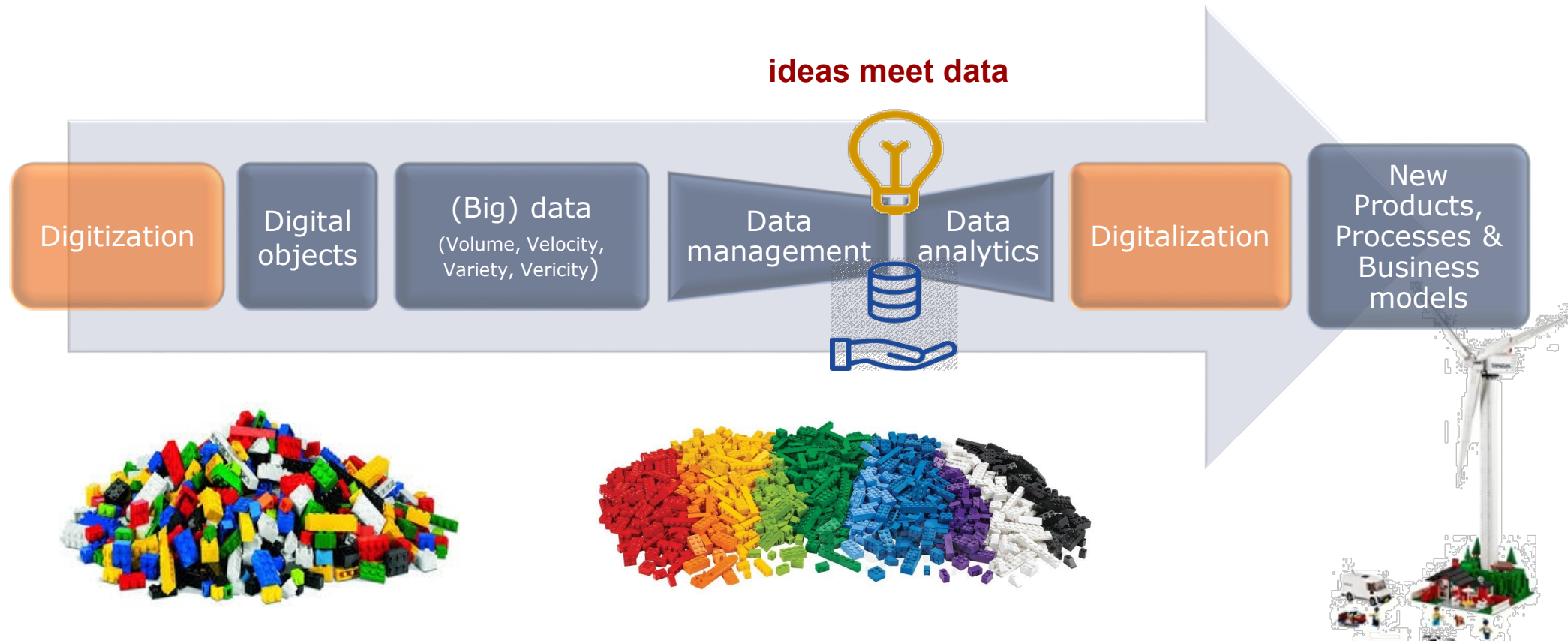
1. Ackerson, M. S. et al. **FAIR Research Principles for research data management and sharing**. *Sci. Data* 10:1000 (2023) <https://doi.org/10.1038/s41598-023-28100-1>

2. Sempreviva, A. M., Vasiljevic, N., Davis, N., Hüser, F., Lavanchy, P., Dimitrov, N., et al. **FAIRification and metadata for 2020-2023 Research & Innovation** (p. 12)

# Context: Digital transformation of Wind Energy sector

## Grand challenge: 2050 be the leading renewable

There is a bottleneck in whatever workflow if data is not properly engineered and managed



# Context: the stakeholders

## The data owner (Most Industry) demands:

- Control on data as assets
- Recognition
- Protection of competitive advantage

But willing to expose his data



## The data user needs:

- Improved efficiency
- Money saving

Willing to sign NDAs & reward

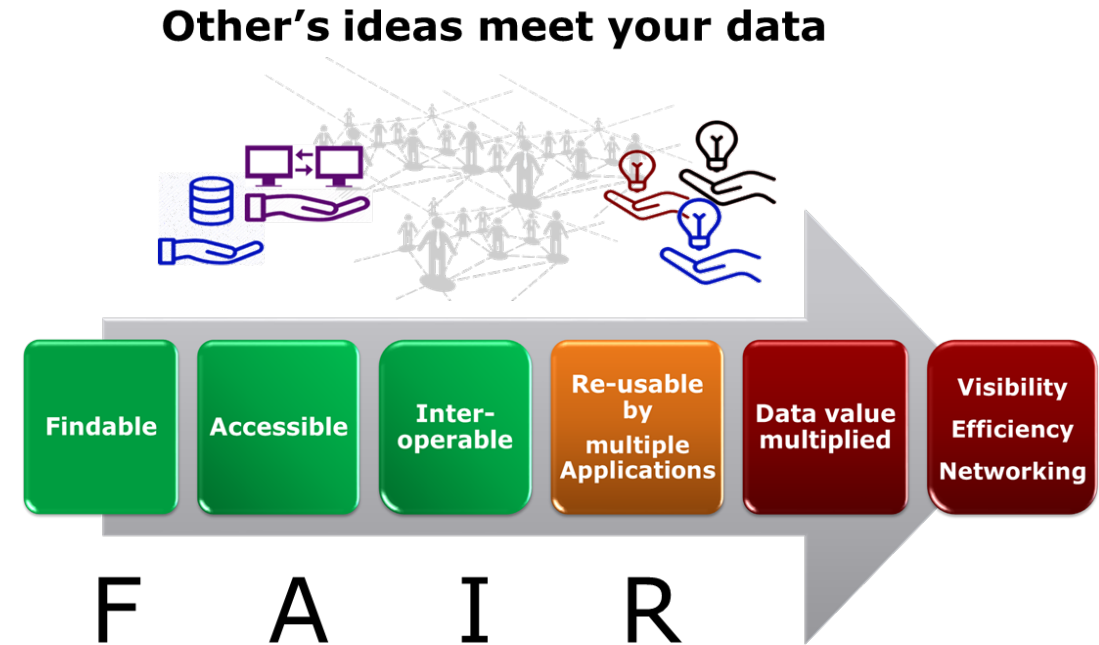
**Visibility – Efficiency – Networking – Control – Recognition – Reward**

# Context: The Funding body, H2020 from Open to FAIR

- 2014 H2020: Open Data .... **Panic!**
- 2016 H2020: FAIR Data Principle(s)\* ..... Milestone:
- **FAIR changes focus: From Available to Findable data**
- Data as open as possible as closed as necessary

**ISSUE:** How to make data findable but safe?

**SOLUTION:** Create a searchable data catalog for **distributed** data



\*Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. doi:10.1038/sdata.2016.18



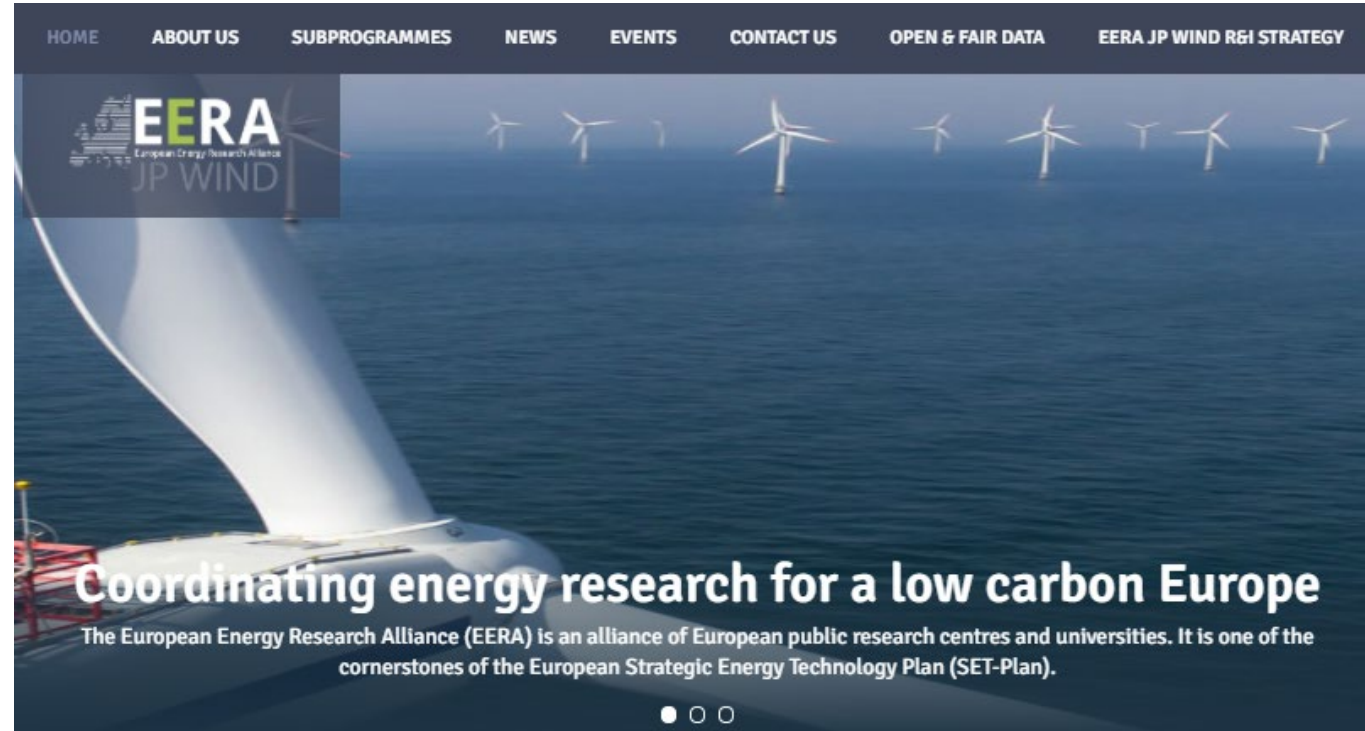
# How it started. A bottom up approach.

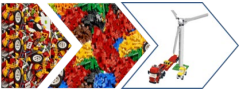
European Energy Research Alliance, Joint  
Programme on Wind Energy,  
EERA JPWind Community  
**50+ European Research Organizations**

FP7 Integrated Research Programme in  
Wind Energy IRPWind project  
**27 EERA JPWind Organization**

**Goal : create a new WP for open data**

**EERA JP Wind** digital object findability





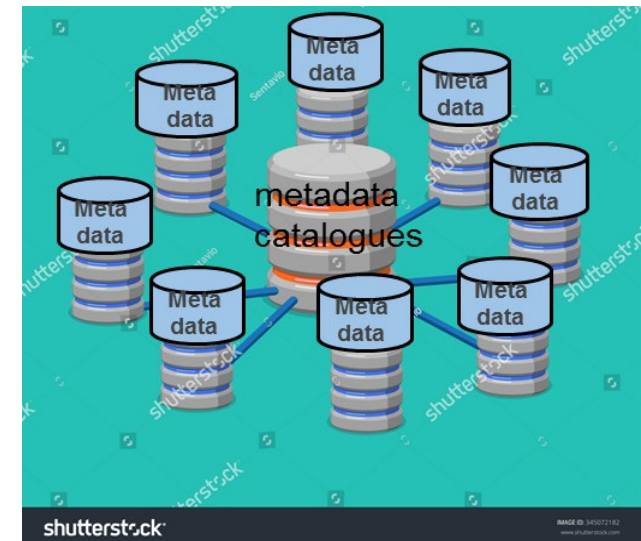
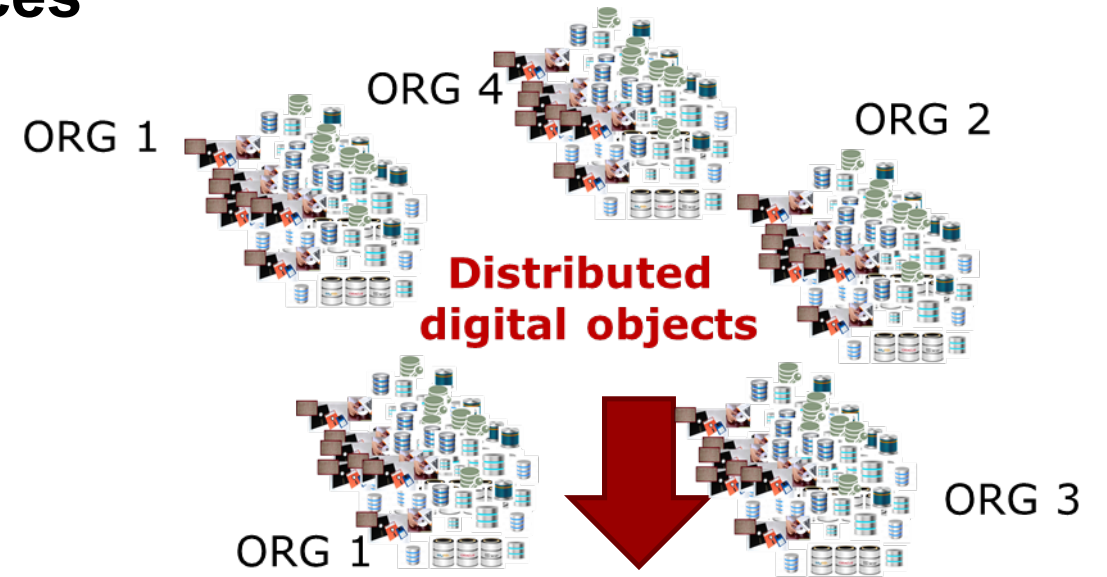
# FIND EERA JPWind distributed resources

## Issue EERA digital object **findability**

- Datasets are distributed in the “cloud”, organized and stored in different ways.
- Datasets often miss **documentation (Metadata)**

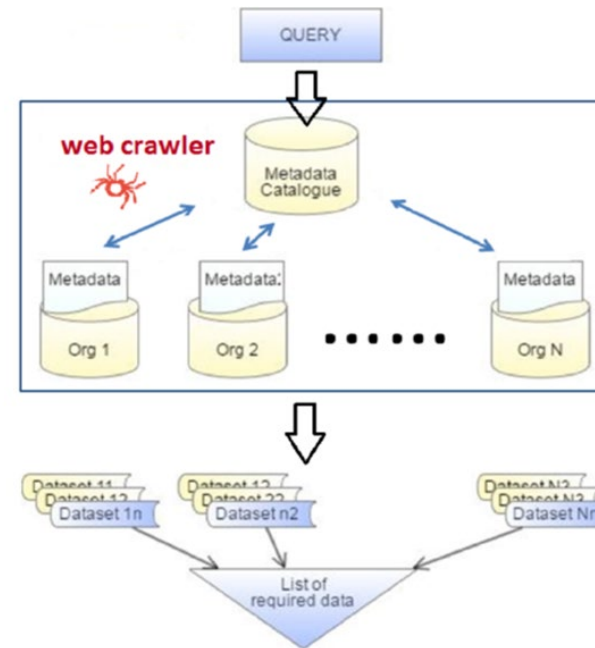
## Action: we need 3 ingredients

- Create a list of **metadata**
- Assign **taxonomies** to metadata: controlled vocabularies to tag data
- Design **a data portal as a Virtual Library with a metadata catalog**



# How virtual distributed data catalogs / libraries work?

## Search Engine



## ADVANTAGES

- Data owner has **control**
- Data are **visible** without being accessible
- No uploading data and no need of storage and data preservation.
- By applying filters, users can accurately locate needed data

Search engine  
Registry

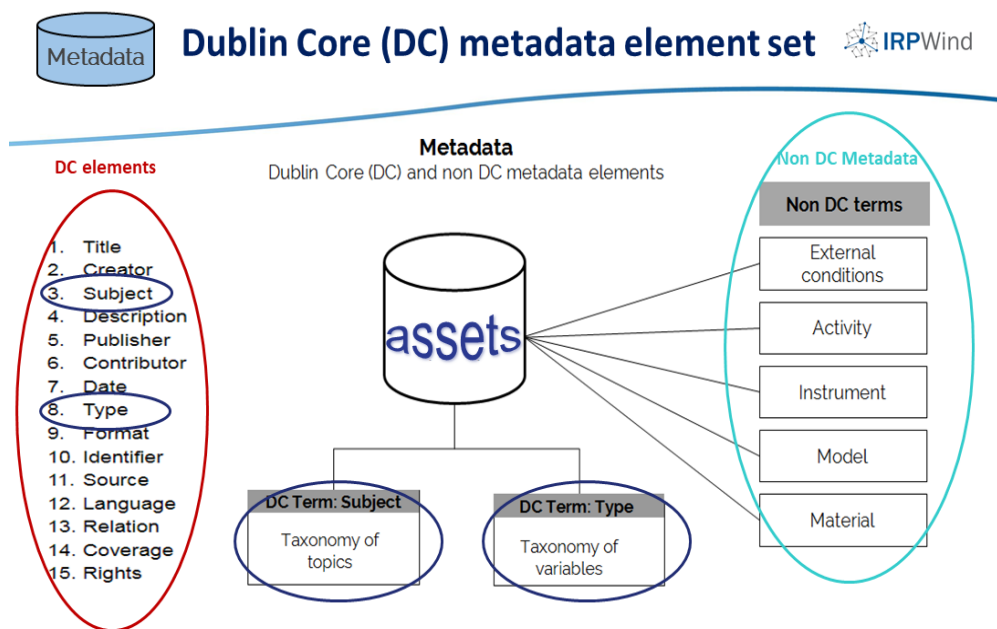


Storage

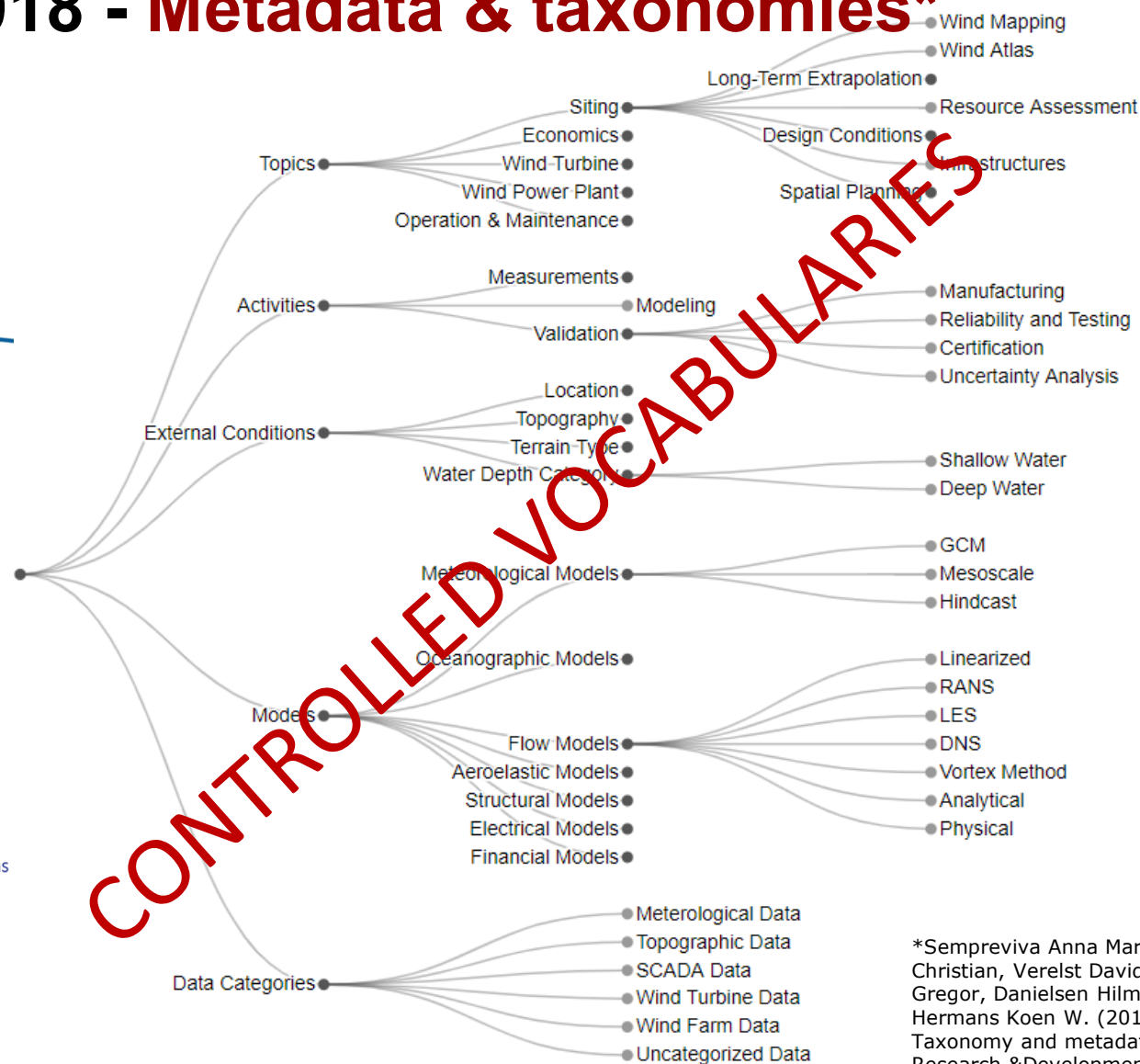




# IRPWind Project 2014-2018 - Metadata & taxonomies\*



- For the 7 metadata elements we defined specific taxonomies as VOCABULARIES with acknowledged terms



\*Sempreviva Anna Maria, Vesth Allan, Bak Christian, Verelst David Robert, Giebel Gregor, Danielsen Hilmar Kjartansson, ... Hermans Koen W. (2017, December 12). Taxonomy and metadata for wind energy Research & Development. Zenodo. <http://doi.org/10.5281/zenodo.1199489>

Anna Maria Sempreviva\*, Nikola Vasiljevski\*, Mattias Andersson\*, Christian Book\*, Gregor Glebe\*, David R. Verelst\*,  
Petr Maule\*, Hilmar Danielsen\*, Allan Vesth\*, Lars Pilgaard Mikkelsen\*,  
Stephan Barth\*, Javier Sanz Rodrigo\*, Pawel Ganoorski\*, Tor Inge Reigstad\*, Hans Christian Bolstad\*,  
Jan Willem Wagenaar\*, Koen Hermans\*

FIND THE DATA



\* DTU Wind Energy, \* ForWind, \* CENER, \* SINTEF, \* ECN

## Wind Energy

Siting

Economics

Wind Turbine

Wind power plant

Operation & Maintenance

### Introduction

Open Access (OA) to data, or Open Data (OD), is widely acknowledged as a fundamental step to support a fast track from research to innovation. In IRPWind and the EERA Joint Programme for Wind Energy, EERA JP Wind, OD and data management are considered important building block for increasing the European collaboration in wind energy research, by making the vast amount of existing and future data findable.

As part of the IRPWind project, researchers from DTU Wind Energy, ECN, FORWIND, CENER and SINTEF have developed a taxonomy of the topics in the wind energy research area and standard metadata to map and structure the wind energy research data. The taxonomy and metadata will subsequently be used to create a so called metadata search portal that will allow researchers to easily find and access relevant research data.

The full work and reports will be available at the beginning of 2018, but we want to already now share with the WE Community the taxonomy of the wind energy topics.

This poster shows the hierarchical structure of the topics in the Wind Energy Research area]

You can learn more by participating to the Slide event "IRPWind Open data and data management" Room: 1 (1.08) Monday at 12:00.

## Economics

Project Finance

LCOE models

Support schemes

Market models

Business models

## Wind power plant

Wind farm

Wind Farm Control

Ancillary Services

Grid Connection

WIND PHYSICS

Wakes

Array Cables

Offshore Substation

Transmission System

## Operation & Maintenance

Short-term prediction

Health & Safety

Maintenance Scheduling

Decommissioning

Re-certification

End-of-life extension

Recycling

Revamping

Repowering

## Wind Turbine

Rotor

Hub

Pitch

Blades

Concept design

Horizontal axis

Vertical axis

Aerial

Gearbox

Generator

Power electronics

Nacelle

Turbine Control

Yaw

Main shaft

Cooling

Tower

Tubular

Lattice

Support structure

Foundation

Mooring lines

Substructure

## Siting

Wind Mapping

Wind Atlases

Long-term corrections

Resources assessment

Design conditions

Infrastructures

Spatial planning

Wind Indices

Shear

Turbulence

Extreme wind

Flow angle

Legal aspects

Environmental Impact

Noise Perception

Nature impacts

Social acceptance

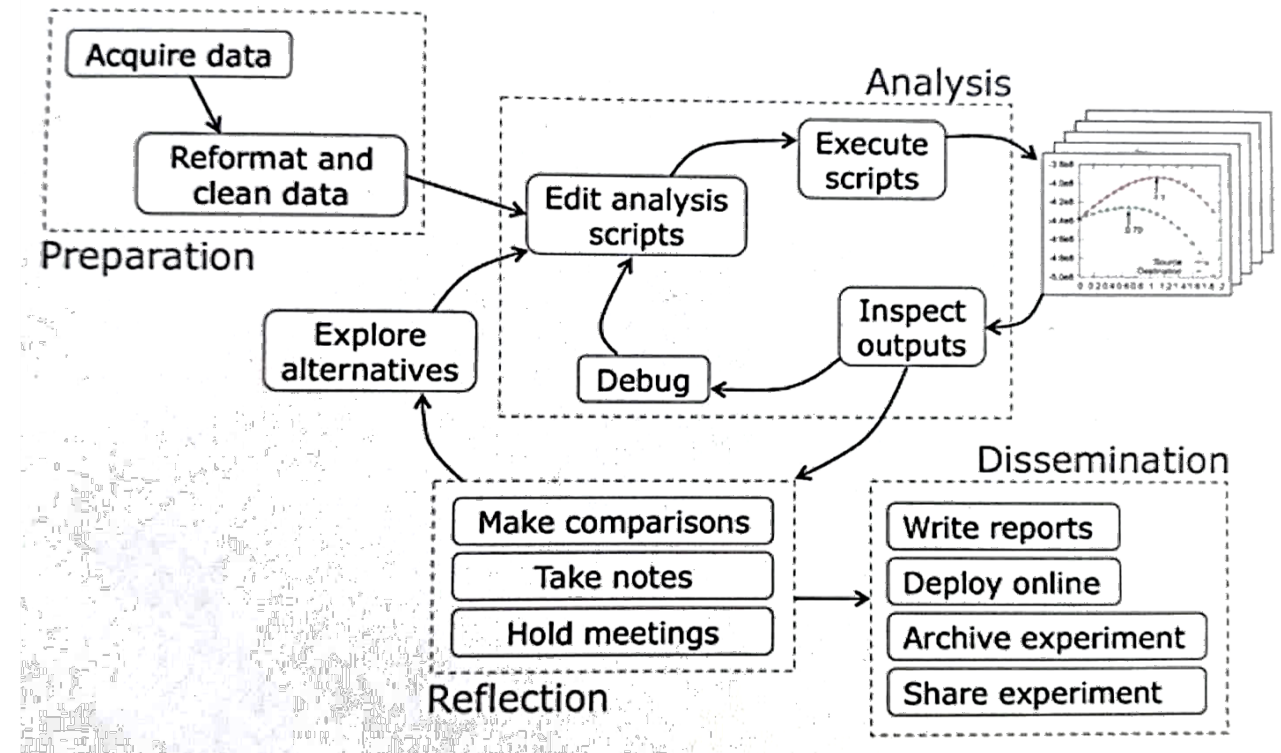
# Creating a taxonomy

## Expert elicitation

- 8 Experts from DTU were engaged in the task of creating taxonomy for WE topics
- Results were submitted and discussed to the IRPWind core group from Forwind, CENER, ECN, SINTEF


# Metadata & Taxonomy: cataloguing workflow components

- Generic work flow can be specialized and automatized using element taxonomies



# IRPWind – EERA JPWind Energy. [www.ShareWind.eu](http://www.ShareWind.eu)

## Data registry and metadata catalogue



Q SEARCH

HELP COMMUNITIES UPLOAD CONTACT PROJECTS

Login

Store and publish your research data

Search in public datasets or register as a user to upload and publish your data!

[Login](#) or [Register](#)

Create Record

Create a new record

Latest Records

High-resolution full-scale wind field measurements of the ECN's 2.5 MW aerodynamic research wind turbine using DTU's 3D WindScanner and SpinnerLidar for IRPWind's and EERA's benchmark (ScanFlow)

9 Dec 2019 by Wagenaar, Jan Willem; Hasager, Charlotte Bay; Alting, Ingmar; Mikkelsen, Torben; Angelou, Nikolas; Sjöholm, Mikael; Pena Diaz, Alfredo ;


In the framework of the IRPWind 1st call for joint experiments DTU and ECN have executed the ScanFlow project. The aim of the project was to establish a unique turbine power performance and induction

Lidars for Wind Tunnels (L4WT) and Lidar Data for Use (LD4U) - Joint experiment and open data set within the IRPWind project

9 Dec 2019 by Sjöholm, Mikael; Vignaroli, Andrea; Angelou, Nikolas;

The IRPWind joint experiment project Lidars for Wind Tunnels (L4WT) was a collaboration between DTU Wind Energy in Denmark, VTT in Finland, NTNU in Norway, and Sintef in Norway, where coherent continu

More Records ...

 Project funded by EERA JP Wind.

HTTP API v2.1.2



# Results – taxonomies

EERA Wind Energy Metadata

Activity	<input type="text"/>		<input type="button" value="Add"/>
Link to the data	<input type="text"/>		
External conditions	<input type="text"/>		<input type="button" value="Add"/>
Geo Locations	<input type="text"/>		
Model	<input type="text"/>		<input type="button" value="Add"/>
Sensors	Other sensor	<input type="text"/>	
	Sensor	<input type="text"/>	<input type="button" value="Add"/>
Sensor supports	List of sensors used for data collection		
	Other support	<input type="text"/>	
	Sensor support	<input type="text"/>	<input type="button" value="Add"/>
Variables	Other variable	<input type="text"/>	
	Variable	<input type="text"/>	<input type="button" value="Add"/>

☐ Submit draft for publication

When the draft is published it will be assigned a PID, making it publicly citable. But a published record's files can no longer be modified by its owner.

The draft is up to date



# Conclusive remarks

**A web data portal** with a data catalog has a two-fold purpose

- To connect safely data owners to users
- Inform on the availability of shared resources NOT necessary direct access

## Data owner /creator

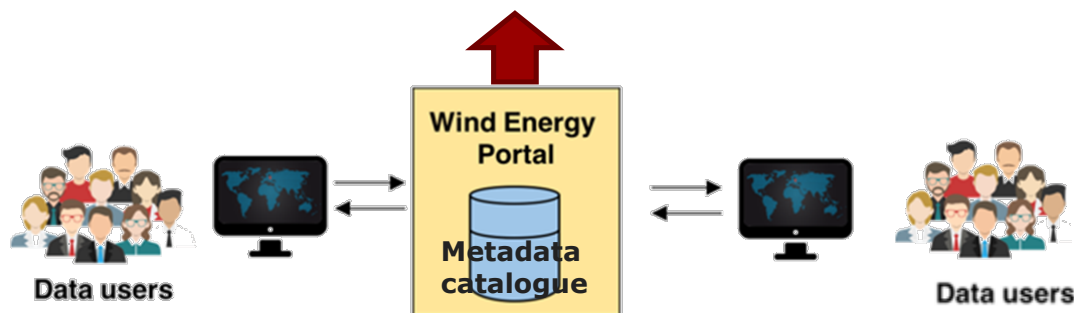
- Can make visible data via metadata
  - without uploading any data, and
  - maintain control on data access
- Multiply value of data

**Data= asset**  
**Market Place?**

€ £ \$?

**Services?**

**Co-creation?**



## Data user

- Can find data accurately by searching the same terms used by the data owner
- Can retrieve information on available data
- Efficient path to results

# Current work: taping into machine-actionability

*Nikola Vasiljević*

- Made taxonomies machine-actionable using FAIR Data Collective tools and workflows e.g., <http://purl.org/neat>
- Creating and exposing machine-actionable metadata templates and theirs instances <https://www.youtube.com/watch?v=60ZRKeUe9D4>
- Harvesting machine-actionable metadata for next generation share-wind portal
- Developing share-wind v2.0, <https://www.share-wind.com> (*in production from Oct 28<sup>th</sup> 2021*)
- Created rapid M4M workshops, which have been successful beyond wind energy domain



Thank you for your attention!