

EUDP IEA Task 41

Deliverable 2.1. Report on the adopted metadata and taxonomies specific for DW and metadata catalogue for data sharing

Department of
Wind Energy
E Report 2020

Author: Anna Maria Sempreviva

Report number: TBA

February 2021

DTU Wind Energy
Department of Wind Energy



Authors: Anna Maria Sempreviva,

Title: Report on the adopted metadata and taxonomies specific for DW

February 2021

Department: DTU Wind Energy

Contributors: Danielle Prezioso, Alice Orrel

Pacific Northwest National laboratories, PNNL, US

Summary:

This report offers the executive summary of the first two deliverables of the WP2 data catalogue within the IEAWind Task 41 Distributed Winds, DW. It describes the process followed in choosing metadata and taxonomies specific for DW and presents the first solution for a metadata catalogue. Metadata support Findable, Accessible, Interoperable and Reusable (FAIR) data principle. Assigning metadata to datasets make data visible for the DW research and innovation communities.

Project no.:

43468 46 EUDP IEA Task 41

Sponsorship:

The EUDP IEA Task 41 project is funded by Energy Technology Development and Demonstration Program (EUDP) of the Danish Energy Agency

Pages: 14

Tables: 2

References: 2

Technical University of Denmark

Department of Wind Energy

Frederiksborgvej 399

Building 118

4000 Roskilde

Denmark

Telephone

www.vindenergi.dtu.dk

Contents

1. Introduction	4
1.1 EUDP IEA Task 41: Distributed Wind	4
2. (Meta)data catalog, Metadata & Taxonomy.....	5
2.1 Metadata and taxonomies.....	5
3. Advancement of the metadata and taxonomies-	5
3.1 Metadata.....	5
3.2 Taxonomy.....	6
3.3 Metadata registry and portal.....	6
4. References.....	8
5. ANNEX1 “Deliverable D10. Work Package 2: Data Catalog Specification” by Danielle Prezioso	11
6. ANNEX II “Deliverable D11. Data Sharing, Storage, Security Protocols, and a Specification of a Potential Data Sharing Portal, by Danielle Prezioso, PNNL; Anna Maria Sempreviva; DTU and Alice Orrell, PNNL	12

1. Introduction

This report provides an overview of the achievements in WP2 *Data Catalog* towards the final goal: create a data catalog finalized to data sharing. The first objective of WP2 is to lay the basis to design a data management system to organize and manage research data and information of use for DW research.

The Team of Technical University of Denmark, Department of Wind Energy, DTU Wind Energy (DTU WE), entered Task 41 on January 1st 2020; nevertheless, DTU WE had been contacted by the project management since the beginning as already had a solution to be adapted by the IAETask41.

The challenge is to make Visible and Findable data distributed in several institutions working in the DW field, as no central, global way currently exists. To be searchable, data must be previously labelled with standard keywords from controlled vocabularies in common between data provider and data user.

The current focus of this effort has been to collect and make available metadata (information about data) not the data itself, sidestepping issues around data sensitivity and intellectual property. It is envisioned that if researchers need specific products, they can consult the catalog by a faceted search, i.e. applying filters, retrieve information about available products, identify the product and apply for access.

The first deliverable of WP2, D10 has been issued in December 2019. It contains:

- The specification of a data-sharing catalog including a review of needs, e.g. what metadata should be collected, and
- Potential options for hosting the catalog. DTU Wind Energy contribution, has been in –kind.

The second deliverable of WP2, D11, had a delay and was completed in January 2020. It looks at the development of data sharing, storage, and if needed security protocols for metadata to be stored on the platform. It also gives specification of the potential data sharing portal that expands on the catalog. As this report summarize and includes the above IEA Task reports, it was delayed as well.

1.1 EUDP IEA Task 41: Distributed Wind

Here is a summary of the IEA Task41. The overall objective of this project is to identify and explore studies of particular Danish interest of Distributed Wind (DW) for cost effective technology development and integration into a continuously evolving energy system. This is done by collaborating and contributing to the IEA Wind TPC Task 41 international activities on DW turbine technology development and assessment in a series of dedicated work packages (WPs). IEA Wind TPC Task 41 is an international network centered on international collaboration and coordination in the field of DW. The purpose is to accelerate the development and deployment of DW technology as one of the leading generation source in global renewable markets, the facilitation of easier and faster DW integration into electrical grids, increasing thus the competitiveness of wind and accelerating the replacement of fossils fuels. The IEA collaboration is enforced partly by exchange of information, sharing of results, and conducting analyses and explorative studies in the form of reports and publications and partly by implementing a strong cross IEA Wind TPC Tasks collaboration effort. The idea of the data catalog is to start actions leading to facilitate the sharing of resources

2. Metadata catalog, Metadata & Taxonomy

This section summarizes the two IEA Task 41 reports:

“ Deliverable D10. Work Package 2: Data Catalog Specification” by Danielle Prezioso.

“ Deliverable D11. Data Sharing, Storage, Security Protocols, and a Specification of a Potential Data Sharing Portal, by Danielle Prezioso, PNNL; Anna Maria Sempreviva, DTU and Alice Orrell, PNNL.

2.1 Metadata and taxonomies

Metadata are the high-level pieces of information that allow to search resources and information to be searched, identified, and accessed. These elements can serve various purposes from describing the content of the resource to who collected the data contained within the resource to the structure or format of the resource (Sempreviva et al. 2017). Some metadata elements should be dictated by controlled taxonomies to force consistent terminology. A controlled taxonomy is a set of predefined terms that standardize the information input into a given metadata element. This enables more accurate search results and data archiving. Metadata elements can be specified by different controlled taxonomies, and metadata cards store the information input for the metadata elements.

3. Advancement of the metadata and taxonomies-

The task force decided to adopt the metadata and taxonomies produced by the IRPWind project as starting point and build up on it.

3.1 Metadata

In its metadata cards, IRPWind uses metadata elements from the Dublin Core Metadata Initiative (DCMI) in addition to several metadata elements that IRPWind developed specific to wind energy research. DCMI is an initiative dedicated to metadata best practices. IEA Wind Task 41 will adopt the metadata elements and wind energy taxonomy (or an expanded version of the wind energy taxonomy) implemented in ShareWind. Specifically, IEA Wind Task 41 will use the following DCMI elements.¹

1 Definitions for metadata elements are contained within the metadata card. Examples and descriptions can also be found at: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#elements-contributor>.

2 These metadata elements will either be dictated by IRPWind's wind energy taxonomy or the IRPWind's wind energy taxonomy will be expanded to fill the gaps needed to address distributed wind and/or meet IEA Task 41 needs.

3 These elements require unique controlled taxonomies separate from IRPWind's wind energy taxonomy. The task will select and implement appropriate taxonomies for these elements.

Table 1 Metadata elements and controlled taxonomy needs.

Element	Controlled Taxonomy
Title	Not Applicable
Creator	Not Applicable
Subject	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Description	Not Applicable
Publisher	Not Applicable
Contributor	Not Applicable
Date	Not Applicable
Type	Specified by the DCMI Type Vocabulary
Format	Identify a controlled taxonomy in next steps
Identifier	Identify a controlled taxonomy in next steps
Source	Identify a controlled taxonomy in next steps
Language	Identify a controlled taxonomy in next steps
Relation	Identify a controlled taxonomy in next steps
Coverage	Identify a controlled taxonomy in next steps
Rights	Not Applicable
Variables	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
External Conditions	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Activity	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Instrument	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Model	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Material	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy

The last seven of these 21 collected metadata elements, have been specified by IRPWind's wind energy taxonomies. The work done in this WP2 has been to expand the last version of the taxonomy including specific to the distributed wind industry. The metadata elements dictated by the wind energy taxonomy include: subject, variables, external conditions, activity, instrument, model, and material. The metadata "Subject" was called Topic" in IRPWind.

The other metadata elements require a controlled taxonomy or specified formatting that will not be dictated by IRPWind's wind energy taxonomy even in an expanded version. This includes type, format, identifier, source, language, coverage, relation, and coverage. DCMI makes recommendations for existing controlled taxonomies for these metadata elements.

The next step of IEA Wind Task 41 is to review available options and determine how to best implement the additionally selected vocabularies. The next report will show the conclusions.

3.2 Taxonomy

The IEA task 41 evaluated the wind energy taxonomy of topics implemented in ShareWind and added supplemental distributed wind terms. Also, controlled taxonomies for four other DCMI elements. Table 1 outlines the metadata elements that need a controlled taxonomy. Figure 1 shows the extended taxonomy of the topics. The red squares indicate the keywords added by Task41, referring to terms appropriated for the DW activities.

3.3 Metadata registry and portal

Next steps also include selecting an appropriate location for the task's catalog bearing in mind this initial review of Tethys, DAP, OpenEI, and ShareWind. PNNL will consider potential host locations that embrace the FAIR principle and can be easily maintained over time.

4. Next step

Next step will be to put the work into practice and provide guidelines for best practices for compiling DW distributed object catalogues. Data Management Plan Template, for Danish actors.

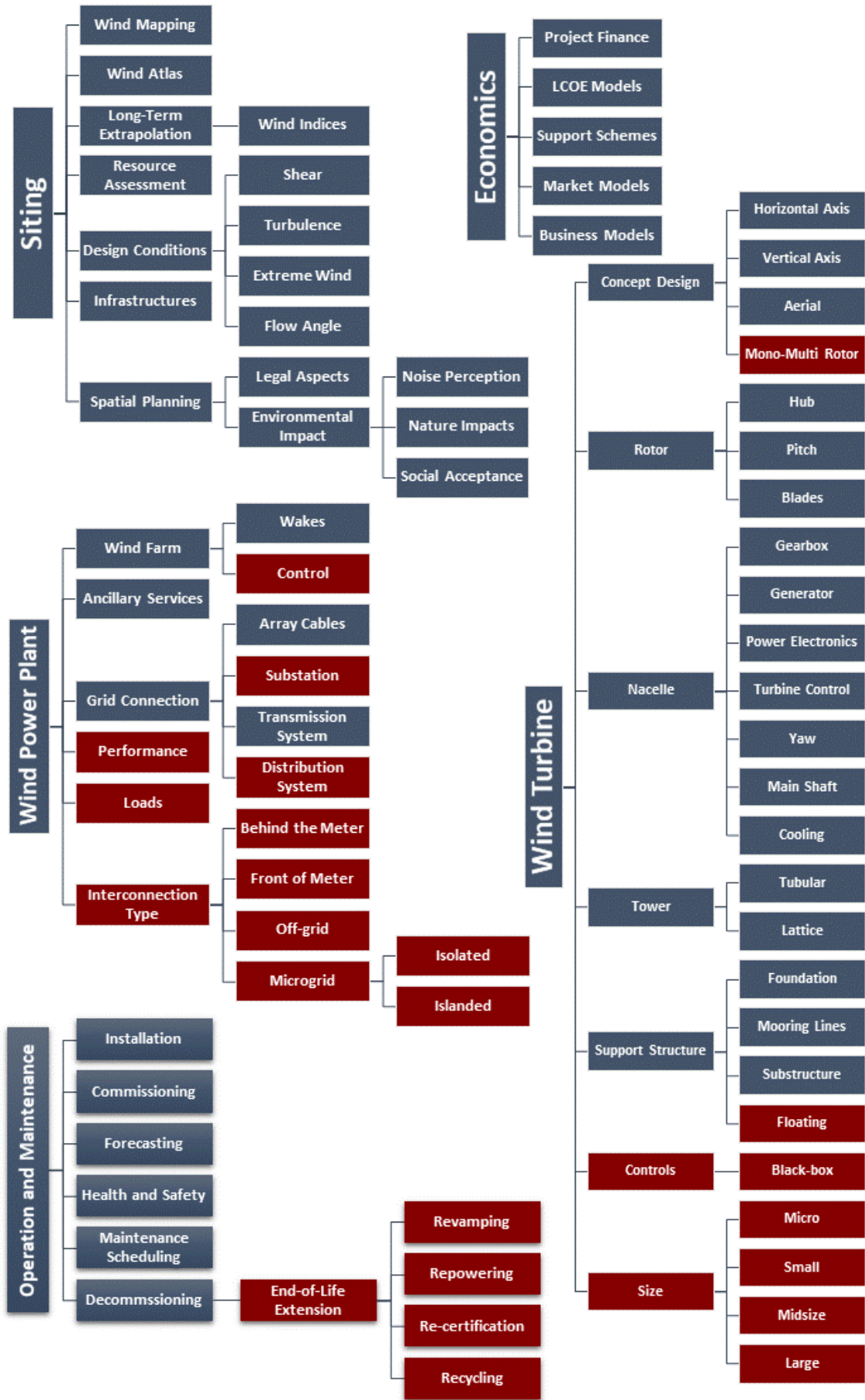


Figure 1. Taxonomy of the topic modified and adopted by IEATask41. Note the red color represents new added topics relevant for task 41 Distributed winds

5. References

Sempreviva, Anna Maria, Allan Vesth, Christian Bak, David Robert Verelst, Gregor Giebel, Hilmar Kjartansson Danielsen, Lars Pilgaard Mikkelsen, Mattias Andersson, and Nikola Vasiljevic. 2017. *Taxonomy and meta data for wind energy R&D*. <https://zenodo.org/record/1199489#.Xefe3ehKhdi>

Gancarski, Pawel. 2019. Sharewind - wind data registry. Presented at the *Wind Energy Science Conference 2019 (WESC)*, Cork: Zenodo. <http://doi.org/10.5281/zenodo.3357860>

Table 2. Taxonomy of 7 descriptive metadata in Table 1

Type	Activities	External Conditions	Models	Materials	Variables	Instrument
Collection	Certification	Coastal	Aeroelastic Models	Aluminium	2D	Ampermeter
Dataset	Condition Monitoring	Complex	Analytical	Aluminium	2D airfoil coefficients	Anemometer
Event	Fatigue Test	Deep Water	BoS	Amorphous Alloy	3D	Capacitance meter
Image	Field Experiment	Depth Category	Contingency Analysis	Cable insulation	3D mesh geometry	Ceilometer
Interactive Resource	Field Test	Escarpment	DNS	Cast iron	3D mesh geometry	Continuous
Moving Image	Full Scale Test	Flat	Dynamic Models	Cobalt-iron	AC voltage control	Cos Phi Meter
Physical Object	Interviews	Forest	Electrical Models	Composite laminate	Active damping	Cup
Data Type	Laboratory Test	Hilly	FEM	Concrete	Active damping	Curve tracer
Meteorological Data	Long-Term Monitoring	Inland	Financial Models	Concrete	Active power	Distortionmeter
SCADA Data	Manufacturing	Location	Flow Models	Copper	Active power control	Drones
Topographic Data	Measurements	Nearshore	Fully Nonlinear	Core	Aerodynamic	Electrical data
Uncategorized Data	Modeling	Offshore	GCM	Drivetrain	AG	Electron microscopy
Wind Farm Data	Questionnaires	Offshore	Harmonic Models	Ferrite	Air Density	ESR meter
Wind Turbine Data	Reliability and Testing	Onshore	Hindcast	Geal coats	Air Pressure	Frequency counter
	Static Test	Ridge	Hydrodynamic	Insulation	Beam properties cross sectional	Hyperspectral camera
	Uncertainty Analysis	Rural	IRR	Laminated steel	Cadaster	Imaging
	Validation	Semi-Urban	LCOE	Magnets	Campbell diagram	Supports
	Wind Tunnel Test	Shallow Water	LES	Nanocrystalline	Control	LCR meter
	Terrain Type		Linear	Nickel-iron	Control loops	Leakage tester
	Terrain Type		Linearized	Powder core	Control loops	Lidars
	Topography		Mesoscale	Powder iron	Converter data	Long-range
	Urban		Meteorological	Sandwich Structure	Copper losses	Masts
			Morison Equation	Silicon steel	cross-section geometry	Moored
			NPV	Steel	Current control	Multimeter
		Oceanographic	Steel	Current control	Nacelle lidar	
		OPF	Tower Materials	Curtailement Power set point	Network analyser	
		Physical	Wood	damping	Ohmeter	

Power Flow
Power Protection Analysis Models
Radiation-Diffraction
RANS
Second Order
Short Circuit Models
Small-Signal Models
State Estimation
Vortex Method
Wave

DC voltage control	Oil sensors
DFIG	Optical microscopy
Dynamics	Oscilloscope
Eigen frequencies	Pitot
Eigen frequencies	Power analyzer
failure modes buckling	Profilers
Filter capacitance	Psophometer
Filter constants	Pulsed
Filter inductance	Radars
Filter type	Radiosondes
Gains	Rass
Generator data	Satellite
Generator speed tracking	short-range
Geo Spatial Data	Signal analyzer
Geology	Signal generator
Grid side converter	Sodars
Heat Fluxes	Sonic
Humidity	Spectrum analyzer
Installed capacity	Strain gauges
Land-use	Sweep generator
mass and inertia	T profiles
Maximum power-point tracking	Thermometer
Mechanics, structure	Transistor tester
mode shapes	Ultrasonic testing
representation nth order system	Vane
Nacelle wind DIR	Vibration sensors
Nacelle wind speed	Voltmeter
Noise	Wattmeter
Noise perception	Waves sensors
No-load losses	Wind Scanners
Pitch	X-ray CT data

6. ANNEX1 “Deliverable D10. Work Package 2:
Data Catalog Specification” by Danielle Prezioso.

7. ANNEX II “Deliverable D11. Data Sharing, Storage, Security Protocols, and a Specification of a Potential Data Sharing Portal, by Danielle Prezioso, PNNL; Anna Maria Sempreviva; DTU and Alice Orrell, PNNL

Deliverable 2.1. Report on the adopted metadata and taxonomies specific for DW and metadata catalogue for data sharing.

DTU Wind Energy is a department of the Technical University of Denmark with a unique integration of research, education, innovation and public/private sector consulting in the field of wind energy. Our activities develop new opportunities and technology for the global and Danish exploitation of wind energy. Research focuses on key technical-scientific fields, which are central for the development, innovation and use of wind energy and provides the basis for advanced education at the education.

We have more than 240 staff members of which approximately 60 are PhD students. Research is conducted within nine research programmes organized into three main topics: Wind energy systems, Wind turbine technology and Basics for wind energy.

Technical University of Denmark
Department of Wind Energy
Frederiksborgvej 399
Building 118
4000 Roskilde
Denmark
Telephone 46 77 50 85

IEA Wind Task 41: Enabling Wind to Contribute to a Distributed Energy Future

Work Package 2: Data Catalog
Specification

December 2019

Danielle Prezioso

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <<https://www.ntis.gov/about>>
Online ordering: <http://www.ntis.gov>

1.0 Introduction

Under *Work Package 2: Data Information Catalog for Distributed Wind Research*, Pacific Northwest National Laboratory (PNNL) is leading the development of a data-sharing catalog for the purposes of disseminating and archiving distributed wind data and information. As part of Deliverable D10 and Milestone M5 of the International Energy Agency (IEA) Wind Task 41 work plan, this memo describes the specifications of a data-sharing catalog, including:

- a review of needs
- catalog hosting options
- metadata needs
- next steps.

2.0 Review of Needs

After preliminary discussions about the data-sharing catalog at the IEA Wind Task 41 Fall 2019 meeting in Boston, a review of needs indicates the following are necessary to develop and enable the success of the IEA Wind Task 41 data-sharing catalog:

- a protocol for collecting the data
- controlled taxonomies and formatting for specific metadata fields
- an understanding of the primary users, contributors, and content to inform the metadata collection process and catalog interface
- assurance that the catalog embraces the FAIR—findable, accessible, interoperable, and reusable—principle to better guarantee catalog users can access the range of research and information contained within
- content prioritization by the research needs of other IEA Wind Task 41 work packages to realize the benefits of the catalog in the near term
- an organizational construct that serves the broader distributed wind research community, if possible, to increase the impact of the work.

3.0 Catalog Hosting Options

Foundational research has been conducted around these needs to lay the groundwork for the catalog. Four existing catalogs were evaluated as potential hosts of the IEA Wind Task 41 catalog, to document lessons learned from previous cataloging efforts, and to develop best practices to satisfy the needs identified above.

3.1 Tethys Environmental

Description: Tethys is a public, online knowledge management tool that PNNL developed in 2009 to support research for the U.S. Department of Energy (DOE). Its primary focus is to disseminate qualitative and quantitative information related to the environmental effects of wind and marine renewable energy to developers, regulatory agency staff, stakeholders, and researchers.

Operation: The Tethys website contains content that is manually curated by researchers at PNNL, which is labor and cost intensive. Tethys does not host the data, but instead points to resource (i.e., either provides a link to where the resource is located online or the contact information of an individual or organization that can provide the resource on a case-by-case basis). The search engine returns a list of metadata corresponding to a user's search criteria. PNNL staff periodically check the contact information to verify website users can access the data.

Hosting IEA Wind Task 41 Catalog: Tethys's scope is limited to environmental impacts while the data catalog scope of IEA Wind Task 41 will be broader, creating a mismatch in content and potentially diluting the focus of Tethys. Using the backend and functionality of the website may be an option but would require a steady funding source.

Primary Takeaways: Manually maintaining a data catalog and its content can be costly and challenging over time. Researchers maintaining the Tethys catalog recommend against developing an independent data catalog and suggest finding a way to collaborate with an existing structure.

3.2 Data Archive and Portal

Description: The Data Archive and Portal (DAP) is a DOE-funded data-management structure hosted on the Atmosphere to Electrons (A2e) website. The DAP currently collects, stores, and preserves large data sets coming out of projects under the A2e research initiative. Its purpose is to afford timely, open, and reliable data access.

Operation: DAP stores large, quantitative data sets on its site. A team of researchers approves metadata submissions and verifies that data sets comply with completeness and compatibility requirements.

Hosting IEA Wind Task 41 Catalog: As DAP currently only hosts quantitative data, utilizing this platform would limit the IEA Wind Task 41 catalog to quantitative data sets.

Primary Takeaways: Metadata are integral to searchable data. Once IEA Wind Task 41 has collected metadata for the desired resources, those metadata must be reviewed to confirm they accurately represent the associated resource to enable searchability.

3.3 Open Energy Information

Description: Open Energy Information (OpenEI) is a wiki platform that serves a broad energy community including policymakers, developers, and researchers. The platform largely focuses on renewable energy and energy efficiency, hosting a range of resources including but not limited to international utility rates, resource data, policy and regulatory information and tools, and geospatial data sets.

Operation: OpenEI is essentially maintained by users as they view, edit, add, and download the data for free. Users are encouraged to edit content for clarity when appropriate. Users can access data in various ways, and the site can store information internally or link to an outside source.

Hosting IEA Wind Task 41 Catalog: OpenEI is a free, easy-access platform that could relatively easily accommodate the IEA Wind Task 41 catalog. However, its practices for maintaining and sharing information lack the standardization associated with implementing controlled taxonomies for metadata and the interface does not lend itself to easily finding resources. These practices do not fully embrace the FAIR principle.

Primary Takeaways: A wiki platform where users directly upload resources would not be as costly for IEA Wind Task 41 to maintain over time; however, OpenEI's metadata practices and interface do not appear to follow the FAIR principle.

3.4 ShareWind

Description: ShareWind is a prototype catalog that came out of the Integrated Research Programme in Wind Energy (IRPWind) with contributions from the European Energy Research Alliance Joint Programme on Wind Energy (EERA JPWind). A major part of their efforts to date has been developing a wind energy taxonomy and metadata forms that the portal will implement. Researchers have applied for more funding to finish developing the portal and its capabilities. ShareWind researchers prioritized the FAIR principle in the development of their portal.

Operation: ShareWind will use a web crawler to find outward-facing metadata cards that have been tagged with the wind energy taxonomy and return those resources when searched. This process will allow data owners to control who can access their data and for what purposes. ShareWind will not host the data, but instead return metadata pointing to where the data can be accessed.

Hosting IEA Wind Task 41 Catalog: As ShareWind only exists as a prototype, it is unlikely capable of hosting the catalog in the near term. However, the efforts that went into developing the wind energy taxonomy and metadata offer a path for collaboration and a viable option for IEA Wind Task 41 to build upon the work of other wind energy researchers who are working to develop a catalog specific to wind energy.

Primary Takeaways: IRPWind and EERA JPWind invested significant time into developing metadata requirements and a controlled taxonomy specific to wind energy and have been distributing their taxonomy for adoption, including to other IEA Wind Tasks. These efforts are a great starting point for IEA Wind Task 41.

4.0 Metadata Needs

Metadata are the high-level pieces of information that allow for resources and information to be searched, identified, and accessed. These elements can serve various purposes from describing the content of the resource to who collected the data contained within the resource to the structure or format of the resource (Sempreviva et al. 2017). Some metadata elements should be dictated by controlled taxonomies to force consistent terminology. A controlled taxonomy is a set of predefined terms that standardize the information input into a given metadata element. This enables more accurate search results and data archiving. Metadata elements can be specified by different controlled taxonomies, and metadata cards store the information input for the metadata elements.

In its metadata cards, ShareWind uses metadata elements from the Dublin Core Metadata Initiative (DCMI) in addition to several metadata elements that IRPWind developed specific to wind energy research. DCMI is an initiative dedicated to metadata best practices. IEA Wind Task 41 will adopt the metadata elements and wind energy taxonomy (or an expanded version of the wind energy taxonomy) implemented in ShareWind. Specifically, IEA Wind Task 41 will use the following DCMI elements.¹

- Title
- Creator
- Subject²
- Description
- Publisher
- Contributor
- Date
- Type³
- Format³
- Identifier³
- Source³
- Language³
- Relation³
- Coverage³
- Rights

IEA Wind Task 41 will also use the additional metadata elements IRPWind developed for ShareWind that were developed for tagging resources specific to the wind industry. This includes the following:

- Variables²
- External Conditions²
- Activity²

¹ Definitions for metadata elements are contained within the metadata card. Examples and descriptions can also be found at: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#elements-contributor>.

² These metadata elements will either be dictated by IRPWind's wind energy taxonomy or the IRPWind's wind energy taxonomy will be expanded to fill the gaps needed to address distributed wind and/or meet IEA Task 41 needs.

³ These elements require unique controlled taxonomies separate from IRPWind's wind energy taxonomy. The task will select and implement appropriate taxonomies for these elements.

- Instrument²
- Model²
- Material²

Of these 21 collected metadata elements for the IEA Wind Task 41 catalog, seven will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy that includes additional terms specific to the distributed wind industry. The metadata elements dictated by the wind energy taxonomy include: subject, variables, external conditions, activity, instrument, model, and material.

Seven other metadata elements require a controlled taxonomy or specified formatting that will not be dictated by IRPWind's wind energy taxonomy even in an expanded version. This includes type, format, identifier, source, language, coverage, relation, and coverage. DCMI makes recommendations for existing controlled taxonomies for these metadata elements. IEA Wind Task 41 will review available options and determine how to best implement the additionally selected taxonomies.

The task will evaluate the wind energy taxonomy implemented in ShareWind to determine if supplemental distributed wind terms are necessary and select different controlled taxonomies for four other DCMI elements.¹ Table 1 outlines the metadata elements that need a controlled taxonomy.

¹ Note that a controlled taxonomy has already been selected for type.

Table 1 Metadata elements and controlled taxonomy needs.

Element	Controlled Taxonomy
Title	Not Applicable
Creator	Not Applicable
Subject	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Description	Not Applicable
Publisher	Not Applicable
Contributor	Not Applicable
Date	Not Applicable
Type	Specified by the DCMI Type Vocabulary
Format	Identify a controlled taxonomy in next steps
Identifier	Identify a controlled taxonomy in next steps
Source	Identify a controlled taxonomy in next steps
Language	Identify a controlled taxonomy in next steps
Relation	Identify a controlled taxonomy in next steps
Coverage	Identify a controlled taxonomy in next steps
Rights	Not Applicable
Variables	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
External Conditions	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Activity	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Instrument	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Model	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy
Material	Will be specified by IRPWind's wind energy taxonomy or an expanded version of that taxonomy

5.0 Next Steps

PNNL will distribute a metadata card with proposed elements to IEA Task 41 participants for feedback. Following distribution, PNNL will lead a discussion with IEA Wind Task 41 participants regarding the status of the data catalog, metadata collection, and upcoming data catalog work. This process will help the task determine the following:

- if the proposed metadata suffice
- what data are easily accessible to participants and what data the task might need to acquire to round out the data catalog
- the existing scope of data that the task can access
- a process for collecting data outside of the task's immediate control
- the primary users, contributors, and content of the catalog.

When the metadata card is distributed, Task participants will also be asked to provide a list of data that they would like to see included in the catalog even if they do not have personally have access to such a data set. This will help prioritize the research needs of the task while also constructing the catalog in a way that can serve the broader distributed wind community if possible.

With respect to controlled taxonomies, PNNL will evaluate the wind energy taxonomy developed by IRPWind to decide if it is sufficient for distributed wind resources or if supplemental terms are necessary to accurately describe IEA Wind Task 41 resources in the seven identified metadata elements. PNNL is developing a list of common vocabulary terms used by task participants. They will compare this vocabulary list to IRPWind's wind energy taxonomy to determine if that taxonomy can be implemented in its existing form for the IEA Wind Task 41 catalog or if IEA Wind Task 41 will supplement the IRPWind taxonomy with terms from the task's vocabulary list.

For the remaining six metadata elements requiring a controlled taxonomy but do not yet have one implemented, existing options will be reviewed and selected. Applying controlled taxonomies to specific metadata elements guarantee the data are findable and accessible per the FAIR principle. Once controlled taxonomies are implemented, the task will begin filling out metadata cards for resources they have.

Next steps also include selecting an appropriate location for the task's catalog bearing in mind this initial review of Tethys, DAP, OpenEI, and ShareWind. PNNL will consider potential host locations that embrace the FAIR principle and can be easily maintained over time.

6.0 References

Sempreviva, Anna Maria, Allan Vesth, Christian Bak, David Robert Verelst, Gregor Giebel, Hilmar Kjartansson Danielsen, Lars Pilgaard Mikkelsen, Mattias Andersson, and Nikola Vasiljevic. 2017. *Taxonomy and meta data for wind energy R&D*.
<https://zenodo.org/record/1199489#.Xefe3ehKhdi>.

Deliverable D11 – Data Sharing, Storage, Security Protocols, and a Specification of a Potential Data Sharing Portal

January 2021

Danielle Prezioso, Pacific Northwest National Laboratory
Anna Maria Sempreviva, Technical University of Denmark
Alice Orrell, Pacific Northwest National Laboratory



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <<https://www.ntis.gov/about>>
Online ordering: <http://www.ntis.gov>

Deliverable D11 – Data Sharing, Storage, Security Protocols, and a Specification of a Potential Data Sharing Portal

January 2021

Danielle Prezioso, Pacific Northwest National Laboratory
Anna Maria Sempreviva, Technical University of Denmark
Alice Orrell, Pacific Northwest National Laboratory

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Acknowledgments

The authors wish to thank Pawel Garcanski (Centro Nacional de Energías Renovables) for sharing his insights on ShareWind and Nikola Vasiljevic (Technical University of Denmark) for discussing metadata schemes and implementation with us.

Acronyms, Abbreviations, and Definitions

catalog	the collection of resource metadata
digital object	an object in digital format; a generalization of data
DTU	Technical University of Denmark
FAIR	Findable, Accessible, Interoperable and Reusable
IEA	International Energy Agency
IRPWind	Integrated Research Programme in Wind Energy
platform	website hosting the portal
PNNL	Pacific Northwest National Laboratory
portal	interface allowing a user to access the catalog of metadata
resource	a reference that could include, but is not limited to, datasets, analysis, tools, and publications

Contents

Acknowledgments.....	ii
Acronyms, Abbreviations, and Definitions	iii
Contents	iv
1.0 Introduction	1
2.0 Data Sharing	2
3.0 Specification of a Potential Data Sharing Portal.....	4
4.0 Storage.....	5
5.0 Security Protocols for Metadata Stored on Platform	6
6.0 Next Steps.....	7
Appendix A – Comparison of the platforms and portals considered for IEA Task 41.	A.1

Figures

Figure 1: Partial screenshot of the metadata collection form distributed to task members.....	3
---	---

1.0 Introduction

Pacific Northwest National Laboratory (PNNL) and Technical University of Denmark (DTU) completed this deliverable as part of Work Package 2: Data Information Catalog for Distributed Wind Research for the International Energy Agency (IEA) Wind Technology Collaboration Programme Task 41: Enabling Wind to Contribute to a Distributed Energy Future. As part of the work plan, Deliverable D11 requires the development of data sharing, storage, and security protocols for metadata to be stored on the platform, if needed. The specification of a potential data sharing portal that expands on the catalog is also required. The deliverable was parsed into five sections to ensure its completion, and the format of the paper follows the same order. Namely, this report addresses:

- Data sharing,
- The specification of a potential data sharing portal,
- Storage,
- Security protocols for metadata, and
- Next steps.

2.0 Data Sharing

A major undertaking for this deliverable was beginning to collect resource metadata from task members to build content for the catalog. This effort included transforming the metadata needs from Deliverable D10 into an Excel form that was distributed to all task members. The form, for which a partial screenshot is shown in Figure 1, includes the metadata elements¹ defined below:

- title: name given to the resource
- creator: entity, or entities, primarily responsible for making the resource; this might be a person, organization, or a service
- publisher: entity responsible for making the resource available; this might be a person, an organization, or a service
- contributor: entity responsible for making contributions to the resource; a secondary figure to the creator; this might be a person, an organization, or a service
- date: date associated with the creation or availability of the resource, such as the publish or release date
- type: nature or genre of the resource
- language: language of the resource
- rights: information about rights held in and over the resource; this could include a statement about various property rights associated with the resource, including intellectual property rights
- description: account of the resource
- subject: key words describing the resource
- coverage: spatial and/or temporal characteristics of the resource, such as a country or time frame relevant to the resource
- variables: parameters measured or tracked within the resource
- external conditions: context in which the resource is relevant
- activity: type of action conducted within the resource
- instrument: device used within the resource
- model: type of model used within the resource
- material: material components of wind turbine
- format: file format or physical medium of the resource
- identifier: unambiguous reference to the resource, such as a hyperlink, ISBN, or publisher number

¹ A future version of the metadata form will include “Free Keywords,” which will allow users to put in any keywords associated with the resource (i.e., there is no controlled taxonomy). This metadata element can help identify new trends in wind energy vocabulary as established in Haakon, Lund, Anna Maria Sempreviva *Semi-automatic taxonomy development for research data collections: the case of wind energy*. Proceedings of ISSI 2019: 2764-2765. zenodo.org/record/3368636#.X_es69hKiUk.

- notes: additional information about the source not yet cataloged through other entries

These metadata elements are a combination of those included in the Dublin Core Metadata Initiative and Integrated Research Programme in Wind Energy (IRPWind) wind energy metadata scheme. Of the aforementioned metadata elements, controlled taxonomies were implemented for type, subject, variables, conditions, activity, instrument, model, and material. The taxonomies selected for subject, variables, conditions, activity, instrument, model, and material draw upon the terms contained in IRPWind taxonomy², and the taxonomy selected for type was developed for Dublin Core Metadata Initiative³. As part of the IRPWind taxonomy, DTU has published wind energy topics⁴ and wind energy parameter taxonomies⁵ based on the simple knowledge organization system⁶. The goal is to create a community-based taxonomy that wind energy researchers can add to over time (e.g., via GitHub).

Task members were asked to record the metadata for one resource per tab in the Excel form and to fill in as much information as possible with the understanding that editing and updates will continue to take place. At the time this report was published, metadata for 15 resources had been cataloged. The metadata forms completed by task members were aggregated into a single Excel file, which largely resembles the structure of the metadata collection form that was distributed to task members. This file is easy to update as we continue to build additional content, and it also enables batch uploads of information once a platform is selected. In the near term, this file will be shared on the task website to enable collaboration across the task.

Metadata Element:	Title	Creator	Publisher	Contributor	Date	Type
Definition	A name given to the resource	An entity, or entities, primarily responsible for making the resource. This might be a person, an organization, or a service.	An entity responsible for making the resource available. This might be a person, an organization, or a service.	An entity responsible for making contributions to the resource, a secondary figure. This might be a person, an organization, or a service.	A date associated with the creation or availability of the resource such as the publish or release date.	The nature or genre of the resource
Example	Vestas V52 Wind Turbine - Dundalk Institute of Technology	Dundalk Institute of Technology	Dundalk Institute of Technology		2006-2019	Dataset
Your Resource						

Figure 1: Partial screenshot of the metadata collection form distributed to task members.

² Sempreviva, Anna Maria, Allan Vesth, Christian Bak, David Robert Verelst, Gregor Giebel, Hilmar Kjartansson Danielsen, Lars Pilgaard Mikkelsen, Mattias Andersson, and Nikola Vasiljevic. 2017. *Taxonomy and meta data for wind energy R&D*. <https://zenodo.org/record/1199489#.Xefe3ehKhdi>.

³ <http://dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/>

⁴ <https://data.windenergy.dtu.dk/ontologies/view/wtax/en/>

⁵ <http://data.windenergy.dtu.dk/ontologies/view/wep/en/>

⁶ <https://www.w3.org/2004/02/skos/>

3.0 Specification of a Potential Data Sharing Portal

Task 41 members assessed existing platforms (websites hosting portals) and portals (user-interfaces to access a catalog) to inform the selection of a platform to host the portal or help determine if the catalog could be embedded into an existing portal. The table in Appendix A compares the three options considered in this assessment: OpenEI, ShareWind, and a least-cost option (i.e., an Excel sheet). OpenEI and ShareWind are both platforms and portals. They each have existing portals on their platforms to which Task 41 could contribute resources, and they could each host a separate portal specifically for the task. We categorized each option's characteristics and practical considerations for selection. The proposed plan for OpenEI in Appendix A includes a custom portal implementation for Task 41. Some of the existing foundation on which OpenEI operates will support an individualized portal for Task 41 using our selected taxonomies and metadata scheme. This does not require data owners to upload their data to a repository, which is traditional for the broader OpenEI portal. Alternatively, ShareWind can implement a community (i.e., a portal) under their existing portal structure or build a custom interface for Task 41 using the platform's source code. Both of these options are described in the ShareWind column of the table. Finally, the least-cost option for the portal is posting a static Excel sheet to either the traditional OpenEI repository or to the Task 41 website that contains the task's collected metadata. This static sheet would be periodically updated and re-uploaded to the chosen platform.

4.0 Storage

While storing data within the catalog or portal was initially discussed, the task has decided against the practice to better ensure the privacy and control of data while also avoiding replicating work. Data will remain stored at the owner's premises or chosen repository. Any burden of preservation, safety and security of digital objects, and preserving the data remains with the data owner. Much of this decision was informed by the findable, accessible, interoperable, and reusable (FAIR) data principle.

In 2016, the European Commission adopted the FAIR data principle in response to industry feedback of the Open Data policy introduced in the Horizon 2020 Framework Programme⁷ that began in 2014. The Open Data policy was created to optimize the impact of publicly funded projects by making data associated with those efforts available across Europe. However, while there is general agreement in the wind energy community that sharing data decreases the time from new ideas to innovation and boosts a digitalization workflow, data is still considered an important asset that can provide a competitive advantage. As such, there is also reluctance to share important data. This created a shift to data findability (the F of FAIR) and subsequently visibility. Increased visibility of data, tools, and workflows leads to increased workflow efficiency within organizations, as it can reduce the amount of repetition of similar analysis tasks.

One approach to this, employed by IRPWind, is to develop a dissemination platform that provides the information to three key types of digital objects available at partner organizations: data sets, tools, and workflows. The FAIR data principle underpins the structure of the database as each digital object is included in a searchable metadata catalog. The catalog is increasingly populated, and ultimately, all data sets and tools available in the linked organizations can become part of a large catalog available to internal users. A selection of data, tools, and workflows with high relevance for external stakeholders are made available to external users.

If task members are interested in posting their datasets for public use in addition to providing metadata for the task catalog, several efforts have already created places to do so. Zenodo and OpenEI are two options. Zenodo was formed under the OpenAIRE project to provide researchers with an open access option to publish their data. Although Zenodo originated in Europe, the site allows researchers across the globe to publish their data and provides a DOI for resources as well⁸. Alternatively, OpenEI largely contains US-focused sources, and includes a limited amount of international data. The site is driven by community input and is simple to use⁹. Both Zenodo and OpenEI, however, lack data standards.

⁷ <https://ec.europa.eu/programmes/horizon2020/en>

⁸ <https://about.zenodo.org/>

⁹ <https://openei.org/wiki/Information>

5.0 Security Protocols for Metadata Stored on Platform

Because a platform has not yet been selected, it is not possible to determine security protocols for the metadata. Once a platform has been selected, security protocols will be reassessed and considered with the platform experts.

6.0 Next Steps

Task 41 members will continue to evaluate platform options in the coming months. In parallel, PNNL and DTU will prepare the collected metadata for the portal, collect additional metadata from task participants, and optimize the content of the catalog. This effort will include implementing updates to the taxonomies and metadata elements to ensure seamless use for portal users.

Appendix A – Comparison of the platforms and portals considered for IEA Task 41.

CATEGORY	OPEN EI https://openei.org/wiki/Main_Page	SHAREWIND https://sharewind.eu/ <small>Gancarski, Pawel. 2019. Sharewind - wind data registry. Presented at the <i>Wind Energy Science Conference 2019 (WESC)</i>, Cork: Zenodo. http://doi.org/10.5281/zenodo.3357860.</small>	LEAST-COST OPTION (EXCEL SHEET ON OPEN EI)
WEBSITE STATUS	Fully operational	Prototype only	Fully operational
TAXONOMIES/METADATA CARDS	Ability to implement any newly created metadata card and selected taxonomies.	Already implemented a version of the wind energy taxonomy and a version of the meta data card. We could create a Distributed Wind community with additions to the existing taxonomies if desired.	Can implement a reduced version of the metadata and taxonomies.
STORING DATA	For each entry in the portal, there is the option to upload and store data and/or track the resource through a metadata entry. Ability to accommodate different resource formats.	Only tracks metadata. Option to store very limited amounts of data.	Possible to upload data to OpenEI or Zenodo separate from the static Excel sheet catalog of resources, if data owner wants to make the resource publicly available.
CONNECTION/COLLABORATION	Working relationship with National Renewable Energy Laboratory and PNNL.	Working relationship with DTU. Looking for collaborations to add content to the service.	Working relationship with National Renewable Energy Laboratory and PNNL.
WEBSITE ACCESS	Option to limit website access to task participants or create a fully public data portal. Flexibility exists to initially start with a private portal and later open it to the public, if desired.	No option to create a private portal for Task 41. We could, however, restrict who posts to the community.	Fully open source.
DATA ANALYTICS	Generates data analytics to inform stakeholders on catalog visitors and behavior.	This does not currently exist, but would be possible to set up with some small budget.	None

CATEGORY	OPEN EI https://openei.org/wiki/Main_Page	SHAREWIND https://sharewind.eu/ Gancarski, Pawel. 2019. Sharewind - wind data registry. Presented at the <i>Wind Energy Science Conference 2019 (WESC)</i> , Cork: Zenodo. http://doi.org/10.5281/zenodo.3357860 .	LEAST-COST OPTION (EXCEL SHEET ON OPEN EI)
CONNECTION WITH OTHER IEA TASKS	None	Working to connect with other IEA tasks.	None
TIMELINE	Estimated nine months for completion (based on original scope of work).	Up to a couple of months. If it takes more time than that to create the portal we envision, then we'd be forcing ShareWind to become something it is not.	Likely only a couple weeks to format the existing file and coordinate the upload
UPFRONT COSTS	Proposed scope of work costs: \$70,000.	Undetermined (Building a community within Sharewind would be a lower cost. Using ShareWind source code to create a user-friendly Task 41 portal would be a higher cost.)	Task members' labor costs only.
MAINTENANCE COSTS	"Minimal" continued maintenance costs.	Undetermined	
EXISTING AUDIENCE	Largely USA/North American	European	Largely USA/North American (OpenEI), international (Task 41 Website)
COLLECTING RESOURCES	Manual data collection and uploads.	Manual data collection and uploads.	Manual data collection and uploads.
LONG-TERM SUPPORT	Yes	No	No
USER INTERFACE	Customizable, intended for browsing the catalog of data.	Implementing a custom interface would be timely. The existing interface is intended to help discover resources, not browse through an entire set.	None. In this option, the Excel sheet would not be directly editable or interactive. Periodic, local updates to the Excel list could be made and then a new file could be uploaded.

CATEGORY	<p style="text-align: center;">OPEN EI</p> <p style="text-align: center;">https://openei.org/wiki/Main_Page</p>	<p style="text-align: center;">SHAREWIND</p> <p style="text-align: center;">https://sharewind.eu/</p> <p style="text-align: center;">Gancarski, Pawel. 2019. Sharewind - wind data registry. Presented at the <i>Wind Energy Science Conference 2019 (WESC)</i>, Cork: Zenodo. http://doi.org/10.5281/zenodo.3357860.</p>	<p style="text-align: center;">LEAST-COST OPTION (EXCEL SHEET ON OPEN EI)</p>
ADDITIONAL CONSIDERATIONS	Can easily communicate via an API, so it could “publish” datasets from other databases, like OpenEI for example.		
EXAMPLES	https://openei.org/wiki/RAPID	https://sharewind.eu/	https://openei.org/datasets/dataset/ane-mometer-data-wind-speed-direction-for-pascua-yagui-az-2003-2004 or https://data.pnnl.gov/ or https://community.ieawind.org/task41/home
NOTES	Action E84-1 Operating Agent working group to propose a strategy on FAIR data. Task 25 has a database deliverable. Task 42 has a data-sharing milestone.		

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov