energies

Article

Topic Taxonomy and Metadata to Support Renewable Energy Digitalisation

Andrea Michiorri [1,*], Anna Maria Sempreviva [2], Sean Philipp [3], Paula Perez-Lopez [4], Alain Ferriere [5] and David Moser [6]

[1] Mines Paris, PSL University, Centre for Processes, Renewable Energy and Energy Systems (PERSEE), 06904 Sophia Antipolis, France
[2] Department of Wind Energy and Energy Systems, Technical University of Denmark, DTU, Risoe Campus, Frederiksborgvej 399, 4000 Roskilde, Denmark
[3] AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria
[4] Mines Paris, PSL University, Centre Observation, Impacts, Energy (O.I.E.), 06904 Sophia Antipolis, France
[5] Laboratoire PROMES-CNRS, 7 Rue Du Four Solaire, 66120 Font-Romeu Odeillo, France
[6] EURAC Research, Viale Druso, 1, 39100 Bolzano, Italy
* Correspondence: andrea.michiorri@minesparis.psl.eu

Abstract: Research and innovation in renewable energy, such as wind and solar, have been supporting the green transformation of energy systems, the backbone of a low-carbon climate-resilient society. The major challenge is to manage the complexity of the grid transformation to allow for higher shares of highly variable renewables while securing the safety of the stability of the grid and a stable energy supply. A great help comes from the ongoing digital transformation where digitisation of infrastructures and assets in research and industry generates multi-dimensional and multi-disciplinary digital data. However, a data user needs help to find the correct data to exploit. This has two significant facets: first, missing data management, i.e., datasets are neither findable because of missing community standard metadata and taxonomies, nor interoperable, i.e., missing standards for data formats; second, data owners having a negative perception of sharing data. To make data ready for data science exploitation, one of the necessary steps to map the existing data and their availability to facilitate their access is to create a taxonomy for the field's topics. For this, a group of experts in different renewable technologies such as photovoltaics, wind and concentrated solar power and in transversal fields such as life cycle assessment and the EU taxonomy for sustainable activities have been gathered to propose a coherent and detailed taxonomy for renewable energy-related data. The result is a coherent classification of relevant data sources, considering both the general aspects applicable to electricity generation from selected renewable energy technologies and the specific aspects of each of them. It is based on previous relevant work and can be easily extended to other renewable resources not considered in this work and conventional energy technology.

Keywords: renewable energy; taxonomy; wind power; photovoltaics; concentrated solar power

check for updates

Citation: Michiorri, A.; Sempreviva, A.M.; Philipp, S.; Perez-Lopez, P.; Ferriere, A.; Moser, D. Topic Taxonomy and Metadata to Support Renewable Energy Digitalisation. *Energies* **2022**, *15*, 9531. https://doi.org/10.3390/en15249531

Academic Editor: Surender Reddy Salkuti

Received: 27 September 2022
Accepted: 10 December 2022
Published: 15 December 2022

1. Introduction

1.1. Context: The Energy and Digital Transformation

During recent decades, research and innovation in Renewable Energy (RE), such as wind and solar, have supported the green transformation of energy systems, the backbone of a low-carbon climate-resilient end society. One of the challenges is to manage the complexity of the grid transformation to allow for higher shares of highly variable renewables while securing the stability of the grid and a stable energy supply. Help comes from the ongoing digital transformation where digitisation of infrastructures and assets in research and industry generates multi-dimensional and multi-disciplinary digital data.

The escalation of data from new sources and the growing complexity of challenges introduce different roadblocks: among them, a data user who must perform data analytics needs help finding the correct data to exploit. This has two significant facets: firstly, missing data management (i.e., datasets are neither findable because of missing community standard metadata and taxonomies) and interoperable data (i.e., missing standards for data formats); secondly, data owners having a negative perception of sharing data.

It is increasingly clear that we need to introduce information science and data science elements to deal with the data challenge. The former is essential for data management to organise and make data ready for data science exploitation. The latter enables the digitalisation process, i.e., transforming data into information and then into value to find innovative solutions and business models, e.g., to cut costs by better planning, optimising processes to increase operational efficiency and by reducing risk to optimise investments.

As the energy sector is central in the actions to mitigate climate change, digitalisation can become one of the effective tools to fight climate change. This can be added to the other central challenges already faced by the energy sector such as the contribution to economic development and pollution reduction. In this framework, data generated and consumed by digital applications in the energy sector have significant importance. They range from meteorological to consumer data, electricity consumption and production data or data relative to the infrastructures and the state of machines. In each case, such data may have a value outside the specific scope for which they have been generated, especially considering new information-centric business models and the potential in research and development.

However, to be exploited, data must be available to potential users. To do so, different paradigms have been proposed such as open or Findable, Accessible, Interoperable and Reusable (FAIR) data which will be discussed in the next section. In any case, data for each topic in a specific field must also be findable; for this, a topic taxonomy for energy-related fields is necessary for two reasons: to standardise the nomenclature, supporting a common understanding amongst the actors, and to classify existing and future data available in this field.

In this paper, we address the information science issue by creating metadata and a taxonomy of the topics of photovoltaics and concentrated solar power based on previous work carried out by the wind energy community. We adopted expert elicitation where a group of experts in different renewable technologies and for cross-cutting fields such as Life Cycle Assessment (LCA) and the European Union (EU) taxonomy for sustainable activities have been gathered to propose a consistent and detailed taxonomy for renewable energy-related data. The result is a coherent classification of relevant data sources, considering the general aspects applicable to electricity generation from selected renewable energy technologies and their specific elements. Metadata and taxonomies can be easily extended to other renewable resources not considered here.

1.2. State of the Art

The knowledge of the existence of data and their prompt access is widely recognised as a method to optimise research impact and several frameworks are proposed to achieve it. Among them, the two more debated ones are currently the open data [1] and the FAIR [2] approaches. In the first approach, data access is free to users with several licences granting different reusability levels. In the second approach, data property and access are maintained by the owners who make data (a) findable by a faceted search and (b) accessible through agreements, e.g., compensation and/or confidentiality clauses.

Independently of whether data are open or FAIR, data should be findable and reusable; to comply with these conditions, data must be identified and described using metadata, i.e., a series of information, e.g., who, what, when, where, how etc. to put data in a context. Furthermore, metadata should be machine-actionable allowing for searchability and usability by machines. With metadata, data are preserved for future reuse. These can be general or discipline-specific, with the first defined, for example, by the Dublin Core (DC) [3] whilst the second must be defined by domain experts.

Within the information included in metadata is the position of the dataset in a relevant taxonomy, which is a method to structure the knowledge related to a discipline into topics. An example is the famous biological classification or the most recent EU taxonomy for sustainable activities [4]. In this work, the most common hierarchical taxonomy approach has been used.

Regarding data availability, it is observed that datasets about renewable energy resources are now largely accessible thanks to long-term work carried out at independent research centres and large multinational organisations. Examples are the numerous wind or solar atlases created for different regions. Among them are Danish Technical University's (DTU) Global Wind Atlas [5] and Solargis' Global Solar Atlas [6], both funded by the World Bank. Regulated grid operators also disclose relevant information through dedicated data platforms such as the European Association for the Cooperation of Transmission System Operators for Electricity (ENTSOE) Transparency Platform [7] or similar portals at the national level. On the other hand, data relative to operational aspects of renewable plants, such as output or maintenance, are much rarer despite a long operational history. Remarkable is the Open Mod Initiative, which aims at collecting access to existing open data related to the energy sector [8]. The landscape can be concluded with the scarcity of data related to emerging aspects of renewable energy such as plants' end of life, to disposal, recycling and reusing aspects or to concepts emerging in importance such as environmental performance evaluated through LCA.

The idea of this work originates from the lack of such an organisation for renewable energy, potentially due to the relative youth of this field in the landscape of science and technology research. Anyway, this work could be built based on previous pioneering works in subaspects of renewable energy and neighbouring fields.

Regarding wind energy, the deepest and most complete taxonomy for wind energy research and development topics has been developed during the European Commission FP7 Project, Integrated Research Programme in Wind Energy (IRPWIND). IRPWIND combined strategic research projects and support activities within the field of wind energy, to leverage the long-term European research potential. To guarantee the reliability of the results, the work used the expert elicitation procedure, gathering experts from the major organisations in wind energy associated with the European Energy Research Alliance, Joint Programme on Wind Energy (EERA JP Wind) whose organisational structure and participation is mirrored in the IRPWIND consortium. This guarantees the reliability of the taxonomies. Details of the work are presented in [9]. This work is also enhanced by the production of a metadata schema and the design and demonstration of the metadata catalogue Share-Wind. The metadata schema includes the list of general Dublin Core metadata completed by wind energy domain-specific metadata and related taxonomies/vocabularies. Data owners can register the metadata of available datasets to populate the data catalogue, and data users can search for needed data. The platform Share-Wind is in the process of being transferred to the domain http://share-wind.net (accessed on 29 October 2021).

Another significant activity on this topic has been carried out at NREL [10] and IEA [11] levels. In the first case, a taxonomy has been developed for studying the cost structure of wind energy to identify potential cost reduction sources. In the second case, the activity is not linked exclusively to cost but to several aspects of wind turbines and plants, including design and operation. A second theme which sparked the creation of taxonomies is the need to completely map activities linked to maintenance and condition monitoring. In this case, it is possible to mention [12] focused on wind turbines and [13] where the ontology is used to represent the knowledge extracted by fault diagnosis analysis. Finally, it is possible to mention [14] where an ontology of wind energy-related topics is created semi-autonomously from the open-source text.

The same types of studies can be found for photovoltaics, where taxonomies have been proposed for both cost and maintenance analysis. In the first group—cost—it is possible to mention the works carried out at the National Renewable Energy Laboratory (NREL) [15,16] for cost benchmarking and [17] focused on soft costs for Photovoltaics (PV). The second group, maintenance, presents a detailed breakdown of aspects related to photovoltaic energy that have been developed in [18] and the Trust PV project [19], to analyse the risk and maintenance aspects in the whole PV value chain. The need for a common nomenclature in PV has also been highlighted in the H2020 project Solar Bankability [20] where a cost factor was added to the Failure Mode and Effect Analysis (FMEA) in PV risk management, but to develop common results from different systems a common taxonomy must be in place. A third topic has been identified in the need for the taxonomy for systems design [21] and planning [22]. Finally, it is worth mentioning two works particularly relevant to this research: in [23], a topic and metadata taxonomy is presented for enabling FAIR PV production data time series, whilst in [24] a topic taxonomy is developed for agri-PV systems, with a deep level of detail.

Regarding concentrated solar power, little activity has been found, except for attempts to structure the knowledge around solar irradiation forecasts, presented in [25,26].

Finally, it is important to mention works related to knowledge organisation in life cycle assessment thinking. This is the notion of going beyond the traditional focus on the manufacturing site to account for the environmental, social and economic impacts over the whole product's life cycle. It is important to use this method to cover all the aspects of renewable project life, rather than focusing on specific phases such as planning or operation. This perspective is described in [27], which focused on the data necessary for life cycle assessment, and [28], which attempted to reduce the ontology to a minimal extent and used a coal power plant as a case study. A comprehensive view of the topics in LCA can be found in [29] and aspects related to uncertainty are detailed in [30]. The Share-Wind initiative started as a data catalogue for wind energy-related datasets and evolved into a fully searchable metadata catalogue [31].

On a general level, relevant work is the above-mentioned EU taxonomy for sustainable activities [4], developed to direct investments towards sustainable projects and clarify to the industry which type of investments can be considered sustainable. This taxonomy covers several topics affecting the level of sustainability of a system, including electricity production from renewable sources, which is the scope of this work. For each of them, information is given on the type of effect that investments have on improving the impact on the environment, whether their nature is permanent or transitional, direct or enabling, etc.

1.3. Conclusions and Contributions

We summarise the current research activity on taxonomies for renewable energy as follows: (1) activity is mainly carried out for wind and PV technology; (2) activity is often carried out at the level of large research bodies or international organisations; (3) the main objectives are: (a) understanding cost structure, (b) classifying maintenance issues, (c) to a lesser extent, understanding system planning; (4) all works are technology-specific and not related to renewable energies in general, except for [4], even if renewable resources share several common aspects; (5) most work focuses on a subset of the life cycle of renewable energy technologies, often planning or operation and maintenance, except for [9,18].

In light of this, the work presented in this paper aims to provide the following main contribution:

- Metadata and topic taxonomies for renewable energy systems, as flexible as possible to be extended to several technologies, with attention to the data relevant to each topic and the whole life cycle.

This is considered important for several reasons, such as the need for generalising the concepts for several renewable technologies. It is clear that different technologies such as wind power or photovoltaics are based on completely different operating principles, but several common points exist. Among them is the dependence on renewable resources, which needs to be assessed at the planning and operation stage, the structure of power plants which is characterised by an array of captors and the types of studies that need to be carried out for renewable projects. For example, a trans-technology taxonomy could compare benchmark practices, performances and costs among different technologies and not only between plants or generations of the same technology. Secondly, the attention to data is due to the fact that this is where information is stored, both that necessary to carry out studies and analysis and that produced by and for the renewable industry and research. Finally, the attention given to the whole life cycle of renewables is necessary because of the current attention also given to phases before and after the visible life of a renewable project, such as the manufacture and the disposal of the equipment.

2. Materials and Methods

2.1. Criteria

Two general criteria have been used in developing this work:

- it must build on existing taxonomies wherever possible;
- it must be able to be expanded to include other renewable energy technologies.

To fulfil the first point, the current taxonomy has been imagined as an expansion of [4], limited to specific nodes relative to the supply of electricity and other energy vectors. Additionally, proposed taxonomies have been used where possible such as in maintenance [32] or LCA [27]. This allows one side to focus the work on the specifically identified gap, while the other improves its generalisation and reusability.

Special attention has been given to [9], considered the most pertinent and detailed work in this field. Its structure and spirit have been integrated almost completely into the wind section and it has been used to shape the sections for other renewable technologies. For this reason, the taxonomy is coupled with a set of metadata able to describe and document the data.

For the second point, data and concepts related to the actual energy converters and their plants, which are technology-dependent, have been carefully separated from more general concepts such as the ones related to facility siting, economics or operation and maintenance. In particular, an attempt has also been made to standardise and generalise these latter concepts to be able to reuse them for other renewable generation technologies not yet included.

2.2. Considerations on the Impact of Life Cycle Assessment and Life Cycle Thinking

As previously mentioned, LCA is a widely accepted method to evaluate the potential environmental impacts of products or systems over their whole life cycle and has been largely applied to energy production systems. The importance of accounting for the whole life cycle of products is highlighted in the documents produced by the Technical Working Group on EU taxonomy.

For the economic activity to be considered sustainable, according to the Taxonomy Regulation, it has to significantly contribute to one environmental objective, among which are climate change mitigation or adaptation, the sustainable use of water and marine resources and the protection and restoration of biodiversity and ecosystems, among others. Robust methodologies and metrics are needed to ensure the applicability of the tool.

In this context, LCA may provide valuable information to improve benchmarking and set targets at different levels, from the portfolio to European level [4]. LCA has the potential to provide results not only in terms of Greenhouse Gas (GHG) emissions contributing to climate change but for a large range of impact categories that include the impacts on human health and ecosystem quality to the depletion of abiotic resources [33]. The current barrier to integrating these aspects is the scarcity of robust life cycle data for many economic activities together with other methodological considerations to meet the needs of the taxonomy framework.

2.3. Clarifications on the EU Taxonomy and the Do No Significant Harm Criteria

With the publication of the Regulation on the establishment of a framework to facilitate sustainable investment, herein referred to as the EU Taxonomy, the EU continues its efforts in establishing a low-carbon economy and investment activities. This taxonomy lists six environmental objectives and contains a set of eligible economic activities, including the criteria that need to be met for an activity to be regarded as sustainable. For each listed activity, technical screening criteria and "Do No Significant Harm" (DNSH) criteria are defined. This ensures that the pursuit of economic activity, even though it may contribute to one specific environmental objective of the EU Taxonomy, does not hamper any other objective at the same time.

To date, two Annexes have been published, stating technical screening criteria and DNSH criteria for eligible economic activities. The Annexes comprise the criteria for the environmental objectives, i.e., climate change mitigation and climate change adaptation. These specify not only technical screening criteria that need to be met but also the conditions for an activity to be regarded as not conflicting with any other taxonomy objective, such as DNSH.

Reporting will, however, depend on the type of company (i.e., financial or non-financial). The latter is, in short, requested to report Key Performance Indicators (KPIs) that represent the share of sustainable activities concerning total turnover, Opex and Capex.

Concerning financial undertakings, the EU distinguishes between the activities of credit institutions and banks, investment firms, asset managers and insurers/reinsurers. For these institutions, various other Key Performance Indicators (KPIs) may be subject to their reporting requirements, depending on the nature of their business and core activities. Such KPIs may include the Green Asset Ratio (GAR), the Green Ratio for Financial Guarantees to Corporates (FinGuar KPI) and the Green Ratio for Assets under Management (AuM KPI).

As a result of the assessment of economic activities under the EU Taxonomy, companies shall disclose in the future the share of sustainable activities by indicators such as Capex, Opex and revenue. This acts as a market signal for investors seeking low-carbon, climate-friendly investment opportunities. In the larger picture, this will promote access to finance and market opportunities for companies conducting sustainable business activities to encourage shifting towards a more rational production system.

2.4. Perimeter of the Work

Choices have also been necessary regarding the perimeter of this work. As mentioned above, the technologies considered have been limited by the ones included in [4], but in this case, the field needed to be restricted further. The version considered, available in February 2022, listed a total of 72 activities with substantial contributions to climate change mitigation. Of these, 61 were considered with their performance in climate mitigation (such as electricity generation from wind), 12 were considered "enabling" (such as infrastructures) and 34 were considered "transition" (such as bioenergy). It was then decided to limit the focus on activities related to the "electricity, gas, steam and air conditioning supply", which were not considered "transition" and which were related to the "production of electricity". Marine power has not been integrated into this work because of the large number of possible designs and the rapid evolution of this field. This can be seen in Table 1. We hope that the present work will be used as a solid base and extended for the other activities on the list.

Table 1. Climate change mitigation activities related to the NACE macro-sector electricity, gas, steam and air conditioning supply and without transitory aspects. From: [4].

Selected	Activity	Own Performance	Enabling	Transition Activity
x	Production of Electricity from Solar PV	X		
x	Production of Electricity from Concentrated Solar Power	X		
x	Production of Electricity from Wind Power	X		
	Production of Electricity from Ocean Energy	X		
	Transmission and Distribution of Electricity	X	X	
	Retrofit of Gas Transmission and Distribution Networks	X		
	District Heating/Cooling Distribution	X		
	Installation and Operation of Electric Heat Pumps	X		
	Cogeneration of Heating/Cooling and Power from Concentrated Solar Power	X		
	Production of Heating/Cooling from Concentrated Solar Power	X		
	Production of Heating/Cooling using Waste Heat	X		

2.5. Metadata and Taxonomies

In this section, we will introduce the definitions of the terms used in this work to have a common understanding with the readers. This will add clarity to the results presented below.

Metadata are information necessary to facilitate the documentation and contextualisation of data, e.g., who, what, when, where and how to put data in a context. A list of metadata, a metadata schema, is necessary to facilitate the retrieval of data by a faceted search to allow the reusability of data by other users. Generally, there are three metadata categories: structural, descriptive and administrative. Structural metadata are used to represent the relationships between components of an object and between different objects. An example would be the page numbers, paragraphs and chapters of a book or a digital document. Descriptive metadata help the user to link to the resource and describe the content and the context and can usually be searched, sorted or filtered. Examples are the title or the abstract of a document. Finally, administrative metadata represents information external to the information content of the resource but gives additional context, such as information about data owner and access rights. In this paper, we use the standard schema proposed in the DC [3].

Taxonomies explain descriptive metadata. They have the double function of creating a representation of the structure of the knowledge and formalising the terms used for the concepts. For the first action, taxonomies use a hierarchical structure, defining concepts and then detailing their differences. The use of a hierarchical structure facilitates the research of information and prevents the creation of non-biunivocal relations. For the second action, the taxonomy associated a concept with a term and a definition, limiting ambiguity as much as possible.

The relation between taxonomies and metadata and how they are utilised in this work are represented in Figure 1, which shows the schema of descriptive metadata explained by taxonomies. It is possible to see how each resource (data) is described by several facets: two from the DC elements and six non-DC elements defined by the domain specialists. For each one of the facets, it is necessary to develop a taxonomy or vocabulary of concepts.

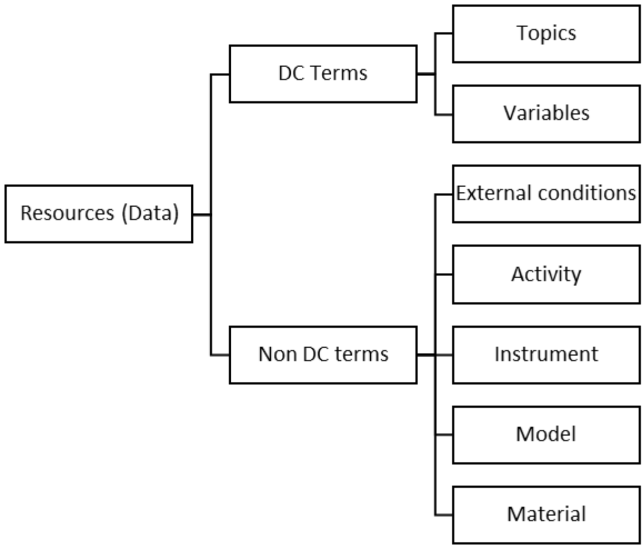

Figure 1. Classification elements as proposed in [9] for wind energy, both DC and non-DC, for the description of the single resource.

3. Results

Level 1

This level, shown in Table 2, includes the parent nodes relative to the three chosen renewable technologies. An alternative would have been to choose a complete technology-agnostic approach, leaving the technology-dependent topics assigned to a specific facet. Finally, the approach considering each technology as a parent node was chosen to facilitate the integration with the existing ones [4].

Table 2. Subset of the EU Sustainable Finance Taxonomy considered in this study.

- Electricity, Gas, Steam and Air Conditioning Supply
  - Production of Electricity from Solar PV
  - Production of Electricity from Concentrated Solar Power
  - Production of Electricity from Wind Power

3.1. Topic Taxonomy for Solar PV

For this technology, the approach followed is based on the existing work described in Section 2.4, which encompasses a deeper tree with more levels.

Level 2

The second level for this technology, in Table 3, is relatively simple; topics are divided according to the three main aspects of renewable energy generation: the resource, the producing assets and other aspects related to the life cycle of the plants.

Table 3. Hierarchy for the Production of Electricity from Solar PV.

- Production of Electricity from Solar PV
  - Resource
  - Components
  - Life Cycle

Level 3 and further

In this first facet, the topics relative to the resource and the productivity assessment are presented. A first differentiation is made between the renewable resource (different aspects of solar irradiance) and the ground-related aspects (from shading to road and power network access) influencing the productivity of the site and the overall cost of the power plant. This is shown in Table 4.

Table 4. Hierarchy for Resource-related topics in the Production of Electricity from Solar PV.

- Resource
    - Land-Related
        - Access to Network
        - Land Use
        - Ground Topography
        - Access to Site
        - Shading
    - (a) Renewable Resource
        - Direct Normal Irradiation
        - Global Solar Irradiation
        - Diffuse Solar Irradiation
        - Reflected Solar Irradiation
        - Temperature
        - Wind Speed
        - Soiling

In this second facet, the topics related to the producing assets are described. A particular distinction is made between aspects related to the conversion technology, in this case, the photovoltaic cells and modules, and the power plant. For cells and modules, data related to the characterisation, such as the I-V curve and the temperature behaviour, are listed. Regarding cell technology, the tree is not expanded further since a large number of possibilities are available, although only a few are currently used. This is shown in Table 5.

Table 5. Hierarchy for Component-related topics in the Production of Electricity from Solar PV.

- Components
    - Plant
        - Meter
        - Inverter
        - Grid
            - Cabling
            - Protection
            - Grounding
        - Monitoring and Control
    - Conversion Technology
        - Cell
            - Technology
            - Characterisation
        - Module
            - Architecture
            - Characterisation

The third facet, based on the assumption described in Section 2.4, groups together aspects related to the life cycle of the plant, from the manufacture of the cells and other equipment to the decommissioning of the plant, passing through the construction phase and operation and maintenance, where data related to the production and faults are present. It is based on the assumptions described in Section 2.4. This is shown in Table 6.

Table 6. Hierarchy for Life Cycle-related topics in the Production of Electricity from Solar PV.

- Life Cycle
  - Manufacture
    - Materials
    - Process
  - Logistics
    - Distance
    - Transport Technology
    - Fill Rate
  - Planning and Construction
    - Planning
      - Sizing
      - Economic
    - Construction
  - Operation and Maintenance
    - Production
    - Prediction
    - States of the Systems
    - Maintenance Tickets
    - Health and Safety
  - Decommissioning
    - Recycling
    - End-of-Life Extension
    - Revamping
    - Repowering

3.2. Topic Taxonomy for Solar Concentrated Solar Power

Level 2

In the case of Concentrated Solar Power (CSP), the more generic structure already seen for PV is used. The details related to the technology are based on the assumptions described in Section 2.4. This is shown in Table 7.

Table 7. Hierarchy for Production of Electricity from CSP.

- Production of Electricity from CSP
  - Resource
  - Components
  - Life Cycle

Level 3 and further

At this level, the first node related to the resource is very similar to the one for solar PV. The only notable difference is the presence of information related to water resources, this being an important factor for the technology. This is shown in Table 8.

Regarding the branch related to the components, a section related to the plant similar to the solar PV has been maintained for taking into account the electric connection. However, the largest part is occupied by the CSP conversion technology, which is more complex than in the case of PV and with several possible designs. This is shown in Table 9.

Table 8. Hierarchy for Resource-related topics in the Production of Electricity from CSP.

- Resource
    - Land-Related
        - Access to Network
        - Land Use
        - Ground Topography
        - Access to Site
        - Shading
    - Renewable Resource
        - Direct Normal Irradiation
        - Global Solar Irradiation
        - Diffuse Solar Irradiation
        - Reflected Solar Irradiation
        - Temperature
        - Wind Speed
    - Soiling

Table 9. Hierarchy for Component-related topics in the Production of Electricity from CSP.

- Components
    - Plant
        - Meter
        - Inverter
        - Grid
            - Cabling
            - Protection
            - Grounding
        - Monitoring and Control
    - Conversion Technology
        - Geometry
            - Point Concentration
            - Linear Concentration
        - Solar Field
            - Mirror
            - Structure
            - Drivers
            - Tracker
        - Solar Receiver
        - Thermal Vectors and Storage
            - Use
            - Nature
        - Power Block
            - Cycle Type
            - Thermodynamic Fluid
            - Component
                - Pre-heater
                - Evaporator
                - Superheater
                - Condenser
                - Turbine
                - Pump
                - Cooling Tower
                - Control
            - Back-Up (Hybrid System)

Regarding the life cycle, the structure is not different from what is already seen in the case of solar PV plants. This is shown in Table 10.

Table 10. Hierarchy for Life Cycle-related topics in the Production of Electricity from CSP.

- Life Cycle
  - ○ Manufacture
    - ■ Materials
    - ■ Process
  - (a) Logistics
    - ■ Distance
    - ■ Transport Technology
    - ■ Fill Rate
  - (b) Planning and Construction
    - ■ Planning
      - Sizing
      - Economic
    - ■ Construction
  - (c) Operation and Maintenance
    - ■ Production
    - ■ Prediction
    - ■ States of the Systems
    - ■ Maintenance Tickets
    - ■ Health and Safety
  - (d) Decommissioning
    - ■ Recycling
    - ■ End-of-Life Extension

3.3. Topic Taxonomy for Wind Power

In this case, as mentioned above, the work from [9] has been used and is reported here only for facilitating the reader's understanding.

Level 2

Differences can be seen between this approach and the ones presented above: mainly, this first level is more detailed with additional subnodes, detailing more fully the concepts related to resource and planning (siting and economics), the components divided already at this level between the turbine and the power plant. In general, the concepts related to the life cycle are summarised in the operation and maintenance section. This is shown in Table 11.

Table 11. Hierarchy for Production of Electricity from Wind Power.

- Production of Electricity from Wind Power
  - ○ Siting
  - ○ Economics
  - ○ Wind Turbine
  - ○ Wind Power Plant
  - ○ Operation and Maintenance

Level 3 and further

The siting node groups the concepts related to both the resource (in this case wind speed) and the site conditions, considering the infrastructures and other zonal planning information). This is shown in Table 12.

Table 12. Hierarchy for Siting-related topics in the Production of Electricity from Wind Power.

- Siting
  - Wind Mapping
  - Wind Atlases
  - Long-Term Corrections
    - Wind Indices
  - Resource Assessment
  - Design Conditions
    - Shear
    - Turbulence
    - Extreme Wind
    - Flow Angle
  - Infrastructures
  - Spatial Planning
    - Legal Aspects
    - Environmental Impact
      - Noise Perception
      - Nature Impacts
      - Social Acceptance

Topics related to the turbine are grouped in the node with the same name, which describes the different aspects of the generation unit. This is shown in Table 13.

Table 13. Hierarchy for Wind turbine-related topics in the Production of Electricity from Wind Power.

- Wind Turbine
  - Rotor
    - Hub
    - Pitch
    - Blades
  - Concept Design
    - Horizontal Axis
    - Vertical Axis
    - Aerial
  - Nacelle
    - Gearbox
  - Generator
    - Power Electronics
    - Turbine Control
    - Yaw
    - Main Shaft
    - Cooling
  - Tower
    - Tubular
    - Lattice
  - Support Structure
    - Foundation
    - Mooring Lines
  - (a) Substructure

The economics node summarises the topics related to the financial aspects of a project, from the financing to the business models and the support schemes. This is shown in Table 14.

Table 14. Hierarchy for Economics-related topics in the Production of Electricity from Wind Power.

- Economics
  - Project Finance
  - Levelised Cost of Energy (LCOE) Models
  - Support Schemes
  - Market Models
  - Business Models

Information related to the power plant, constituted by an array of wind turbines, is described in this node, where the aspects related to the grid connection and the interaction between turbines (wake) emerge. This is shown in Table 15.

Table 15. Hierarchy for Wind power plant-related topics in the Production of Electricity from Wind Power.

- Wind Power Plant
  - Wind Farm
    - Wakes
  - Wind Farm Control
  - Ancillary Services
  - Grid Connection
    - Array Cables
    - Offshore Substation
    - Transmission System

Finally, aspects related to the operation of the plant, including maintenance, are included in this last node, which also takes into account the topic related to the end of life of the power plant, from the life extension to the recycling. This is shown in Table 16.

Table 16. Hierarchy for Operation and maintenance-related topics in the Production of Electricity from Wind Power.

- Operation and Maintenance
  - Short-Term Prediction
  - Health and Safety
    - Recertification
  - Maintenance Scheduling
  - Decommissioning
    - End-of-Life Extension
      - Revamping
      - Repowering
    - Recycling

3.4. Metadata Taxonomies

The following is a series of topic taxonomies for the metadata which do not belong to the DC. These metadata, already presented in Figure 1, are shown in Table 17.

Table 17. Hierarchy of top-level metadata.

- Metadata
    - Activity
    - Instrument
    - Material
    - Models
    - Variables
    - External Conditions

It is worth noticing that, influenced by the work in [9], small amendments and additions have been carried out, although attention has been paid to maintaining the compatibility as much as possible.

Activities

With the topics under this node, the user can identify the data necessary for a specific activity, such as certification or performing a life cycle assessment, or originating from an activity, such as testing. This is shown in Table 18.

Table 18. Hierarchy for Activity-related Metadata.

- Activities
    - Certification
    - Do Not Do Significant Harm
    - Environmental Impact
        - Biodiversity
        - Pollution
        - Visual Impact
    - Social Impact
    - Life Cycle Assessment
        - Greenhouse Gas Emissions
        - Land Use
        - Water Consumption
        - Mineral Resource Depletion
        - Toxicity-Related Categories
            - Acidification
            - Eutrophication
            - Ionising Radiation
    - Manufacturing
    - Modelling
        - Resource Estimation
        - Uncertainty Analysis
        - Power System Analysis
    - Monitoring
        - Condition Monitoring
        - Long-Term Monitoring
    - Reliability And Testing
    - Tests
        - Dynamic Tests
        - Fatigue Tests
        - Field Tests
        - Full-Scale Tests
        - Laboratory Tests
        - Reduced Scale Tests
            - Static Tests

Instruments

This section lists the most used instruments in wind and solar power for measurements of the resource or the components. This is shown in Table 19.

Table 19. Hierarchy for Instrument-related Metadata.

- Instruments
  - Ceilometer
  - Electrical Measuring Instruments
    - Frequency Counter
    - Multimeter
    - Real and Reactive Power (PQ) Meter
    - Oscilloscope
    - Spectrum Analyser
  - Imaging
    - Hyperspectral Camera
    - Electron Microscopy
    - X-Ray CT Data
    - Optical Microscopy
  - Instrument Support
    - Drones
    - Satellite
    - Masts
    - Moored Instrument
  - Oil Sensors
  - Profilers
    - Lidars
  - Strain Gauges
  - Solar Radiation
    - Pyranometer
    - Pyrheliometer
    - Reference Cells
  - Temperature
  - Temperature Profilers
    - Rass
  - Ultrasonic Testing
  - Vibration Sensors
  - Waves Sensors
  - Wind Direction
    - Vane
  - Wind Speed
    - Cup
    - Sonic
    - Pitot

Materials

This section describes the materials most commonly used in renewable energy conversion. This is shown in Table 20.

Table 20. Hierarchy for Material-related Metadata.

- Materials
  - Blades
    - Composite Laminate
    - Gel Coats
    - Sandwich Structure
  - Towers and Structures
    - Aluminium
    - Concrete
    - Polymer
    - Steel
    - Wood
  - Drivetrain
    - Aluminium
    - Cable Insulation
    - Cast Iron
    - Magnets
    - Steel
  - Electric
    - Aluminium
    - Copper
    - Ferromagnetic Core
    - Insulation
  - PV Cells
    - Cadmium Telluride (Cdte)
    - Copper Indium Gallium Diselenide (CIGS)
    - Gallium Arsenide (Gaas)
    - Perovskite
    - Silicon Amorphous
    - Silicon Crystalline
  - Polymers
  - Glass
  - CSP Mirrors
    - Glass
    - Silver
    - Polymer
  - Thermal Fluid and Storage CSP
    - Water
    - Thermal Oil
    - Molten Salts
    - Organic
    - Inorganic
    - Gas
    - Particle Suspension

Models

Data are often used as input for models or calculated from them. In such a case, the following nodes can help the user identify the data needed. This is shown in Table 21.

Table 21. Hierarchy for Model-related Metadata.

- Models
  - Meteorological
    - General Circulation Model (GCM)
    - Mesoscale
    - Reanalysis
    - Hindcast
  - Physical
  - Multi-physics
    - Hydrodynamics
    - Structural Dynamics
    - Aerodynamics
    - Control
    - Mechanics
    - Hydraulics
  - Financial Models
  - Electrical Models
    - Power Flow
    - Optimal Power Flow
    - Small-Signal Method
    - Dynamic Models
    - Short Circuit
    - State Estimation
    - Power Protection Analysis
    - Contingency Analysis
    - Harmonic
  - CSP Power-Specific
    - Thermodynamics
    - Heat Exchange
    - Thermal Model
    - Soiling
    - Optics
  - PV Power-Specific
    - Equivalent Circuits
    - Irradiance
    - Thermal Model
    - Soiling
  - Wind Power-Specific
    - Computational Fluid Dynamics
    - Experimental Fluid Dynamics
    - Wake

Variables

A metadata node is introduced to classify variables. This is shown in Table 22.

Table 22. Hierarchy for Variable-related Metadata.

- Variables
  - Weather
    - Air Density
    - Air Pressure
    - Heat Fluxes
    - Humidity
    - Rain
    - Seaspray
    - Sea Surface Temperature
    - Solar Irradiance
    - Stability
    - Temperature
    - Waves
    - Wind Direction
    - Wind Speed
  - Geo-Spatial
    - Cadaster
    - Geology
    - Land Use
    - Roughness
    - Sea Depth
    - Seafloor
    - Terrain Orography
  - SCADA
    - Active Power
    - Available Active Power
    - Blade Pitch
    - Current
    - Curtailment
    - Module Orientation
    - Module Tilt
    - Nacelle Yaw
    - Reactive Power
    - Rotor Speed
    - Solar Irradiance
    - Voltage
    - The Wind Direction at Nacelle
    - Wind Speed at Nacelle
  - Turbine
    - Aerodynamic
    - Campbell Diagram
    - Dynamics
    - Installed Capacity
    - Mechanics Structure
    - Power Curve
    - Power Loss
    - Power Production
    - Wakes
  - PV Modules
    - Current–Voltage (I-V) Curve
    - Temperature
  - CSP Plant
    - Temperature
    - Pressure
    - Mirror Orientation
    - Mirror Tilt

Table 22. *Cont.*

- Inverter
    - Rated Power
    - Rated Voltage
    - Rated Frequency
    - Filter Type
    - Filter Inductance
    - Filter Capacitance
- Transformer
    - Rated Power
    - Rated Voltage High Voltage (HV)
    - Rated Voltage Low Voltage (LV)
    - Rated Frequency
    - Winding Connection
    - No-Load Losses
    - Copper Losses
    - Short Circuit Impedance
- Controls
    - Type
    - Gains
    - Filter Constants

External conditions

Finally, data may be influenced by the conditions in which they were recorded or by what they refer to. This is shown in Table 23.

Table 23. Hierarchy for External condition-related Metadata.

- External Conditions
    - Offshore
    - Onshore
    - Coastal Onshore
    - Coastal Offshore
    - Terrain Type
        - Complex
        - Flat
        - Forest
        - Rural
        - Semi-Urban
        - Urban
    - Geographical Location
        - Coordinates
        - Administrative Boundaries

4. Discussion and Conclusions

The work presented in this manuscript represents a first attempt to classify information related to renewable energy technologies. It is believed that this will help to accelerate research and innovation in the field of energy digitalisation, in two different ways. On one side, such taxonomy can be applied to classify existing and future datasets, facilitating their diffusion and reuse. On the other hand, it can help to identify the availability of existing datasets, or the areas where available datasets are less common.

The analysis, limited to selected technologies for electricity production from renewable resources, identified more than 400 unique terms. They are divided into topics specific to the three technologies chosen: photovoltaics, concentrated solar power and wind power, and six categories for metadata: activity, instrument, material, models, variables and external conditions. The latter in particular are expected to facilitate users' data searches, thereby contributing to better utilisation.

This work is expected to bring the following contribution to the academic community:

- A coherent nomenclature system for renewable energy.
- Facilitation of communication within research and industry, also facilitating system interoperability.
- A contribution to the development of a FAIR data ecosystem for renewable energy-related data.

The possibility to extend the approach to other renewable energy generation and support technologies is considered fundamental for this work. It also represents a step further from previous attempts which were focused on individual technologies (wind, PV) but would not share datasets for a similar topic. An example can be resource estimation which is based on meteorological and climatological assessments for many renewable technologies, but treated differently in technology-specific taxonomies. Other examples are aspects related to grid connection, planning and operation or LCA.

The standardisation allowed with the approach proposed is expected to alleviate two main problems: (1) it can facilitate communication and hence research and development, and (2) facilitate system interoperability on the industrial level. Topic nomenclature is in fact currently suffering from the use of different definitions on the company or sector level.

The development of FAIR data environment is considered necessary to facilitate and accelerate research and development. In particular, it can improve research relevance by preventing analysis from being carried out on small and partial datasets and can facilitate study replicability. These two problems are among the main ones plaguing the current research community and decisive actions must be taken in the coming years.

Finally, it is possible to mention further research areas opened by this work that can be summarised as follows: (1) the extension to other renewable and sustainable energy technologies included in [4], which should be carried out in parallel with (2) extending the harmonisation and deepening the detail for ancillary aspects such as grid connection and support functions, (3) extending the work to software and (4) using the taxonomy and metadata to structure a catalogue for facilitating user searches of classified data and finally (5) developing automated tools to verify the quality of the datasets proposed and suggesting the most relevant tags. It is believed that the development of these steps will considerably facilitate research and innovation in sustainable energy.

Author Contributions: Conceptualisation, A.M. and A.M.S., methodology, A.M. and A.M.S.; data curation, A.M.S., S.P., P.P.-L., A.F. and D.M.; writing—original draft preparation, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: The activity leading to the wind energy metadata and taxonomies has received funding from the European Union Seventh Framework Programme under the agreement 609795.

Data Availability Statement: An xlsx file with the full taxonomy and metadata described in this paper is available at the following link: https://cloud.minesparis.psl.eu/index.php/s/KqUCPI8yLuMPCsc (accessed on 26 September 2022).

Acknowledgments: The authors wish to thank Anna Spoden from EUREC and Klemens Marx from VIRDAD for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Murray-Rust, P. Open Data in Science. *Nat. Prec.* **2008**. [CrossRef]
2. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Mons, B.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
3. DCMI Metadata Terms. Available online: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (accessed on 26 September 2022).
4. TEG Final Report on the EU Taxonomy. European Commission—European Commission. Available online: https://ec.europa.eu/info/files/200309-sustainable-finance-teg-final-report-taxonomy_en (accessed on 26 September 2022).
5. Global Wind Atlas. Available online: https://globalwindatlas.info (accessed on 26 September 2022).
6. Global Solar Atlas. Available online: https://globalsolaratlas.info/map?c=11.609193,8.4375,3 (accessed on 26 September 2022).
7. ENTSO-E Transparency Platform. Available online: https://transparency.entsoe.eu/ (accessed on 26 September 2022).
8. Openmod—Open Energy Modelling Initiative. Available online: https://openmod-initiative.org/ (accessed on 26 September 2022).
9. Sempreviva, A.M.; Vesth, A.; Bak, C.; Varelst, D.R.; Giebel, G.; Danielsen, H.K.; Mikkelsen, L.P.; Andersson, M.; Vasiljevic, N.; Barth, S.; et al. *Taxonomy and Metadata for Wind Energy Research & Development*; Zenodo; FP7 collaborative project, IRPWIND Grant agreement no 609795, Work Package 2—Deliverable D2.3 Lead Beneficiary; DTU Wind Energy: Roskilde, Denmark, 2017. [CrossRef]
10. Forsyth, T.; Jimenez, T.; Preus, R.; Tegan, S.; Baring-Gould, I. *The Distributed Wind Cost Taxonomy*; NREL/TP-5000-67992, 1349551, mars 2017; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2017. [CrossRef]
11. Dykes, K.L.; Zahle, F.; Merz, K.; McWilliam, M.; Bortolotti, P. *IEA Wind Task 37: Systems Modeling Framework and Ontology for Wind Turbines and Plants*; (No. NREL/PR-5000-68864); National Renewable Energy Lab. (NREL): Golden, CO, USA, 2017. Available online: https://www.osti.gov/biblio/1375625 (accessed on 26 September 2022).
12. Papadopoulos, P.; Cipcigan, L. Wind turbines' condition monitoring: An ontology model. In Proceedings of the 2009 International Conference on Sustainable Power Generation and Supply, Nanjing, China, 6–7 April 2009; pp. 1–4. [CrossRef]
13. Zhou, A.; Yu, D.; Zhang, W. A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. *Adv. Eng. Inform.* **2015**, *29*, 115–125. [CrossRef]
14. Küçük, D.; Arslan, Y. Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles. *Renew. Energy* **2014**, *62*, 484–489. [CrossRef]
15. Ramasamy, V.; Feldman, D.; Desai, J.; Margolis, R. *US Solar Photovoltaic System and Energy Storage Cost Benchmarks: Q1 2021*; NREL/TP-7A40-80694, 1829460, MainId: 77478; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2021. [CrossRef]
16. Fu, R.; Remo, T.W.; Margolis, R.M. *2018 U.S. Utility-Scale Photovoltaics-Plus-Energy Storage System Costs Benchmark*; NREL/TP–6A20-71714, 1483474; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2018. [CrossRef]
17. Beck, A.L.; Rai, V. Solar soft cost ontology: A review of solar soft costs. *Prog. Energy* **2019**, *2*, 012001. [CrossRef]
18. Moser, D.; Del Buono, M.; Jahn, U.; Herz, M.; Richter, M.; De Brabandere, K. Identification of technical risks in the photovoltaic value chain and quantification of the economic impact: Economic impact of technical risks in photovoltaic plants. *Prog. Photovolt. Res. Appl.* **2017**, *25*, 592–604. [CrossRef]
19. Risk Matrix—TRUST PV. Available online: https://trust-pv.eu/reports/risk-matrix/ (accessed on 26 September 2022).
20. Solar Bankability Project. Available online: http://www.solarbankability.org/home.html (accessed on 26 September 2022).
21. Abanda, F.H.; Tah, J.H.; Duce, D. PV-TONS: A photovoltaic technology ontology system for the design of PV-systems. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1399–1412. [CrossRef]
22. Khosrojerdi, F.; Gagnon, S.; Valverde, R. Proposing an Ontology Model for Planning Photovoltaic Systems. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 582–600. [CrossRef]
23. Nihar, A.; Curran, A.J.; Karimi, A.M.; Braid, J.L.; Bruckman, L.S.; Koyutürk, M.; Wu, Y.; French, R.H. Toward Findable, Accessible, Interoperable and Reusable (FAIR) Photovoltaic System Time Series Data. In Proceedings of the 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), Fort Lauderdale, FL, USA, 20–25 June 2021; pp. 1701–1706. [CrossRef]
24. Toledo, C.; Scognamiglio, A. Agrivoltaic Systems Design and Assessment: A Critical Review, and a Descriptive Model towards a Sustainable Landscape Vision (Three-Dimensional Agrivoltaic Patterns). *Sustainability* **2021**, *13*, 6871. [CrossRef]
25. Piazza, A.; Faso, G. Concentrated Solar Power: Ontologies for Solar Radiation Modeling and Forecasting. In *Advances onto the Internet of Things*; Gaglio, S., Lo Re, G., Eds.; Springer International Publishing: Cham, Switzerland, 2014; Volume 260, pp. 325–337. [CrossRef]
26. Kantamneni, A.; Brown, L.E. An Ontology for Solar Irradiation Forecast Models. In Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Seville, Spain, 18–20 September 2018; pp. 263–270. [CrossRef]
27. Kuczenski, B.; Davis, C.B.; Rivela, B.; Janowicz, K. Semantic catalogs for life cycle assessment data. *J. Clean. Prod.* **2016**, *137*, 1109–1117. [CrossRef]
28. Janowicz, K.; Krisnadhi, A.A.; Hu, Y.; Suh, S.; Weidema, P.; Rivela, B.; Cheatham, M. A minimal Ontology Pattern for Life Cycle Assessment Data. In Proceedings of the 6th Workshop on Ontology and Semantic Web Patterns (WOP 2015), Bethlehem, PA, USA, 11 October 2015.

29. Peças, P.; Götze, U.; Henriques, E.; Ribeiro, I.; Schmidt, A.; Symmank, C. Life Cycle Engineering—Taxonomy and State-of-the-Art. *Procedia CIRP* **2016**, *48*, 73–78. [CrossRef]
30. Saxe, S.; Guven, G.; Pereira, L.; Arrigoni, A.; Opher, T.; Roy, A.; Arceo, A.; Von Raesfeld, S.S.; Duhamel, M.; McCabe, B.; et al. Taxonomy of uncertainty in environmental life cycle assessment of infrastructure projects. *Environ. Res. Lett.* **2020**, *15*, 083003. [CrossRef]
31. ShareWind. Available online: http://www.share-wind.net/ (accessed on 20 September 2022).
32. Khazraei, K.; Deuse, J. A strategic standpoint on maintenance taxonomy. *J. Facil. Manag.* **2011**, *9*, 96–113. [CrossRef]
33. LCI-Webteam. Life Cycle Considerations within the EU Taxonomy for Sustainable Finance—Life Cycle Initiative. 5 May 2020. Available online: https://www.lifecycleinitiative.org/life-cycle-considerations-within-the-eu-taxonomy-for-sustainable-finance/ (accessed on 26 September 2022).