

DTU



25/08/2020

Mini-Symposium: MaRINET2 Round Robin Testing Results

WESC 2021

Nikola Vasiljevic

# Accessing data and FAIR data principles

# Disclaimer

- *This presentation contains personal opinions which may not reflect the DTU Wind Energy opinions*

# Data access

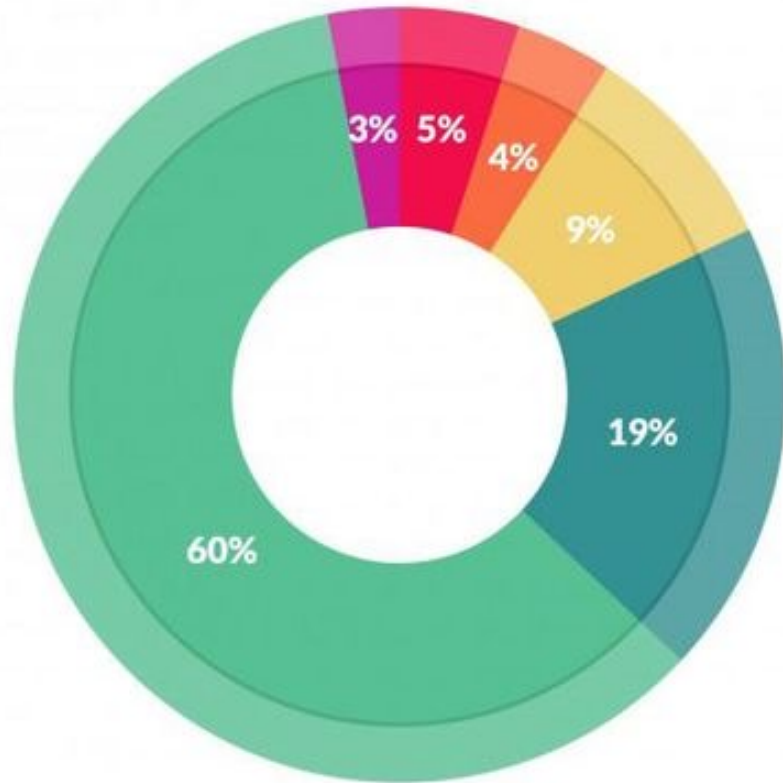
- Only one dataset is published (Boulogne-Sur-Mer):  
<https://www.seanoe.org/data/00471/58265/>
- For other datasets contact data owners and request data

# Why more data are not readily available?

It is rather hard to 'sell' idea of FAIR principles and good Data Management, since data owners find work of producing FAIR data:

- **hard**, as they need to learn new skills,
- **difficult**, there is lack of tooling and training
- **underfunded**, funded projects often do not allocate enough resources for data management, funding agencies do not provide additional funding to deal with data management,
- **underappreciated**, DOI isn't yet a sufficient currency
- **uninteresting**, 80% of people find this activity boring

# Uninteresting

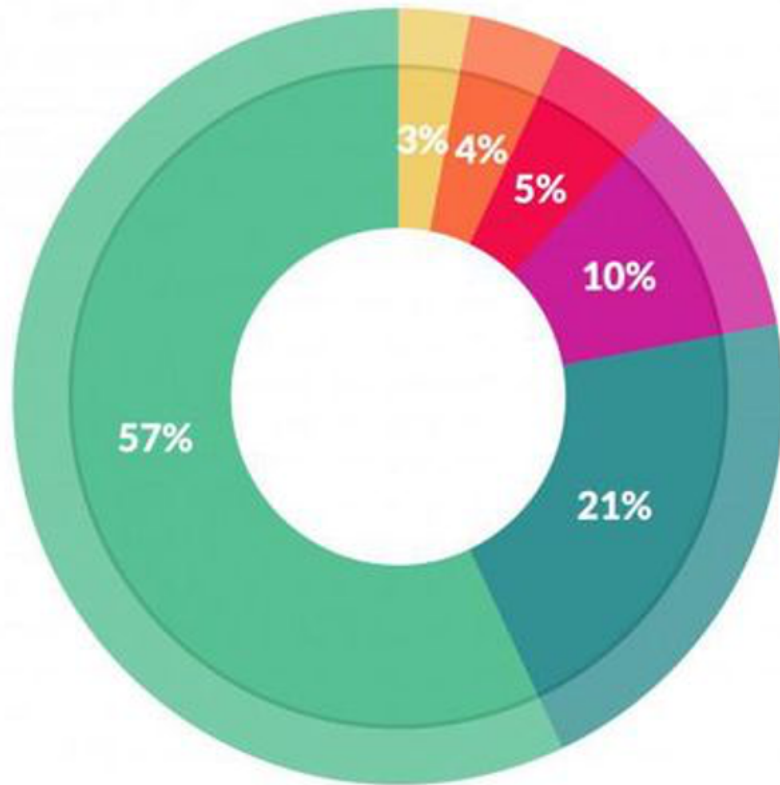


## What data scientists spend the most time doing

- 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](#)

# Uninteresting



## 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](#)

# Underappreciated

- At best one can get DOI for data they publish
- As we know, DOI does not improve researchers' h-Index.
- In the current academic setting and funding landscape, h-Index is the prime KPI for advancing researchers' careers (**publish or perish**)
- There is a growing opinion that researchers should not deal with producing FAIR data. This activity should be handled by data engineers that ideally have transitioned from science to technology.

# Underfunded

The screenshot shows a web browser displaying a Nature article. The browser's address bar shows the URL: <https://www.nature.com/articles/d41586-020-00505-7>. The page header includes the Nature logo, navigation links like 'View all journals', 'Search', and 'Login', and a secondary navigation bar with 'Explore content', 'Journal information', 'Publish with us', 'Subscribe', 'Sign up for alerts', and 'RSS feed'. The article breadcrumb is 'nature > world view > article'. The main heading is 'WORLD VIEW · 25 FEBRUARY 2020' followed by the title 'Invest 5% of research funds in ensuring data are reusable', where the first part is highlighted with a red box. Below the title is a quote by Barend Mons: 'It is irresponsible to support research but not data stewardship, says Barend Mons.', with the first part of the quote highlighted in red. A profile picture of Barend Mons and a contact icon are shown below the quote. Social media sharing icons for Twitter, Facebook, and Email are present. The article's first sentence is: 'Many of the world's hardest problems can be tackled only with data-intensive, computer-assisted research. And I'd speculate that the vast majority of research data are never published. Huge sums of taxpayer funds go to waste because such data cannot be reused.' A blue box at the bottom right of the article states: 'You have full access to this article via DTU Library'.

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons

Many of the world's hardest problems can be tackled only with data-intensive, computer-assisted research. And I'd speculate that the vast majority of research data are never published. Huge sums of taxpayer funds go to waste because such data cannot be reused.

You have full access to this article via DTU Library

## However in Marinet2 we had WP6 which...

- WP6 offered and provided to data owners:
  - Training
  - Tooling
  - Help desk

# Training, Tooling and Help desk

- Organized **Bring Your Own Data** (BYOD) Workshop
- Workshop took place on 23rd June 2020
- The workshop focus was to train Marinet2 Round robin data owners to convert their data to Marinet2 metadata and data standard
- Slides and video recordings available at:  
<https://github.com/Marinet2/byod-workshop-2020>
- After the workshop we established Slack help desk where data owners could get support in the data conversion process
- **2** data owners used the help desk

# Conclusions

- Everything starts and ends with measurement data
- Everyone needs data
- Modern AI revolution started with ImageNet, the canonical image-recognition open database
- A carrot did not work, but the whip will

*Fund data stewardship and impose penalties for projects that do not comply to promises*

# Thank you