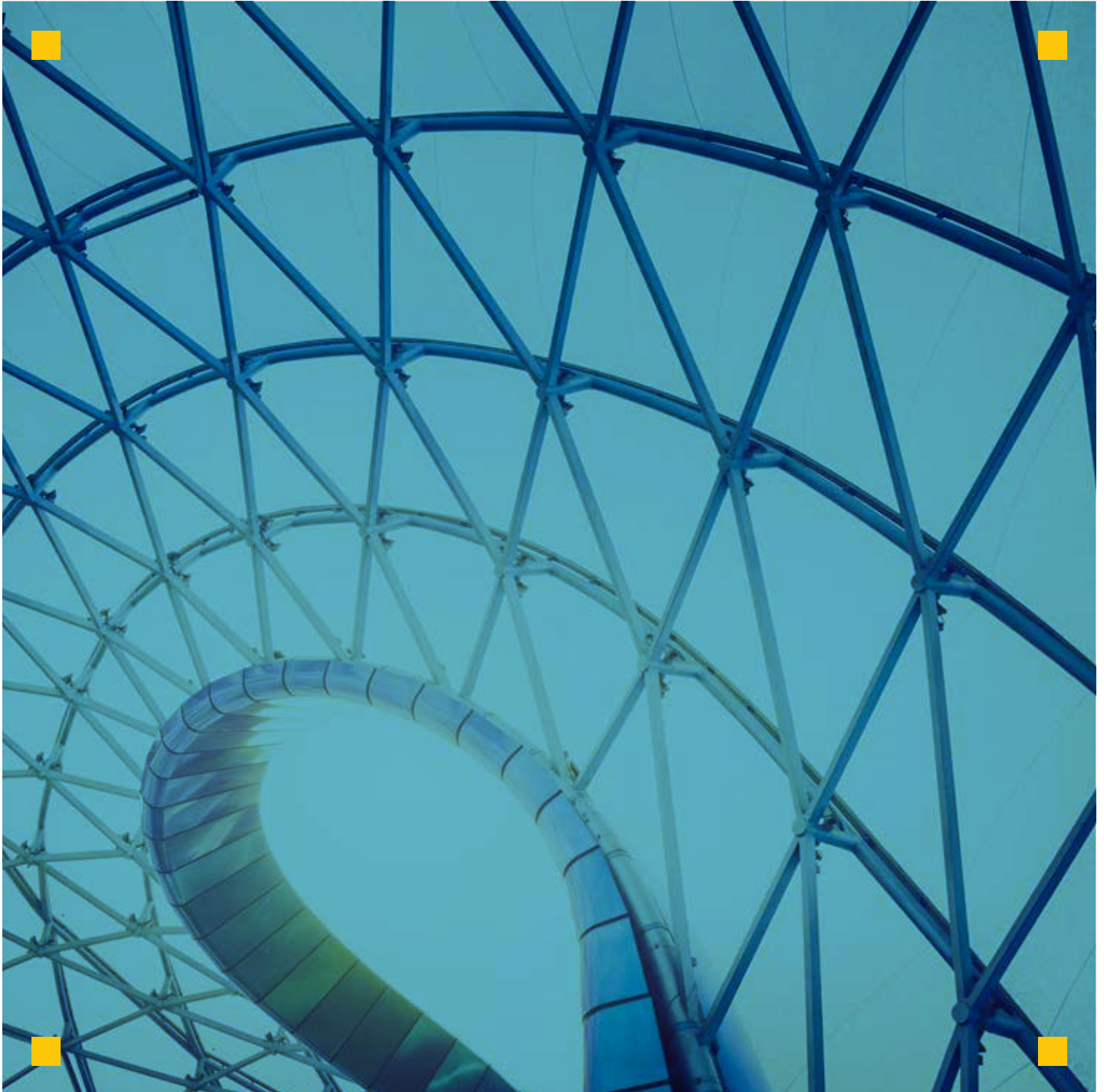


IMPLEMENTATION

STORIES



FAIRSFair
Fostering Fair Data Practices in Europe



THEME
5

**DEFINING DATA
INTEROPERABILITY FRAMEWORKS**

IMPLEMENTATION STORIES

THEME **5** DEFINING DATA
INTEROPERABILITY FRAMEWORKS

Making metadata FAIR for climate research

Authors
Josefine Nordling (CSC) and Angus Whyte (Digital Curation Centre)

From an interview with
Rene Belsø, Anders Sparre Conrad, and Hannah Mihai (DeiC)



Introduction

Implementing the 'I' in the FAIR principles means applying data and metadata standards that are appropriate to the specific research domain. There are many disciplinary standards available, but what if researchers are working in a relatively new or cross-disciplinary domain and wish to share existing data, or do not have a standard to properly describe the data they are working with? In the summer of 2020, this was the problem facing two Danish research groups assisted by the Danish e-Infrastructure Cooperation (DeiC).

Danish researchers in the Wind Energy domain together with the Danish node of the European research infrastructure AnaEE (Infrastructure for Analysis and Experimentation on Ecosystems) were assisted by metadata experts from GO-FAIR, and Stanford University's Center for Expanded Data Annotation and Retrieval (CEDAR). DeiC set out to address the metadata challenge through organising and financing a series of 'Metadata 4 machines' (M4M) workshops.

After several trial-runs to refine the workshop format, by early 2021 the outcomes included FAIRification roadmaps for DTU Wind Energy and AnaEE and a template for making FAIR metadata, tailored to the researchers needs.

FAIRsFAIR recommendation

"Describe research outputs using agreed terminologies and metadata standards to make data FAIR"

[FAIRsFAIR Recommendations on practice to support FAIR principles](#)

■ Initiating the M4M Workshops

During a national event on research data management held in early 2020, a need was identified for engaging in more practical collaborations with researchers in order to avoid the tendency for circular discussions. "This was the key moment where we felt the need to reach out more to research communities," explains Anders Conrad. "We sent out enquiries to various research communities regarding the need for practical hands-on support measures shortly after this event, and the ESFRI infrastructure 'The Infrastructure for Analysis and Experimentation on Ecosystems' (AnaEE) was the first community to respond."

The DeIC team decided to reach out to GO-FAIR to discuss organising Metadata 4 Machines (M4M) events for the Danish research communities. The Danish AnaEE team, coordinated by Klaus Steenberg Larsen from the University of Copenhagen's Department of Geosciences and Natural Resource Management, was keen on participating in such an event and there was funding available to work with them. GO-FAIR became interested in setting the wheels in motion, with support from the CEDAR group. An additional factor motivating DeIC's engagement was a green light from their Steering Committee.

A five-day data stewardship training event, in three modules over a few months was initiated and co-sponsored by DeIC in Nov 2019. At this virtual event, Nikola Vasiljevic, digitisation consultant in the Danish Technical University department of Wind Energy (one of the world's largest) raised the idea of doing a series of M4M workshops. DeIC engaged with the idea and provided funding. This gave the green light to the M4M workshops to further engage with the relevant research communities.

The screenshot shows the 'How to FAIR' website interface. At the top, there are navigation links: 'What is FAIR', 'Why FAIR', 'How to FAIR' (with a dropdown arrow), 'About', and a 'Quiz' button. The main content area features a quiz question: 'What can you do if there is no standard in your field?'. Below the question, there is explanatory text about metadata standards and a specific example from Nikola Vasiljević's work on wind energy taxonomy. To the right of the text is a hierarchical diagram titled 'Wind Energy Taxonomies' which branches into categories like 'Topics', 'Activities', 'External Conditions', 'Models', 'Data Categories', and 'Materials', each with further sub-items.

The DeIC How to FAIR site explains the context¹

1. Extract from How to FAIR: <https://howtofair.dk/how-to-fair/metadata/>

IMPLEMENTATION STORIES

THEME 5 DEFINING DATA
INTEROPERABILITY FRAMEWORKS

Ingredients of success when organising an M4M workshop

The first ingredient was that both groups had done some prior work on metadata. AnaEE Denmark had already taken some steps to create some simple meteorological metadata. This metadata was not based on publicly declared metadata definitions from a schema or linked open data vocabulary, so it was not really machine actionable, but the AnaEE researchers had an appetite for more, and decided to actively participate in a further series of M4M workshops. The approach aimed for a simple, straightforward and manageable baseline for working with meteorological data, to lower the barriers to working with machine actionable metadata and to find a new way to meet common goals for working with it. Meanwhile the DTU team had been engaging with FAIR since 2017. They started by developing taxonomies for wind energy topics, activities, study conditions and models; to tag their datasets they drafted a metadata application profile based on Dublin Core.

The second ingredient was the complementary strengths of the research teams. They established a shared set of goals recognising the added value that inter-connected and richly described data resources would offer their research agenda. This came about through bringing together the DTU Wind Energy team's understanding of the technical requirements for automating the processing of metadata, and the AnaEE team's understanding of international teamwork on metadata.

The third ingredient was the availability of funding that could be put toward the workshops. This was essential to cover the participation of those involved with highly specialised Data Stewardship competences, in what was a short but very intensive exercise, requiring five days of effort from each of the research participants, together with that expert facilitation of metadata experts from CEDAR and GO-FAIR.

■ Co-creating the Rapid M4M process

Investigating current practice and challenges to FAIRness

One of the first challenges arose before the M4M workshops, in establishing consensus on the preconditions – i.e., efforts to describe data in the communities, and their view of the challenges ahead. Among the Wind Energy researchers there was a realisation that, even among three Danish universities, measurements were being made according to different definitions of variables. Some steps had been taken to agree on common terms, using one vocabulary from the ICOS research Infrastructure, and another produced by AnaEE France. With further steps towards integration of AnaEE, the research teams were finding they could no longer live with the discrepancies in the terms they were using in their metadata. These first steps to investigate current practice were instrumental to getting the research teams to agree to take part in the M4M workshops.

Getting shared understanding of expected workshop outcomes

Community engagement in the FAIRification process provided by the M4M workshops, was perhaps the biggest challenge DeIC encountered, and managed to overcome by playing to the complementary strengths of the two research communities, and by some judicious re-working of the M4M workshop format. Reflecting on the engagement aspect, Rene Belsø explained that the quality and credibility of the products of such initiatives is doubtful, if the research groups that are intended to benefit are not fully involved. Getting their motivation to take part in the M4M events was a real concern, as he recalls.

'Generally, many research groups have one respected data 'nerd', an expert whose time is dedicated to metadata work, but in the M4M workshops we managed to get the whole research team to contribute.'

Rene Belsø

What made this possible was the effort put in aligning expectations between these two research teams. That in turn was only possible as they were involved from the start. It was not initially obvious that both communities (AnaEE and DTU Wind Energy) should be involved in the same M4M events. Each community had their own expectations about the technical outcomes they wanted, and differed in how precise these expectations were. For the DeIC team, it was critical to the overall success that mutual understanding about expectations emerged early in the workshop.

A Hands-on approach

The hands-on approach adopted in the M4M workshops was also key to maintaining the research teams' involvement. The workshop format was changed to enhance that focus, and through various iterations the length was reduced from 5 half-days spread across weeks, down to the 2-day lunch-to-lunch 'Rapid M4M' format.

This reduction in event length was achieved by stripping out the conceptual or theoretical aspects of FAIR and providing these as pre-workshop reading assignments. DeIC's Hannah Mihai worked with materials (these are available from the DeIC website)⁵. "In the first workshop series the planned technical outcomes were explained in a very theoretical way, which was not very productive," recalls Rene Belsø.

In the revised format that was agreed with GO-FAIR and CEDAR, the technical aspects were made less discursive and more accessible. This was achieved by having the participants do some home assignments ahead of the workshops, so tooling and larger discussions around reasoning were thus avoided during the events. Video guides were also produced for the exercises, which concisely explained the rationale and building blocks of the FAIRification process, without the workshop participants having to necessarily understand all the theory behind it. It was key to have everything prepared ahead of the workshops in order to make it easier to get to the point of discussion of the scientific concepts. All user accounts were pre-set up, along with documentation.

GO-FAIR's Erik Schultes and CEDAR's John Graybeal led the participants through the steps needed to use the various tools that were applied. "Erik and John were able to explain in simple terms with concrete examples and hands-on exercises about what can happen with the data," says co-organiser Anders Sparre Conrad. "There was an 'ah-ha' moment when the researchers could really bring the discussions back to their level and realise the value they could get from this," adds Rene.

Everyone knew what data they were going to deal with, as they had pre-selected the datasets for the workshops. They deliberately chose common meteorological data to use. This helped to focus on the challenges in thinking about standards and terminologies, and working with them. Once the pieces were put together, there was excitement about being able to deposit the metadata template into a community resource – Biportal – whose life-science origins were no restriction to the research participants once they realised its more general value as a research tool. Nevertheless the precarious funding of Biportal during the period of the workshops was an issue, reflecting the need for sustained resources of this critical infrastructure.

■ Outcomes and Impacts

For the two research communities involved, the workshop process had tangible immediate outcomes, as well as a path for improving FAIRness in the longer term, and a shared understanding of what this would mean for doing science.

For AnaEE Denmark the main outcomes were the shared learning about the FAIRification process, using the CEDAR toolkit to create a template to produce three 'FAIRified' meteorological datasets, and to produce machine-readable and reusable metadata. AnaEE conducts experimental studies in research stations, and needs to be able to publish the observations collected over many years. Getting more reuse across communities is fundamental to the climate research advances they seek. So AnaEE sees it as part of their scientific endeavour to facilitate metadata automation and especially important in the context of ENVRI, the cluster of environmental research infrastructures.

IMPLEMENTATION STORIES

THEME 5
DEFINING DATA
INTEROPERABILITY FRAMEWORKS

The wind energy researchers had a similarly broad aim, to produce interconnected and richly described resources. This was exemplified in the workshop focus on a shared use case – the essential need to help people find locations of wind turbines. Working with long data series, they saw the value of automating the production of analytics derived from this location metadata, for use in the modeling and simulation studies they perform. The workshops were instrumental in extending their application profile to make this metadata machine-actionable, resulting in a generic Dataset Metadata Template, accessible in CEDAR Workbench.

Equipped with a clearer vision of how to make FAIRification work for them, both groups decided to create roadmaps setting out their approaches and strategies.

Next steps for DeIC and key takeaways

This workshop series was a step towards competence building for DeIC. Among the likely next steps are a train-the-trainer program, including some technical training on elements of the rapid M4M format. This would address the resources bottleneck caused by lack of librarians skilled in providing the technical facilitation. Key takeaway points include:

- Work with research groups to facilitate bottom-up approaches, building and testing solutions that fit their science by reusing and adapting what 'front runners' have done.
- Making the workshop process simple and practical is key, especially to streamline the 'tooling'. For the RapidM4M this was reduced to a three-step process.
- Research groups do not have equal opportunities to develop tailored solutions to their needs. Building collaborations between unequal groups can address the dilemma that mid-size research groups usually have only one person who is really keen on driving this forward. In bigger research groups connected to a Research Infrastructure this expertise may already be available. The solution is therefore to collaborate across scales, and do so internationally.
- Funders could help by incentivising and rewarding researchers for effort on metadata framework definitions. DeIC expects Horizon Europe to be a major driver for change in the future, and in Denmark the national strategy on research data management steers the direction.

■ Further information

AnaEE FAIRification Roadmap:

<https://www.deic.dk/sites/default/files/documents/PDF/AnaEE%20DK%20FAIR%20data%20roadmap%20201009.pdf>

DTU NEST FAIRification Roadmap:

https://www.deic.dk/sites/default/files/documents/PDF/DTU_NEST_FAIRification_RoadMap_old.pdf

GO-FAIR 'Metadata for Machines':

<https://www.go-fair.org/how-to-go-fair/metadata-for-machines/>

CEDAR Center for Expanded Data Annotation and Retrieval:

<https://more.metadatacenter.org/>

■ *About FAIRsFAIR Implementation Stories*

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report [Turning FAIR into Reality](#) and the [FAIRsFAIR Recommendations on practice to support FAIR principles](#). FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

■ *FAIRsFAIR - Fostering FAIR Data Practices in Europe*

European Commission Grant Agreement No. 831558
H2020-INFRAEOSC-2018-4
www.fairsfair.eu - support@fairsfair.eu

■ *Acknowledgements*

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

■ *Disclaimer*

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

■ *Copyright rests with the authors*



This work is released under a Creative Commons Attribution License, version 4.0.
For details please see <https://creativecommons.org/licenses/by/4.0/>

JOIN OUR COMMUNITY!



@FAIRsFAIR_eu



/company/fairsfair

www.fairsfair.eu



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.