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

January 2021

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

Acknowledgments

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

Figures

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 elements[1] 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 taxonomy[2], and the taxonomy selected for type was developed for Dublin Core Metadata Initiative[3]. As part of the IRPWind taxonomy, DTU has published wind energy topics[4] and wind energy parameter taxonomies[5] based on the simple knowledge organization system[6]. 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 Programme[7] 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 well[8]. 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 use[9]. 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/ 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.	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 *Wind Energy Science Conference 2019 (WESC)*, 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	OPEN EI https://openei.org/wiki/Main_Page	SHAREWIND https://sharewind.eu/ 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.	LEAST-COST OPTION (EXCEL SHEET ON OPEN EI)
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/anemometer-data-wind-speed-direction-for-pascua-yaqui-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