

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>.