Best Practice Recommendations for improving FAIR data maturity in wind energy

DOI: -

Authors: Sarah Barber[1], Sirko Schindler[2], Catherine Jones[3], Pablo Rodríguez-Sánchez[4], Yuriy Marykovskiy[1], Wind Energy Community Standards WG

[1] Eastern Switzerland University of Applied Sciences, Switzerland
[2] Institute of Data Science, German Aerospace Center (DLR), Germany
[3] Science and Technology Facilities Council, UK Research and Innovation, UK
[4] Netherlands eScience Center, The Netherlands

Published: ?? ?? 2025

Version: 1.1.0

Language: English

RDA webpage: https://www.rd-alliance.org/groups/wind-energy-community-standards-wg/outputs/

Citation:
S. Barber, S. Schindler, C. Jones, P. Rodríguez-Sánchez, Y. Marykovskiy. (2025). Best Practice Recommendations for improving FAIR data maturity in Wind Energy. Research Data Alliance.

Table of Contents

Executive Summary

The RDA Wind Energy Community Standards Working Group has created these new best practice recommendations for improving FAIR data maturity in wind energy in practice. It addresses key challenges in the sector, which are to foster data exchange and re-use as well as to create more value from data by increasing the efficiency of data cleaning and analytics processes. The Working Group first reviewed and evaluated the current data sharing and publishing landscape in wind energy, finding that there is no widely recognised community ontology at the sector level, and alignment with upper-level ontologies is largely absent. A review of existing open wind energy data led us to the conclusion that the platform itself tends to dictate the FAIRness of a published dataset. It was therefore decided to analyse the potential FAIRness of a range of different platforms (divided into catalogues, repositories, knowledge hubs, organisation websites and dashboards) relevant to the wind energy sector. After shortlisting 15 relevant data sources covering these different types, we evaluated potential FAIRness assessment tools, and chose to work with the FAIR Data Maturity Model due to previous positive experience with it. However, the criteria were adapted slightly for our needs, including reducing the evaluation scale to a Boolean "1/0" choice, rather than using the suggested scale. We also added some specific criteria, including the capability to link to defined wind energy specific terms, to access a preview plot of the data and to access a preview of the file structure. The results showed that the repositories all score similarly, mostly between 74% and 83%, the knowledge hubs also score slightly lower (69-71%), with the company websites (43% and 0%) and the dashboards (34-66%) the lowest. A more detailed analysis of the different criteria allowed us to define recommendations for wind energy data publishing platform developers for wind energy data publishers and for the general wind energy community. Finally, the project allowed us to define some lessons learned, both in terms of community building and interdisciplinary collaboration, as well as in the application of the FAIR Data Maturity Model.

1 Introduction

Digitalisation is one of the five "megatrends" of wind energy technology[1], and one of the main challenges of fostering digitalisation in wind energy is to increase the FAIRness of data [5]. Therefore, the RDA Wind Energy Community Standards Working Group was formed in order to create best practice recommendations for improving FAIR data maturity in wind energy. Ultimately, this will reduce data management overhead within and between organisations in the sector. The goal was achieved by first reviewing and evaluating the current data sharing and publishing landscape in wind energy, based on a recent review paper about knowledge engineering in wind energy [9]. Next, rather than assessing the FAIRness of specific published data sets, we decided to assess the potential FAIRness of different platforms used to publish wind energy data. This process was thought to be more flexible to enable the creation of recommendations for a wide range of stakeholders. We chose a range of diverse data publishing platforms relevant to wind energy through an analysis of 41 different options. We classified them based on the type of ownership, whether F-UJI (an automated tool for assessing data FAIRness, see Subsection 2.3) identified them as a data source, and whether it is indexed by Google Dataset Search, and then chose 15 platforms for detailed analysis. In the next step, we evaluated the use of different FAIR data assessment tools for our purpose, and chose the FAIR Data Maturity Model, with some small amendments. Finally, we applied the chosen tool and analysed the results, allowing us to define a set of recommendations for data platform developers, data publishers, and the general wind energy community. The approach is described in Section 2, the results in Section 3, the resulting recommendations in Section 4 and the lessons learned in Section 5.

2 Approach

In this section, the following four steps are described:

1. Review and evaluation of the current data sharing and publishing landscape
2. Choice of data publishing platforms
3. Choice and application of analysis tool
4. Creation of recommendations

2.1 Review and evaluation of the current data sharing and publishing landscape

This step involved first reviewing and evaluating existing semantic artefacts and data publishing standards, and then reviewing existing data publishing platforms, as described below.

2.1.1 Existing semantic artefacts and data publishing standards

Domain-specific semantic artefacts (also referred to as data models, conceptual models, ontologies, taxonomies, controlled vocabularies, or schemas) are essential for publishing structured (meta)data on the Web, as discussed in RDA's "Guidelines for publishing structured metadata on the web" [12]. A systematic review of semantic artefacts in the wind energy domain was carried out as described by Marykovskiy et al. [9] (Section 5). This analysis showed that the community-driven development and adoption of semantic artefacts in the wind energy sector has been limited up until now. There is no

[1] https://etipwind.eu/files/reports/Flagship/fit-for-55/ETIPWind-Infographic-Mega-Trends.pdf

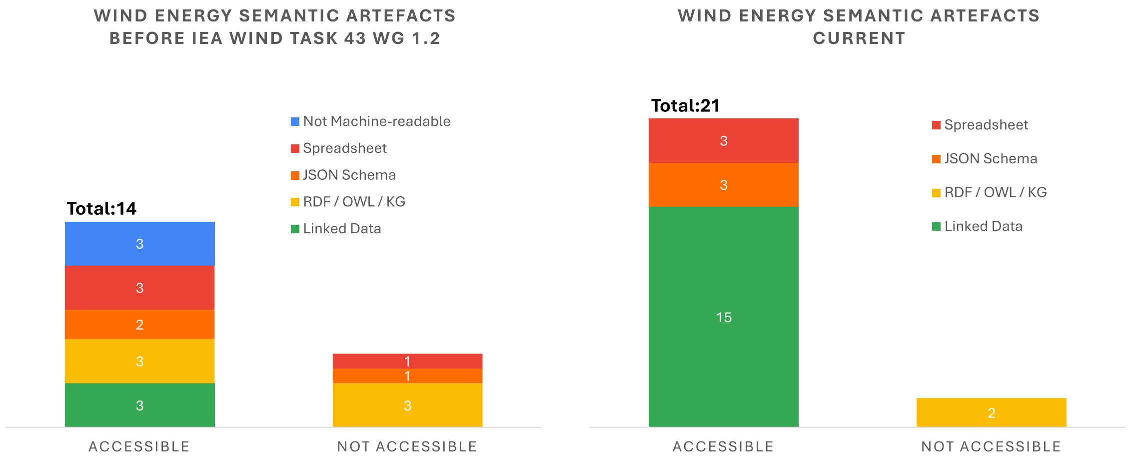

Figure 1: Total number of semantic artefacts grouped by format and accessibility before (left) and after (right) IEA Wind Task 43 WG1.2.

widely recognised community ontology at the sector level, and alignment with upper-level ontologies is largely absent, as is reuse of semantic artefacts from neighbouring domains such as meteorology or environmental monitoring (even where clear synergies exist). Additionally, it was found that most existing efforts have focused on modelling concepts related to the operational stage of the wind farm life cycle, particularly on failure and reliability analysis. No semantic artefacts were identified for later stages such as end-of-life management. Finally, most existing semantic artefacts take the form of taxonomies or vocabularies that have not been formalised in standard modelling languages, leaving many data-related activities dependent on manual processing (see Figure 1 (left)). As a result, IEA Wind Task 43 on Digitalisation in Wind Energy formed a working group (WG1.2) on the WeDoWind platform to formalise and improve accessibility of existing semantic artefacts, as well as develop new semantic artefacts. While the situation has improved (see Figure 1 (right)), a need for greater harmonisation and sustained community adoption of semantic practices in wind energy still remains. This is evidenced by the fact that TechnoPortal[2], an ontology hosting portal for engineering and technology domains, currently hosts less than fifty ontologies (15 of them pertaining specifically to the wind energy) with only a couple thousand mappings between concepts. At the same time, its equivalent in the biomedical domain – Bioportal[3] – hosts more than a thousand ontologies with more than a hundred thousand mappings between individual concepts, which enables data annotation with rich metadata.

2.1.2 Data sources on the web

Data is published on the web in a variety of diverse ways including using specialist platforms. Very often the platform itself dictates the FAIRness of a published dataset by providing (or lacking) a set of functionalities to the user. In this work, we consider the following five types of platform:

Data catalogue (also: Open data platform, Metadata catalogue, Data discovery platform) is an

[2] https://technoportal.hevs.ch/
[3] https://bioportal.bioontology.org/

organised inventory of datasets that provides metadata (such as descriptions, formats, sources, licensing, and access policies) to help users discover and understand available data. It does not store the actual datasets but acts as a directory that points to where the data can be accessed, often linking to multiple repositories or storage systems. Data catalogues are commonly used by governments, research institutions, and organisations to improve data discoverability. Data catalogues often index data repositories automatically, but unlike data repositories, pure data catalogues do not offer the possibility to archive the data to the users. Sometimes there is a possibility to manually create a pure metadata record, by adding an entry to the catalogue. Most notable examples of catalogues that auto-index many major repositories are Google Dataset Search[4] (Global) and OpenAIRE[5] (Europe).

Data repository (also: Digital archive, Data archive, Research data repository, Institutional repository) is a digital storage system that holds, manages, and preserves datasets for long-term access and reuse. Unlike a data catalogue, a repository stores the actual data using data hosting services, ensuring it is curated, versioned, and accessible under specific policies. Generally, a user can submit their data for storage in line with the repository policies. Some repositories are similar to data catalogues as they can offer a possibility to create a pure metadata record, and link to data that is stored on a different repository, or with an external data hosting service. One of the most notable examples of a data repository highly adopted by academia is Zenodo[6]. Many universities manage their own repositories, which usually host publications and other digital information in addition to datasets.

Data dashboard (also: Web application, Visualisation tool, Analytics dashboard) is an interactive web-based tool that presents data through charts, tables, and basic analyses to help users explore patterns and monitor indicators. Dashboards typically aggregate data from one or more sources and provide features such as filtering, drill-downs, and alerts. Unlike a data repository or catalogue, a dashboard usually does not act as the authoritative host for datasets nor provide complete archival services; instead, it consumes data and focuses on presentation and decision support. Many dashboards allow users to download or export the data shown, link to underlying sources or APIs, and document update frequencies, though the level of metadata varies widely. Access can range from fully open to role-restricted, and update schedules may be real-time or periodic. Notable examples include Statista[7] and World Bank Open Data[8].

Knowledge hub (also: Knowledge platform, Knowledge portal, Virtual research environment) is an integrated, often domain-specific platform that brings together datasets (or links to them), documentation, analysis code and tools, dashboards, learning materials, and publications in one place. Its aim is to support end-to-end knowledge workflows often with guidance on standards, policies, and best practices. Unlike a pure data catalogue, a Knowledge Hub provides rich context, narratives, and community features (e.g., how-tos, forums, or curated collections), and may host some assets locally while indexing others in external repositories. Submission and governance can be open or curated, with varying levels of quality control, metadata requirements, and versioning. Hubs frequently promote persistent identifiers and citation, and may offer APIs or notebooks for reproducible research. Notable examples include NASA/ESA

[4] https://datasetsearch.research.google.com/
[5] https://www.openaire.eu/
[6] https://zenodo.org/
[7] https://www.statista.com/
[8] https://data.worldbank.org/

Earth data portals[9] and Open Energy Data Initiative[10] (OEDI).

Organisation/Company website (also: Corporate/Institutional portal) is a general-purpose web presence used to communicate with stakeholders via pages, posts, documentation, reports, and announcements, and may occasionally provide files or data for download. Unlike catalogues, repositories, or knowledge hubs, these websites typically do not offer structured metadata, persistent identifiers, standardised licensing, or long-term preservation guarantees for datasets. Content can change without version histories, links may move, and access may be gated or subject to terms of use that are not data-reuse licences. Organisation websites often link out to the organisation's catalogue, repository, or knowledge hub where formal data management occurs. Examples include ministerial or agency portals, NGO project sites, and corporate websites of energy companies such as ORSTED[11], EDP Group[12], or Vattenfall[13].

It should be noted that the distinction introduced above is not strict, as there are platforms that can act primarily as catalogues from a user perspective and, at the same time, choose to self host some datasets, thus acting as repositories, making this distinction fuzzy. In this document, if a data catalogue is highly curated or self-hosts datasets in addition to external references, we refer to such a platform as a repository.

Some platforms may claim to be knowledge hubs, while the majority of the data remains isolated, making these platforms much more similar to data repositories. At the same time, some repositories, especially ones managed by academic institutions, do an excellent job connecting data to publications, code, and projects.

All the platforms listed above usually use some kind of data hosting service (also: Cloud Storage Service, Data Storage Infrastructure), which provides raw storage space for dataset files on their infrastructure, often in the cloud or on-premises These services do not necessarily provide structured metadata management or curation like repositories do, but they offer scalable, secure, and high-performance storage for large datasets. They are typically used by data repositories or by individual researchers, businesses, and organisations that need to store and process large volumes of data efficiently. For example, Zenodo repository[14] stores data with CERN's Data Centre. OEDI[15], IEEE Dataport[16], and many others repositories use Amazon S3 Cloud Object Storage[17]. Other popular services include Amazon Elastic File System[18], and Google Cloud Storage Buckets[19].

2.2 Choice of data publishing platform

Rather than assessing the FAIRness of specific published data sets, we decided to assess the potential FAIRness of different platforms and data sources used to publish wind energy data. This process was thought to be more flexible to enable the creation of recommendations for a wide range of stakeholders. For this, a list of the data sources commonly used for wind energy data was compiled manually. In addition to platform types defined in the previous section, we classified all the

[9] https://www.earthdata.nasa.gov/ and https://earth.esa.int/eogateway/catalog
[10] https://data.openei.org/
[11] https://orsted.com/en/what-we-do/renewable-energy-solutions/offshore-wind/offshore-wind-data
[12] https://www.edp.com/en/innovation/data
[13] https://group.vattenfall.com/our-operations/our-energy-sources/wind-power
[14] https://zenodo.org/
[15] https://data.openei.org/
[16] https://ieee-dataport.org/
[17] https://aws.amazon.com/s3/
[18] https://aws.amazon.com/efs/
[19] https://cloud.google.com/storage/docs/buckets

Table 1: List of data sources assessed in this work.

Platform	Domain	Type	Ownership / Management	Indexed by Google?	Parsed by F-UJI?
Zenodo	General	Repository	Intergovernmental	✓	✓
DTU Data	General	Repository	Academic Inst.	✓	✓
4TU.ResearchData	General	Repository	Academic Inst.	✓	✓
ETH Zurich Research Collection	General	Repository	Academic Inst.	✓	✓
Wind Data Hub	Wind Energy	Repository	Governmental	✓	✓
Marine Data Exchange	Offshore Industry	Repository	Company	✗	✗
ENERGYDATA.INFO	Energy	Repository	Intergovernmental	✓	✓
Data One	Earth and Environment	Repository	Company	✓	✓
WindLab Knowledge and Data Hub	Wind Energy	Knowledge hub	Intergovernmental	✓	✓
Open Energy Data Initiative	Energy	Knowledge hub	Governmental	✓	✓
Ørsted	Energy	Company website	Energy company	✗	✗
EDP*	Energy	Company website	Energy company	✗	✗
Scenario Viewer	Energy	Dashboard	Governmental	✗	✗
Renewable Energy Materials Properties Database (REMPD)	Energy	Dashboard	Governmental	✗	✗
Statista	General	Dashboard	Company	✓	✗

* Data and metadata taken down by the company during analysis, and is no longer available

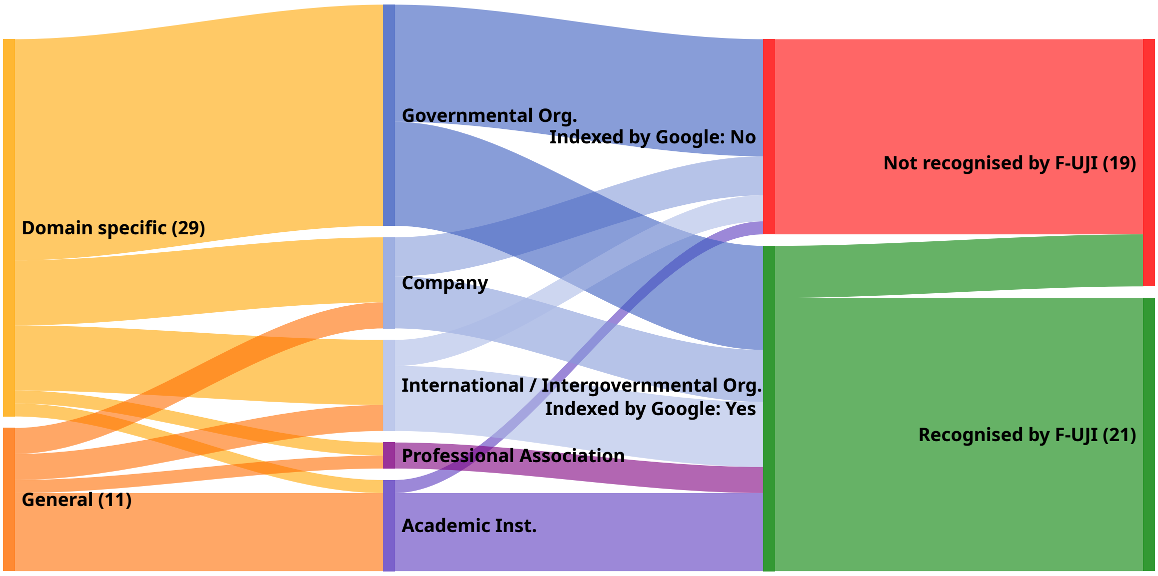

Figure 2: Sankey diagram of data sources grouped by management / ownership, domain, Google Dataset Search indexing, and possibility to be processed by F-UJI.

data sources based on the type of ownership/management such as: governmental organisation, academic institution, international/intergovernmental, and companies, including both for-profit and non-profit. Moreover, we identified whether the data source is general or domain specific, whether F-UJI (an automated tool for assessing data FAIRness, see Section 2.3) [6] identified the data source as such, and whether it is indexed by Google Dataset Search. We have not analysed pure data catalogues, such as national data catalogues, as the FAIRness of the data would be related to the original source/repository, unless the catalogue is highly curated and provides additional metadata, which is not available at the source.

We analysed a total of 40 sources with the criteria mentioned above and compiled the result in a tabular form in [2]. In Figure 2, the results are presented in a visual form of the Sankey diagram. As can be observed from the diagram, the majority of the data sources are domain specific, consisting mainly of energy and environmental sciences domains. These data sources are mostly managed by various governmental and intergovernmental organisations. Meanwhile, among data sources analysed which are not domain specific, most are managed by academic institutions. In terms of indexing by Google Dataset Search most sources had almost a 50/50 split with a notable exception for the ones managed by academic institutions. Moreover, datasets hosted on sources indexed by Google Dataset Search were also recognised by F-UJI automated fair assessment tool, with an exception of datasets indexed by Google Dataset Search via the Statista platform.

We then selected 15 data sources, maintaining the mix of different classification facets, as shown in Table 1. As mentioned before, we have not analysed pure data catalogues.

2.3 Choice and application of analysis tool

In order to decide which analysis tool to apply to this case, a choice of tools were compared and evaluated as shown in Table 2. The tools were taken from the FAIRassist full list of assessment tools[20].

Table 2: Assessment of analysis tools.

Name	Description	Evaluation
FAIR Data Maturity Model	Excel file to fill out with scores for different FAIR criteria	Some group members already familiar with it. Easy to use and compare results.
F-UJI Automated FAIR Data Assessment Tool	Tool analysing data sets with an URL or PID	One of the most extensive evaluations, full automation, and ability to export analysis results make it an excellent resource. However, as the automation relies on some basic interoperability being present, the tool didn't recognise datasets as such on almost half of the sources..May get different score if using DOI rather than platform identifier if the platform is not as machine readable than DataCite
ARDC FAIR data self-assessment tool	Web tool with a checklist for FAIR indicators	Similar to, but more difficult to compare results than the FAIR Data Maturity Model
FAIRsharing FAIR evaluator		Didn't work on all of the sources; Results of the evaluation are presented in a way that is difficult to interpret for non-experts.
FAIR Enough	Web tool for automatic evaluation of FAIRness	Didn't work even on the most popular sources like Zenodo.

Due to its ease of use, expected reliability of the results, and ease of comparison between platforms, we chose the FAIR Data Maturity Model [1]. As the FAIR Data Maturity Model was developed to assess the FAIRness of data itself, and not "potential FAIRness" of data published on a specific platform, in the best-case scenarios we introduced the following modifications:

- The criteria were interpreted to ask the question "Does the repository/platform have the option to ... in the context of wind energy?"
- "Metadata" was understood to refer to the information on the website, rather than to additional metadata files provided with the data set.

[20] https://fairassist.org/tools

- The criteria were assessed on a Boolean "1" and "0" scale (where "1" meant "yes" and "0" meant "no"), rather than on the suggested scale from 1-4, due to scaling being dependant on the platform developer / manager implementation plans, which are generally not known to wider public.
- The following criteria were disregarded: RDA-I1-01D, RDA-I1-02D, RDA-I2-01D, RDA-I3-01D, RDA-I3-02D, RDA-I3-04M due to the fact that these criteria assess FAIRness of data itself and do not reflect capabilities of a certain platform.)

The following additional specific criteria were added:

- Defined wind energy specific terms can be linked to (Does the provider make data owners choose from a list of existing terms, such as CF conventions[21]?)
- Data analysis code can be linked (Is there a specific relationship that connects code to a dataset?)
- A plot of the data is available (A graph of a sample of data such as power vs. time)
- A preview of the data structure is available (A preview of the columns and rows of the data)
- A preview of the file structure is available (A graphic or list of the different files included in a certain record)

2.4 Creation of recommendations

The results of the assessment were analysed and discussed within the working group, which allowed a set of recommendations for platform developers, data providers and the wind energy community in general to be created. The results are presented in Section 4.

3 Results

The full results of the analysis can be found in [2]. In Figures 3-8 below, the scores have been added up for each data source for the following criteria:

- Figure 3: Only the original RDA FDMM criteria (without the additional criteria)
- Figure 4: Only the Findable (F) criteria
- Figure 5: Only the Accessible (A) criteria
- Figure 6: Only the Interoperable (I) criteria
- Figure 7: Only the Reusable (R) criteria.
- Figure 8: Only the additional criteria

[21] https://cfconventions.org/

In the figures, the data source type is marked by colour (blue = repository, green = knowledge hub, purple = company website, red = dashboard). The scores are given in percent relative to the total available score for each platform, where 100% means that all the criteria are fulfilled.

The results for the original RDA FDMM criteria show that the repositories all score similarly, from 74% to 83%, with the exception of the Marine Data Exchange, which scores lower (46%). The knowledge hubs also score slightly lower (69-71%), with the company websites (43% and 0%) and the dashboards (34-66%) the lowest. The EDP website scores 0% because the data and metadata were removed from the company's web page during the analysis. The highest scoring repositories, Zenodo, DTU Data, 4TU.ResearchData, and ETH Zurich, score 88% on the Findable criteria, 100% on Accessible and Interoperable criteria, and 50% on the Reusable criteria. For the Findable criteria, they do not obtain a full score because the RDA FDMM creators considered that in addition to metadata, the data itself should have a persistent identifier (RDA-F1-01D), which is not a current practice for any of the platforms we have analysed. For the Reusable criteria, they score "0" for RDA-R1.2-01M "Metadata includes provenance information according to wind energy community-specific standards", RDA-R1.3-01M "Metadata complies with a wind energy community standard", RDA-R1.3-01D "Data complies with a wind energy community standard", RDA-R1.3-02M "Metadata is expressed in compliance with a machine-understandable wind energy community standard", and RDA-R1.3-02D "Data is expressed in compliance with a machine-understandable wind energy community standard" because the wind energy community standards do not exist.

The plots showing the individual FAIR criteria provide insights into the areas responsible for the differences between the data publishing platforms:

For the Findable criteria, the lower-scoring platforms score lower due to the lack of DOI (RDA-F1-01M "Metadata is identified by a persistent identifier") (Statista, Marine Data Exchange, Ørsted website, Scenario Viewer, REMPD, WindLab), and lack of API access (RDA-F4-01M "Metadata is offered in such a way that it can be harvested and indexed") (Marine Data Exchange, Ørsted website, Scenaria Viewer, REMPD, OEDI).

For the Accessible criteria, the lower-scoring platforms score lower due to the lack of a metadata record that is accessible from the website (RDA-A1-03M "Metadata identifier resolves to a metadata record") (Statista, Marine Data Exchange, Ørsted website, Scenario Viewer, REMPD), the lack of automatic access without human intervention (RDA-A1-05D "Data can be accessed automatically (i.e. by a computer program") (Marine Data Exchange, Scenario Viewer), and lack of guarantee that the metadata will remain available after data is no longer available (RDA-A2-01M "Metadata is guaranteed to remain available after data is no longer available") (Wind Data Hub, Marine Data Exchange, Ørsted, Scenario Viewer, WindLab, OEDI). Failure to satisfy RDA-A1-03M usually results in difficulties of automated indexing of datasets by data catalogues, while RDA-A1-05D hampers the viability of automated workflows, integration of data with code, and other (meta)data.

It is clear that the Interoperable criteria are the most variable, with the Marine Data Exchange (17%), the Ørsted website (0%) and the Scenario Viewer (17%) scoring particularly low. These three data sources do not use knowledge representation expressed in standardised format, or machine-readable knowledge representation, for the metadata, and they do not use FAIR-compliant vocabularies for the metadata. The Ørsted website does not allow references to other data to be included.

For the Reusable criteria, many platforms have a score of 50%, due to the complete lack of wind energy community standards mentioned above. The lower-scoring platforms score lower due to the lack of a plurality of accurate and relevant attributes are provided to allow reuse (RDA-R1-01M) (Ørsted website, Scenario Viewer, REMPD), the lack of information about the licence under which the data can be reused in the metadata (RDA-R1.1-01M) (Marine Data Exchange, Scenario Viewer), lack of reference to a standard machine-understandable reuse licence in the metadata (RDA-

R1.1-02M and RDA-R1.1-03M) (Statista, Marine Data Exchange, Ørsted website, Scenaria Viewer), and lack of provenance information according to a cross-wind energy community language in the metadata (RDA-R1.2-02M) (Scenario Viewer).

Looking at the additional criteria, a large variation can be seen across the platforms. ENERGY-DATA.INFO[22] scores particularly highly (80%), the only criteria missing being the ability to link to wind energy specific terms. The Ørsted website scores 0%, as it cannot link to defined wind energy specific terms or to data analysis code, and no plot of the data, preview of the data structure, or preview of the file structure is available.

3.1 Limitations of the analysis

The "1-0" analysis means that we did not assess how well a given platform fulfilled a criteria, perhaps missing some nuances. Among the criteria that we found most challenging to assess are the functionalities which can greatly vary in terms of their realisation such as RDA-F2-01M, RDA-I2-01M, and RDA-R1-01M:

- For RDA-F2-01M Rich metadata is provided to allow discovery it is not clear what exactly would constitute rich metadata and what granularity of metadata would satisfy the criteria.
- For RDA-I2-01M Metadata uses FAIR-compliant vocabularies assigning "1-0" does not communicate whether all vocabularies are FAIR compliant (including rich metadata) or only a subset of them are FAIR (for example, if the platform is using DC-core, or DataCite Schema)
- For RDA-R1-01M Plurality of accurate and relevant attributes are provided to allow reuse, the question of accuracy and relevancy is use-case dependent. I.e. dictated by the intended use of the data, which is virtually impossible to predict. For example, date/time information is important for historic analysis or for analysis that requires fusion with other data sources. Meanwhile, many analyses in wind energy can be performed in isolation.

Assessment of some criteria requires specific technical knowledge and access to the platform source code, which wind energy domain experts can lack. For example, RDA-F4-01M "Metadata is offered in such a way that it can be harvested and indexed".

For some criteria we have found that taking a platform user perspective would be more useful for assessing data FAIRness:

- RDA-A1-03M "Metadata identifier resolves to a metadata record": For a convenient and automated use of metadata records it is especially important that metadata is not only provided in a human-readable format but also in machine-interpretable ones. In practice, this can be facilitated via content negotiation allowing to retrieve metadata in the format most suitable for further use.
- RDA-A1.2-01D "Data is accessible through an access protocol that supports authentication and authorisation": We believe it is much more important whether the platform (repository) offers the option for the data to have restricted access, rather than whether the protocol supports it in principle.

[22] https://energydata.info/

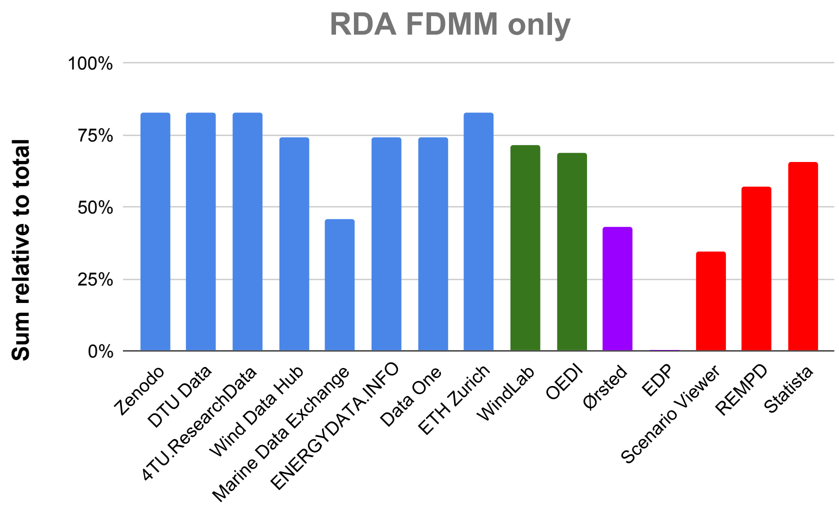

Figure 3: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the chosen RDA FAIR Data Maturity Model criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

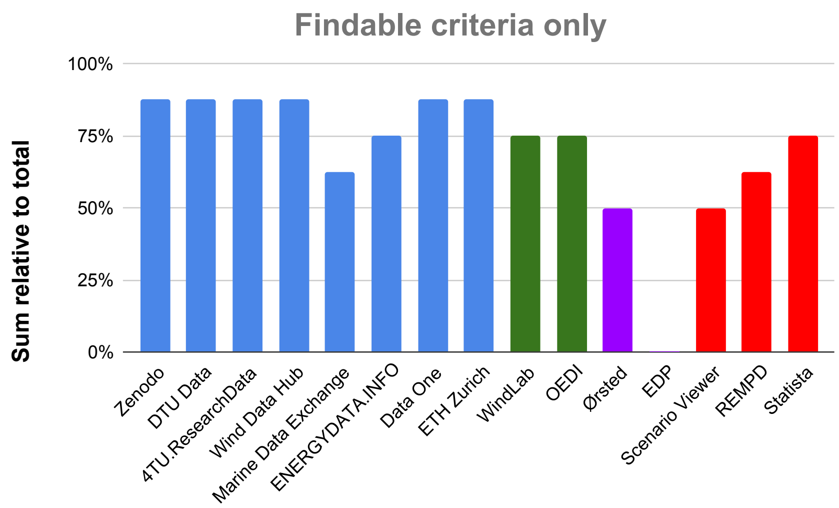

Figure 4: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the Findable criteria from the chosen RDA FAIR Data Maturity Model criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

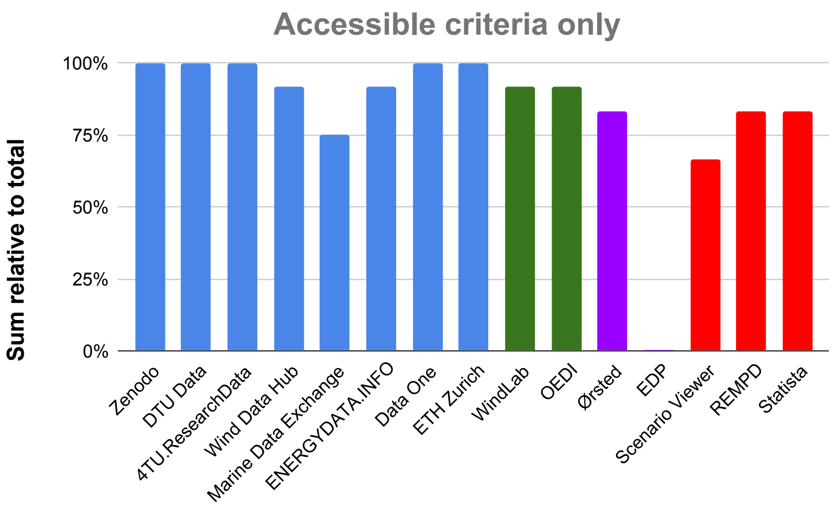

Figure 5: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the Accessible criteria from the chosen RDA FAIR Data Maturity Model criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

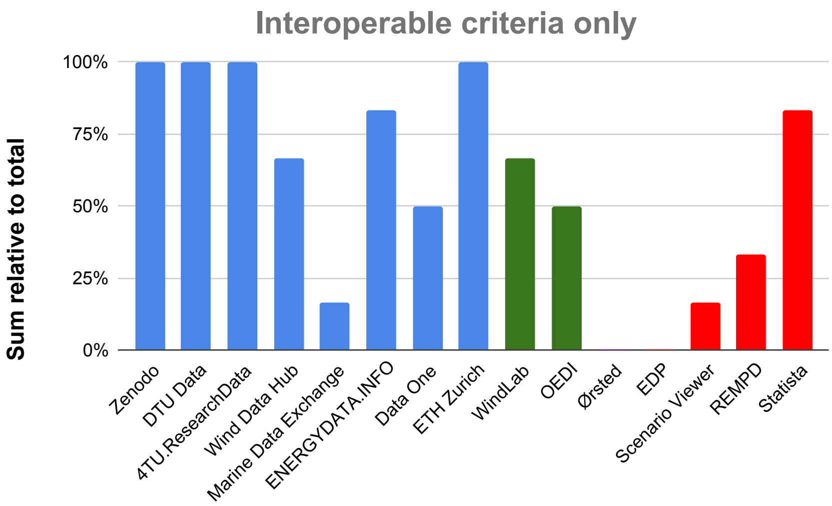

Figure 6: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the Interoperable criteria from the chosen RDA FAIR Data Maturity Model criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

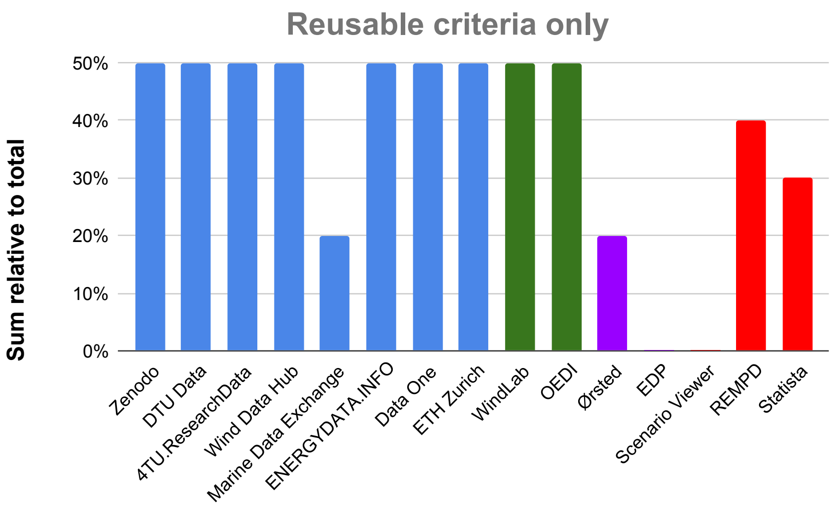

Figure 7: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the Reusable criteria from the chosen RDA FAIR Data Maturity Model criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

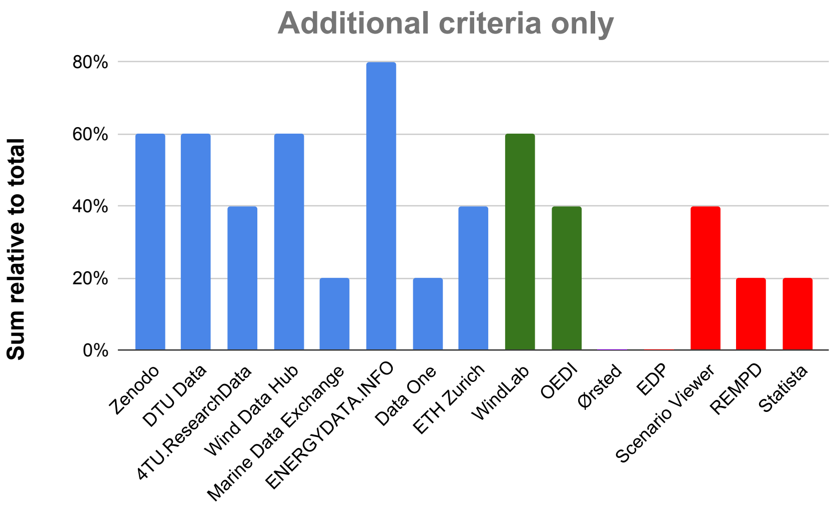

Figure 8: Sum of FAIR criteria scores relative to the total available score for each data publishing platform for the additional criteria (blue = repository, green = knowledge hub, purple = company website, red = dashboard).

4 Recommendations

The recommendations are split into three categories, depending on the role of the reader:

Wind energy data publishing platform developers: recommendations aimed at people involved in the creation, development and maintenance of wind energy specific data publishing platforms.

Wind energy data publishers: recommendations aimed at people in the wind energy sector involved in publishing data, either from measurement campaigns, simulations or numerical campaigns, or from operating assets.

The wind energy community: recommendations aimed at people involved in the wind energy community, who are able to contribute to international collaborations and industry-wide initiatives.

4.1 For wind energy data publishing platform developers

R1.1: Platforms should converge on how to include rich metadata, especially in the case of controlled vocabularies, in metadata forms (free text search or controlled vocabularies tags). As well as adopting general metadata standards (such as Dublin Core[23] and/or DCAT[24]), they should adopt specialised domain standards and work with the community to develop them where they do not exist.
Examples: Zenodo uses a "Keywords and subjects" field that can be either controlled vocabulary terms or custom tags, with supported controlled vocabularies including EuroSciVoc[25] *and MeSH*[26]*. The Open Energy Platform*[27] *uses the domain-specific Open Energy Metadata standard together with the Open Energy Ontology. For a wind energy platform, instead of providing a free-text field for 'Turbine Model', a domain standard should be adopted, allowing users to select from a controlled list (e.g., Vestas V112, Siemens Gamesa SG 14-222 DD).*

R1.2: Wind energy data publishing platform developers should perform an evaluation of the data FAIRness on their platforms with models like the RDA-DMM or automated tools like the F-UJI Automated FAIR Data Assessment Tool. Since these tools look at the FAIRness of the data, not the platform, these analyses can be run on their "best-case" example of data. To demonstrate the trustworthiness and reliability of the platform itself, developers should also pursue certification as a trusted digital repository, such as CoreTrustSeal[28].
Example: Datasets2Tools[29] *enables users to rapidly find datasets, tools, and canned analyses through an intuitive web interface, a Google Chrome extension, and an API. Furthermore, Datasets2Tools provides a platform for contributing canned analyses, datasets, and tools, as well as evaluating these digital objects according to their compliance with the FAIR principles.*

R1.3: Platforms should have FAQ pages, which explain how users can use these functionalities, and perhaps point to the "exemplary datasets".

[23] https://www.dublincore.org/specifications/dublin-core/
[24] https://www.w3.org/TR/vocab-dcat-3/
[25] https://op.europa.eu/en/web/eu-vocabularies/euroscivoc
[26] https://www.ncbi.nlm.nih.gov/mesh/
[27] https://openenergyplatform.org/
[28] https://www.coretrustseal.org/
[29] https://maayanlab.cloud/datasets2tools

Example: The Datasets2Tools platform mentioned in R1.2 has a details Help Center, helping users to get started, to understand the FAIR principles, and to contribute to the community.

R1.4: Wind energy data publishing platform developers should investigate awarding users badges, in order to incentivise them to submit FAIR data, by essentially gamifying the process.
Example: Scientific journals such as APA use Open Science badges[30].

R1.5: New data publishing platforms should build on existing data publishing platforms, rather than be created from scratch. This can either be done by forking open source platform codes or by initiating collaborative projects with existing platform developers. This is especially true for ambitious projects like developing knowledge hubs. As our analysis showed, these are not potentially more FAIR than platforms which do not claim to be "knowledge hubs".
Example: The Open Energy Platform provides an ecosystem of tools and vocabularies targetted at energy system research (reference in R1.1 above).

R1.6: Wind energy data publishing platform developers should meet regularly on a virtual platform to discuss common challenges and collaboration opportunities in order to avoid inefficiencies.
Example: The international collaboration IEA Wind Task 43[31] *meets regularly to discuss common challenges and collaboration opportunities. The open innovation ecosystem WeDoWind*[32] *includes a digital platform where members of the community can organise themselves into working groups.*

R1.7: Wind energy data publishing platforms should ensure that their records can be harvested by data catalogues, such as Google Dataset Search, and the records can be parsed by automated FAIR assessment tools like F-UJI.
Example: Using schema.org[33] *annotations embedded as RDFa*[34]*, Microdata*[35]*, or JSON-LD*[36] *on the landing page for a dataset provides structured, semantically annotated information about the dataset for systems to automatically process.*

R1.8: Wind energy data publishing platform developers should engage with diverse domain experts active in digitalisation efforts, to inquire about availability of community standards, which can be included in the platform development roadmaps.
Example: Currents efforts that can be joined for doing this are WeDoWind and IEA Wind Task 43, referenced in R1.6.

R1.9: For central search interfaces, metadata harvesting schemes should be provided to allow federated data infrastructures.
Example: This can be facilitated by supporting the OAI-PMH protocol[37] *that allows to automatically collect metadata from several sources to be provide a common, central interface. Another existing, successful model is be Europe's INSPIRE*[38] *initiative that provides multiple geodata-focused portals aggregating data from the local up to the European level. A user on a central portal like INSPIRE could search for 'German offshore wind data' and get results*

[30] https://www.apa.org/pubs/journals/resources/open-science-badges
[31] https://iea-wind.org/task43/
[32] https://www.wedowind.ch/
[33] https://schema.org/
[34] https://www.w3.org/TR/rdfa-primer/
[35] https://html.spec.whatwg.org/multipage/microdata.html
[36] https://www.w3.org/TR/json-ld/
[37] https://www.openarchives.org/pmh/
[38] https://knowledge-base.inspire.ec.europa.eu/index_en

from both the BSH's data platform and an industry partner's platform, all in one list. This is possible because both platforms provide an OAI-PMH endpoint for harvesting.

4.2 For wind energy data publishers

For wind energy data publishers, we recommend the following:

R2.1: Wind energy data publishers should use the highest scoring available repositories such as Zenodo to publish their data, where possible. The data set should be submitted to a relevant community, to ensure data curation. For wind energy data publishers with academic affiliations, institutional repositories are recommended. These repositories are dedicated to data FAIRness principles and are managed by experts in the topic, as reflected by high scores in our analysis.
Example: A university researcher publishes their data on Zenodo. After uploading, they also submit it to the 'WeDoWind Zenodo Community[39]*. A data steward from that community then checks the metadata for completeness before it is formally published, improving its quality and findability for other wind energy experts.*

R2.2: If organisations prefer to publish datasets on their website rather than on a data catalogue, a good practice is to archive them in a recognised repository and reference them from the website.
Example: The companies Cubico and RES have published datasets openly on Zenodo and advertised them on their own websites [10, 11, 4].

R2.3: Data publishers should use a repository that relies on controlled vocabularies or even ontology terms to provide metadata attributes like subject or keywords.
Example: Zenodo uses a "Keywords and subjects" field that can be either controlled vocabulary terms or custom tags, with supported controlled vocabularies including EuroSciVoc and MeSH. The Open Energy Platform uses the domain-specific Open Energy Metadata standard together with the Open Energy Ontology.

R2.4: Data publishers should select terms from subject, and other, controlled vocabularies when available within the platform. If the platform allows for free text tagging, then use terms from an established vocabulary rather than creating new ones.
Example: A platform provides a free-text field for keywords. Instead of inventing a new term like 'turbine breakdown,' a publisher should use a term from an established vocabulary, such as one from the IEA Wind Task 43, like "component failure".

4.3 For the wind energy community

For the wind energy community in general, we recommend the following:

R3.1: Wind energy experts, computer scientists and data stewards should contribute to the development of community information models, preferably as part of an international cooperation.
Example: The community could develop an information model for

Wind Turbine Wakes. This model would formally define the relationships between concepts, stating that a "Wake" is caused by a "Turbine" has properties like "velocity deficit" and "added

[39] https://zenodo.org/communities/wedowind/

turbulence", and is measured by an "Instrument" (e.g., LiDAR). Many existing international cooperations for developing information models collaboratively in wind energy can be accessed on the WeDoWind Information Modelling Community webpage[40]*.*

R3.2: Wind energy experts, computer scientists, and data stewards should contribute to existing efforts to develop a unifying Wind Energy Domain Ontology.
Example: The existing community effort of developing the Wind Energy Domain Ontology can be accessed on the WeDoWind platform[41]*.*

R3.3: Wind energy experts, computer scientists and data stewards should initiate and contribute to the development of a taxonomy of topics for publishing wind energy data, which should be connected to current efforts to the Wind Energy Domain Ontology.
Example: Such a taxonomy is, e.g., developed by the Association for Computing Machinery (ACM) and is used to classify publication via the ACM Computing Classification System[42]*.*

R3.4: International cooperations such as IEA Wind TCP, or organisations such as WindEurope should run regular FAIRness benchmarks (using tools like F-UJI) and publish public dashboards disseminating the results.
Example: The FAIR-IMPACT project ran a FAIRness Assessment Challenge[43]*.*

R3.5: Wind energy domain experts should approach platform developers to include community-accepted standards in their platform.
Example: The biomedical community ensured that Zenodo provides functionality to add metadata terms from MeSH[44]*.*

R3.6: Members of the wind energy community should create and share open and training material that connects general Research Data Management good practices with the specifics of the wind energy domain. This should especially include tutorials on how ontologies and taxonomies can be applied in practice.
Example: The Research Data Management toolkit for Life Sciences[45] *includes a community-built, open online handbook providing general RDM guidance and domain-focused solutions for life science subfields. The wind energy community could start with providing a one-hour webinar titled 'How to Use the WeDoWind Ontology to Annotate Your LiDAR Data.' This provides a practical, hands-on guide for a specific task.*

5 Lessons Learnt

5.1 Community building and interdisciplinary collaboration

One of the key lessons learned in this work was the importance of community engagement and commitment to creating, aligning, and maintaining semantic artefacts. At the same time, we discovered how difficult it is to motivate practitioners to provide the necessary contributions. It may be beneficial to apply community engagement methods from the social sciences, such as the Nominal Group Technique [7], Ethnography [3] and Ethnomethodology [8] in order to address this.

[40] https://community.wedowind.ch/spaces/17168189/page
[41] https://community.wedowind.ch/spaces/17701604/page
[42] https://dl.acm.org/ccs
[43] https://fair-impact.eu/support-offer-1-fairness-assessment-challenge-datasets-and-semantic-artefacts
[44] https://help.zenodo.org/docs/deposit/describe-records/keywords-and-subjects/
[45] https://rdmkit.elixir-europe.org

Furthermore, we learned the importance of balancing ambition with practicality, both in terms of the resources available within the working group, and in terms of the digital maturity of the sector. In terms of the resources available within the working group, we quickly realised that a fully detailed analysis of all the data sources and platform types was not possible. In terms of the digital maturity of the sector, we realised that a stepwise approach and gradual capacity building is essential to bringing the necessary stakeholders onboard. For example, the introduction of controlled vocabularies formalised through SKOS[46] would allow a first incremental step towards ontology alignment to be made.

Finally, we learned the importance of introducing and maintaining flexible processes, due to the dynamic nature of digitalisation and data standards in the sector. This can be done on a central community platform such as WeDoWind.

5.2 The FAIR Data Maturity Model

In applying the FAIR Data Maturity Model, we learned that a binary "1–0" assessment, while easier for us to apply, may have overlooked some important nuances. As well as this, several criteria proved difficult to interpret consistently, such as what qualifies as "rich metadata" (RDA-F2-01M), how to judge partial versus full use of FAIR-compliant vocabularies (RDA-I2-01M), or what constitutes "accurate and relevant" attributes for reuse (RDA-R1-01M), which are highly use-case dependent. In some cases, such as RDA-F4-01M, meaningful assessment would require technical expertise or access to platform source code beyond our capabilities (and the capabilities of most domain experts). We also observed that a user-centric perspective may be more informative for some criteria than a purely technical one, such as focusing on whether repositories provide usable options for restricted access for RDA-A1.2-01D, rather than whether the underlying protocol theoretically supports it.

6 Acknowledgements

The Wind Energy Community Standards working group has received support from the RDA TIGER programme funded under the Horizon Europe Grant agreement ID: 101094406[47].

We are also grateful to Carlos Martínez Ortiz[48], from the Netherlands eScience Center, for his useful feedback and support.

[46] https://www.w3.org/TR/skos-reference/
[47] https://cordis.europa.eu/project/id/101094406
[48] https://orcid.org/0000-0001-5565-7577

References

[1] Christophe Bahim, Carlos Casorrán-Amilburu, Makx Dekkers, Edit Herczog, Nicolas Loozen, Konstantinos Repanas, Keith Russell, and Shelley Stall. "The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments". In: *Data Science Journal* 19 (2020). ISSN: 1683-1470. DOI: 10.5334/dsj-2020-041.

[2] Sarah Barber, Sirko Schindler, Catherine Jones, Pablo Rodríguez-Sánchez, and Yuriy Marykovskiy. *Best Practice Recommendations for improving FAIR data maturity in wind energy - accompanying data*. Zenodo, Oct. 2025. DOI: 10.5281/zenodo.17255305.

[3] John D. Brewer. *Ethnography*. repr. Understanding social research. Includes bibliographical references and index. Buckingham [u.a.]: Open Univ. Press [u.a.], 2008. 211 pp. ISBN: 0335202691.

[4] Alex Clerc and Elizabeth Lingkan. *Hill of Towie wind farm open dataset*. 2025. DOI: 10.5281/ZENODO.14870023.

[5] Andrew Clifton, Sarah Barber, Andrew Bray, Peter Enevoldsen, Jason Fields, Anna Maria Sempreviva, Lindy Williams, Julian Quick, Mike Purdue, Philip Totaro, and Yu Ding. "Grand challenges in the digitalisation of wind energy". In: *Wind Energy Science* 8.6 (June 2023), pp. 947–974. ISSN: 2366-7451. DOI: 10.5194/wes-8-947-2023.

[6] Anusuriya Devaraju and Robert Huber. *F-UJI - An Automated FAIR Data Assessment Tool*. 2025. DOI: 10.5281/ZENODO.6361400.

[7] Nichole Harvey and Colin A Holmes. "Nominal group technique: An effective method for obtaining group consensus". In: *International Journal of Nursing Practice* 18.2 (Mar. 2012), pp. 188–194. ISSN: 1440-172X. DOI: 10.1111/j.1440-172x.2012.02017.x.

[8] John Heritage. "Ethnomethodology". In: *Social theory today* (1987), pp. 224–272.

[9] Yuriy Marykovskiy, Thomas Clark, Justin Day, Marcus Wiens, Charles Henderson, Julian Quick, Imad Abdallah, Anna Maria Sempreviva, Jean-Paul Calbimonte, Eleni Chatzi, and Sarah Barber. "Knowledge engineering for wind energy". In: *Wind Energy Science* 9.4 (Apr. 2024), pp. 883–917. ISSN: 2366-7451. DOI: 10.5194/wes-9-883-2024.

[10] Charlie Plumley. *Kelmarsh wind farm data*. en. Zenodo, 2022. DOI: 10.5281/ZENODO.5841834.

[11] Charlie Plumley. *Penmanshiel Wind Farm Data*. en. Zenodo, 2022. DOI: 10.5281/ZENODO.5946808.

[12] Mingfang Wu, Nick Juty, RDA Research Metadata Schemas WG, Julia Collins, Ruth Duerr, Chantel Ridsdale, Adam Shepherd, Chantelle Verhey, and Leyla Jael Castro. *Guidelines for publishing structured metadata on the web*. Version 3.1. Oct. 2021. DOI: 10.15497/RDA00066.