Integrated Research Programme on Wind Energy

Project acronym: **IRPWIND**
Grant agreement nº 609795
Collaborative project
Start date: 01st December 2013
Duration: 4 years

Taxonomy and meta data for wind energy R&D
Work Package 2 - Deliverable D2.3

Lead Beneficiary: DTU
Delivery date: 12 December 2017
Dissemination level: PU

The research leading to these results has received funding from the European Union Seventh Framework Programme under the agreement 609795.

Supported by:

Author(s) information (alphabetical):		
Name	**Organisation**	**Email**
Anna Maria Sempreviva	DTU	anse@dtu.dk
Allan Vesth	DTU	alve@dtu.dk
Christian Bak	DTU	chba@dtu.dk
David Robert Verelst	DTU	dave@dtu.dk
Gregor Giebel	DTU	grgi@dtu.dk
Hilmar Kjartansson Danielsen	DTU	hkda@dtu.dk
Lars Pilgaard Mikkelsen	DTU	lapm@dtu.dk
Mattias Andersson	DTU	matan@dtu.dk
Nikola Vasiljevic	DTU	niva@dtu.dk

Aknowledgements/Contributions:		
Name	**Name**	**Name**
Hans Chr. Bolsted	Koen Hermans	Tor Inge Reigstad
Jan Willem Wagenaar	Pawel Gancarski	
Javier Sanz Rodrigo	Stephan Barth	

Document Information

Version	Date	Description		
		Prepared by	**Reviewed by**	**Approved by**
01	12-12-2017	Anna Maria Sempreviva	Malene Pogager	Mattias Andersson

Table of Contents

Executive Summary

The Open Access (OA) to knowledge is a principle established by the European Commission, underlying the H2020 EU Framework Programme for Research and Innovation. OA aims at optimizing the impact of publicly funded projects, by making information openly available and reusable to everyone in Europe. The Open Data (OD) policy is part of the OA strategy and is widely acknowledged as a fundamental step to support a fast track from research to innovation. Although there is a general acknowledgment for the need of OD, a mindset similar to the "not-in-my-backyard" holds back the scientific and industrial communities to implement a joint OD policy. This is partly due to the fear that sensitive and proprietary data could be misused.

To overcome this problem, the European Commission posed an important milestone by declaring that data must be at the same time “Findable, Accessible, Interoperable and Re-usable (FAIR)” [1] and “as much open as possible, and as closed as necessary”[1].

The , Joint Programme on Wind Energy of the European Energy Research Alliance (EERA JPWind)[2], represents the largest public European scientific community in the Wind Energy (WE) sector. JP WIND recognises the necessity of implementing an OD plan by setting the goal to create a data portal. The data portal will a) collect information on data from “cloud distributed” data centers, b) catalogue the collected information and c) provide end-users with tools to find data for their needs.

In this report, we focus on the first phase that lays the basis for the implementation of a Data Web Portal i.e. the information architecture to make data Findable and Interoperable.

The first phase relates to making data “Findable” and “Interoperable” helping data owners to describe the data and end-users to accurately locate and retrieve the needed data. There are two components for this task: (i) Metadata (data tagging) and (ii) taxonomy for the WE sector topics, the topic related data and descriptive types of metadata.

(i) Metadata

To accurately locate specific datasets, they should be tagged with a series of information, metadata, using so-called metadata cards. Besides preserving the information on data for a future re-use, metadata are used for indexing datasets to refine their findability. Metadata are classified into three categories: descriptive, administrative and structural. Descriptive metadata provide information on e.g. what (associated topic, type of variables, etc.), where data were collected (external conditions or geographical location, etc.) or how data were collected (instruments, activity type). Administrative metadata provide information on e.g. who collected the data (data owner), access rights, links to data, etc. Structural metadata provide information on e.g. data format. In this task, we use standard metadata defined in the Dublin Core metadata element set [3].

(ii) Taxonomy

Taxonomy is the descriptive type of metadata containing terms that assign textual information to the data. In a broad sense, it is any means of organizing concepts of knowledge.. The

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[2] https://www.eera-set.eu/eera-joint-programmes-jps/wind-energy/

classification of disciplines into e.g. Environment, Climate, Agriculture, Engineering etc is an example of a taxonomy. In a narrow sense, taxonomy is a hierarchical classification or categorization system as we know from e.g. the classification of species. In this report, taxonomy is used to put data into the correct context by defining and hierarchically classifying the WE research area topics and organize data within topics.

A good taxonomy enables users to immediately grasp the overall structure of the knowledge domain and the associated data. Practically, taxonomy terms are used as a controlled wind energy vocabulary by data owners for tagging data in the metadata card and by end-users as “facets” to filter content progressively via a “faceted search”. Furthermore, the taxonomy insures Interoperability[2].

The main deliverable of this task is a set of taxonomies: the taxonomy of the topics distinctive of the WE sector and the taxonomy of the data type relevant to different topics and taxonomies of other facets. The first step to create this was to choose the number of hierarchical levels with top-topics and sub-topics. To keep the topic taxonomy structure simple, the development of taxonomy levels ended as soon as the next narrower level reached the “data” dimension. The following case is given as an example: the topic “Siting” includes, amongst others, “wind mapping” for prospective sites. The “wind mapping” activity needs time series of wind speed and direction, and terrain roughness and orography data.

Other taxonomies were created for facets to describe data: External Conditions, Activities, Instruments, Models, and Materials. The following case is given as an example: to perform the resource assessment offshore in Denmark, wind speed and directions from long-term observations using a wind lidar are needed. The search would be:

Siting (Topic),
└Wind Resources (Subtopic),
└Offshore, (External conditions),
└Long-term monitoring (Activity type),
└Wind lidar (Instrument) and
└Wind speed and direction (Data type).

Conclusions

With metadata cards, describing data made available by each organization, data can be searched through a data portal containing a metadata catalog updated by a web crawler, i.e. a program continuously harvesting metadata cards. The data itself resides on the data owner domain and security and data management issues remain in the hand of the data owner.

A user will access the portal to submit a query containing keywords from the established vocabularies from the taxonomy of the metadata. The system will return an optimized list of available data. Data can be accessed either directly via provided download links in the metadata card or by contacting data owners.

This approach has a two-fold purpose: to make data owners feel more comfortable in sharing data by maintaining the control on data access and data use, while end-users will access information on datasets needed for a specific goal optimising time and funding. Both data owners and end-user will have the opportunity to start or reinforce collaboration activities.

Introduction

Openness of data is nowadays one of the most demanded key indicators by worldwide funders when issuing calls for public funding applications. To support this, the European Commission has established the concept of Open Access (OA) to knowledge as the general principle for the H2020 with the aim of making information, paid by public money, open and available to all in Europe. The main goal is to optimize the impact of funded projects, thereby providing a large community access to information for several purposes e.g. test new ideas, to verify and re-use produced results. .

The Open Data (OD) policy is part of the Open Access strategy and is widely acknowledged as a fundamental step to provide answers to scientific, technological, social and economic challenges and make informed decisions. Despite the general acknowledgment of the importance of Open Data, there is still a mindset similar to the "not-in-my-backyard" that holds back the scientific and industrial communities to implement a joint Open Data policy. One of the major issues is the fear that the several types of sensitive and proprietary data would be openly distributed or even worse, misused. The European Commission clearly acknowledges the problem and posed an important milestone by declaring that data[i] must be at the same time Findable, Accessible, Interoperable and Re-usable, following the *FAIR Guiding Principles* [1] and "*as much open as possible, and as closed as necessary*"[1], while imposing a data management plan to all H2020 projects.

The EERA JP WIND[3], as the largest public European scientific community in the wind energy sector, recognisesthe necessity of implementing an open data strategy and has taken a first step toward the design of an Open Data strategy for the sector within the Work Package 2 *Integration activities* of the FP7 project *Integrated Research Programme in Wind Energy, IRPWind.*

A wind farm life cycle involves different activities in different phases: planning, financing, environmental reporting, selecting the right turbines, designing a wind farm with its components, installing, commissioning, operating, maintaining, decommissioning and recycling. All these activities demand informed decisions that can only be taken with the right information at hand. Data are the building blocks for information: to create information, we need to find, access, process and interpret the data. However, data has no meaning if outside a context. Therefore, it is necessary to organize data in topics and accurately describe data in order to be found and used.

The Wind Energy sector "Big Data" is web distributed with high need of data organization at both intra- and inter-institution levels. In this view, there is high demand for an easier access to data and information, and improved internal and external collaboration. However, although the wind energy research community is very dynamic and multidisciplinary, there is a strong awareness of IP rights being projected from the industry. This makes the tasks of standardization, collaboration, integration and knowledge exchange challenging. Bearing this in mind, a roadmap defining guidelines, for paving the way to free flow of information is required in IRPWind and EERA JP WIND.

[3] European Energy Research Alliance Joint Programme, EERA JP https://www.eera-set.eu/eera-joint-programmes-jps/wind-energy/

The overall goal of the IRPWind (and EERA JP WIND) initiative on Open Data is consequently to create and demonstrate a web portal to search and locate cloud distributed data relevant for wind energy.

The first phase relates to making data "Findable" and "Interoperable".

To allow a search engine to precisely find data, data owners must tag datasets. This is achieved by associating to datasets a series of information "Metadata", e.g. what (associated topic(s), variable(s)), where data were collected (geographical location) or where they are stored, when, how long, who, format, access rights, etc. This information is included in a metadata card that contains a list of pre-established element that can be used as "facets" to filter content progressively by users.

The challenge here is to create standard vocabularies to function as a common understanding between data owners and users; common vocabularies will guide data owners to accurately fill the metadata card and user to accurately search for the right data. Common vocabularies are generated and organized by a topical classification i.e. the Taxonomy. A taxonomy is a descriptive type of metadata answering to the "what " tag and it is needed to put data into the correct context.

A taxonomy of the topics of a research area enables an immediate understanding of the overall structure of the knowledge domain and allows a classification and allocation of associated data. A good example of the application of taxonomy is an organizational chart showing the structure of an organization in department and sections each covering the main activities the sector and each in need of specific data and models.

The taxonomy of the topics combined with taxonomies of other descriptive metadata elements such as data type, instruments, external conditions etc. will allow the data owner to allocate data in a research area while helping data users to accurately find data for their objective.

The international wind energy sector community has no yet jointly established neither metadata standards nor a taxonomy classifying the research topics for scientific purpose.

A specific taxonomy is "The Distributed Wind Cost Taxonomy"[4] by the National Renewable Energies Laboratory, NREL, published in March 2017. NREL classified the component of a wind energy project in main categories related to the cost of planning, building and operating a wind farm. NREL followed a system engineering approach applied to estimate the cost of wind energy, splitting the phases of the life cycle of wind farms in components each with an associated cost model.

The first top levels are:

1. Wind Turbine System Equipment in CapEx: Capital Expenditures
2. Balance Station in CapEx (Including the cost of all activities for "getting the turbine in the ground" (resource assessment, etc)
3. Operation and Maintenance (OpEx: Operation Expenditures)

In this report, we focus on the process followed to decide metadata elements and create the taxonomy of the wind energy sector topics and set of taxonomies for focusing on data, and the steps to design and implement a Wind Energy data portal.

[4] http://www.nrel.gov/docs/fy17osti/67992.pdf

1. Need for a data portal in the Wind Energy sector

Wind energy is a well-defined sector with well-defined data need with a broad variety of data types from geospatial to meteorological variables to turbine power curves and curtailment etc; activities in different environments (external conditions) and using different type of instruments and platforms. However, this enormous amount of data is spread over several groups at several institutions and industry and mostly invisible to search and protected behind firewalls. Therefore, for finding data, we need to go beyond the classic concept of databases, where datasets are uploaded by data owner in predetermined formats, minimally described. Though the classical database platform has the advantage to be thematic and collects data suitable for a specific project, it has the following disadvantages:

- *Data owner's "inertia".* There is a slow response by data owners because it needs an extra effort to extract and structure data according to agreed standards and update data according to specific project needs.
- *Missing the "Big Picture".* Stored data are usually a subset of available data, so the "Big Picture" of the dataset context that could unveil a possible inter-disciplinarily aspect of the data can be overlooked.
- It has to be *maintained and updated continuously.*

An *alternative approach* is to leave datasets web distributed (cloud) at the data owner premises like the WMO Arctic Web Portal[5], GEOSS Discovery Web Broker[6] [2]etc..

If data are organized and cataloged according to common metadata standards e.g. INSPIRE[ii]-Dublin Core[iii], data can be searched through a "Distributed Data Archive Portal" DDAP with a user-friendly interface.

The DDAP contains a metadata catalog, updated by a program continuously harvesting metadata cards (web crawler) that allows an accurate search for relevant data; at the other end, the requirements for the owners is to compile and expose metadata cards describing datasets, by filling a metadata card using taxonomy (common vocabularies). This approach has the following advantages:

- Data that answer the need for a specific project can be found together with administrative information, such as access rights and eventually retrieved on a collaboration agreement if not directly available.
- Confidentiality and security issues are the responsibility of the data owner that would administrate the access right. One of the metadata elements will be the access information, e.g. open access, constrained, and confidential. Each data owner is free to choose a suitable repository to preserve the data.
- The data portal can be linked and searchable by other organizations or providers outside EERA Wind Energy e.g. EU (EUData, Zenodo) and International (WMO, Met.no Arctic data center, US (NCAR/UCAR)). In this way, data will be multi-purpose serving to design inter-sectoral, inter-disciplinary projects as well.

[5] http://arcticdata.met.no/metamod/search
[6] http://www.geodab.net/

- Data owners would feel more comfortable in sharing data by maintaining the control on data access and how data are used; on the other hand, end-users will access information on datasets needed for a specific goal optimising time and funding. Both data owners and end-user will have the opportunity to start or reinforce collaboration activities.

Data portal for a distributed database

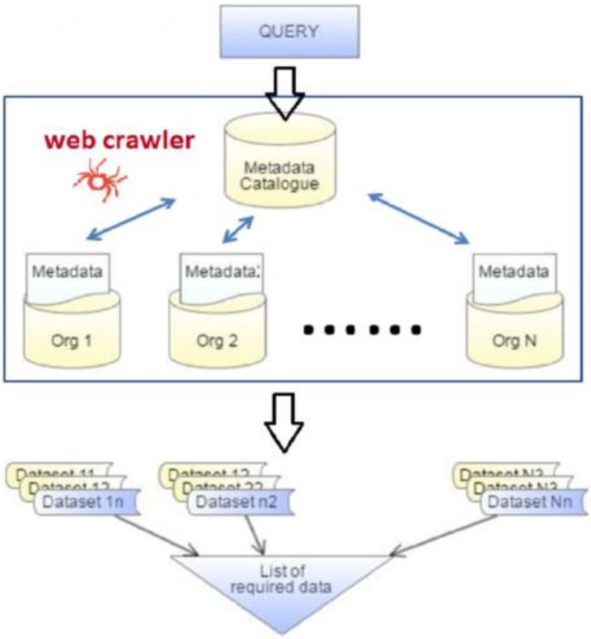

Figure 1 Sketch of a portal for distributed data.

A distributed database is made of datasets residing in the web at the premises of data owners. It includes a metadata catalog containing the metadata describing the available datasets.

Data owners expose up-to-date metadata cards describing each dataset.

A web crawler continuously harvests metadata cards ensuring that the catalog is up-to-date

Access to data

A user will access the DDAP and submit a query containing keywords from the established taxonomy and metadata catalogs.

The system will return a list of available data with the needed information to evaluate

Data can be accessed either directly or indirectly by requesting access to the data owners.

2. Preparing metadata and taxonomy of the Wind Energy sector

The first phase to lays the basis for the implementation of a Data Web Portal is the information architecture to make data Findable and Interoperable.

To help data owners to describe the data and end-users to accurately locate and retrieve the needed data, we need to define the metadata elements and the taxonomies.

In this section, we describe the process followed to develop taxonomies.

2.1 IRPWind Core Group for Open Data

DTU has been tasked to suggest standard metadata elements as well as to develop a taxonomy for wind energy topics, data types and other facets for data tagging.

The DTU task has been led by Anna Maria Sempreviva who coordinated a working group including Scientists from the DTU Wind Energy Department expert on different topics. This work was supported by Mattias Andersson, leader of IRPWind WP2, and Nikola Vasiljevic, leader of the task of designing a data portal.

R&D area experts:

- Anna Maria Sempreviva Resource assessment and wind conditions
- Christian Bak Aerodynamic design
- David Robert Verelst; Wind turbine loads & control
- Gregor Giebel Integration & Planning
- Hilmar Kjartansson Danielsen; Materials science and characterization
- Petr Maule , Informatics
- Lars Pilgaard Mikkelsen Composites and Materials Mechanics
- Nikola Vasiljevic, Meteorology and Remote Sensing
- Allan Vesth, Test and Measurements

The working group performed the following activities:

1) A desk review of documents and a*nalysis of the current state of the art of taxonomies and metadata used in Wind Energy.*

2) *Analysis of existing international activities on data management and needs of users such as access to information and knowledge within and outside the organizations.*

3) *Incremental development of a first draft of the Wind Energy sector taxonomy for topics and for the related data in the form of hierarchical keyword lists defining a common vocabulary.*

4) *Analysis of current search engines in use and suggest requirement specifications needed to improve search capacities.*

The suggested taxonomy has been presented to and commented by the test members of the core group including:

- Stephan Barth from ForWind - University of Oldenburg (GE),
- Javier Sanz Rodrigo and Pawel Garcarski from CENER (SP),
- Tor Inge Reigstad and Hans Christian Bolsted from SINTEF (NO), and
- Jan Willem Wagenaar and Koen Hermans from ECN (NL).

2.2 Metadata

Metadata is information on data, needed to assure that the data is documented for future re-use and put into the correct context.

There are three types of metadata: descriptive (e.g. topic, datatype, abstract, etc.), structural (e.g. format) and administrative (access rights, data owner, etc.).

Besides preserving the information on data for a future re-use, metadata are also used for indexing datasets to assure their findability. To this end, data is described by a so-called metadata card, generally an XML or JSON file, containing the list of metadata elements. Metadata elements must follow standards for interoperability purpose. Figure 2 shows an example of a metadata card and the 15 standard core metadata defined in the Dublin Core[7], DC, Metadata element set, chosen in this task.

Figure 3 shows the descriptive metadata assigned to data that will be used as facets to filter the data search: 2 elements are Dublin Core and 5 are not DC elements. For these 7 elements, taxonomies specific for wind energy must be created. We will come back to this issue in Sections 0 and 2.3.3.

Bibliography as an example of a metadata card

The standard for references in a bibliography, is an example of metadata card, for a library, i.e. The reference *Smith, A.C. Jones B.D., 2017, Where is the data? J. Open Science, 2.0. 00, 1-N p.*, i.e. contains the following elements: Authors, Year, Topic, Journal, Issue number, and Page.
To locate an article/book/reports etc. without a standard reference including the necessary information (metadata), a search on Smith A.C. could return hundreds of papers.

Building a metadata card might seem a tedious unnecessary step; however, if we think about how many datasets have been lost because stored without documentation, we can realize that, in a long-term, documenting data can pay off. The NetCDF files, used in the modeler community are the closest we find as an example of data with associated metadata. They are compressed files containing metadata (attributes); however, there is not yet an agreement on standard NetCDF metadata nor are they fast searchable since it is a compressed data format.

[7] http://dublincore.org/

Dublin Core metadata elements

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights


```
{
 "created": "2017-12-07T15:34:47.588179+00:00",
 "files": [
  {"bucket": "4db38d7b-6d24-4e72-bd71-160dc1b87357",
   "checksum": "md5:c8afdb36c52cf47278366669019e69222",
   "ePIC_PID": "http://hdl.handle.net/0000/myfile",
   "key": "myfile",
   "size": 9,
   "version_id": "58d452ce-e46d-44f1-8113-324d41d7f873" }],
 "id": "a1c2ef96a1e446fa9bd7a2a46d2242d4",
 "links":
          { "files": "https://35.195.185.165/api/files/4db38d7b-6d24-4e72-bd71-160dc1b87357",
            "self": https://35.195.185.165/api/records/a1c2ef96a1e446fa9bd7a2a46d2242d4 },
 "metadata":
          {  "$schema": "https://35.195.185.165/api/communities/4ba7c0fd-1435-4313-9c13-
4d888d60321a/schemas/0#/json_schema",
  "DOI": "http://doi.org/XXXX/b2share.a1c2ef96a1e446fa9bd7a2a46d2242d4",
  "community": "4ba7c0fd-1435-4313-9c13-4d888d60321a",
  "community_specific": { "fa52bec3-a847-4602-8af5-b8d41a5215bc":
          { "activity":               [ "Field experiment" ],
            "external_conditions": ["Onshore"],
            "instrument":           [ "\u2192WindScanners" ],
            "model":                         ["\u2192Fullscale"]           } },
  "contact_email": "x@example.com",
  "descriptions":
          [ {   "description": "Abstract about the dataset",
                "description_type": "Abstract" } ],
  "ePIC_PID": "http://hdl.handle.net/0000/a1c2ef96a1e446fa9bd7a2a46d2242d4",
  "keywords":
          [ "\u2192Wind farm",
            "\u2192Resource assessment" ],
  "open_access": true,
  "owners": [ 1 ],
  "publication_state": "published",
  "resource_types":
          [ { "resource_type_general": "Type1" } ],
  "titles":
          [ { "title": "Example dataset" } ]
 },
 "updated": "2017-12-07T15:34:47.588187+00:00"
}
```


Figure 2 The Dublin core standard metadata and an example of metadata card.

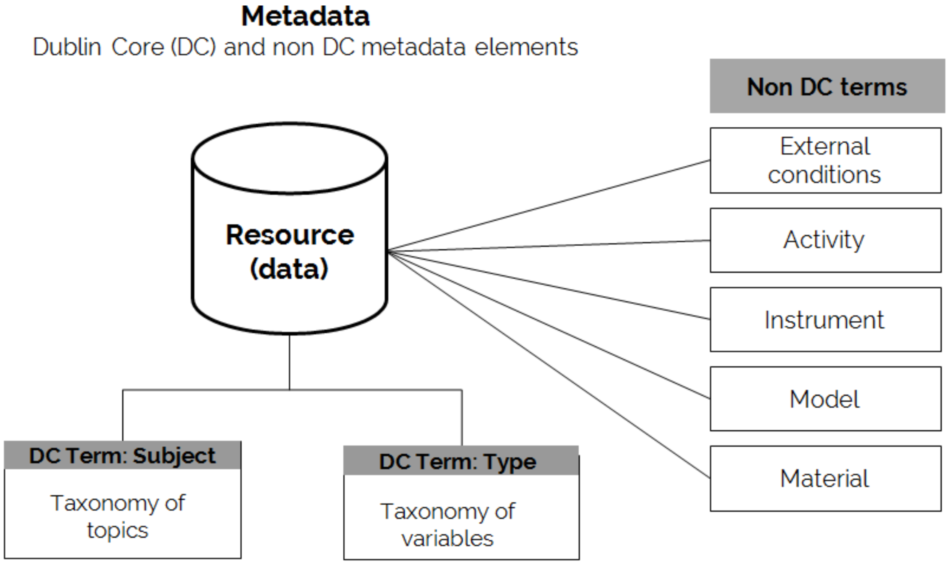

Figure 3 Data as resource is described by two Dublin Core elements and 6 non DC elements defined as facets. Taxonomies must be created for this 8 descriptive metadata.

2.3 Taxonomy

Taxonomy[3] is the descriptive type of metadata containing terms that assign textual information to the data. In a broad sense, it is any means of organizing concepts of knowledge (e.g. Environment, Climate, Agriculture, Engineering, etc.).

In a narrow sense, it is a hierarchical classification or categorization system as we know from, e.g. the classification of species, such as the Linnaean taxonomy of species [4]

In a hierarchical tree structure, it is crucial to define the top terms correctly, because it will determine how the branches of sub-topics develop.

For this report, we built several types of taxonomies. The identified terms will be a common vocabulary to choose the right terms to describe the data and be used as filters in data search:

- The taxonomy of the topics distinctive of the Wind Energy sector. This will help to tag datasets according to topics and putting data in a correct context.
- The taxonomy of the Data Types (Variables)
- The taxonomy of Activity types
- The taxonomy of External Conditions
- The taxonomy of Instruments/Measurement platforms
- The taxonomy of Materials
- The taxonomy of Models

Creating more metadata element with specific taxonomies for describing data, adds dimensions to the search allowing multiple filters and, at the same time, keeps the structure of the topic taxonomy simple. Within the taxonomy community, these are called “facets” i.e. qualifying general terms.

Example

To estimate the resource assessment in complex terrain in Spain, long-term wind speed and directions time series, at least 5 years are needed. Long-term time series can either be created by mesoscale models or by observations, better if from a cup anemometer.
The search would be:

Siting (Topic),
- Wind Resources (Subtopic),
  - Complex terrain (External conditions),
    - Wind speed and Direction (Variables),
      - Cup anemometer (instrument)
        - in Spain (geographical location)

If such observations are not found, outputs from models can be chosen under "modeling" activity and then choose the type of models.

2.3.1 Process to draft a taxonomy

There are several ways of creating taxonomies. Two classical methods follow the Expert elicitation (Top-down) and Text analysis (Bottom-up) approaches[5]:

Expert elicitation. In this approach, experts in the relevant topics of a sector are gathered for collaborating in defining the hierarchy of terms (top topics) defining the main branches of the hierarchical tree starting from the Broad Term i.e. "Wind Energy" down to the Narrower Terms.

Bottom-up approach: A bottom-up approach for building taxonomies can be undertaken in several ways, but does usually include text analysis. The text might be the collection of authors-generated "free keywords" entries in journals and/or Journal abstracts. Following this approach, the topics for inclusion in the taxonomy are selected by analyzing the clusters of keywords

In this task, we used the expert elicitation approach.

2.3.2 IRPWind topic taxonomy

The first step has been to evaluate how many hierarchical levels to create. There can be many layers in the hierarchy tree, e.g. from "wind turbine" down to the smallest component such as joints. Details depend on the purpose of the taxonomy but to keep simple the structure of the taxonomy is recommended.

For tagging data, we came to the understanding that we could stop the topic taxonomy as soon as the details reached the "data" dimension.

As an example: "Siting" includes e.g. wind mapping for prospecting sites. Wind mapping methodologies, e.g. WAsP, need time series of wind speed and direction and files with roughness and orography of the terrain. In this view, we do not need to go beyond wind mapping as a topic. Data variables needed for wind mapping will be included in the metadata category "Data Type" and a taxonomy will be developed.

It should be noted that one term should only be placed under one topic, in order to avoid ambiguities associated with data markup. A straightforward consequence is that organizing topics according to disciplines, is not the preferred way, since disciplines e.g. meteorology, aerodynamics, aeroelastics, are not exclusively relevant for Wind Energy but might be cross-cutting i.e. relevant for (i) different topics and (ii) different sectors. That is: (i) Meteorology is cross-cutting for siting, operation, and maintenance, wind integration. (ii) Meteorology is also relevant to other sectors such as Solar Energy, Wave Energy, Environment, and Climate.

There have been a series of meetings where the approach and results were discussed. From the meetings, an agreement was reached amongst the experts on the structure of the taxonomy.

Furthermore, the taxonomy of the topics was presented at the Annual IRPWind Conference in Amsterdam 25-26 September 2017.

The final Taxonomy for Wind Energy topics is shown in the following figures.

Figure 4 shows the Top level of Narrow Terms. Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9 show the development of the taxonomy of the first narrow terms Siting, Economics, Wind Turbines, Wind Power Plants, and Operation and Maintenance, respectively.

Topic taxonomy

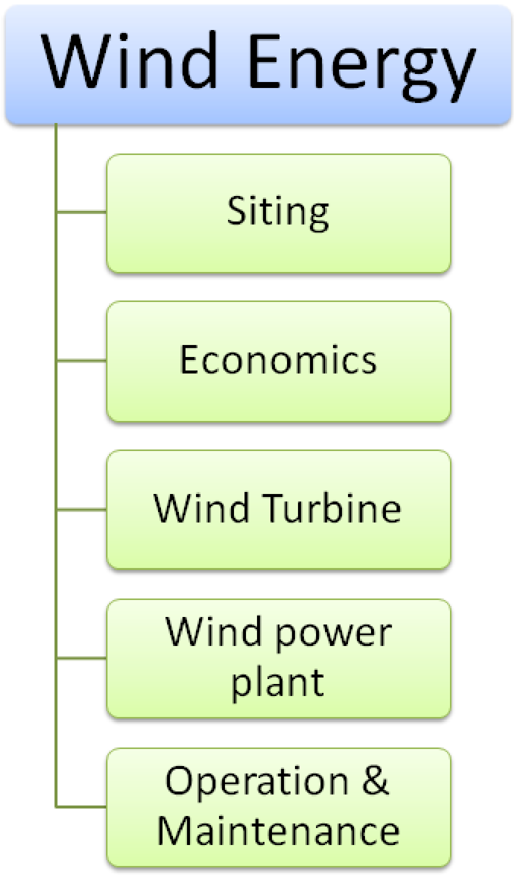

Figure 4 Top term 'Wind Energy' and first- level Narrow Terms in the taxonomy.

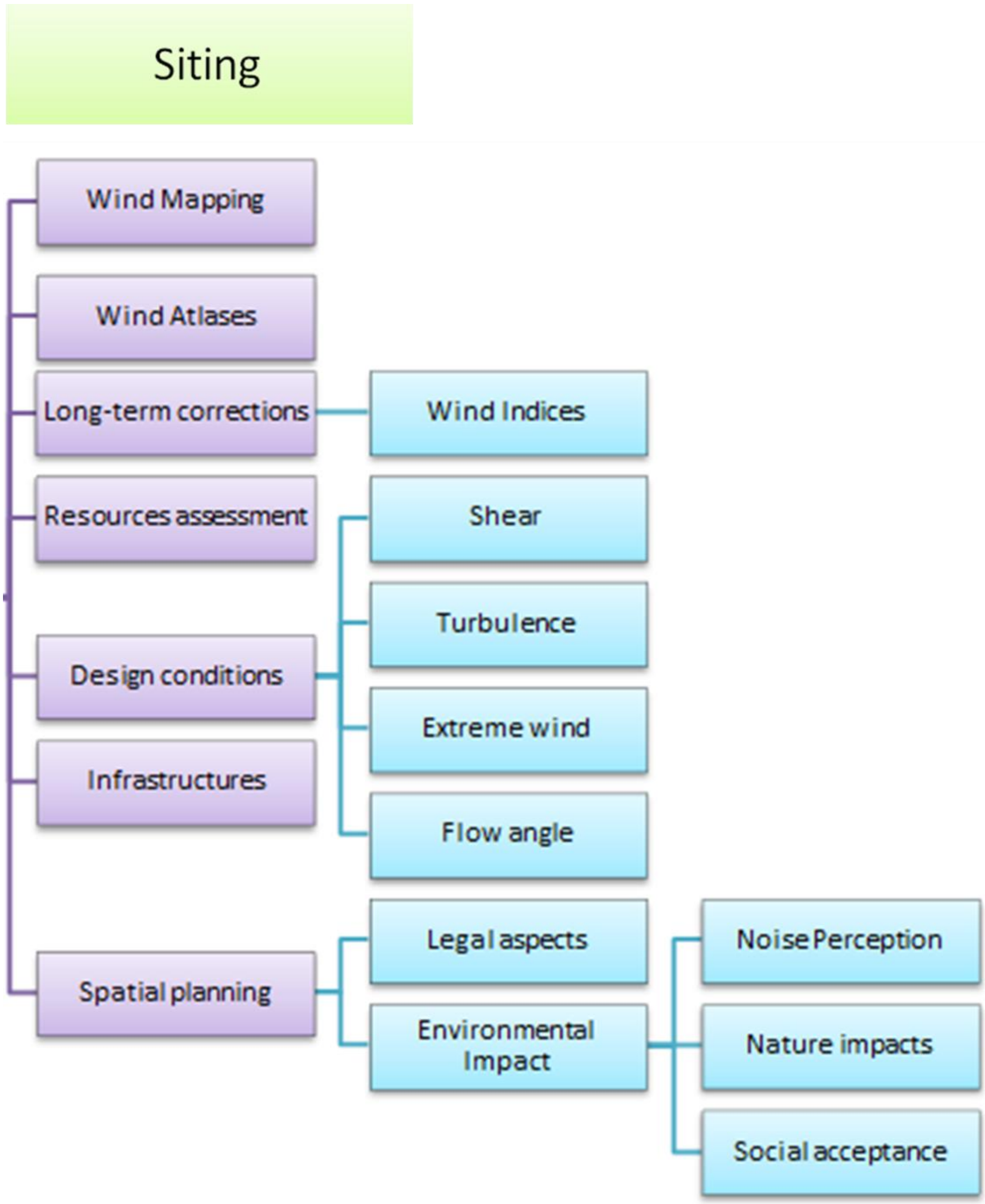

Figure 5 Narrow Terms hierarchy for Siting.

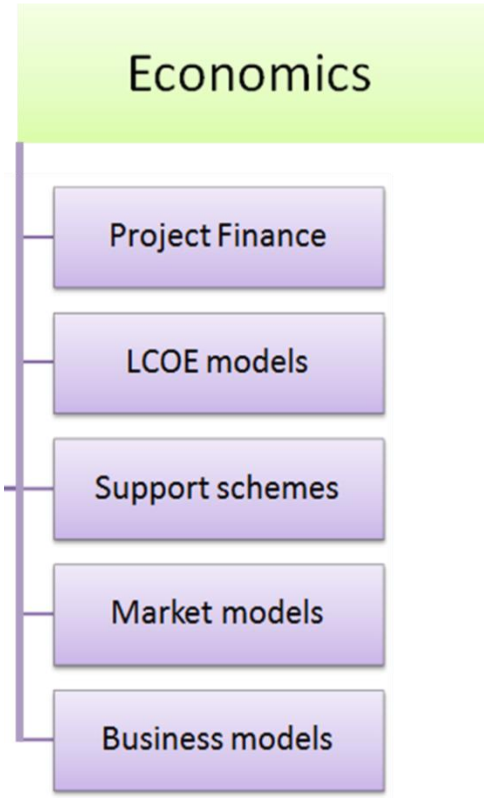

Figure 6 Narrow Terms hierarchy for Financial Aspects/Economics.

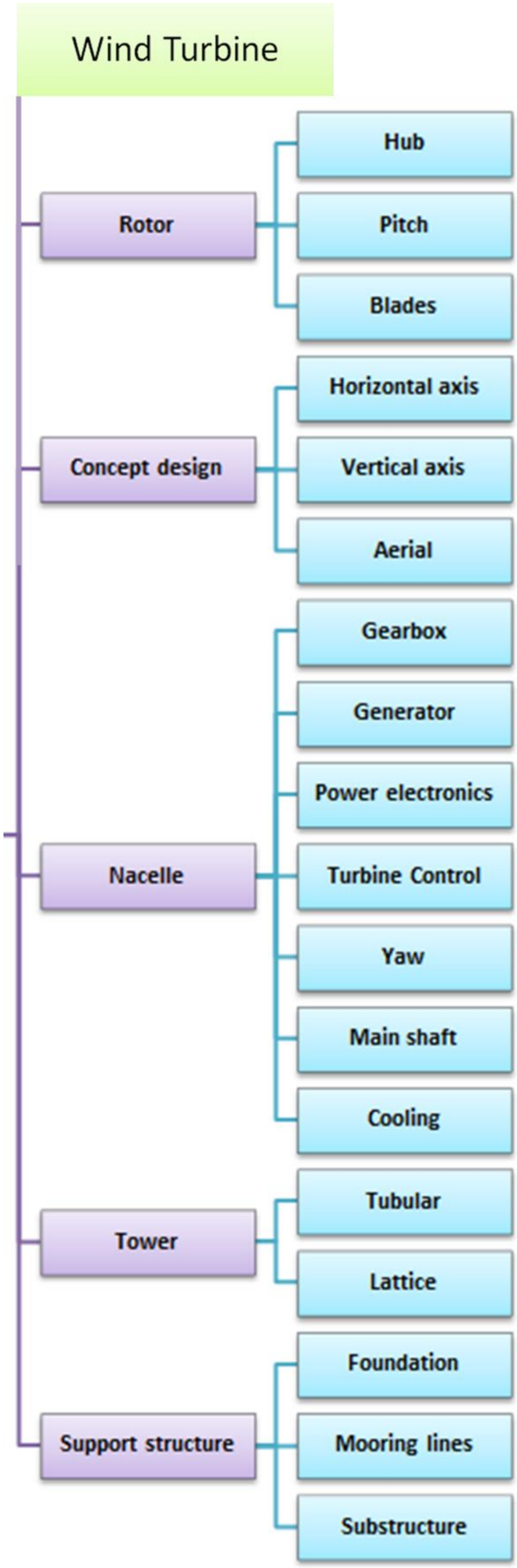

Figure 7 Narrow Terms hierarchy for Wind Turbine.

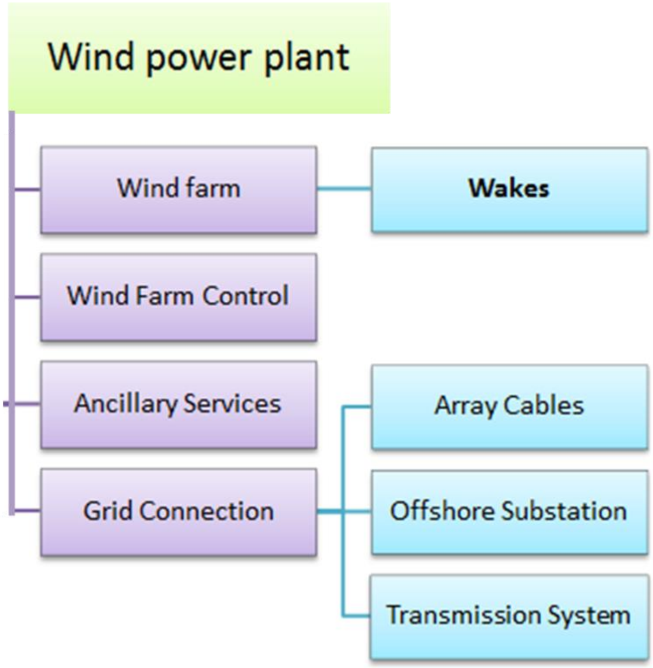

Figure 8 Narrow terms hierarchy for Wind Power Plant.

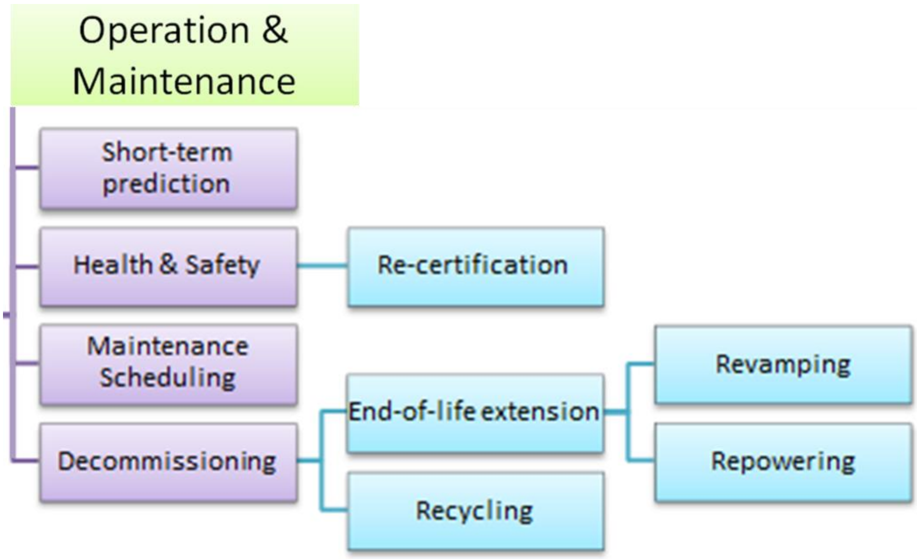

Figure 9 Narrow terms hierarchy for Operation & Maintenance.

2.3.3 Taxonomy of data type and other metadata elements

To add further filters, we have developed the following taxonomies for the following metadata elements:

- Data type (Dublin Core)
- External Conditions (No Dublin Core)

- Activities (No Dublin Core)
- Instruments (No Dublin Core)
- Models (No Dublin Core)
- Materials (No Dublin Core)

The attached file *DataTaxonomy_2017-05-22.xlsx* contains the taxonomy of the metadata elements that will be used as facets for finding data. The list is not exhaustive as we expect input from other colleagues and that more data type will be created and more measuring techniques will be adopted.

3. Conclusive remarks and future development

This report describes the process followed to develop a set of taxonomies for the Wind Energy sector, of the related data types and other facets. The work is the first towards the information architecture for a data portal for making the WE sector data Findable, Accessible, Inter-operable and Re-usable, 'FAIR', adhering to the Open Data strategy in the H2020 Programme. The developed taxonomies will be used as common vocabularies for tagging data in the metadata cards assigned to each dataset.

With metadata cards, and common vocabularies describing available data, eused by each organization, data can be searched through a data portal containing a metadata catalog updated by a web crawler, i.e. a program continuously harvesting metadata cards. The data itself resides on the data owner domain and security and data management issues remain in the hand of the data owner.

A user will access the portal to submit a query containing keywords from the established vocabularies from the taxonomy of the metadata. The system will return an optimized list of available data. Data can be accessed either directly via provided download links in the metadata card or by contacting data owners.

This approach has a two-fold purpose: to make data owners feel more comfortable in sharing data by maintaining the control on data access and data use, while end users will access information on datasets needed for a specific goal optimising time and funding. Both data owners and end-user will have the opportunity to start or reinforce collaboration activities.

In the topic taxonomy, each term in the tree branches is like a box to allocate data. The lower we move into the hierarchy tree the more we find detailed topics that might contain specialized data. Allowing the "free keyword" metadata element, data owners are enabled to insert new terms; however, the terms will be allocated in the specific topics. New terms will be updated in the relevant taxonomies. This will enable an analysis the frequency of the use of the new terms to explore whether the new term might indicate a new research niche. Niches from different sectors/fields might be integrated to produce innovative products (methodologies, measuring techniques etc.). This second use will be the subject of future development.

4. Acknowledgements

We thank Ass.Prof. Haakon Lund from the Royal School of Library and Information Science at the University of Copenhagen, for his valuable comments on information retrieval issues.

5. References

[1] M. D. Wilkinson, M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. . 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. a. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, p. 160018, 2016.

[2] G. Giuliani, S. Nativi, A. Obregon, M. Beniston, and A. Lehmann, *Spatially enabling the Global Framework for Climate Services: Reviewing geospatial solutions to efficiently share and integrate climate data & information*. 2017.

[3] M. Hlava, "The Taxobook: History, Theories, and Concepts of Knowledge Organization," *Synth. Lect. Inf. Concepts, Retrieval, Serv.*, vol. 6, no. 3, pp. 1–80, 2014.

[4] G. Miller, "TAXONOMY: Linnaeus's Legacy Carries On," *Science (80-.).*, vol. 307, p. 1038a–11039, 2005.

[5] H. Hedden, "What Are Taxonomy," *Accid. Taxon.*, pp. 1–37, 2010.

6. APPENDIX 1. OTHER FACETS

6.1 Activities

Uncertainty analysis
Field experiment
Long-term monitoring
Aircraft
Cruise
Modeling
Manufacturing
Reliability and testing
Laboratory test ≡ scaled ≡ Controlled environment
Wind Tunnel test ≡ scaled ≡ Controlled environment
Field test ≡ Full scale test
Certification
Full scale test
Condition Monitoring
Fatigue tests
Static tests
Environmental impact Note: Species Pollution People
Questionnaires Note: Pubblic acceptance
Interviews Note: Pubblic acceptance
Electrical analysis

6.2 External conditions

Offshore
Onshore
Coastal Onshore
Coastal Offshore
Complex
Flat
Forest
Urban
Rural
Semi-urban

6.3 Instruments

- Wind speed
  - Cup
  - Sonic
  - Pitot
- Wind Direction
  - Vane
- Profilers
  - Lidars
    - Wind Scanners
    - Continuous
    - Pulsed
    - Long-range
    - short-range
    - Nacelle lidar

	Sodars
	Radars
	Radiosondes
Temperature	Thermometer
Temperature profiles	
	Rass
Ceilometer	
Imaging	
	Hyperspectral camera
	Electron microscopy
	X-ray CT data
	Optical microscopy
Straing gauges	
Vibration sensors	
Oil sensors	
Ultrasonic testing	
Instrument support	
	Drones
	Satellite
	Masts
	Moored instrument
Waves sensors	
https://en.wikipedia.org/wiki/List_of_electrical_and_electronic_measuring_equipment	
Electrical measuring instruments	
	Power analyser
	Oscilloscope
	Ampermeter
	Capacitance meter
	Curve tracer
	Cos Phi Meter
	Distoritionmeter
	ESR meter
	Frequency counter
	Leakage tester
	LCR meter
	Multimeter
	Network analyser
	Ohmeter
	Psophometer
	Signal analyser
	Signal generator
	Spectrum analyser
	Sweep generator
	Transistor tester
	Wattmeter
	Voltmeter

6.4 Materials

Blades materials

Composite laminate
Sandwich Structure
Geal coats

Tower Materials

Steel
Concrete
wood

Drivetrain

Steel (gearbox, main shaft)
Aluminium (cables)
Cable insulation
Magnets
Cast iron
Concrete

Existing databases

https://www.wmc.eu/optimatblades_optidat.php (x-ray data from fatigue damaged wind turbine materials)
https://zenodo.org/record/154714#.WRI3ZeuGOHt database for fatigue test data

Material for transformers/inductors

Cobber
Aluminium
Insulation
Core
- Laminated steel sheets
- Silicon steel
- Nickel-iron
- Cobalt-iron
- Amorphous Alloy
- Nanocrystalline
- Powder core
- Powder iron
- Ferrite

Material for capacitors

6.5 Models

Meteorological

GCM reanalysis hindcast
Mesoscale

Computational

Rans
CFD
LES

Wake

Physical

Scaled

	Full scale		
Multi-Physics			
	Hydrodynamics		
	Structural dynamics		
	Aerodynamics		
	Control		
	Mechanics		
	Hydraulics		
Aerodynamics modeling			
	Computational methods		
		RANS	Note: also used in atmosferic physics
		LES	
		CFD	
		Vortex methods	
		Engineering methods	
		Finite Element Models	
	Experimental methods		Note: also included in 'activities'
		Wind Tunnel tests	
		Full scale test	
	Electrical models		
		Power flow	
		OPF	
		Small-signal models	
		Dynamic models	
		Short circuit models	
		State estimation	
		Power protection analysis models	
		Contigency analysis models	
		Harmonic models	
Financial models			
Grid System models			

6.6 Variables

Wind speed				
Wind Speed components				
Wind Direction				
Temperature	Sensible	Virtual		
Humidity	Relative	Absolute	Specific	
Air Density				
Rain				
Sea Spray				
Waves				
Sea Surface Temperature				
Heat Fluxes				
Air Pressure				
Stability				

Geo Spatial Data

Terrain Orography
Roughness

Land-use
Surface roughness
Sea Depth
Cadaster
Geology
Sea Floor

Scada

Curtailement — Power set point
Active power
Reactive power
Yaw
Pitch
Rotor speed
Nacelle wind speed
Nacelle wind direction

Turbine data, model input data, properties — Including Wind Farms

Power Curve
Power production
Power loss
Installed capacity
Campbell diagram
Wakes

Aerodynamic
- 2D airfoil coefficients
- 3D mesh geometry

Mechanics, structure
- beam properties — cross sectional properties
- eigenfrequencies
- structural damping
- mass and inertia
- ultimete strength
- yield strength
- mode shapes
- failure modes — buckling
- cross-section geometry
- 3D mesh geometry

2D
3D

Dynamics
- eigenfrequencies
- model representation — nth order system
- damping

MODEL INPUT DATA
aero-elastic model
- airfoil cofficients
- beam data
- eigenfrequencies

……………

Control

- Tuning
- Gains

Noise perception
Noise measurements

Generator data

- PMGS
  - Rated power
  - Rated voltage
  - …..
- DFIG
  - Rated power
  - Rated voltage
  - …..
- SG
  - Rated power
  - Rated voltage
  - …..
- AG
  - Rated power
  - Rated voltage
  - …..

Converter data

- Grid side converter
  - Rated power
  - Rated voltage
  - Rated frequency
  - Filter type
  - Filter inductance
  - Filter capacitance
  - Control loops
    - DC voltage control
      - Gains
      - Filter constants
    - AC voltage control
      - Gains
      - Filter constants
    - Reactive power control
      - Gains
      - Filter constants
    - Current control
      - Gains
      - Filter constants
    - Active damping
      - Filter constants
    - PLL
      - type
      - Gains
      - Filter constants
- PMGS converter
  - Rated power
  - Rated voltage
  - Rated frequency
  - Control loops
    - Active power control
      - Gains

- Filter constants
- Current control
  - Gains
  - Filter constants
- Active damping
  - Filter constants

Transformer data

- Rated power
- Rated voltage HV
- Rated voltage LV
- Rated frequency
- Winding connection
- No-load losses
- Copper losses
- Short-circuit impedance

Turbine control

- Rotor speed control/pitch control
  - Gains
  - Filter constants
- Generator speed tracking
  - Gains
  - Filter constants
- Maximum power-point tracking
  - Gains
  - Filter constants

---

[i] In this document, data is defined as in Wikipidia "a set of values of qualitative or quantitative variables " generated by a research, technological and innovation activity; e.g. time series of atmospheric variables, generated either from observations or model activities, geospatial data, such as orography maps, reports on tests, images, etc. specimens, models for the cost of energy, statistics.

[ii] http://inspire.ec.europa.eu/index.cfm

[iii] http://dublincore.org/documents/dces/