



Knowledge engineering for wind energy

Yuriy Marykovskiy^{1,2}, Thomas Clark³, Justin Day⁴, Marcus Wiens⁵, Charles Henderson⁶, Julian Quick⁷,
Imad Abdallah¹, Anna Maria Sempreviva⁷, Jean-Paul Calbimonte⁸, Eleni Chatzi¹, and Sarah Barber²

¹Institute of Structural Engineering (IBK), Swiss Federal Institute of Technology (ETH Zurich),
Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

²Institute for Energy Technology, Eastern Switzerland University of Applied Sciences (OST),
Oberseestrasse 10, 8640 Rapperswil, Switzerland

³Octue Ltd, British Antarctic Survey, High Cross, Madingley Road, CB3 0ET Cambridge, UK

⁴Pacific Northwest National Laboratory, Richland, Washington, USA

⁵Fraunhofer Institute for Wind Energy Systems IWES, Am Luneort 100, 27572 Bremerhaven, Germany

⁶Stacker Group, 708 Altavista Ave., Charlottesville, VA 22902, USA

⁷Department of Wind and Energy Systems, Technical University of Denmark, Risø Campus
Frederiksborgvej 399, 4000 Roskilde, Denmark

⁸Institute of Informatics, University of Applied Sciences and Arts of Western Switzerland (HES-SO),
Technopôle 3, 3960 Sierre, Switzerland

Correspondence: Yuriy Marykovskiy (yuriy.marykovskiy@ost.ch)

Received: 17 December 2023 – Discussion started: 3 January 2024

Revised: 12 February 2024 – Accepted: 13 February 2024 – Published: 12 April 2024

Abstract. With the rapid evolution of the wind energy sector, there is an ever-increasing need to create value from the vast amounts of data made available both from within the domain and from other sectors. This article addresses the challenges faced by wind energy domain experts in converting data into domain knowledge, connecting and integrating them with other sources of knowledge, and making them available for use in next-generation artificial intelligence systems. To this end, this article highlights the role that knowledge engineering can play in the digital transformation of the wind energy sector. It presents the main concepts underpinning knowledge-based systems and summarises previous work in the areas of knowledge engineering and knowledge representation in a manner that is relevant and accessible to wind energy domain experts. A systematic analysis of the current state of the art on knowledge engineering in the wind energy domain is performed with available tools put into perspective by establishing the main domain actors and their needs, as well as identifying key problematic areas. Finally, recommendations for further development and improvement are provided.

1 Introduction

1.1 Extracting value from data

In the wind energy sector, it is becoming increasingly important to create value from data (Veers et al., 2019). To this end, vast amounts of data generated by various sources, including sensors and other monitoring systems, need to be effectively structured and represented in a way that can be easily understood and processed by both artificial intelligence (AI) systems and humans. The digitalisation of the wind energy sector is one of the key drivers for reducing costs and

risks over the whole wind energy project life cycle (Klonari et al., 2021). The digitalisation process encompasses solutions such as digital twins, decision support systems, and AI systems, some of which need to still be developed, in order to contribute to reducing operation and maintenance costs, for increasing the amount of energy delivered, and for maximising the efficiency of wind energy systems. In this context, the term “knowledge-based systems (KBS)” refers to AI systems that formalise knowledge as rules, logical expressions, and conceptualisations (Akerkar and Sajja, 2009; Davis, 1986). Such systems can be realised as AI-enabled digital twins or

decision support systems that rely on databases of knowledge (also referred to as knowledge bases or knowledge graphs), which contain machine-readable facts, rules, and logics about a domain of interest, to assist with problem-solving and decision-making (Hogan et al., 2021).

1.2 The need for managing data

Currently, the stage for the digital transformation in wind energy is set by the democratisation of computing, technological maturity of AI systems, and the reduction in costs of data storage and sensing technologies. Along with this, a necessity to structure, organise, manage, and make use of substantial amounts of operational and synthetic data has emerged (Naghieb et al., 2022). However, it is often the case in industrial settings that data are not treated as an asset. Even though the importance of efficient data management has already been recognised by major stakeholders both in industry and academia (Veers et al., 2019), only a few organisations can afford to have a person dedicated to oversee data-related activities (Clifton et al., 2023). This has left many domain experts one-on-one with the problems related to the actual, practical use of data (Barber et al., 2023c). The FAIR (findable, accessible, interoperable, and reusable) data approach introduced by Wilkinson et al. (2016) provides general data management guiding principles. However, FAIR has mostly been applied in academic settings, and there is a disconnect between conceptual or descriptive guidelines and concrete implementations or defined prescriptions and practices. Several groups such as GO FAIR¹, the Data Readiness Group², and the Research Data Alliance³ have emerged in recent years in an effort to provide practical implementation recommendations and solutions for increasing the FAIRness of data. Nevertheless, creating FAIR data frameworks still remains one of the major challenges in the digitalisation process (Wierling et al., 2021).

1.3 The challenge addressed in this paper

Wind energy experts facing the challenge of managing their data will most likely find themselves overwhelmed by unfamiliar terms such as “data schema”, “relational data model”, and “metadata”. They may ask questions such as “what are the differences between a Structured Query Language (SQL) database, graph database, and an object store?”, “which one would fit best to my data types?”, or “how do I publish my data on the web so that it conforms to the FAIR principles?”. The same holds true during practical development of AI-enabled systems. In this context, a wind energy domain expert is increasingly expected to grasp concepts such

as schema/ontology development, logic, and semantic networks, among others. Moreover, they often have to interact with a rather complex technology stack that includes data formats like Extensible Markup Language (XML), JavaScript Object Notation (JSON), or Semantic Web technologies⁴ such as Resource Description Framework (RDF) (Schreiber and Raimond, 2014), SPARQL Protocol and RDF Query Language (SPARQL) (Group, 2013), and Web Ontology Language (OWL) (Hitzler et al., 2009), among others. These issues are particularly acute in the wind energy sector due to the fact that the industry is relatively new, the systems are highly multidisciplinary, and the relevant disciplines are currently highly siloed (Clifton et al., 2023). Additionally, the modelling and measurement uncertainties are largely due to the high complexity of wind energy systems and difficulties in measuring operational data. As a result of all the factors mentioned above, wind energy data are often hidden but are, even worse, ill managed, missing documentation and context, uncertain, or incomplete. Recent efforts that have started to address the problem of data and knowledge management in the wind energy domain have not yet gained traction in the community. This is due to difficulties of cross-domain interactions, knowledge silos, lack of awareness among stakeholders, and other cultural and organisational factors (Heidenreich and Mattes, 2022; Clifton et al., 2023; Kirkegaard et al., 2023). The need for holistic knowledge-based systems, however, is increasingly providing the necessary external pressure for the natural evolution and emergence of commonly accepted and adopted paradigms.

1.4 Contribution of this paper

To address the aforementioned challenges, this paper presents the main concepts and summarises previous work in the areas of knowledge engineering and knowledge representation in a manner that is relevant and accessible to wind energy domain experts. The insights presented in this article are not only beneficial for the wind energy sector, but also applicable to other domains undergoing digitalisation.

The paper is structured as follows: Sect. 2 presents the scope of knowledge engineering activities and the common roles in the overall context of digitalisation. Section 3 presents a conceptual overview of the knowledge representation problem in general and introduces the basic concepts, or vocabulary, of the knowledge engineering domain. Section 4 discusses practical technological implementations that enable the adoption of knowledge representation solutions, with a specific focus on web technologies. Section 5 discusses how the knowledge representation technologies and knowledge engineering techniques enable the development of AI systems, in particular, AI-enabled digital twins. Section 6 presents a systematic and methodological overview of

¹<https://www.go-fair.org/> (last access: 10 September 2023)

²<https://datareadiness.eng.ox.ac.uk/> (last access: 15 September 2023)

³<https://www.rd-alliance.org/> (last access: 16 September 2023)

⁴https://www.w3.org/2001/sw/wiki/Main_Page (last access: 25 March 2024)

the current initiatives by the wind energy community in the knowledge engineering domain. Recommendations for fostering a healthy and thriving wind energy knowledge engineering ecosystem are introduced in Sect. 7, followed by the concluding remarks in Sect. 8.

2 Knowledge engineering: scope and activities

Knowledge engineering refers to activities related to the development of AI systems capable of processing, interpreting, and performing logical operations on structured data (Studer et al., 1998). Knowledge representation refers to representing, or structuring, real-world information in a way that renders this exploitable by AI systems. This involves choosing an appropriate representation language or formalism and determining how to map knowledge from the real world to the chosen representation.

Knowledge engineering activities often overlap with data management. In particular, the creation of conceptual data models (also referred to as semantic data models), which, conventionally, fall under the umbrella of data management, are also instrumental in the development of KBSs. When designing a KBS, conceptual data models are used to represent the knowledge needed by the said system. In a broader context, data management also involves activities related to the storage and maintenance of this knowledge. This includes defining how the knowledge is structured and stored, how it is accessed and updated, how its quality is ensured, and how it is integrated with other systems. While data modelling activities involve the structuring of data, in a similar manner to knowledge representation, the focus of data management is slightly different from that of knowledge engineering. Knowledge engineering is focused on capturing, representing knowledge, and logical reasoning and inference. Data management is focused on the overall process of collecting, storing, and using data within an organisation. As part of this process, conceptual data modelling is focused on the identification and organisation of key concepts and relationships.

In Fig. 1 some typical roles and activities are presented. It is important to note that, in practice, it might not always be possible to clearly distinguish between these actors in a given organisation and there is no uniquely agreed upon classification. For example, roles like data modeller and/or database designer are often considered to reflect a narrower role of a data manager, as opposed to standalone positions. Terms like data steward are in some cases used interchangeably with data manager, while others are used to denote more specific roles like the ones related to data governance. For the purpose of this paper, we distinguish between the following roles: data manager, data engineer, data scientist, and knowledge engineer. A wind energy domain expert involved in the process of digitalisation is likely to interact with all these types of actors. This requires understanding of the jargon, work-

flows, and methodologies used in each respective domain. For this reason, some of the concepts from data science, data management, and data engineering are also discussed in the present work. However, given the vastness of these topics, this paper cannot be an exhaustive source on all these matters but rather a simplified overview for a domain scientist of important concepts, directing the readers to the relevant works.

3 Knowledge representation: conceptual overview

In order to understand the practical application of knowledge engineering, it is important to be familiar with the concepts of knowledge representation and formal systems in general. The starting point of the discussion is human-centric. It revolves around the human perception, interpretation, and understanding of the world. The question of how humans model the world is an open-ended one, and, consequently, the question of how the knowledge should be represented and shared does not have a unique answer. This discussion about the nature and organisation of the world is the area of interest of a branch of philosophy called ontology. In a knowledge engineering context, the underlying assumption is that the world, or domain of interest, consists of entities, relationships, and concepts.

Knowledge representation deals with the problem of capturing the meaning of facts (i.e. the aforementioned entities, relationships, and concepts) from a certain domain of interest in a formal way as structured data. For example, consider the following informal text:

An IET group at OST acquired an Aventa AV-7 wind turbine, located in Winterthur, and has instrumented one of the blades with a novel pressure measurement system. The dataset produced by the system for the month of July is now available upon request. Additionally, during the measurement campaign wind turbine SCADA data were acquired and the inflow characteristics were measured with a lidar.

A wind energy domain expert, especially a researcher with a background in experimental measurements, would not find it particularly difficult to understand and interpret this description. As a result, they can infer some additional information about the mentioned dataset. That is to say, the text provides contextual information about measurements that a domain expert uses to assign meaning to particular data. However, there are several limitations to this representation of knowledge. First, this information is meant to be processed by a human (as opposed to some automated algorithm) with some command of the English language. Secondly, the reader must be a domain expert to infer the context and purpose of the text as well as resolve the inherent ambiguity of some of the statements. The underlying assumption is that a do-

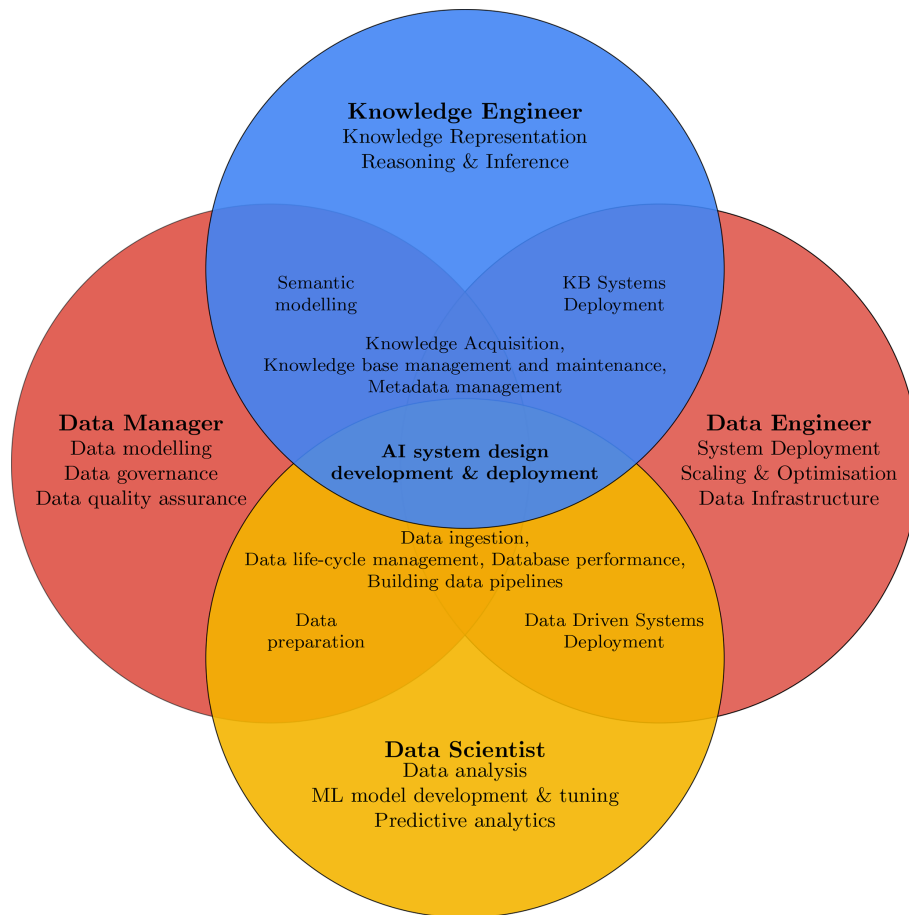


Figure 1. Roles and activities overlap during AI system design, development, and deployment. A data manager and data engineer often provide a supporting role for a data scientist or a knowledge engineer.

main expert will rely on some informal logical framework (Groarke, 2022) and personal domain knowledge.

Sidenote. The term “formal” is mostly used as in a “formal system” (The Editors of Encyclopaedia Britannica, 2012). However, since AI systems are computational systems, i.e. machines, performing data manipulation based on a set of instructions (i.e. algorithms), in a knowledge engineering context, “formal” may also mean machine-interpretable. “Informal” is mostly used to denote something outside of such formal setting, for example, assertions made using natural language (Johnson and Blair, 2002).

In the example provided, some inferences are trivial for a human, while others are more complicated. For instance, it is clear that Aventa AV-7 is a type of a wind turbine. A domain expert will also have an understanding of what SCADA data may contain or what the lidar measurements may look like. Meanwhile, other information is not as straightforward. It is not clear what IET or OST is, especially if the text is presented by itself and not as a blog post on the university web

page. Similarly, the month of July is not enough to pinpoint the time period without any additional context like the time and date when the text was written. Some information is fully absent, such as how the data are structured, what are the units used, whether the dataset is free, and under what licence it is distributed. This example brings up the concepts of semantics, pragmatics, context, metadata, language, logic, open-world vs. closed-world assumption, and ontology, which will be discussed below.

3.1 Understanding representation: semantics, pragmatics, context, and metadata

The process of understanding and interpreting a particular representation involves semantics, pragmatics, and context. Semantics is the study of meaning in language and is concerned with the relationship between words or symbols and their counterparts in the real world (Cann et al., 2009). In the knowledge representation context, comprehensive semantics ensures that the terms used to describe data and information are unambiguous and clearly defined. Pragmatics, on the other hand, is concerned with the social and cultural fac-

tors that influence the use of language (Andersen and Aijmer, 2011). In knowledge representation, pragmatics ensures that the meaning of a term or concept is understood in the appropriate social and cultural context. This can be particularly important when working with data or information from different disciplines or cultures.

In the realm of knowledge representation, context is pivotal as it impacts the interpretation of information and ultimately its meaning. Context may encompass a variety of factors, including the data source, the conditions under which the data were collected, their intended use, or their relationship to other data. Metadata is data that provide this context as structured information about a dataset. It is a form of formalised context, to make the representation not only human-interpretable, but also machine-interpretable. For example, when publishing aerodynamic measurements and SCADA data on the web, additional context can be provided by wind turbine characteristics. This contextual information can be expressed in a natural language (like English), in a form of technical specification sheets provided by the producer, or ultimately as some kind of formal representation. Such metadata would enable a data scientist to draw more meaningful conclusions while performing data analysis. For instance, the knowledge of the location along with historical weather data can provide understanding that a clustered group of measurements is due to an icing event or the specification of wind turbine status codes can link measurements to a certain wind turbine component failure. The relevant question in this case is how to represent the knowledge about a particular wind turbine in a formal way. This is explored further as we discuss modelling languages, their expressive power, and formalising representations.

3.2 Expressing representation: language

In the example text given in the introduction to this section, the authors relied on the English language as a means of knowledge expression. Similarly, knowledge engineers and data managers rely on modelling languages for knowledge representation and data structuring. On a fundamental level, the formal basis for modelling languages is provided by logic. Any conceptual model can be specified using some kind of logical language.

Definitions. “Modelling languages” are formal languages that express information, knowledge or systems in a structure that is defined using a certain syntax (i.e. a consistent set of rules). In this paper, we will discuss the most notable knowledge modelling languages in the context of knowledge engineering, computer sciences, and information technologies. However, in the broader context of expressing knowledge, systems, and processes many other modelling language exist such as Unified Modeling Language (UML), Integration Definition (IDEF) languages, and Petri net, to name a few.

“Logic languages” are formal languages that provide a way to express logical statements and reason about them. Logic languages include syntax rules and a set of semantics that allow users to formally define and manipulate logical statements. Examples of logic languages include predicate logic, description logics (DL), first-order logic (FOL), and fuzzy logic.

For example, to represent the fact that Aventa AV-7 is a type of a wind turbine, it is possible to use first-order logic (FOL) expressions:

$$\forall x : AventaAV7(x) \rightarrow WindTurbine(x).$$

This statement can be read in English as “for all x , if x is a AventaAV7, then x is a wind turbine”. It is also possible to express similar semantics using description logics (DL) expressions:

$$AventaAV7 \sqsubseteq WindTurbine.$$

This statement can be read as “all AventaAV7s are wind turbines”. These two statements in two different languages use different syntax to convey similar (but not exactly the same) semantics.

The use of the logic languages in the context of information systems is rather impractical. As can be seen from the example above, the construction of a rather simple fact using FOL is often verbose and complex. This verbosity and complexity can result in misunderstandings, errors, and increased difficulty in managing and manipulating the data. Hence, in the domain of knowledge engineering and data modelling, specialised modelling languages are used, as discussed in Sect. 4.3.

3.3 Representation complexity: language expressive power

The choice of representation language depends on the desired semantics of the statements and on their complexity. For example, a formal dataset description published on the web that focuses only on the information presented, without a connection to other concepts or attributing additional semantics to the relationships with other entities, does not require high expressivity. The metadata in such description might include fields for the specific turbine model (Aventa AV-7) and the location (Winterthur). However, consider two statements about the OST-WindTurbine:

locatedIn(OST-WindTurbine, Winterthur),

locatedIn(OST-WindTurbine, Switzerland).

A more expressive language can define “locatedIn” to be a transitive relationship. If the system performing automated reasoning also has access to the fact that Winterthur is located in Switzerland, then a second statement can be automatically inferred without it being explicitly defined.

Even though, in general, most of the facts about the world can be described using FOL, it may be more practical to use languages based on DL for certain representations. For example, DL is especially powerful in situations where the goal is to represent knowledge in a structured and formalised way, such as in the creation of ontologies (see Sects. 3.5 and 4.3). At the same time, both FOL or DL are impractical for representing certain facts and knowledge, such as complex or dynamical systems. While FOL can express relationships between objects and properties, it does not offer a good mechanism for expressing compositionality, causality, observations of the states of the system, and related uncertainties. To describe these in a formal way, one would need to use a representation with different underlying theory. Recently, category theory formalisms have been proposed by Spivak and Kent (2012) for knowledge representation to provide more expressive power in terms of compositionality, which allows domain experts to describe how constitutive parts of complex systems are interrelated and combined together. Additionally, applied category theory and type theory have been proposed as basis for various applications such as modelling dynamic systems (Spivak, 2020; Jaz Myers, 2021; Lavore et al., 2024; Shapiro and Spivak, 2023; St. Clere Smithe, 2023), formalising co-design problems (Zardini et al., 2021), data management (Spivak, 2012; Johnson et al., 2012), and creation of digital twins (Qi et al., 2022). For sequential decision-making, the value of information theory and partially observable Markov decision process (POMDP) model has been used to express relationship between an agent and its dynamic system environment along with related uncertainties (Papakonstantinou and Shinozuka, 2014; Andriotis et al., 2021), with applications in a wind energy context (Morato et al., 2022; Liang et al., 2022; Hlaing et al., 2022). It should be noted that many of the above-mentioned formalisms have not yet received widespread adoption and therefore often lack practical technological implementations (see Sect. 4), as opposed to DL and FOL. Moreover, in many use cases, wind energy domain experts still rely on ad hoc algorithms and models as discussed Sect. 6.

3.4 Representation assumptions: open world vs. closed world

The choice of a logical language for knowledge representation can be influenced by the open-world and closed-world assumptions (OWA and CWA) (Magee, 2011). In the CWA, it is assumed that everything not known to be true is false. This assumption is used in some logical languages such as FOL. In this context, the goal is to explicitly state all the necessary information about a domain and derive logical consequences based solely on this information. The CWA is useful in situations where the domain is well defined and the data are complete. In contrast, the OWA states that everything not known to be true is simply unknown. This assumption is used in some logical languages such as DL. In this context, the

objective is to define a set of axioms and a set of incomplete data. The logical consequences derived from these axioms and data are considered true until proven false. The OWA is useful in situations where the domain is complex, dynamic, and the data are incomplete.

Consider the task of creating a common representation of a wind turbine. When describing a wind turbine it is reasonable to include information about the sensors and the control systems installed on it. However, as new sensing technologies and control strategies are developed, wind turbines are upgraded to improve their performance and management. A particular wind turbine instance, at a given time, might not have been upgraded. In this case, the absence of data about the sensors is ambiguous: the wind turbine in question might not have sensors yet installed, the upgrade is not possible, or the information was not available when the data were compiled. Interestingly, this question is related to the admission of “null” values in databases. It can be argued that if the information is incomplete, it should not be in the database in the first place. Of course, when dealing with a complex domain such as wind energy and complex systems such as wind power plants, the idea of complete representation is rather ludicrous.

Sidenote. Consider the statement $\exists x(\text{WindTurbine}(x) \wedge \text{ratedPower}(x) > 20\text{MW})$. To evaluate this statement and obtain a true or false answer, one has to impose a certain restriction and assume complete knowledge of all the wind turbines and their power ratings in existence (or at least in the domain of interest). In an open-world assumption, on the other hand, the question of existence of a wind turbine that is rated for more than 20 MW of power generation (just like the existence of pink elephants and unicorns) remains, well, open.

It is important to note that neither assumption is inherently superior – the choice between the two depends on the specific application and the nature of the data and knowledge being represented. As mentioned above, some logical languages are designed to work with the CWA, while others are designed to work with the OWA. For example, adopting an OWA- and FOL-based representation would lead to undecidability. That is to say there is no algorithm to formally prove the “truthfulness” of the statements made using this logic. The FOL is the underlying logic for Structured Query Language (SQL) databases (see Sect. 5.3). Thus these databases usually operate under CWA. The DL is the basis for Web Ontology Language (OWL), which is discussed in Sect. 4.3.

3.5 Formalising representations: ontology

Ontology is a broadly used term that can take on different meanings. As mentioned in the introduction to this section, it can be used to denote a branch of philosophy. In the context of knowledge engineering and KBS, however, ontology

has been defined as “explicit specification of a conceptualisation” by Gruber (1993). Here, “explicit” means that the types of concepts used, and the constraints on their use, are explicitly defined. That is to say, each concept, attribute, relationship, and rule in the ontology is precisely articulated, often through formal semantics. This explicitness avoids ambiguity and fosters understanding, allowing the ontology to serve as a shared and common description of a domain that can be communicated across people and systems. In addition, being explicit in an ontology also means that it is machine-interpretable. This is crucial for automated processing, reasoning, and interoperability in computer systems. With explicit ontologies, computers can process the semantic meaning of data, enabling more advanced and flexible uses of the data, such as inference and knowledge discovery.

“Conceptualisation”, in Gruber’s definition, refers to an abstract view or model of the world, i.e. the types of objects, concepts, and other entities that are assumed to exist in a domain of interest and their associated properties and relationships. An ontology, then, serves as a specific and concrete representation of that conceptualisation, allowing the underlying assumptions about the domain to be made explicit and facilitating their communication and processing.

Sidenote. It is possible to think of an ontology as a directed labelled graph, where each “concept” (equivalently “type” or “class”) is represented as a node, and the edges represent the relationships (equivalently “properties”). In fact, ontologies are often presented visually as a graph and expressed using graphical languages, as exemplified in Fig. 2.

Given this general definition, controlled vocabularies, formal taxonomies, and schemas are also ontologies. Similarly, conceptual (semantic), logical, and physical data models can also be thought of as ontologies. Generally, the difference in the use of these terms relates to the complexity (or “expressiveness”) of the specification, with the term “ontology” typically being used to denote more expressive one (Lassila and McGuinness, 2001). In an effort to avoid ambiguity, some communities adopt an umbrella term – “semantic artefact” (Le Franc et al., 2020) – to denote conceptualisations with various degrees of expressiveness and reserve the term “ontology” only when referring to the most expressive conceptualisations.

Definitions. In this paper we use different terms to describe various types of semantic artefacts.

A “controlled vocabulary” is used to describe a set of terms or phrases that have been pre-selected and authorised for use in a particular domain or context. It provides a standardised way of naming and describing concepts, which helps to improve consistency and accuracy in indexing, searching, and retrieval of information. In some contexts the term may be used to refer to a specific

set of terms or concepts without any explicit relationships between them. However, in some fields the term may be used to denote a taxonomy.

A “formal taxonomy” is a hierarchical classification conceptualisation that organises concepts or objects based on their relationships to one another. It is possible to view taxonomy as an ontology that includes only a subsumption relationship between classes. In practical terms it means that it possible to express a relationship of a type `isSubclassOf` (or equivalently `isA`) between the classes, thus modelling class hierarchy. Some taxonomies may, however, include other types of hierarchical relationships. For example, the Simple Knowledge Organisation System (SKOS) data model for taxonomy description defines semantic relationships of the type `skos:broader`, `skos:narrower`, `skos:related`, etc. We use the term “taxonomy” in the latter, broader definition, thus including SKOS taxonomies.

A “schema” defines the relationships between different concepts and entities. In the knowledge engineering community, the term is used to refer to simple conceptualisations, like the ones written using less expressive knowledge representation languages such as RDF Schema⁵. In the data management community, a schema is usually used to indicate a blueprint or framework that defines the structure and content of a particular type of data or information (e.g. JSON Schema for JSON data). It may specify the types of data elements that are allowed, their relationships to one another, and the rules for encoding or validating them. We use term schema mostly in this last sense.

An “ontology” written in an expressive ontology language (see Sect. 4.3 for more discussion about ontology languages) can represent rich, complex knowledge about concepts and their interrelationships. For example, ontologies expressed with OWL-DL can define transitive, inverse, reflexive, and irreflexive relationships; impose cardinality restrictions; and so on.

Another difference relates to the way and the context in which the specification is defined – a representation of a piece of (meta)data in a JSON Schema would be called a schema, whilst the same representation expressed in Web Ontology Language (OWL) would likely be referred to as an ontology. In the context of database design, a specification describing a structure of a database would more likely be referred to as database schema or logical data model or logical schema, rather than as an ontology (Spyns et al., 2002). At the same time, an ontology that is populated with instances is often referred to as a knowledge graph or a knowledge base

⁵<https://www.w3.org/wiki/SchemaVsOntology> (last access: 1 October 2023)

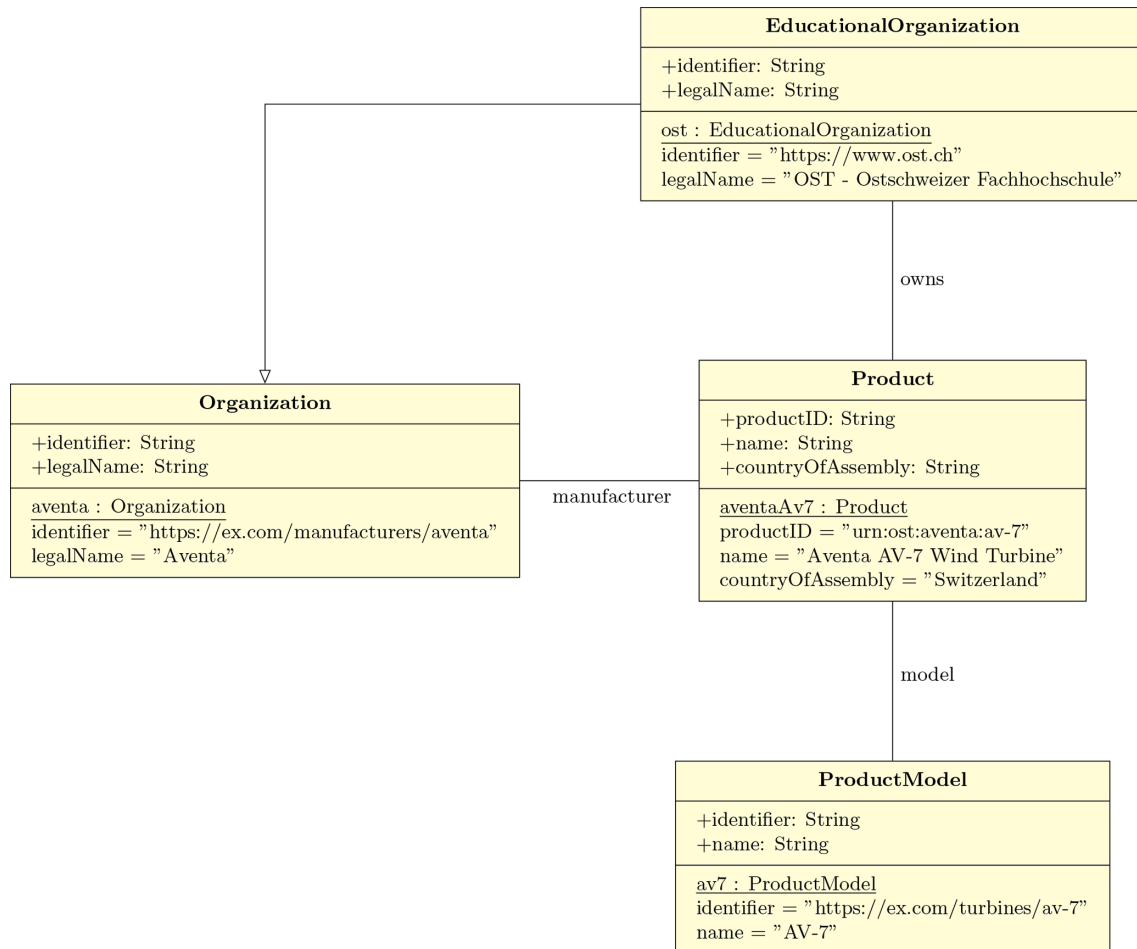


Figure 2. Information about a specific wind turbine expressed with an UML diagram. A part of Schema.org ontology contains instantiated concepts of “EducationalOrganisation”, “Organisation”, “Product”, and “Product Model”. The concepts are connected by “isSubclassOf”, “owns”, “model”, and “manufacturer” relationships.

rather than as a database (Heist et al., 2020). This is done to distinguish these representations from relational databases (see Sect. 5.3). In this paper, the main focus is on the most expressive side of the ontology spectrum, as these can describe and formalise more complex relationships, which can facilitate the creation of the type of AI system needed for the digitalisation of wind energy, as discussed more in Sect. 5.

Using the dataset publishing example, it is possible to demonstrate how semantic expressiveness increases when moving from controlled vocabulary to ontology. Controlled vocabulary can include concepts (equivalently “terms”) like “wind turbine”, “pressure measurement system”, “SCADA”, and “lidar”. Using a controlled vocabulary, it is possible to identify and label these key concepts in the paragraph, but there is no explicit representation of the relationships between these concepts or their properties. A taxonomy can include subsumption relationships between the concepts in the paragraph. For example, “lidar” and “pressure measurement system” can be subsumed by “measurement system”.

It is also possible to include other relationships, such as part of relationships between the wind turbine and its blades or between the measurement system and the turbine. A schema can include additional information about the properties of the concepts and their relationships. For example, it can specify the expected attributes of the wind turbine, such as its manufacturer, capacity, and location. Moreover, a schema may provide constraints that specify the relationships between these concepts: the wind turbine’s location is specified by a pair of latitude and longitude coordinates. Finally, an ontology can further specify the meaning and relationships of the concepts in a formal and machine-interpretable way. An aligned ontology will enable knowledge to be connected from different domains. For instance, the use of an ontology of geographical names can help to connect wind turbine data with specific locations and regions. Moreover, a formal ontology enables automated reasoning and inference of the concepts and their relationships; thus, new knowledge can be derived from the information provided. This can form a base

for KBS such as digital twins or decision support systems, which are becoming increasingly important in wind energy in order to reduce costs and increase deployment.

In addition to semantic expressiveness, ontologies can be differentiated based on their scope or level of generality. Here it is common to distinguish top-level, domain, application, and task ontologies (Guarino, 1998).

Definitions. “Top-level ontologies” provide a broad framework for organising concepts and relationships that are applicable across multiple domains or applications. These are useful for inter-domain knowledge exchange.

“Domain ontologies” are aimed to capture domain-specific knowledge. They contain concepts and relationships that are relevant to a specific domain such as medicine or engineering. Adoption of these ontologies can ensure that the terminology used within a particular domain is consistent and clear.

“Application ontologies” are designed to support a specific software application or system. These ontologies provide a more detailed and specialised vocabulary that is tailored to the needs of the application.

“Task ontologies” are focused on the specific tasks or activities that need to be performed within a particular domain or application.

Developing, publishing, and using ontologies needs an effective collaboration among the different actors introduced in Sect. 2 and including domain experts, stakeholders, and target users, each with specific competencies and interests. Ontology Development 101 by Noy and McGuinness (2001) is a good starting point to familiarise oneself with the concept of ontologies, terminology used, and development methods. Another ontology development methodology particularly well suited for application and task ontologies and knowledge base development was proposed by De Nicola et al. (2005). This method focuses on collaboration between domain experts and knowledge engineers during ontology development.

3.6 Common representation: standard

An ontology that is accepted and enforced by a certain community can be included in a standard. In the context of knowledge representation, a standard is a set of guidelines or specifications that prescribe how to represent and organise knowledge in a consistent and interoperable way. Standards ensure that knowledge representations can be shared, reused, and understood by different systems and applications, regardless of their implementation or environment. For example, the Dublin Core metadata schema, which defines essential metadata elements for the web publishing task (e.g. creator, publisher, abstract), has been formally standardised as ISO 15836.

Standard conceptualisations allow for standard data generation and transformation procedures. A central organisation publishes a comprehensive set of standard semantic artefacts, which can be updated based on community feedback, though any revision process will inherently be slow-moving and filled with compromises. It can be expected that organisations will develop and publish their own terminologies, schemas, or ontologies based on their specific needs and use cases.

The development of standards can be affected by whether a OWA or CWA paradigm is adopted by the standard designers. An OWA standard semantic framework allows for modularity between different conceptualisations, where an ecosystem of different semantic artefacts can develop (Villegas et al., 2014; Chah, 2018). This system allows anybody to define or iterate on an ontology or its sub-elements. This OWA framework is particularly useful for developing big-data insights (Rogushina and Gladun, 2020). As an example, turbine Supervisory Control and Data Acquisition (SCADA) data are complex, involving hundreds of thousands of data variables, where the same metadata vocabulary can have different meanings between turbine models. The naming scheme for most of these terms is defined in the IEC 61400-25 standard. However, this standard does not include a machine-readable formalisation as part of its specifications, which would be needed for the digitalisation process (see Sect. 6). A SCADA ontology conforming to an OWA standard semantic framework would allow for standard data transformation procedures while maintaining model-specific semantic heterogeneity. Examples of data transformation standards include fault codes, power curve measurement, and damage estimation.

4 Knowledge representation: technologies

In this section, we discuss the practical aspects of knowledge representation. We will explore various technologies that have been developed to implement the theoretical concepts of knowledge representation we have discussed so far. This includes the Semantic Web’s vision for a web of data and how it interrelates with the FAIR principles, the Resource Description Framework as a way to express graph-based data, and the use of ontology and schema languages for expressing knowledge structures.

4.1 Semantic Web

The Semantic Web is an ambitious extension of the world wide web proposed by the World Wide Web Consortium (W3C) that seeks to create a web of data to make data more machine-readable and interoperable. The core idea of the Semantic Web is made possible through technologies discussed further below, such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL), which allow data to be annotated and related in a machine-understandable

way. At the same time, conceptualisations and abstractions form the foundation of the Semantic Web technology stack (Fig. 3). The web of data vision carries a wealth of practical advantages such as creating knowledge bases (Noy et al., 2019) or providing the necessary technological foundations for the development of decision support systems (Tsalapati et al., 2018; Pease et al., 2020). This has shown early dividends across various industries. For instance, the e-commerce industry provides a valuable precedent for how the Semantic Web can deliver tangible benefits. Online retailers like Amazon and eBay use structured data to enrich product descriptions, enhancing product discoverability and improving customer experience. At the same time, online advertisement companies like Google rely on structured metadata and microdata⁶ for search engine optimisation (SEO) by including them in their knowledge bases⁷. These structured data also allow for better integration with suppliers and logistics providers, creating a more seamless and efficient e-commerce ecosystem.

The web of data is poised to become transformative for the wind energy sector as well, helping to address key challenges around data use. Some of the significant benefits of the Semantic Web include the following.

- *Intelligent data discovery.* Semantic Web improves data discoverability by enabling search engines and applications to understand the context, content, and relationships of data. This can speed up data-driven investigations, like root cause analysis of turbine faults, by helping engineers quickly find relevant data and information.
- *Data interoperability.* The Semantic Web allows for seamless integration of different data formats and sources. This becomes particularly beneficial in the context of the wind energy industry where heterogeneous data – ranging from wind speed measurements and power output statistics to maintenance records and weather forecasts – need to be integrated and analysed for effective decision-making. By structuring and interlinking data on wind turbines, weather conditions, maintenance activities, and grid demand, wind farm operators can create a rich, machine-readable data environment. This can enable intelligent applications, like AI-based predictive maintenance systems and decision support, enhanced operational efficiency, and ultimately decreased levelised cost of energy. Additionally, increased data interoperability facilitates data sharing and collaboration within the wind energy community. Development of shared ontologies for wind energy data

can improve data exchange between different stakeholders – from wind farm operators and maintenance providers to equipment manufacturers and researchers. Such a collaborative approach could accelerate innovation and efficiency gains across the industry.

- *Automation and AI readiness.* The machine-readable nature of Semantic Web data lays the groundwork for automation and AI applications. For the wind energy industry, this means the potential for advanced analytics, predictive maintenance, and automated optimisation of wind farm operations with AI-augmented systems such as digital twins.
- *Data reusability.* Semantic Web encourages the use of standardised schemas and ontologies, making data readily reusable across different contexts and applications. In the wind energy industry, this can facilitate cross-project and cross-site analytics, increasing the confidence in the analysis results and enhancing the understanding of wind turbine performance and reliability.

Interrelation with FAIR principles

A reader might notice a strong similarity between the benefits of the Semantic Web and FAIR (findable, accessible, interoperable, and reusable) principles, proposed by Wilkinson et al. (2016). While the discussion of FAIR principles falls more into the data management domain, the Semantic Web's vision, and specifically the concept of linked data (LD) (Heath and Bizer, 2011), is intertwined with the ultimate goals of FAIR approach. According to W3C themselves, "Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web" (W3C, 2023). At the same time, technologies that enable LD also enable data FAIRness. In fact, the practical recommendations for increasing data FAIRness, such as publishing structured metadata on the web, refer to Semantic Web technologies and LD (Wu et al., 2021). For more discussion about how LD can enable FAIR data see Appendix B.

While there is a significant overlap between the LD and FAIR principles in terms of their instrumental values, the fulfilment of one set of these principles does not generally imply the other. In fact, FAIR principles are descriptive in nature and are technology independent. Moreover, while LD focuses on the interoperability aspect and data openness, FAIR data principles are not restricted to open data. Additionally, FAIR principles introduce requirements of metadata persistence and adherence to community standards. To illustrate the difference in the two perspectives, we can consider how LD and FAIR data are evaluated. A common way to evaluate LD is the "5 star linked data" specification⁸. This concept relies heavily on the use of RDF and other Semantic Web

⁶<https://html.spec.whatwg.org/#microdata> (last access: 25 March 2024)

⁷<https://www.blog.google/products/search/introducing-knowledge-graph-things-not/> (last access: 10 September 2023)

⁸<https://www.w3.org/community/webize/2014/01/17/what-is-5-star-linked-data/> (last access: 10 September 2023)

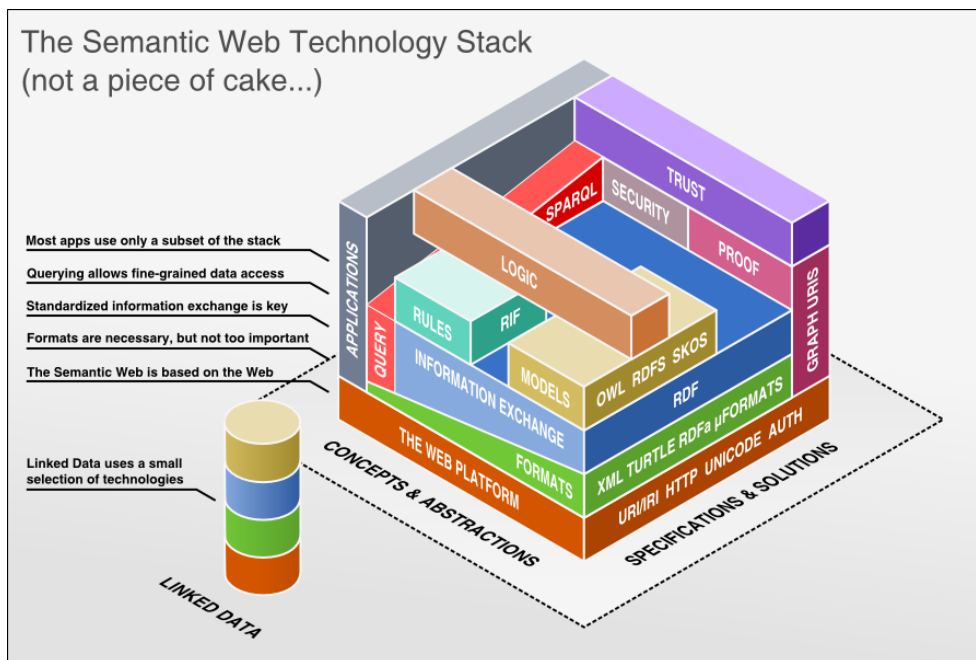


Figure 3. Semantic Web stack (source: Nowack, 2009, available under CC-BY).

technologies. On the other hand, a structured approach to assess the FAIRness of data was proposed by FAIR Data Maturity Model Working Group (2020) with their data maturity model. This model seeks to create a standard understanding of FAIR principles across diverse stakeholder groups. However, it does not dictate the exact means of evaluation or the specific technical solutions. Instead, it offers a degree of flexibility while assessing data FAIRness. This is indicative of the model's recognition of the diverse contexts in which data can exist and the different standards that may apply in different fields or sectors.

4.2 Resource Description Framework

In the context of knowledge engineering, the key technological foundation for ontological representation and information exchange is provided by the Resource Description Framework (RDF). RDF is a W3C standard for representing knowledge in the form of a graph-based data model. It was initially designed as a metadata model for describing resources on the web. RDF serves as one of the fundamental layers in the Semantic Web technology stack. In RDF, data are represented as triples, consisting of subject–predicate–object expressions. The subject is a resource, typically identified by a Uniform Resource Identifier (URI), representing the entity being described or related to another entity. The predicate represents a relationship between the subject and the object, typically identified by a URI as well. The object can be either another resource or a literal value. These triples form a directed graph that can be queried and reasoned about using

various technologies such as SPARQL, a query language designed for RDF, and RDFS (RDF Schema) or OWL ontology languages (discussed further below in Sect. 4.3) built on top of RDF. The following is an example of an RDF statement using Turtle (see Appendix A2 for more information about serialisation formats) syntax.

```
@prefix ex: <http://example.com/resource/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
ex:Aventa_AV-7 rdf:type ex:WindTurbine .
```

Here, Aventa AV-7 (subject) is linked to the concept of a wind turbine (object) by a type (predicate) relationship. As discussed before, an ontology can be visualised as a labelled graph. RDF triplets are a natural way to describe a graph with the subject as a starting node, the predicate indicating a label of the edge, and the object as a target node.

4.3 Ontology and schema languages

Any ontology or schema must be expressed using a language. A schema (or ontology) language is a combination of syntax and semantics (particular to each language) allowing the user to express the structure and content of data. A variety of languages exist to do this. Such languages can be more or less “feature-complete”; i.e. their ability to express complex relations and semantics vary.

Generally, ontology languages are more oriented (in terms of their features and abilities) toward OWA data representation and relation, whilst schema languages are more oriented toward defining and validating CWA data structures. These two schools of thought are converging as the languages them-

selves evolve. For example, JSON Schema (which emerged for validation of closed data coming through web APIs) is increasingly moving toward a full ontological language (Angele and Angele, 2021), while Shapes Constraint Language (SHACL) provides closed-world validation (in the manner of JSON Schema) on open graphs described by RDF.

Sidenote. The language does not have to be a text-based language. For example, IDEF5 is a graphical language that can be used to express an ontology.

A variety of schema languages were reviewed for the purpose of describing CWA data by Clark (2022). The summary of some commonly adopted ontology/schema languages and other data representation technologies is presented in Table 1. A more detailed description and the examples of simple statements made using these languages can be found in Appendix A1.

5 Knowledge engineering: knowledge-based systems

Knowledge-based systems (KBS) are a class of intelligent systems that utilise knowledge engineering techniques to capture, represent, store, and manipulate domain-specific knowledge to solve complex problems, support decision-making, and enable advanced applications. This section explores how the next generation of AI systems, such as cognitive digital twins (CDTs) (Zheng et al., 2021) or autonomous-management digital twins (Wagg et al., 2020), can combine recent developments in machine learning (ML), uncertainty quantification (UQ), verification and validation (V&V), Bayesian approaches, and decision support systems (DSS) with classical rule-based KBS. These hybrid systems are enabled through knowledge integration and interoperability as we discuss in Sect. 5.2. Additionally we touch upon data management and data engineering aspects of the creation of such systems in Sect. 5.3.

5.1 Digital twins

The digital twin (DT) conceptual model was initially introduced in the context of product life cycle management by Grieves (2002) and later adopted for a wide range of applications in various domains, including wind energy. The basis of the DT model is the concept of duality and strong similarity between the physical world and its digital representation (Grieves, 2022). Practical manifestations of DT instances come in a variety of types, depending on the actual realisation of the digital object and the extent to which the strong similarity is achieved. Recently, several attempts at classification of the DT types have been made (van der Valk et al., 2021; Pronost et al., 2021; Uhlenkamp et al., 2022; Marykovskiy et al., 2023).

Sidenote. In their joint position paper, the American Institute of Aeronautics and Astronautics and the Aerospace Industries Association proposed a general definition of a DT as a “virtual representation of a connected physical asset”. Moreover, examples and the added value of 17 different DT types are proposed (AIAA and AIA, 2020).

As digital representation is at the core of the DT concept, knowledge representation and knowledge engineering methods can be, and often are, employed in the development of DT instances. Such DTs can also employ data science methods (Ding, 2019) and DSS (Seyr and Muskulus, 2019) to offer advanced functionalities including integration of heterogeneous data sources, prediction of unmeasured and future quantities based on historical data, and capability to produce actionable insights from updatable models.

In the wind energy domain, DTs can be implemented at various system levels (components, assemblies, wind turbines, wind farms, and grid) and throughout the asset’s life cycle starting from the design phase and ending with the decommissioning. Data integration or ontologies on a higher level provide the backbone for the functional capabilities. Heterogeneous interfaces of single systems can be connected with others by describing the system with a knowledge graph. Thereby, the orchestration of the interactions between subsystems and processes is enabled (Wagg et al., 2020). Semantic technology enables the verification of existing metadata, knowledge inference, and the creation of new knowledge via rule-based reasoners, thus providing cognitive capabilities for CDT-type systems (Zheng et al., 2021; Arista et al., 2023). Additionally, ontologies can be used to describe model interfaces for simulations used in digital twins. In this case, the structure and variables of model inputs and outputs are described and can be utilised in the automated setup of a modular model (Wiens et al., 2021).

Knowledge engineering is crucial in developing digital twins as it integrates heterogeneous data, automates data management and data science workflows, and facilitates connections with other digital twins or models in larger systems. For DTs which include DSS, knowledge engineering provides the ability to perform complex queries, as well as reasoning and inference capabilities. Overall, knowledge engineering methods enhance the functionality and effectiveness of digital twins.

5.2 Knowledge integration and interoperability

Knowledge integration and interoperability lies at the core of knowledge engineering. Ontology-based data integration (OBDI) has emerged as a powerful solution to consolidate and interoperate heterogeneous data sources, utilising ontologies as shared (or aligned) semantic schemas (De Giacomo et al., 2018). Through the use of ontologies, OBDI enables the harmonisation of diverse data sources into a coher-

Table 1. Overview of knowledge representation languages.

Technology	What is it?	Most common use case
RDFS	Data modelling language, for definition of a basic ontology	Creation and publication of ontologies and knowledge bases
OWL	An ontology language built on RDF and RDFS	Creation and publication of ontologies, allowing definition of more expressive relationships (e.g. transitivity, reflexivity).
SHACL	A schema language for describing RDF data	Validation of data in an RDF graph
JSON-LD	A syntax for serialisation of ontologies and linked data (lightweight linked data format)	Adding context to properties within JSON data in order to attach semantic meaning from a linked ontology
JSON Schema	A schema language for describing JSON (or JSON-serialisable) data	Communicating requirements and validating data at boundaries between applications/services (like APIs)/organisations
YAML Schema	A schema language for description of YAML data	Linting and validating text-based configuration files
XML Schema	A schema language	Validating data at boundaries between applications/services (e.g. APIs), typically for legacy systems
Avro schema	A schema language for use with the Avro serialisation utility	Validating event data in high-data-rate event-driven systems (e.g. usually high volumes of small events with low complexity in data structure)
HDF5 schema	A schema language for describing HDF5	Validating data in saved HDF5 data files
Protobuf	A mechanism for serialising typed and structured data	Validating event data in high-data-rate event-driven systems (e.g. usually high volumes of small events with low complexity in data structure)

ent, query-able whole, promoting knowledge discovery and inference across systems that may otherwise remain isolated. In the wind energy sector, OBDI could integrate disparate data sources such as weather forecasts, energy production logs, and maintenance records, promoting a comprehensive, multi-perspective analysis of wind turbine performance, reliability, and optimisation.

Ontology evaluation and alignment are crucial for interoperability and OBDI. Ontology evaluation ensures the suitability and quality of a given knowledge base. Methods for ontology evaluation may differ from one context to another. Vrandečić (2009) proposes assessing the quality of an ontology by evaluating such properties as accuracy, adaptability, clarity, completeness, computational efficiency, conciseness, constituency, and organisational fitness. Ontology alignment identifies semantically equivalent entities from different ontologies, enabling the harmonisation of heterogeneous data sources. In practice, this can be implemented by connecting different concepts using OWL `owl:sameAs` or SKOS `skos:exactMatch` relations. This can significantly benefit the wind energy industry by allowing disparate systems and databases to interact and exchange information seamlessly, promoting a more efficient and effective operational workflow. For example, different organisations perform reliability and failure analysis of using their own taxonomies

of wind turbine parts. Aligning these taxonomies between themselves not only allows a more comprehensive analysis, but also significantly increases the amount of available data, resulting in higher confidence in analysis results.

Ontology reuse is another important aspect of knowledge integration. Reusing existing ontologies can reduce the effort and complexity involved in developing new ontologies from scratch and promote interoperability by using shared semantic artefacts. An important tool for ontology reuse is the ontology-hosting services. The hosting and sharing of ontologies requires the use of platforms and repositories allowing the discovery, search, versioning, and interconnection of the semantic models. While ontology hosting has been initially performed for specific communities and domains, there are several common functionalities (search, identification, alignment, annotation, etc.) that are orthogonal to domain-specific aspects. An example of such application is the OntoPortal Alliance, a consortium constituted of multiple research institutions dedicated to the development and maintenance of the OntoPortal platform (Graybeal et al., 2019), available as open-source code. Based on this common platform, different instances of the portal are made available to specific communities, as for example BioPortal (Noy et al., 2009), AgroPortal (Jonquet et al., 2018), and EcoPortal (Kechagioglou et al., 2021). Compared to other platforms

and initiatives for ontology hosting, the OntoPortal platform provides not only the most comprehensive set of features, but also the widest adoption in different domains (Jonquet et al., 2023). We discuss the possibility of ontology reuse in the wind energy domain in Sect. 6.

5.3 Data storage and management for knowledge-based systems

The topic of data storage and management for KBS is where knowledge engineering overlaps heavily with the data management and data engineering domains. In terms of data management, the relational model based on FOL (Codd, 1990), usually in the form of SQL databases, has been widely adopted since the early 1990s across all industries as a solution for creating and managing structured data. In recent years, non-relational systems such as not only SQL (NoSQL) databases have gained popularity, as they provide for more flexible database expansion, allow for multiple data structures, and offer better performance (in terms of computing) when scaling up to deal with large datasets (Lourenço et al., 2015). In the context of knowledge engineering, triple stores (such as Ontotext⁹) and graph databases (such as Neo4j¹⁰) are well suited to provide a technological foundation for the development of Semantic Web applications (Soussi and Bahaj, 2019). However, before selecting a database for storing and managing ontology-based data, several considerations should be taken into account, as described in this section.

Relational and SQL databases excel at organising data in a structured, tabular format. They are particularly powerful when dealing with large amounts of structured data that need to be queried with complex logic, given their ability to perform reliable and robust transactions (Haerder and Reuter, 1983). SQL databases can be very efficient for look-ups and queries that involve tabular type data. Nevertheless, fitting ontology-based data, which are more graph-like in nature, into the format of a SQL database can pose significant challenges. As discussed before, the fundamental data model for ontology is graph-based, whereas a tabular data structure is typically relational. These two differing structures often do not align seamlessly, leading to issues in data management. The term “impedance mismatch” is used to denote the issues that surface when a system tries to transform one type of data structure into another. Specifically, when data are mapped from a graph-like or an object-oriented model to a relational model, a mismatch arises due to the structural differences between these representations. Over time, a variety of strategies, often referred to as “object-relational mapping” methods, have been developed to address this mismatch. These methods focus on transitioning data from object-oriented models (based on classes and objects) into a format suitable

for storage in relational databases (based on tables and relations).

Regardless of these developments, the fundamental divide in modelling approaches remains. Ontology-based modelling focuses on concepts or objects and describing the relationships between these concepts. While SQL databases have introduced some object-oriented features, they have not yet introduced rich modelling semantics that are seen in ontological approaches. Hence, in cases when data have complex relationships or when the relationships themselves are inherently valuable, NoSQL graph databases offer more efficiency in terms of query speeds.

Object stores and NoSQL databases were developed as solutions to certain limitations that traditional SQL databases had, particularly in two aspects: handling larger amounts of data and dealing with a variety of data types. To understand the first aspect, it is important to understand the notion of scaling. In simple terms, “scaling” refers to the increasing capacity to handle more data or requests. There are two main ways of doing this: scaling up (also known as vertical scaling) and scaling out (also known as horizontal scaling). Scaling up refers to improving the capacity of a single server, such as by adding more memory or a faster processor. However, there are physical limitations to the extent by which a single server can be upgraded. On the other hand, scaling out involves adding more servers to a system and distributing the data and workload among them. This can provide greater increases in capacity and allows for more flexibility and resilience because if one server fails, others can take over its workload. The second aspect, data variety, refers to the shift from storing data in tables – as SQL databases do – to storing data in more flexible formats, such as documents, which NoSQL databases are designed to handle. In recent years, there has been progress in improving the ability of relational databases (like Aurora¹¹) to scale out, which is traditionally a strength of NoSQL databases. However, in some high-scale environments, where a large amount of data needs to be managed, object stores and NoSQL databases are still often required because they offer a purer form of horizontal scaling.

Triple stores and graph databases provide a natural fit for storing and managing ontology-based data like RDF and OWL (Besta et al., 2023). While SQL database tables can encode RDF triples and the expressive power of FOL enables specification of almost any conceptualisation, the semantics of the SQL as a means of performing graph queries is often limited as compared with a dedicated/purpose-built graph database or triple store. Additionally, as mentioned above, NoSQL databases greatly benefit from scaling-out approaches. Triple stores are databases designed specifically for storing RDF triples (an example of an RDF triplet shown before in Sect. 4.2). They typically support SPARQL, a query language for RDF, allowing for efficient querying and ma-

⁹<https://www.ontotext.com> (last access: 10 September 2023)

¹⁰<https://neo4j.com/> (last access: 10 September 2023)

¹¹<https://aws.amazon.com/rds/aurora/> (last access: 10 September 2023)

nipulation of the stored RDF data. Triple stores provide the physical technological support for the practical implementation of Semantic Web applications and services (like the ones discussed in Sect. 4.1), providing efficient storage and retrieval of RDF data. Graph databases, on the other hand, are more general-purpose databases that use graph structures to store data. Each entity (or node) and relationship in the database can have an arbitrary number of attributes, allowing for rich and complex data models. Some graph databases support RDF and SPARQL, making them suitable for Semantic Web applications, while others use proprietary query languages. Compared to triple stores, graph databases may provide more flexibility and performance optimisations for certain types of queries and data models.

Database selection and integration

When selecting a database for storing and managing ontology-based data, it is more important to consider not only the storage of data (whether a SQL database can or cannot store JSON data) but as important or more importantly also the semantics of the data (how the data are typed and queried). Almost any database can store the RDF data or document (there is usually a mapping of some kind). The more important question is whether the query and type language give themselves to this mapping. If the mapping is forced, one may be able to store data, but it may be very difficult to query or to enforce constraints. For example, one may be able to store data as JSON in a SQL column, but can one impose constraints over the structure of the data in the JSON column? And can one more easily query the data using SQL language or using a graph-based query language? The questions one may wish to consider when adopting a database are the following. First, does the type system enable you to model data of your domain and enforce constraints? And second, does the query language fit the shape of your domain and are the queries easy to write and understand once written? Quite often, it is possible to store data using a poorly fitted databases; but cracks emerge in data that are poorly constrained and queries that are hard to read or understand.

6 Knowledge engineering: wind energy domain review

In this section, we review and evaluate existing knowledge-engineering-related efforts and initiatives in the wind energy sector. This work was centred on the following four questions, which are discussed in more detail in the next sections. (Q1) Who are the data users and producers in the wind energy domain? (Q2) Which semantic artefacts relevant to these data users and producers in the wind energy domain already exist? (Q3) What are the gaps and overlaps in existing semantic artefacts and to what extent have existing artefacts gained domain or industry adoption? (Q4) What types of digital twins and decision support systems have been developed so far in

the wind energy domain and how can these systems be improved by applying knowledge engineering methods?

6.1 (Q1) Who are the data users and producers in the wind energy domain?

6.1.1 Scope of the domain

This part began with establishing the scope of the wind energy domain. While we expected that the semantic artefacts we would find would focus on the fields of engineering and atmospheric science, we wanted to be aware of user communities and stakeholders outside these fields, who might use data to inform their decisions. Consequently, we were purposeful towards being inclusive of all the roles and touch points with the domain. An understanding with taking this approach is that there would be semantic artefacts and data models from other domains that could interact with ones specific to wind energy. It was important to us that we consider the multidisciplinary nature of activities undertaken by various stakeholders and their interactions. Recognising these relationships could influence how ontologies are designed in areas where gaps or overlaps exist. Moreover, these aspects shape ontology reuse and alignment activities. In the absence of universally accepted classifications for the roles and activities within the wind energy domain, we bounded our scope to specific stages in the life cycle of wind energy assets. In particular, we have adopted the same stages as the ones used by Barber et al. (2023c) in their analysis of various stakeholder “pain points” related to the digitalisation process in wind energy:

- A. wind turbine design;
- B. wind farm planning;
- C. wind farm operation;
- D. project selling/buying;
- E. end of life;
- F. general.

This classification is sufficiently top level to include activities with a narrower scope such as wind resource assessment or wind turbine maintenance. Roles and activities related to wind turbine design and wind farm planning were selected as the initial bound to the domain of inquiry, as these provided a definite point in time in which data are present in the life cycle. To close our scope, we selected roles related to the end-of-life stage. When looking at other power alternatives such as hydro, nuclear, and fossil, the decommissioning stage creates new data such as the impact on the industry and environment. The same should be expected for wind energy as turbines’ age and build materials and designs are enhanced, climate change impacts the atmospheric conditions at existing sites, and other energy generation technologies come to market.

It is important to note that some of the roles and activities present in the resulting search space are not entirely wind energy specific, such as environmental reviews and power consumption analysis. The search for semantic artefacts related to such roles and activities was not as extensive.

6.1.2 Type of data users

For defining the data users and consumers in our scope of the wind energy domain, we supplemented our own domain expertise with information from Hamilton and Liming (2012) (United States Bureau of Labor Statistics) that described careers in wind energy and sources describing stakeholder analysis in real and theoretical wind energy scenarios (Bremere and Indriksone, 2017; de Vivero, 2023). These scenarios were of applications of onshore and offshore wind energy in Europe. A search in the Scopus database for the query (“wind energy” AND “stakeholder analysis”) provided only seven results, which focused mostly on socioeconomic effects (Vicuña and Pérez, 2020; Huesca-Pérez et al., 2016), decision-making in a regulated industry (Rosenberg, 2019), synergy with agriculture sector (Markovska et al., 2013), and marine biodiversity and aquaculture implications (Wever et al., 2015; Aschenbrenner and Winder, 2019; Weber and Köppl, 2022). There were no results with a focus on wind turbine design, wind farm planning and operation, project selling, or end of life of wind energy assets.

The report by the United States Bureau of Labor Statistics mentioned in the previous paragraph included jobs that can be mostly attributed to the OEMs, wind power project developers, and energy producers (see Table 2). A more inclusive (but not exhaustive) classification provided by a stakeholder analysis for wind energy project assessment and planning phases in a European context (Bremere and Indriksone, 2017) included the following:

1. public authorities;
2. energy producers;
3. investors;
4. experts (consultants);
5. environmental NGOs;
6. professional associations;
7. citizen/societal groups;
8. land owners;
9. wind turbine producers (OEMs);
10. wind power project developers;
11. electricity grid owners;
12. universities (academia).

The majority of these stakeholders continue to interact with wind energy domain data well through the later stages of the wind energy asset’s life cycle, up until the end of life.

6.2 (Q2) Which semantic artefacts relevant to these data users and producers in the wind energy domain already exist?

To build a collection of semantic artefacts to review and analyse, we solicited the IEA Wind Task 43 Working Group 1 participants, a group comprised of industry, academic, and government collaborators with interests and experience in wind energy metadata¹². A search of the SCOPUS database using the query (“taxonomy” OR “schema” OR “ontology” OR “knowledge base”) AND (“wind energy” OR “wind turbine” OR “wind plant” OR “wind power plan”)) provided 202 results from scholarly literature. Chosen results from this query were selected based on having a primary focus of describing the development of a semantic artefact or presented a clear application of a semantic artefact in an applied setting. A search using the same query in the web search engines presented trade literature, technical reports from government agencies, and wind energy domain semantic artefacts of various degrees of expressiveness and generality. The processed results of these searches are presented hereafter. Brief summaries of wind-energy-domain-specific artefacts are presented in Tables 3 and 4. Meanwhile, cross-domain and wind-energy-related semantic artefacts that appeared in search queries are briefly summarised in Table 5.

6.2.1 Review methodology

To perform a methodological overview of the relevant semantic artefacts, we have evaluated them with the following criteria: (1) context and purpose for the semantic artefact development, (2) target audience/role, (3) associated activity, (4) associated life cycle stage of the wind energy assets, (5) semantic artefact type, (6) alignment with other semantic artefacts, and (7) technologies used. Additionally, we have assessed semantic artefacts according to semantic expressiveness, generality, and granularity. Next, we prepared a matrix that mapped roles and activities we identified within our scope of the wind energy domain with the conceptualisations that were found. Identifying the intersections between role and ontology was based on qualitatively analysing the purpose and applicability of the ontology as described by the resource or by analysis of the terms in the ontology and the list of roles. A role that did not have an identified semantic artefact would indicate a potential gap. On the other hand, a role with several distinct ontologies serving a similar purpose would indicate a lack of community adoption and dialogue. These cases required a further investigation into the possible

¹²<https://www.ieawindtask43.org/wg1> (last access: 1 October 2023)

Table 2. Wind energy sector jobs per life cycle phase according to United States Bureau of Labor Statistics.

Life cycle phase	Occupation	Job titles
Manufacturing phase	Engineering	Aerospace engineers, civil engineers, electrical engineers, electronics engineers, environmental engineers, health and safety engineers, industrial engineers, materials engineers, mechanical engineers, engineering technicians, drafters
	General manufacturing	Machinists, computer-controlled machine tool operators, assemblers, welders, quality-control inspector, industrial production managers
Project development	Management and legal	Project managers, asset managers, land acquisition specialists, logisticians
	Scientists	Atmospheric scientists, wildlife biologists, geologists, environmental scientists
	Construction occupation	Construction equipment operators, crane operators, electricians
Operation and maintenance phase	General operation	Plant supervisors, service technicians

causes of this lack in sustainable development. To add context to these inquiries, we have classified all semantic artefacts according to the following criteria: (1) level of adoption, (2) stakeholder type (academia/industry/government/standardisation body), (3) availability of the semantic artefact for download in some kind serialisation or as linked data, and (4) continued development and maintenance.

6.2.2 Limitations in the analysis

Results from our academic literature and internet searches do not include proprietary semantic artefacts. This may constrain the conclusions of the analysis of gaps and overlaps in these conceptualisations because we are unsure of the extent that industry has identified these and developed solutions. We are also unaware of the impact of these gaps and overlaps to their data needs and operations. An exhaustive search for semantic artefacts that included multiple energy sources in addition to wind was not performed. The assumption was that upper-level conceptualisations would not have the specificity of terms or architecture demanded by the roles in our scope of wind energy.

6.2.3 Wind-energy-specific semantic artefacts

This group of semantic artefacts obtained from SCOPUS and web search engine queries are fully wind energy domain specific as defined per the scope outlined above. The IRPWIND¹³ project initiative (Sempreviva et al., 2017) marked an initial endeavour to furnish wind energy datasets with comprehensive metadata and high-level taxonomies within the framework of web publishing for data sharing. The objec-

tive was to extend the Dublin Core metadata model by incorporating seven wind-energy-specific metadata fields. These additions aimed to accommodate additional controlled vocabularies for contextualising datasets, thereby facilitating their search and retrieval. The original plan entailed establishing a metadata catalogue for the dataset-distributed data among participating institutions. By tagging data with the terms from these taxonomies, institutions could enhance data visibility and findability for potential users, who could employ the same terms as filters to locate relevant data (Michiorri et al., 2022). The IEA Wind data models were developed as part of different IEA Wind Tasks. The lidar ontology developed as a part of Task 32 had as a goal to facilitate analysis and exchange of data produced during measurements with various lidar Systems. IRPWIND and Task 32 are the only initiatives that published their semantic artefacts as linked data. The WindIO ontology was developed by Bortolotti et al. (2022) within the Task 37 group for definition of the inputs and outputs for systems engineering multidisciplinary design optimisation (MDAO) of wind turbine and plants. This resulting ontology is formalised as a YAML Schema and is used to describe the structure of YAML input files for Wind-plant Integrated System Design and Engineering Model (WISDEM) software. Recently, WindIO ontology was also suggested by the IEA Wind Task 55 work group as a basis for a more general use ontology describing technical specifications and characteristics of wind turbines and power plants. This ontology will be used to define reference wind turbines and plants for the purposes of V&V, benchmark testing, and impact assessment of novel technologies on wind plants. The WRA data model developed in Task 43 standardises how properties of a wind resource measurement station (e.g. latitude, longitude, anemometer serial number, installation height, logger slope, logger offset) are structured and

¹³<https://www.irpwind.eu/> (last access: 2 February 2024)

Table 3. Description of wind-energy-domain-specific semantic artefacts.

		Brief description
IRPWIND initiative	ASPECT	Controlled vocabulary of variables, parameters, and constants used in the wind energy community.
	EXTRACT	Classification of external conditions in which a wind farm operates.
	IDEM	Classification of models used in wind energy.
	NEAT	Taxonomical organisation of research topics in wind energy which follows a typical life cycle of wind farm development.
	WEAR	Classification of wind turbine materials.
	WEAVE	Classification of activities in which data are produced.
IEA Wind data models	Task 32: wind lidar ontology	Ontological representation of wind lidar supporting development of modular tools and processes for wind lidars.
	Task 37: WindIO	Wind turbine and wind power plant schemas defining input and output variables for systems engineering MDAO.
	Task 43: WRA data model	A JSON Schema to describe and verify wind resource assessment data.
Published as academic article	WPP ontology	Ontology built on the basis of wind power plant information model with OWL by Zhu et al. (2008).
	Offshore wind ontology (OWO)	Offshore wind domain ontology model based on the IEC 61400-25 standard by Nguyen et al. (2014).
	Intelligent fault diagnosis of wind turbines	Wind turbine fault diagnosis application ontology based on failure mode, effects, and criticality analysis (FMECA) and a knowledge base by Zhou et al. (2015).
	WT components taxonomy	A unified taxonomy by Artigao et al. (2018) of wind turbine components (including Reliawind and Reder et al., 2016) for the purpose of reliability analysis.
	Condition monitoring of WT	Wind turbine condition monitoring application ontology and a knowledge base by Papadopoulos and Cipcigan (2009).
	WT operational states	An ontology and a knowledge base of wind turbine operational states by Bunte et al. (2018).
	WPP spatial database	Conceptual schema for a SQL database by Lungu et al. (2012).
	WPP expert system	Expert system for wind power plant's equipment diagnosis by Duer et al. (2017).
	Onshore WT maintenance	Wind turbine maintenance task ontology for onshore wind turbines by Strack et al. (2021).
	OntoWind	Wind energy domain ontology and a knowledge base by Küçük and Küçük (2018).
WESgraph	Top-level ontology and a knowledge base for the wind farm domain, implemented as a graph database by Quaeghebeur et al. (2020).	
Industry developed	Reliawind taxonomy	Taxonomy of wind turbine components for the purpose of reliability and FMECA analysis.
	Power curve schema	A JSON Schema to describe and verify WT power curve data serialised as JSON (Clark, 2023).
	ENTR Alliance OpenOA	Controlled vocabulary for SCADA data and a schema based on IEC61400-25 standard describing renewable energy variables.
	Global Wind Data Tag-List	A comprehensive controlled vocabulary of data tags used in wind energy SCADA systems, based on the IEC 61400-25 standard; maintained by SCADA International and published as a spreadsheet.
	GADS WT reporting	CSV file schema for Generating Availability Data System (GADS) wind turbine generation data reporting used by North American Electric Reliability Corporation.

Table 4. Wind energy domain semantic artefacts and stakeholder use cases.

Name	Life cycle stage	Context/purpose
IRPWIND	General	Data management: FAIR practices for data; publishing datasets on the web; data search and retrieval
IEA Task 32: wind lidar ontology	Wind farm planning	Siting: lidar data processing
IEA Task 37: WindIO	Wind turbine design	Economics: LCOE optimisation; WISDEM inputs definition
IEA Task 43: WRA data model	Wind farm planning	Siting: wind resource assessment (WRA); WRA data sharing, processing, and interoperability
WPP ontology	Wind farm operation	Operations and maintenance: WPP information systems development
Offshore wind ontology (OWO)	Wind farm operation	Operations and maintenance: WPP management; data integration and knowledge sharing
Diagnosis of wind turbines	Wind farm operation	Operations and maintenance: FMECA analysis; DSS
WT component taxonomy, Reliawind taxonomy	Wind farm operation	Operations and maintenance: reliability analysis; failure localisation on component basis
Condition monitoring of wind turbines	Wind farm operation	Operations and maintenance: condition monitoring; DSS
WT operational states	Wind farm operation	Operations and maintenance: WPP operation and control
WPP spatial database	Wind farm planning	Siting: WRA; spatial planning; querying GIS database; DSS
WPP expert system	Wind farm operation	Operations and maintenance: WPP operation and control; expert system development
Onshore WT maintenance	Wind farm operation	Operations and maintenance: condition-oriented maintenance; maintenance reports digitisation and SCADA information integration
OntoWind	General	Knowledge management: wind energy KB development
WESgraph	General	Knowledge management: wind energy KB development
Power curve schema	General	Data management: automating workflows; application development
ENTR Alliance OpenOA	Wind farm operation	Operations and maintenance: WT supervision and control; condition monitoring and operational analysis; SCADA information management
Global Wind Data Tag-List	Wind farm operation	Operations and maintenance: SCADA information management
GADS WT reporting	Wind farm operation	Operations and maintenance: energy generation availability data reporting

Table 5. Description of cross-domain and wind energy overlapping domains semantic artefacts.

	Brief description	
Upper level and mid-level	BFO	Basic Formal Ontology (BFO) is a top-level ontology that provides a foundational framework for organizing and structuring domain-specific ontologies. It aims to promote interoperability and integration across different domain ontologies by providing a common set of basic categories and relationships.
	CCO	Common Core Ontologies (CCO) is a collection of 12 ontologies that comprise mid-level extension of BFO. The CCO provide semantics for concepts and relations that are used in most domains of interest, such as concepts from units-of-measure ontology, event ontology, and time ontology. Many domain ontologies are aligned with CCO. Among them are aircraft ontology, occupation ontology, and sensor ontology.
	Dublin Core	Dublin Core is a simple, flexible, and extensible metadata standard for describing a wide range of resources, including digital and physical assets. It consists of a set of 15 core elements (e.g. title, creator, subject) that can be used to describe resources in a consistent and structured manner, facilitating resource discovery and interoperability.
	Schema.org	A shared vocabulary for structured data markup on web pages to improve search engine results and discoverability.
W3C	SKOS	A W3C data model recommendation for expressing controlled vocabularies, taxonomies, and thesauri.
	PROV-O	The PROV Ontology (PROV-O) is an ontology that provides a vocabulary for expressing provenance information.
	SSN-XG	Semantic Sensor Network (SSN-XG) ontology describes sensors, actuators and observations, and related concepts. Domain concepts, such as time and locations, are intended to be included from other ontologies via OWL imports.
W3C example ontologies	AWS	Agriculture meteorology example, Ontology for Meteorological Sensors showcasing the ontology developed by the W3C Semantic Sensor Networks incubator group (SSN-XG).
	CF	Climate and Forecast (CF) metadata conventions vocabulary encoded with OWL.
	WEATHER	Linked Sensor Data – Weather station is an example ontology for weather data publication on the LOD using SOSA/SSN ontology.
	WM30	An example of the SSN-XG sensor ontology used to describe a specific sensing device, the Vaisala WM30, which measures wind speed and wind direction.
Observations and measurements	SciData	SciData is a data model for scientific data that provides an ontologically defined framework for organizing and linking (with JSON-LD) both the data and metadata from scientific experiments, calculations, and theories.
	OBOE	The Extensible Observation Ontology (OBOE) is a formal ontology for capturing the semantics of scientific observation and measurement. The ontology supports researchers in adding detailed semantic annotations to scientific data, thereby clarifying the inherent meaning of scientific observations.
	I-ADOPT	An ontology designed to facilitate interoperability between existing variable description models (including ontologies, taxonomy, and structured controlled vocabularies).
	OM	Ontology of units of Measure (OM) is an ontology focused on units, quantities, measurements, and dimensions relevant to scientific research.
Domain level	SWEET	Semantic Web for Earth and Environmental Terminology (SWEET) is a foundational ontology that contains over 6000 concepts organised in 200 ontologies expressed in OWL. SWEET is a highly modular, general-purpose ontology suite designed to represent Earth and environmental science concepts and their relationships.
	CF metadata – conventions	Climate and Forecast (CF) standardised set of metadata elements for describing climate and forecast data stored in netCDF files. The conventions aim to facilitate data sharing, discovery, and interoperability in the climate and forecasting communities.
	SLACKS	Semantic LAMinated Composites Knowledge management System (SLACKS) (Premkumar et al., 2014) based on suite of ontologies for laminated composite materials and design for manufacturing (DFM).
Energy domain	OEO	Open Energy Ontology (OEO) (Booshehri et al., 2021) is an energy system modelling domain ontology.
	EKG	Energy Knowledge Graph (EKG) (Chun et al., 2018) is an upper-level ontology for the integration of knowledge resources in energy systems.
	EDF PPO	Electricity of France (EDF) power plant ontology and knowledge base (Dourgnon-Hanoune et al., 2010).
	GCIEO	Global City Indicator Energy Ontologies (GCIEO) (Komisar and Fox, 2020) is a standard ontology for Semantic Web based representations of general knowledge for the Energy Theme indicators (ISO 37120).

serialised as a JSON file. This data model is described with JSON Schema. The majority of semantic artefacts presented in academic literature are not available for download. Among these are various taxonomies of wind turbine components for reliability analysis such as WT component taxonomy (Artigao et al., 2018). These taxonomies are often not formalised in any modelling language. In a similar context of reliability and failure analysis, more expressive ontologies were proposed by various authors – in particular, the ontology for the failure mode, effects, and criticality analysis (FMECA) proposed by Zhou et al. (2015) and the one for condition monitoring proposed by Papadopoulos and Cipcigan (2009). These ontologies relied on OWL for knowledge representation. More general, domain-level ontologies and knowledge bases such as WPP ontology (Zhu et al., 2008) and OWO (Nguyen et al., 2014) attempted to comprehensively capture wind-energy-related concepts. These authors of these knowledge bases also opted for a Semantic Web technology stack, with a notable exception of WESgraph, which relied on the Neo4j graph database for data storage and querying. As a consequence, the underlying top-level ontology for WESgraph is not formalised with any of the commonly used ontology languages.

Lastly, some attempts were undertaken by the industry at creating controlled vocabularies of terms used for wind turbine system SCADA and reporting data. ENTR Alliance and SCADA International created controlled vocabularies of SCADA terms in accordance with the guidelines presented in IEC 61400-25. In North America, wind turbine generation data reporting for Generating Availability Data System (GADS) follows the schema enforced by North American Electric Reliability Corporation (NERC). The conceptualisations for these three semantic artefacts are specified as lists of terms, stored along with term descriptions in a tabular form which is serialised using CSV or XLS formats.

6.2.4 Cross-domain and wind-energy-activity-related domains

This group of semantic artefacts, which is not entirely contained within the scope of domain of interest, appeared among the results of the SCOPUS and web search engine queries due to their cross-domain nature and applications in wind energy (and as result would match to keywords like “wind turbine” or “wind energy”). These semantic artefacts can be attributed to the domains that overlap with wind energy such as environment and meteorology, sensing, structural health monitoring, material sciences, and energy. As mentioned before, no targeted search and review was performed for each of the overlapping domains; hence, the presented list is not exhaustive. For example, Semantic Laminated Composites Knowledge management System (SLACKS) was developed specifically for the wind turbine blade design use case. However, besides this specific case, a multitude of material ontologies and knowledge bases exist

(De Baas et al., 2023). Table 5 presents a summary of the semantic artefacts reviewed for this work. Most these have been adopted by various communities and are instrumental for inter-disciplinary collaborations. The table does not include semantic artefacts that have not seen the widespread adoption, either due to their “in-development” status or when superseded by more recent efforts. For example, a structural health monitoring (SHM) ontology was recently proposed by Tsialiamanis et al. (2021) to facilitate knowledge sharing, application, and reusability for SHM projects. However, it has not been yet validated and published. At the same time some renewable energy domain ontologies such as OpenWatt (Lamanna and Maccioni, 2014) are no longer supported as the knowledge has been subsumed by knowledge bases such as the Open Energy Ontology (OEO) (Booshehri et al., 2021).

For the sake of completeness, several upper- and mid-level ontologies have been reviewed, such as Basic Formal Ontology (BFO) and Common Core Ontologies (CCO), as many domain-specific semantic artefacts developed by communities outside of wind energy tend to align with some upper-level ontology. It should be noted that there are many upper-level ontologies that have been developed by various authors with different focus. For example, Dublin Core was developed in the context of metadata standards and description of web resources, such as publications, datasets, and images. Another example is Schema.org: a commonly adopted ontology for describing resources on the web, initially developed for the e-commerce scope by a consortium of Google, Microsoft, Yahoo, and Yandex. A comprehensive overview and evaluation of upper-level ontologies was performed by Partridge et al. (2020). Ontologies and data model recommendations developed by W3C such as Simple Knowledge Organisation System (SKOS), PROV-O, and Semantic Sensor Network (SSN-XG) have gained widespread adoption due to the pioneering work of W3C on the web of data and its role in Semantic Web technology stack development. SKOS provides a system for creation of taxonomies, controlled vocabularies, and thesauri creation within this technology stack, following the principles of linked data. In a wind energy context, the IRPWIND initiative used SKOS in the creation of their taxonomies. The PROV Ontology (PROV-O) is an ontology that provides a vocabulary for expressing provenance information, which can be important in a wind energy context for data governance purposes given the multiplicity of stakeholders and complexity of the systems producing the data. Several examples of Semantic Sensor Network (SSN-XG) ontology have been developed specifically for meteorological sensors, which clearly overlaps with types of data generated in the context of wind energy activities. More generally, data generated during observation and measurement activities can be formalised with data models like SciData and described with ontologies like Extensible Observation Ontology (OBOE), Ontology of units of Measure (OM), and I-ADOPT. A comprehensive overview and eval-

uation of various ontologies for units of measurements was recently performed by Keil and Schindler (2019). In addition, semantic artefacts such as Semantic Web for Earth and Environmental Terminology (SWEET) and Climate and Forecast (CF) metadata conventions from weather and environmental domains share a significant terminological overlap for describing observations and measurements related to siting activities. Several energy domain ontologies and knowledge bases like Open Energy Ontology (OEO), Energy Knowledge Graph (EKG), Global City Indicator Energy Ontologies (GCIEO), and Electricity of France (EDF) power plant ontology include some wind energy concepts. As a result, these semantic artefacts appeared among the searches performed for this review.

6.3 (Q3) What are the gaps and overlaps in existing semantic artefacts and to what extent have existing artefacts gained domain or industry adoption?

The results of the search queries and their subsequent analysis has revealed the following.

- The majority of existing semantic artefacts pertain to the wind farm operation life cycle stage and especially to the activities related to failure and reliability analysis of wind turbines.
- There appears to be no semantic artefacts developed specifically within the context of project selling/buying or end-of-life stages.
- Existing semantic artefacts have not gained high adoption by domain experts and there is no common domain-level ontology that is accepted by the community.
- There is no alignment to upper-level ontologies or between semantic artefacts within the wind energy domain. Similarly, there is no alignment or reuse of semantic artefacts from domains overlapping with wind energy in their data generation and producing activities.
- There is a significant corpus of taxonomies and vocabularies that has not been formalised with any modelling language. Many activities still rely on manual data processing.

These conclusions can be illustrated with the example of taxonomies developed for the purpose of failure analysis. Artigao et al. (2018) manually unified and aligned 13 different wind turbine component taxonomies (none of which was made available using some kind of standard formalisation). Following this trend, one of the more recent wind turbine failure analyses performed by Sanchez-Fernandez et al. (2023) once again manually mapped failure and maintenance records to a new WT taxonomy based on the Reference Designation System for Power Plants (RDS-PP). This lack of alignment and reuse is also highlighted by Leahy et al.

(2019), suggesting that the absence of unified standards for turbine taxonomies, alarm codes, SCADA operational data, and maintenance and fault reporting significantly hinders the wind turbine condition monitoring and reliability analyses. Such a situation can be partially attributed to the fact that the existing conceptual models in wind energy and related domains are not maintained and are not published following the LD principles, as can be observed from the Sankey diagram in Fig. 4.

Out of 19 reviewed existing wind energy domain semantic artefacts, six were downloadable in some kind of serialisation, and three were not even available. Such a situation results in low adoption and a lack of further development in a negative feedback cycle. This issue is not unique solely to the wind energy domain but is also an issue for many technological sciences. Meanwhile, this is generally not the case for cross-domain and top-level semantic artefacts. Such artefacts are widely used by the few communities spearheading open-science principles adoption such as the biomedical sciences community. Hence, there is a strong need for a holistic approach: a framework for community development and maintenance of semantic artefacts, in addition to a platform for semantic artefact hosting and usage, which is discussed more in Sect. 7.

6.4 (Q4) What types of digital twins and decision support systems have been developed so far in the wind energy domain and how can these systems be improved by more widespread adoption of common semantic artefacts?

For the purpose of this section, a systematic review of the publications regarding decision support systems and digital twins in the wind energy domain has been performed. A search of the Scopus database using the query (“decision support system” OR “expert system” OR “digital twin”) AND (“wind energy” OR “wind turbine” OR “wind plant” OR “wind power plant”) yielded 532 results on 1 August 2023. After removing false positives, i.e. papers that did not actually relate to any of the queried topics, and selecting the results relevant to the question posed, the remaining 181 results have been compiled in Fig. 5 and classified based on the modelled component or assembly, as well as the functional level of the DT system (supervisory, operational, simulation prediction, intelligent learning, autonomous management) (Wagg et al., 2020). The levels differ in the integration of datasets, starting from signal conditioning, including metadata, up to using ontologies. Further distinctions are made by, for example, the level of integration of numerical models.

Out of the 181 results, 111 of them cover topics related to digital twin implementations, and the remaining are related to decision support systems. Most digital twin implementations were found to belong to the functional levels “supervisory” (26 out of 111), “operational” (22 out of 111), or

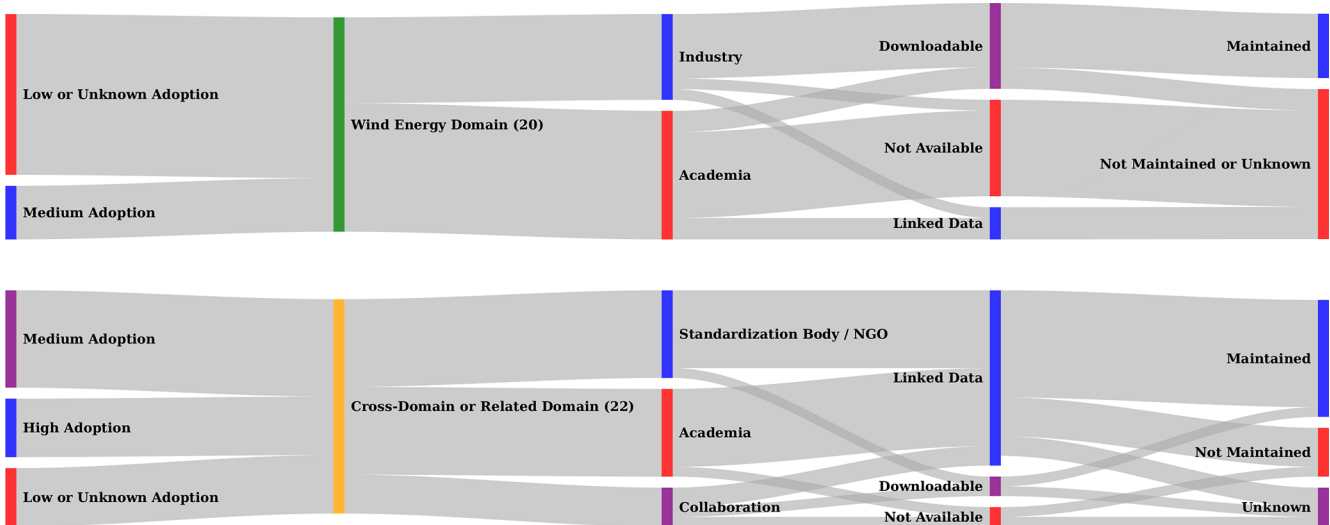


Figure 4. Analysis of semantic artefact adoption levels. Low adoption levels in the wind energy domain can be attributed to low availability and a lack of active development.

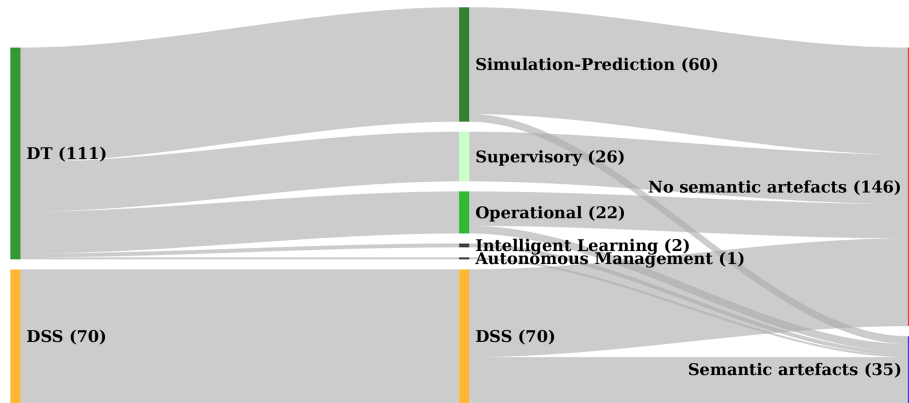


Figure 5. Semantic artefact adoption in digital twins and decision support systems based on literature review.

“simulation prediction” (60 out of 111). Only three papers belong to the functional levels “intelligent learning” (Chatterjee and Dethlefs, 2020; Li et al., 2021) and “autonomous management” (Chavero-Navarrete et al., 2019).

Generally, it can be seen that there is a lack of adoption of semantic artefacts in the research of digital twin and decision support systems, reflected by the low number of papers that use them (35 out of 181). The research of digital twins is focused on the details of the analytic methodologies that are implemented in the digital twin rather than focusing on utilising semantic artefacts. Therefore it can be concluded that there is a low level of adoption of semantic artefacts and of digital twins with intelligent or autonomous features in the wind energy sector.

7 Recommendations

As digitalisation within the wind energy sector matures, we anticipate the current corpus of ontologies, schemas, and data models to develop and grow. The wind energy knowledge engineering ecosystem, including semantic artefacts, tools and applications, and actors, will evolve to enable and support comprehensive data management and analysis throughout the wind energy sector. There are, however, several essential requisites for this ecosystem to be healthy and thriving, which could be identified in this work and which are described below. They are divided into three categories: (1) organisation and diversity, (2) productivity, and (3) resilience.

7.1 Organisation and diversity

The digitalisation process should cater to the distinct needs of the manifold stakeholders in the wind energy landscape.

- *Stakeholder analysis.* A thorough taxonomy of wind energy stakeholders and their activities will aid in discerning use cases. This builds on the groundwork laid by Barber et al. (2023c) but broadens the scope to incorporate entities like public groups, NGOs, and governments. Currently, based on our review, the digitalisation process seems to overlook interests and use cases of the most of stakeholder groups with the exception of wind turbine producers (OEMs) and energy producers.
- *Inclusive stakeholder engagement.* Creation of public forums and collaborations are vital for the community growth and development. Examples of such activities are the WeDoWind framework, which incentivises data sharing via challenges set by data providers (Barber et al., 2022, 2023a, b), and IEA Wind Task 43, which aims to accelerate digital transformation in the wind sector by acting as a catalyst of open collaboration¹⁴.
- *Comprehensive digitalisation of the entire wind energy sector.* It is pivotal to consider every stage of the wind energy project life cycle. Currently, emphasis largely revolves around the operation and maintenance phase. Task and application ontologies catering to other stages, including end-of-life and financial activities, must be developed.
- *Balancing expressiveness with simplicity.* Depending on the use case, the semantic artefacts might require different expressiveness or different paradigm (e.g. OWA vs. CWA) adoption. Not all activities require a definition of a fully developed ontology. In many cases, a controlled vocabulary of terms, a taxonomy, or a schema would be more adequate. This should be assessed by the community during the initial stages of the semantic artefact development.

7.2 Productivity

Productivity encapsulates the community's prowess in generating and innovating new semantic artefacts and tools that can respond to ever-evolving sector needs. The ability to continually produce these new resources ensures that the sector not only remains at the forefront of technology, but also proactively addresses emergent challenges, adding significant value to all stakeholders. In addition to creation of new semantic artefacts and AI tools, a productive ecosystem should streamline existing workflows and maintain effectiveness and efficiency of data and knowledge management processes.

- *Generation of new semantic artefacts and standards.* Development of new semantic artefacts of various degrees of expressiveness based on stakeholder use cases

(or pain points) is paramount for successful digitalisation. This is aligned with IEA Wind Task 43 digitalisation activities, within which the creation of ontology and schema development groups is envisioned. For reference, BioPortal hosts 1065 semantic artefacts, which is almost 2 orders of magnitude more than the number of wind energy domain semantic artefacts reviewed in this work. Some of the ontologies and schemas already proposed for development within IEA Wind Task 43 are reported in Table 6.

- *Cross-pollination.* The increased productivity can be achieved by utilising expertise from diverse groups within the wind energy domain, as well as other industries that are successfully undergoing digitalisation, such as biomedical sciences and e-commerce through ontology reuse and alignment. As highlighted by this review, currently there are no efforts that seek to reuse and align semantic artefact within the wind energy domain. The wind energy community should consider top-level, cross-domain, and related domain ontologies that are already well established and accepted within respective communities for ontology reuse. The infra-domain alignment of wind energy semantic artefacts can significantly improve the efficiency and quality of data analysis. For instance, the alignment of various taxonomies of wind turbine parts can be performed with relatively minimal effort, while offering immediate payback. The alignment of wind energy domain semantic artefacts with relevant ontologies outside of the wind energy domain also may offer significant benefits. For example, alignment with GCIEO can be useful for use cases relative to public and government types of stakeholders.
- *Information access and transparent decision-making.* Among possible solutions for increased information access and more transparent decision-making is the development of web resources and applications that interact with wind energy domain knowledge bases and provide various stakeholders with information of interest. Here the topics such as ontology-based data integration and data management play an important role.
- *Automation.* Development of new tools and practical technological solutions for AI systems is essential for more comprehensive and widespread automation in the wind energy domain. This, as well, is in line with the IEA Wind Task 43 roadmap. Additionally, workflow development with subsequent formalisation and grounding in suitable logical frameworks is the requisite for reducing manual and ad hoc solutions currently prevalent in the wind energy sector practices.

¹⁴<https://www.ieawindtask43.org/> (last access: 1 October 2023)

Table 6. Semantic artefacts proposed for development within IEA Wind Task 43.

Name	Expressiveness	Description and purpose
Vocabulary of SCADA terms	Vocabulary	Controlled vocabulary of terms created with Semantic Web technologies and published online. Assigns URIs to the SCADA terms in accordance with 61400-25. These URIs can be used to describe SCADA data with RDF.
Wind turbine system components	Taxonomy	A classification of wind turbine components using the SKOS data model. This classification can be adopted for various tasks and applications (e.g. reliability analysis) requiring minimal ontological commitment.
Wind energy project life cycle stages	Taxonomy	A classification of a wind energy project life cycle stages. This classification can be used in any context that requires an organisation of data based on life cycle stage criteria.
Wind energy stakeholders	Taxonomy	A classification of wind energy domain actors and roles. Forms a basis for a comprehensive analysis of stakeholder needs. Additionally can be used for Dublin Core Audience ¹ types.
Wind energy activities	Taxonomy	Improved and more comprehensive taxonomy of wind energy activities based on WEAVE. Can be used for use-case-driven analysis of stakeholder needs and in the definition of the Dublin Core Subject ² property.
Blade damage	Taxonomy	A classification and controlled vocabulary to describe types of wind turbine blade damage. Possible use cases include but are not limited to uniform monitoring and maintenance reporting, data tagging (e.g. photographic data, SHM data), and context-aware data analysis.
Airfoil data model	Schema	A schema specifying wind turbine airfoil characteristics. The primary purpose is serialisation, validation, and exchange of data. This can be used in the development of software applications that model airfoil aerodynamics and workflows that involve aerodynamic data.
Wind turbine system specification	Schema	A schema specifying wind turbine characteristics used for serialisation, validation, exchange of data, and software application development.
Wind turbine system sensors	Schema	Description of sensors that are installed on wind turbines for monitoring purposes. This metadata can be included alongside the data generated by the sensor in machine-readable format.
Power curve data	Schema	Specification of wind turbine power curve data with JSON Schema. The primary use of the schema is during serialisation, validation, and exchange of data.
Wind turbine coordinate systems	Schema/ontology	Formal description and URIs for coordinate systems defined for a generic wind turbine in accordance with IEC/TS 61400-13. These coordinate systems can be used to define the location of the wind turbine components, damages, or installed sensors.
Wind turbine system ontology (WTSO)	Ontology	Ontological representation of a wind turbine. Creation of knowledge bases containing information about wind turbine systems. Lower-level conceptualisations can be aligned with this ontology.

¹ <http://purl.org/dc/terms/audience> (last access: 25 March 2024).

² <http://purl.org/dc/elements/1.1/subject> (last access: 25 March 2024).

7.3 Resilience

Resilience ensures that the ecosystem will adapt and evolve in the face of challenges, ensuring its longevity and relevance. For the wind energy sector, it involves creating robust and flexible knowledge infrastructures that can accommodate technological advancements, shifting stakeholder needs, and external disruptions.

- *FAIR principles.* Adherence to FAIR principles can be facilitated by use of Semantic Web technology stacks and linked data. An example of such an effort, which would directly benefit the wind energy community, is the creation of an ontology-hosting catalogue for the technology sciences.

- *Maintenance.* Adoption of Free Open Source Software (FOSS) community practices and technologies (such as Git) can aid significantly in ensuring long-term support and the sustainable development of knowledge engineering applications for the wind energy sector. This is vital, as many of the existing semantic artefacts reviewed are still under development, while others would benefit from further development and improvement after methodological evaluation and assessment.

Aspects related to culture and competition are also important to consider for a healthy and thriving wind energy knowledge engineering ecosystem. These aspects relate to some of the key challenges in the digitalisation of wind energy recently introduced in a review paper by Clifton et al.

(2023). The topic of culture involves, for example, developing and maintaining collaborative organisational cultures, combining staff skills in new ways, enhancing communication skills, developing change processes, and increasing diversity. The topic of competition involves enabling cooperation, collaboration, and competition between organisations. This means working together to create marketplaces or business opportunities that would not otherwise exist and that are mutually beneficial. A discourse on these elements, however, transcends the ambit of this knowledge engineering review.

8 Conclusions

The wind energy sector is amidst the global transformative phase of increased automation and rapid digitalisation. While the digital transformation is paving the way for advancements such as AI-powered digital twins and decision support systems, in the wind energy domain, challenges remain, particularly in the realm of converting raw data into meaningful domain knowledge that is both humanly and machine understandable. A significant part of this challenge is the lack of widespread expertise and tools in data management and knowledge engineering, leading to underutilised, undervalued, and fragmented data often void of context. The current work has attempted to bridge this knowledge gap by shedding light on the relevance and utility of knowledge engineering for the wind energy domain. It has presented a coherent synthesis of existing works in knowledge engineering and representation, tailored for wind energy experts. Through a systematic review, this study also underscores the pressing need for an inclusive approach that caters to a wide range of stakeholders, for creation of new semantic artefacts and data management tools, and for a robust infrastructure with a focus on sustainable development to ensure resilience. However, true progression can only be realised when collaborative efforts within the wind energy community are intensified. This involves not just internal coordination but also leveraging insights from other sectors that have already navigated their digital transformation and have effectively utilised knowledge engineering methods and technologies. Existing efforts such as IEA Wind Task 43 are commendable initiatives in this direction, providing a foundational starting point. The semantic artefacts proposed for development in this review, once published, will be found on the IEA Wind Task 43 GitHub page¹⁵ and dedicated ontology-hosting portal. Embracing these initiatives and fostering collaboration will undoubtedly steer the wind energy sector towards a future that maximises the potential provided by digital transformation.

¹⁵<https://github.com/IEA-Task-43> (last access: 25 March 2024)

Appendix A

A1 Modelling languages: examples

Here we provide several examples of statements made in modelling languages that are commonly used in a knowledge engineering context.

A1.1 RDF Schema (RDFS)

RDF Schema provides a basic type system for RDF. It introduces the concept of classes and properties, allowing for the definition of vocabularies and a limited form of structure to be added to RDF data. For instance, using RDFS, it is possible to define hierarchies of classes and properties, specify the domain and range of properties, and declare subclasses or subproperties. For example, the statement

```
ex:Aventa_AV-7 rdf:type ex:WindTurbine
```

made in RDF can be semantically enriched by stating that wind turbine is a class.

```
@prefix ex: <http://example.com/resource/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

```
ex:WindTurbine rdf:type rdfs:Class .
ex:Aventa_AV-7 rdf:type ex:WindTurbine .
```

A1.2 Web Ontology Language (OWL)

Web ontology language is a formal language based on the description logic representation formalism. Developed by the W3C, OWL is built on top of RDF and extends its expressiveness by providing additional vocabulary for defining complex relationships, classes, properties, and restrictions. OWL enables a higher level of semantic expressiveness compared to RDF and RDF Schema (RDFS), allowing for more sophisticated reasoning and inferencing capabilities. For example, we can express rated power as a datatype property and constrain it to be a float.

```
@prefix ex: <http://example.com/resource/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
ex:WindTurbine rdf:type owl:Class .
ex:Aventa_AV-7 rdf:type ex:WindTurbine .
ex:ratedPower rdf:type owl:DatatypeProperty ;
  rdfs:domain ex:WindTurbine ;
  rdfs:range xsd:float .
ex:Aventa_AV-7 ex:ratedPower "6.2"^^xsd:float .
```

There are several sublanguages of OWL with varying levels of expressiveness and computational complexity, including OWL Lite, OWL DL, and OWL Full. OWL DL, which is based on description logic, offers a balance between expressiveness and computational tractability, making it suitable for many applications.

A1.3 Shapes Constraint Language (SHACL)

Shapes Constraint Language (SHACL) is a World Wide Web Consortium (W3C) specification for validating and describing RDF graphs. SHACL allows for the definition of constraints that can be used to validate RDF data against a set of conditions. SHACL's validation capability makes it particularly suited for ensuring that data adhere to a particular shape or structure, hence the name. In addition to validation, SHACL can be used for data modelling and to guide the process of data transformation and integration. For example, the Aventa AV-7 wind turbine described before can be validated by a SHACL shape as in the following.

```
@prefix ex: <http://example.com/resource/> .
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

ex:WindTurbineShape a sh:NodeShape ;
  sh:targetClass ex:WindTurbine ;
  sh:property [
    sh:path ex:ratedPower ;
    sh:datatype xsd:float ;
  ] .
```

This SHACL shape defines that every instance of the class “WindTurbine” must have a property “ratedPower” which has to be of the datatype float. This way, SHACL helps to ensure data integrity and consistency by providing a mechanism for enforcing data constraints.

A1.4 JSON Schema

JSON Schema represents a vocabulary permitting annotation and validation of JSON data. Unlike RDF, JSON stores information as attribute–value pairs. This type of data structure, when nested, can be visualised as a tree rather than a labelled graph. JSON Schema, hence, defines the structure of JSON data and validates JSON data against defined schemas. It supports various constraints, such as data types, enumerations, pattern matching, optional/required properties, and array item uniqueness. For example, the information about Aventa AV-7 wind turbine can be stored in JSON format as follows.

```
{"WindTurbine":{"model":"Aventa AV-7","ratedPower":6.2}}
```

The related schema would look like the following.

```
{
  "schema": "http://json-schema.org/draft-07/schema#",
  "title": "Wind Turbine",
  "description": "Schema for basic wind turbine attributes",
  "type": "object",
  "properties": {
    "WindTurbine": {
      "type": "object",
      "properties": {
        "model": {
          "type": "string",
          "description": "The model of the wind turbine"
        },
        "ratedPower": {
          "type": "number",
          "description": "The rated power in kilowatts"
        }
      }
    }
  },
  "required": ["model", "ratedPower"]
}
```

```
},
"required": ["WindTurbine"]
}
```

Although primarily designed for JSON data validation, the use of JSON Schema for more intricate data modelling tasks has been increasing, indicating its evolution towards a comprehensive ontology language. It is important not to confuse JSON Schema with JSON-LD. Unlike JSON Schema, the later is an extension of the JSON data format for the serialisation of RDF-based ontology languages.

A1.5 YAML Schema

YAML Schema is a tool dedicated to defining the structure of YAML documents. YAML, a human-friendly data serialisation standard, is extensively used in configuration files and applications where data storage or transmission is involved. YAML Schema bears several similarities to JSON Schema, but it is designed specifically for the YAML data format. This schema validates YAML documents, ensuring compliance with a predefined structure and specific criteria. YAML and YAML Schema are commonly used when data and schema readability is paramount. Below is a YAML Schema serialisation of the same information as the one described before by JSON Schema.

```
schema: http://json-schema.org/draft-07/schema#
title: Wind Turbine
description: Schema for basic wind turbine attributes
type: object
properties:
  WindTurbine:
    type: object
    properties:
      model:
        type: string
        description: The model of the wind turbine
      ratedPower:
        type: number
        description: The rated power in kilowatts
    required:
      - model
      - ratedPower
  required:
    - WindTurbine
```

A1.6 XML Schema

XML Schema, also known as XSD (XML Schema Definition), is a World Wide Web Consortium (W3C) recommendation that prescribes formal descriptions of elements in an Extensible Markup Language (XML) document. It serves to describe and validate the structure and content of XML data. XML Schema supports namespaces, complex data types, inheritance (through extension and restriction), and constraints on values and relationships between elements. Its wide use spans several industries, including publishing, telecommunications, and e-commerce. However, due to complexity of XML syntax, this schema language is much less intuitive and is not as easily humanly readable as compared to JSON or YAML Schema.

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```

<xs:element name="WindTurbine">
  <xs:annotation>
    <xs:documentation>
      A representation of a Wind Turbine.
    </xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:attribute name="model" type="xs:string">
      <xs:annotation>
        <xs:documentation>
          The model of the Wind Turbine.
        </xs:documentation>
      </xs:annotation>
    </xs:attribute>
    <xs:attribute name="ratedPower" type="xs:float">
      <xs:annotation>
        <xs:documentation>
          The rated power in kilowatts.
        </xs:documentation>
      </xs:annotation>
    </xs:attribute>
  </xs:complexType>
</xs:element>
</xs:schema>

```

A2 Serialisation formats

OWL and RDFS ontologies can be serialised for storage using a variety of formats. These formats enable the representation of knowledge in a machine-readable and standardised way. RDF, as the foundation of the Semantic Web, can be serialised in various formats such as RDF/XML, Turtle (Terse RDF Triple Language), N-Triples, and JSON-LD.

RDF/XML was the first standardised serialisation format for RDF, but its verbosity and complexity led to the development of other formats such as Turtle and N-Triples, which offer more human-readable syntax. JSON-LD (JSON for Linked Data) has gained popularity as it combines the simplicity and widespread use of JSON with the ability to express RDF data. OWL, being an extension of RDF, can also be serialised using the aforementioned RDF serialisation formats. However, OWL has its own serialisation formats as well, such as OWL/XML and functional-style syntax (also known as OWL Functional Syntax). OWL/XML is an XML-based syntax specifically designed for expressing OWL ontologies, while functional-style syntax is a human-readable, text-based format that closely follows the structure of the OWL 2 specification. The choice of serialisation format depends on factors such as readability, compatibility with existing tools, and ease of parsing and processing.

A2.1 JavaScript Object Notation for Linked Data (JSON-LD)

JSON-LD is a lightweight data interchange format that extends JSON to provide a means for encoding linked data using standard JSON conventions. JSON-LD is designed to be easy to read and write by humans, as well as simple to parse and generate by machines. It is developed by the World Wide Web Consortium (W3C) and provides a way to represent the RDF data model in JSON. JSON-LD is particularly useful for web developers who want to incorporate structured data into web applications and APIs while leveraging the existing JSON tools and libraries. JSON-LD introduces a notion of

a context, which allows defining short aliases for long IRIs (Internationalized Resource Identifiers) used in RDF, simplifying the representation of RDF triples in JSON. It also supports the definition of data types, language tags for string values, and nested JSON objects to represent complex relationships and structures.

The following is an JSON-LD serialisation example.

```

{
  "@context": {
    "schema": "https://schema.org/",
    "geo": "http://www.w3.org/2003/01/geo/wgs84_pos#"
  },
  "@type": "schema:EducationalOrganization",
  "@id": "https://www.ost.ch",
  "schema:legalName": "OST - Ostschweizer Fachhochschule",
  "schema:owns": {
    "@type": "schema:Product",
    "@id": "urn:ost:aventa:av-7",
    "schema:name": "Aventa AV-7 Wind Turbine",
    "schema:manufacturer": {
      "@type": "schema:Organization",
      "@id": "https://ex.com/manufacturers/aventa",
      "schema:name": "Aventa"
    }
  },
  "schema:model": {
    "@type": "schema:ProductModel",
    "@id": "urn:ost:aventa:av-7",
    "name": "AV-7"
  },
  "geo:location": {
    "@type": "geo:Point",
    "geo:lat": "47.52000",
    "geo:long": "8.68236"
  }
}

```

Appendix B

Table B1 presents Semantic Web technology enablers for FAIR data practices. The definition of specific FAIR principles can be found on the GO FAIR web page¹⁶.

Table B1. FAIR principles and Semantic Web technology enablement.

FAIR principle	Semantic Web technology enablement
Findable	
F1	URIs, especially HTTP URIs, are globally unique and can be persistent. This requirement is similar to the first LD principle, even though persistence is not specifically addressed by LD.
F2	Rich metadata models are developed and adopted by communities for their specific domain. These models can be published on the web as LD and referred to when describing the data.
F3	RDF triples connect unique identifiers (URIs) with well-defined properties to their respective values.
F4	RDF data can be stored in triplestores which are optimised for semantic queries using SPARQL. This requirement along with F2 and F3 is similar to the third LD principle.
Accessible	
A1.1	Semantic Web technologies are built on open standards such as HTTP and RDF. This requirement is similar to the second linked data principle.
A1.2	Web protocols like HTTPS include provisions for authentication and authorisation.
A2	Persistent URIs and versioning methodologies ensure that metadata remains accessible. Metadata models can be published on third-party websites like schema.org; more expressive semantic artefacts can be published on ontology-hosting websites like the ones powered by OntoPortal. This requirement has no equivalent in LD principles.
Interoperable	
I1	RDF, RDFS, and OWL are formal languages that are broadly applicable for knowledge representation.
I2	Semantic Web enables the use of vocabularies and taxonomies through standard data models such as SKOS.
I3	RDF's linking capability enables metadata to include qualified references to other metadata. This requirement is similar to the fourth LD principle.
Reusable	
R1.1	Metadata in RDF can include licensing information using appropriate vocabularies. In general, LD has evolved into the direction of open data, meanwhile FAIR principles can be applied to data subject to any explicitly defined license.
R1.2	Provenance information can be recorded using cross-domain, community-accepted vocabularies such as PROV-O.
R1.3	The Semantic Web supports the use of domain-specific ontologies, which can be developed according to community standards. This requirement is similar to the third LD principle.

¹⁶<https://www.go-fair.org/fair-principles/> (last access: 10 September 2023)

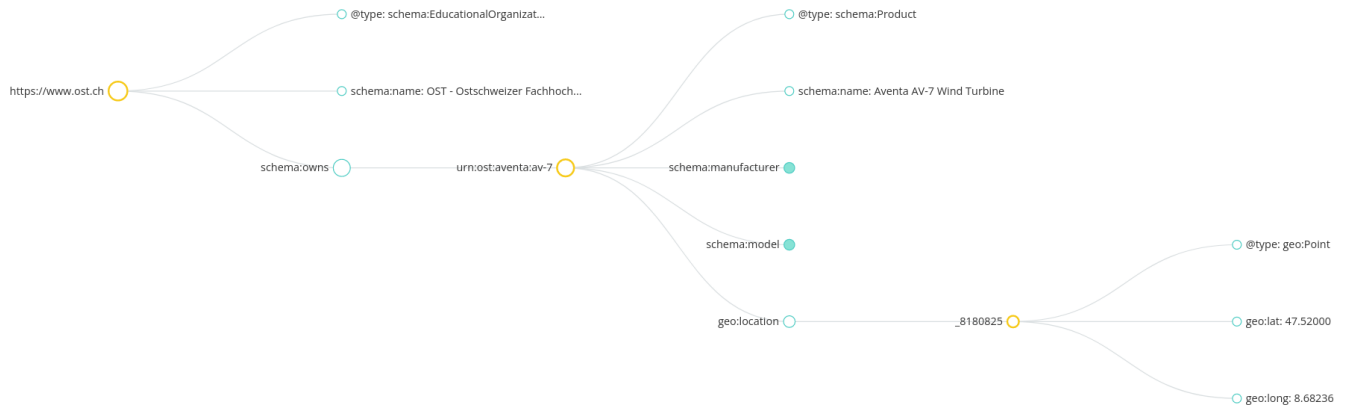


Figure B1. Aventa AV-7 described using JSON-LD and visualised as a graph.

Data availability. No data sets were used in this article.

Author contributions. Conceptualisation, YM and JD; methodology, YM and JD; investigation, YM, TC, JD, MW, JQ, and AMS; data curation, YM; writing – original draft preparation, YM; writing – review and editing, YM, TC, JD, MW, CH, JQ, AMS, IA, JPC, EC, and SB; visualisation, YM; supervision, EC and SB; project administration, SB; funding acquisition, SB. All authors have read and agreed to the published version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. A portion of this work is funded by the BRIDGE Discovery programme of the Swiss National Science Foundation and Innosuisse (project no. 40B2-0_187087).

A portion of this work was supported by the Wind Data Hub funded by the US Department of Energy Office of Energy Efficiency and Renewable Energy’s Wind Energy Technologies Office operated and maintained by the Pacific Northwest National Laboratory at <https://a2e.energy.gov> (last access: 25 March 2024).

Financial support. This research has been supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant no. no. 40B2-0_187087).

Review statement. This paper was edited by Weifei Hu and reviewed by two anonymous referees.

References

- AIAA and AIA: Digital Twin: Definition & Value – An AIAA and AIA Position Paper, <https://www.aiaa.org/advocacy/Policy-Papers/Institute-Position-Papers> (last access: 1 October 2023), 2020.
- Akerkar, R. and Sajja, P.: Knowledge-based systems, Jones & Bartlett Publishers, ISBN: 9780763776473, 2009.
- Andersen, G. and Aijmer, K.: Pragmatics of society, Vol. 5, Walter de Gruyter, <https://doi.org/10.1515/9783110214420>, 2011.
- Andriotis, C. P., Papakonstantinou, K. G., and Chatzi, E. N.: Value of structural health information in partially observable stochastic environments, *Struct. Saf.*, 93, 102072, <https://doi.org/10.1016/j.strusafe.2020.102072>, 2021.
- Angele, K. and Angele, J.: JSON towards a simple Ontology and Rule Language, in: Proceedings of the 15th International Rule Challenge, 7th Industry Track, and 5th Doctoral Consortium @ RuleML+RR 2021, edited by: Soylu, A., Nezhad, A. T., Nikolov, N., Toma, I., Fensel, A., and Vennekens, J., Vol. 2956 of CEUR Workshop Proceedings, CEUR, Leuven, Belgium, 8–15 September 2021, virtual, <http://ceur-ws.org/Vol-2956/#paper8> (last access: 25 March 2024), 2021.
- Arista, R., Zheng, X., Lu, J., and Mas, F.: An Ontology-based Engineering system to support aircraft manufacturing system design, *J. Manuf. Syst.*, 68, 270–288, <https://doi.org/10.1016/j.jmsy.2023.02.012>, 2023.
- Artigao, E., Martín-Martínez, S., Honrubia-Escribano, A., and Gómez-Lázaro, E.: Wind turbine reliability: A comprehensive review towards effective condition monitoring development, *Appl. Energ.*, 228, 1569–1583, <https://doi.org/10.1016/j.apenergy.2018.07.037>, 2018.
- Aschenbrenner, M. and Winder, G. M.: Planning for a sustainable marine future? Marine spatial planning in the German exclusive economic zone of the North Sea, *Appl. Geogr.*, 110, 102050, <https://doi.org/10.1016/j.apgeog.2019.102050>, 2019.
- Barber, S., Lima, L. A. M., Sakagami, Y., Quick, J., Latiffanti, E., Liu, Y., Ferrari, R., Letzgs, S., Zhang, X., and

- Hammer, F.: Enabling Co-Innovation for a Successful Digital Transformation in Wind Energy Using a New Digital Ecosystem and a Fault Detection Case Study, *Energies*, 15, 5638, <https://doi.org/10.3390/en15155638>, 2022.
- Barber, S., Hammer, F., and Henderson, C.: Can data sharing really provide added value? Practical data sharing recommendations for the wind energy sector, *J. Phys. Conf. Ser.*, 2507, 012003, <https://doi.org/10.1088/1742-6596/2507/1/012003>, 2023a.
- Barber, S., Izagirre, U., Serradilla, O., Olaizola, J., Zugasti, E., Aizpurua, J. I., Milani, A. E., Sehnke, F., Sakagami, Y., and Henderson, C.: Best Practice Data Sharing Guidelines for Wind Turbine Fault Detection Model Evaluation, *Energies*, 16, 3567, <https://doi.org/10.3390/en16083567>, 2023b.
- Barber, S., Sempreviva, A. M., Sheng, S., Farren, D., and Zappalá, D.: A use-case-driven approach for demonstrating the added value of digitalisation in wind energy, *J. Phys. Conf. Ser.*, 2507, 012002, <https://doi.org/10.1088/1742-6596/2507/1/012002>, 2023c.
- Besta, M., Gerstenberger, R., Peter, E., Fischer, M., Podstawski, M., Barthels, C., Alonso, G., and Hoefler, T.: Demystifying Graph Databases: Analysis and Taxonomy of Data Organization, System Designs, and Graph Queries, *ACM Comput. Surv.*, 56, 31, <https://doi.org/10.1145/3604932>, 2023.
- Booshehri, M., Emele, L., Flügel, S., Förster, H., Frey, J., Frey, U., Glauer, M., Hastings, J., Hofmann, C., Hoyer-Klick, C., Hülk, L., Kleinau, A., Knosala, K., Kotzur, L., Kuckertz, P., Mossakowski, T., Muschner, C., Neuhaus, F., Pehl, M., Robinius, M., Sehn, V., and Stapper, M.: Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis, *Energy and AI*, 5, 100074, <https://doi.org/10.1016/j.egyai.2021.100074>, 2021.
- Bortolotti, P., Bay, C., Barter, G., Gaertner, E., Dykes, K., McWilliam, M., Friis-Moller, M., Molgaard Pedersen, M., and Zahle, F.: System Modeling Frameworks for Wind Turbines and Plants: Review and Requirements Specifications, Tech. rep. no. NREL/TP-5000-82621, Office of Scientific and Technical Information (OSTI), <https://doi.org/10.2172/1868328>, 2022.
- Bremere, I. and Indriksone, D.: Regional stakeholder maps and analyses of decision flows, WP3.1, Baltic Energy Areas – A Planning Perspective (BEA-APP) project, https://www.balticenergyareas.eu/images/achievements/wp3.1_regional_stakeholder_maps_and_analyses_of_decision_flows.pdf (last access: 15 July 2023), 2017.
- Bunte, A., Li, P., and Niggemann, O.: Mapping Data Sets to Concepts using Machine Learning and a Knowledge based Approach, in: Proceedings of the 10th International Conference on Agents and Artificial Intelligence, Funchal, Madeira, Portugal, 16–18 January 2018, SCITEPRESS – Science and Technology Publications, 2, 430–437, <https://doi.org/10.5220/0006590204300437>, 2018.
- Cann, R., Kempson, R., and Gregoromichelaki, E.: *Semantics: An Introduction to Meaning in Language*, Cambridge University Press, ISBN: 9780521525664, 2009.
- Chah, N.: OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google’s Knowledge Graph, arXiv [preprint], <https://doi.org/10.48550/arXiv.1805.03885>, 22 May 2018.
- Chatterjee, J. and Dethlefs, N.: Temporal Causal Inference in Wind Turbine SCADA Data Using Deep Learning for Explainable AI, *J. Phys. Conf. Ser.*, 1618, 022022, <https://doi.org/10.1088/1742-6596/1618/2/022022>, 2020.
- Chavero-Navarrete, E., Trejo-Perea, M., Jáuregui-Correa, J. C., Carrillo-Serrano, R. V., and Ríos-Moreno, J. G.: Expert Control Systems for Maximum Power Point Tracking in a Wind Turbine with PMSG: State of the Art, *Appl. Sci.-Basel*, 9, 2469, <https://doi.org/10.3390/app9122469>, 2019.
- Chun, S., Jung, J., Jin, X., Seo, S., and Lee, K.-H.: Designing an integrated knowledge graph for smart energy services, *J. Supercomput.*, 76, 8058–8085, <https://doi.org/10.1007/s11227-018-2672-3>, 2018.
- Clark, T.: How to communicate and collaborate on data: easy-to-use tools and techniques for eliminating overwhelm, confusion and ambiguity, *Zenodo*, <https://doi.org/10.5281/zenodo.7928816>, 2022.
- Clark, T.: Powerful Power Curves: A production-proven, open-source schema for wind turbine power curves, *Zenodo*, <https://doi.org/10.5281/zenodo.7940068>, 2023.
- Clifton, A., Barber, S., Bray, A., Enevoldsen, P., Fields, J., Sempreviva, A. M., Williams, L., Quick, J., Purdue, M., Totaro, P., and Ding, Y.: Grand challenges in the digitalisation of wind energy, *Wind Energ. Sci.*, 8, 947–974, <https://doi.org/10.5194/wes-8-947-2023>, 2023.
- Codd, E. F.: *The Relational Model for Database Management: Version 2*, Addison-Wesley Longman Publishing Co., Inc., USA, ISBN 0201141922, 1990.
- Davis, R.: Knowledge-based systems, *Science*, 231, 957–963, 1986.
- De Baas, A., Nostro, P. D., Friis, J., Ghedini, E., Goldbeck, G., Paponetti, I. M., Pozzi, A., Sarkar, A., Yang, L., Zaccarini, F. A., and Toti, D.: Review and Alignment of Domain-Level Ontologies for Materials Science, *IEEE Access*, 11, 120372–120401, <https://doi.org/10.1109/ACCESS.2023.3327725>, 2023.
- De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., and Rosati, R.: *Using Ontologies for Semantic Data Integration*, Springer International Publishing, Cham, 187–202, https://doi.org/10.1007/978-3-319-61893-7_11, ISBN 978-3-319-61893-7, 2018.
- De Nicola, A., Missikoff, M., and Navigli, R.: A Proposal for a Unified Process for Ontology Building: UPON, in: *Database and Expert Systems Applications*, edited by: Andersen, K. V., Debenham, J., and Wagner, R., Springer Berlin Heidelberg, Berlin, Heidelberg, 655–664, ISBN 978-3-540-31729-6, 2005.
- de Vivero, J. L. S.: An exercise in Stakeholder Analysis for a hypothetical offshore wind farm in the Gulf of Cadix, Tech. rep. no. 00000909, University of Seville, https://spicosa-inline.databases.eucc-d.de/files/documents/00000908_Stakeholder%20Exercise_28%20ene%2008.pdf (last access: 25 March 2024), 2023.
- Ding, Y.: *Data Science for Wind Energy*, Chapman and Hall/CRC, <https://doi.org/10.1201/9780429490972>, 2019.
- Dourgnon-Hanoune, A., Dang, T., Salaun, P., and Bouthors, V.: An ontology for I&C knowledge using trees of porphyry, in: 2010 8th IEEE International Conference on Industrial Informatics, Osaka University Nakanoshima Center Osaka, Japan, 13–16 July 2010, IEEE, 86–92, <https://doi.org/10.1109/indin.2010.5549458>, 2010.
- Duer, S., Wrzesień, P., and Duer, R.: Creating of structure of facts for the knowledge base of an expert system for wind

- power plant's equipment diagnosis, E3S Web Conf., 19, 01038, <https://doi.org/10.1051/e3sconf/20171901038>, 2017.
- FAIR Data Maturity Model Working Group: FAIR Data Maturity Model: specification and guidelines, Version 1.0, Zenodo, <https://doi.org/10.15497/RDA00050>, 2020.
- Graybeal, J., Jonquet, C., Fiore, N., and Musen, M. A.: Adoption of BioPortal's Ontology Registry Software: The Emerging OntoPortal Community, 13th Research Data Alliance Plenary Meeting (RDA P13), Philadelphia, United States, April 2019, poster session, <https://hal-lirmm.ccsd.cnrs.fr/lirmm-02360625> (last access: 25 March 2024), 2019.
- Grieves, M.: SME Management Forum Completing the Cycle: Using PLM Information in the Sales and Service Functions, SME Management Forum, Troy, MI, 31 October 2002, https://www.researchgate.net/publication/356192963_SME_Management_Forum_Completing_the_Cycle_Using_PLM_Information_in_the_Sales_and_Service_Functions (last access: 25 March 2024), 2002.
- Grieves, M.: Intelligent digital twins and the development and management of complex systems, *Digital Twin*, 2, 8, <https://doi.org/10.12688/digitaltwin.17574.1>, 2022.
- Groarke, L.: Informal Logic, in: *The Stanford Encyclopedia of Philosophy*, edited by: Zalta, E. N. and Nodelman, U., Metaphysics Research Lab, Stanford University, Winter 2022 edn., ISSN 1095-5054, 2022.
- Group, W. S. W.: SPARQL 1.1 overview, W3C recommendation, World Wide Web Consortium (W3C), <http://www.w3.org/TR/sparql11-overview/> (last access: 25 March 2024), 2013.
- Gruber, T. R.: A translation approach to portable ontology specifications, *Knowl. Acquis.*, 5, 199–220, <https://doi.org/10.1006/knac.1993.1008>, 1993.
- Guarino, N.: *Formal Ontology in Information Systems: Proceedings of the 1st International Conference*, Trento, Italy, 6–8 June 1998, IOS Press, NLD, 1st edn., ISBN 9051993994, 1998.
- Haerder, T. and Reuter, A.: Principles of Transaction-Oriented Database Recovery, *ACM Comput. Surv.*, 15, 287–317, <https://doi.org/10.1145/289.291>, 1983.
- Hamilton, J. and Liming, D.: *Careers in Wind Energy: U.S. Bureau of Labor Statistics*, <https://www.bls.gov/green/wind-energy/> (last access: 15 July 2023), 2012.
- Heath, T. and Bizer, C.: *Linked Data*, Springer International Publishing, <https://doi.org/10.1007/978-3-031-79432-2>, 2011.
- Heidenreich, M. and Mattes, J.: Knowledge generation and diffusion in the German wind energy industry, *Ind. Corp. Change*, 31, 1285–1306, <https://doi.org/10.1093/icc/dtac022>, 2022.
- Heist, N., Hertling, S., Ringler, D., and Paulheim, H.: Knowledge Graphs on the Web – an Overview, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2003.00719>, 12 March 2020.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., and Rudolph, S.: *OWL 2 Web Ontology Language Primer*, 2nd edn., W3C Recommendation, World Wide Web Consortium (W3C), 11 December 2012, <https://www.w3.org/TR/owl2-primer/> (last access: 25 March 2024), 2009.
- Hlaing, N., Morato, P. G., Nielsen, J. S., Amirafshari, P., Koliou, A., and Rigo, P.: Inspection and maintenance planning for offshore wind structural components: integrating fatigue failure criteria with Bayesian networks and Markov decision processes, *Struct. Infrastruct. E.*, 18, 983–1001, <https://doi.org/10.1080/15732479.2022.2037667>, 2022.
- Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Rula, A., Sequeda, J., Zimmermann, A., Hogan, A., Gutierrez, C., Cochez, M., de Melo, G., Kirrane, S., Polleres, A., Navigli, R., Ngomo, A.-C. N., Rashid, S. M., Schmelzeisen, L., Staab, S., Blomqvist, E., d'Amato, C., Gayo, J. E. L., Neumaier, S., Rula, A., Sequeda, J., and Zimmermann, A.: Knowledge graphs, *ACM Comput. Surv.*, 54, 1–37, 2021.
- Huesca-Pérez, M. E., Sheinbaum-Pardo, C., and Köppel, J.: Social implications of siting wind energy in a disadvantaged region – The case of the Isthmus of Tehuantepec, Mexico, *Renew. Sust. Energy. Rev.*, 58, 952–965, <https://doi.org/10.1016/j.rser.2015.12.310>, 2016.
- Jaz Myers, D.: Double Categories of Open Dynamical Systems (Extended Abstract), *Electronic Proceedings in Theoretical Computer Science*, 333, 154–167, <https://doi.org/10.4204/eptcs.333.11>, 2021.
- Johnson, M., Rosebrugh, R., and Wood, R. J.: Lenses, fibrations and universal translations, *Math. Struct. Comp. Sci.*, 22, 25–42, <https://doi.org/10.1017/S0960129511000442>, 2012.
- Johnson, R. and Blair, J.: Informal logic and the reconfiguration of logic, in: *Handbook of the Logic of Argument and Inference*, edited by: Gabbay, D. M., Johnson, R. H., Ohlbach, H. J., and Woods, J., *Studies in Logic and Practical Reasoning*, Elsevier, 1, 339–396, [https://doi.org/10.1016/S1570-2464\(02\)80010-6](https://doi.org/10.1016/S1570-2464(02)80010-6), 2002.
- Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M. A., Pesce, V., and Larmande, P.: AgroPortal: A vocabulary and ontology repository for agronomy, *Comput. Electron. Agr.*, 144, 126–143, <https://doi.org/10.1016/j.compag.2017.10.012>, 2018.
- Jonquet, C., Graybeal, J., Bouazzouni, S., Dorf, M., Fiore, N., Kechagioglou, X., Redmond, T., Rosati, I., Skrenchuk, A., Vendetti, J. L., and Musen, M.: Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology, in: *Lecture Notes in Computer Science*, Springer Nature Switzerland, 38–58, https://doi.org/10.1007/978-3-031-47243-5_3, 2023.
- Kechagioglou, X., Vaira, L., Tomassino, P., Fiore, N., Basset, A., and Rosati, I.: EcoPortal: An Environment for FAIR Semantic Resources in the Ecological Domain, in: *Joint Ontology Workshops*, Bolzano, Italy, 11–18 September 2021, Vol. 2969, ISSN 1613-0073, <https://ceur-ws.org/Vol-2969/paper6-s4bioidiv.pdf> (last access: 25 March 2025), 2021.
- Keil, J. M. and Schindler, S.: Comparison and evaluation of ontologies for units of measurement, *Semant. Web*, 10, 33–51, <https://doi.org/10.3233/SW-180310>, 2019.
- Kirkegaard, J. K., Rudolph, D. P., Nyborg, S., Solman, H., Gill, E., Cronin, T., and Hallisey, M.: Tackling grand challenges in wind energy through a socio-technical perspective, *Nature Energy*, 8, 655–664, <https://doi.org/10.1038/s41560-023-01266-z>, 2023.
- Klonari, V., Papachristos, G., and Fraile, D.: Wind energy digitalisation towards 2030, <https://windeurope.org/intelligence-platform/product/wind-energy-digitalisation-towards-2030/> (last access: 1 October 2023), 2021.
- Komisar, A. and Fox, M. S.: An Energy Ontology for Global City Indicators (ISO 37120), *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2008.04070>, 19 July 2020.

- Küçük, D. and Küçük, D.: *OntoWind: An Improved and Extended Wind Energy Ontology*, arXiv [preprint], <https://doi.org/10.48550/arXiv.1803.02808>, 7 March 2018.
- Lamanna, D. D. and Maccioni, A.: *Renewable Energy Data Sources in the Semantic Web with OpenWatt*, in: *Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014)*, CEUR Workshop Proceedings, Athens, Greece, 28 March 2014, edited by: Candan, K. S., Amer-Yahia, S., Schweikardt, N., Christophides, V., and Leroy, V., CEUR-WS.org, 1133, 128–133, <http://ceur-ws.org/Vol-1133/paper-20.pdf> (last access: 25 March 2025), 2014.
- Lassila, O. and McGuinness, D. L.: *The Role of Frame-Based Representation on the Semantic Web*, Tech. Rep. KSL-01-02, Stanford University, Stanford, <http://www.ida.liu.se/ext/epa/ej/etai/2001/018/01018-etaibody.pdf> (last access: 25 March 2025), 2001.
- Lavore, E. D., Leal, W., and de Paiva, V.: *Dialectica Petri nets*, arXiv [preprint], <https://doi.org/10.48550/arxiv.2105.12801>, 14 February 2024.
- Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväsliho, H., Coen, G., and Staiger, C.: *D2.2 FAIR Semantics: First recommendations*, Zenodo, <https://doi.org/10.5281/zenodo.3707985>, 2020.
- Leahy, K., Gallagher, C., O'Donovan, P., and O'Sullivan, D. T.: *Issues with data quality for wind turbine condition monitoring and reliability analyses*, *Energies*, 12, 201, <https://doi.org/10.3390/en12020201>, 2019.
- Li, F., Li, L., and Peng, Y.: *Research on Digital Twin and Collaborative Cloud and Edge Computing Applied in Operations and Maintenance in Wind Turbines of Wind Power Farm*, in: *Advances in Transdisciplinary Engineering*, IOS Press, <https://doi.org/10.3233/atde210263>, 2021.
- Liang, Y., Sun, L., and Zhao, X.: *Reinforcement Learning-Based Inertia and Droop Control for Wind Farm Frequency Regulation*, in: *2022 IEEE Power & Energy Society General Meeting (PESGM)*, Denver, CO, USA, 17–21 July 2022, IEEE, <https://doi.org/10.1109/pesgm48719.2022.9917075>, 2022.
- Lourenço, J. R., Abramova, V., Vieira, M., Cabral, B., and Bernardino, J.: *NoSQL Databases: A Software Engineering Perspective*, in: *New Contributions in Information Systems and Technologies*, edited by: Rocha, A., Correia, A. M., Costanzo, S., and Reis, L. P., Springer International Publishing, Cham, 741–750, ISBN 978-3-319-16486-1, 2015.
- Lungu, I., Velicanu, A., Bâra, L. A., Botha, I., Mocanu, A. M., and Tudor, A.: *Spatial databases for wind parks*, *Econ. Comput. Econ. Cyb.*, Vol. 6, https://www.researchgate.net/publication/290734023_Spatial_databases_for_wind_parks (last access: 25 March 2024), 2012.
- Magee, L.: *8 – Contemporary dilemmas: tables versus webs*, in: *Towards a Semantic Web*, edited by: Cope, B., Kalantzis, M., and Magee, L., Chandos Publishing, 215–234, <https://doi.org/10.1016/B978-1-84334-601-2.50008-8>, ISBN 978-1-84334-601-2, 2011.
- Markovska, N., Krkoleva, A., Taseska, V., Borozan, V., and Pop-Jordanov, J.: *Enabling an environment for solar and wind energy deployment in the Macedonian agricultural sector*, *J. Renew. Sustain. Ener.*, 5, 041804, <https://doi.org/10.1063/1.4812996>, 2013.
- Marykovskiy, Y., Abdallah, I., Barber, S., and Chatzi, E.: *Extended Taxonomy of Digital Twins, Version v2*, Zenodo, <https://doi.org/10.5281/zenodo.8021787>, 2023.
- Michiorri, A., Sempreviva, A. M., Philipp, S., Perez-Lopez, P., Ferriere, A., and Moser, D.: *Topic Taxonomy and Metadata to Support Renewable Energy Digitalisation*, *Energies*, 15, 9531, <https://doi.org/10.3390/en15249531>, 2022.
- Morato, P., Papakonstantinou, K., Andriotis, C., Nielsen, J., and Rigo, P.: *Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes*, *Struct. Saf.*, 94, 102140, <https://doi.org/10.1016/j.strusafe.2021.102140>, 2022.
- Naghieb, A., Jafari Navimipour, N., Hosseinzadeh, M., and Sharifi, A.: *A comprehensive and systematic literature review on the big data management techniques in the internet of things*, *Wirel. Netw.*, 29, 1085–1144, <https://doi.org/10.1007/s11276-022-03177-5>, 2022.
- Nguyen, T. H., Dunlap, R., Mark, L., Prinz, A., Østgren, B. M., and Friisø, T.: *Offshore Wind Metadata Management*, *International Journal of Metadata, Semantics and Ontologies*, 9, 333–349, <https://doi.org/10.1504/IJMSO.2014.065445>, 2014.
- Nowack, B.: *The Semantic Web – Not a piece of cake...*, <http://web.archive.org/web/20220628120341/http://bnode.org/blog/2009/07/08/the-semantic-Web-not-a-piece-of-cake> (last access: 15 July 2023), 2009.
- Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., and Taylor, J.: *Industry-Scale Knowledge Graphs: Lessons and Challenges*, *Commun. ACM*, 62, 36–43, <https://doi.org/10.1145/3331166>, 2019.
- Noy, N. F. and McGuinness, D. L.: *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford, <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf> (last access: 10 March 2023), 2001.
- Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., and Musen, M. A.: *BioPortal: ontologies and integrated data resources at the click of a mouse*, *Nucleic Acids Res.*, 37, W170–W173, <https://doi.org/10.1093/nar/gkp440>, 2009.
- Papadopoulos, P. and Cipcigan, L.: *Wind turbines' condition monitoring: an ontology model*, in: *2009 International Conference on Sustainable Power Generation and Supply*, Nanjing, China, 6–7 April 2009, IEEE, 1–4, <https://doi.org/10.1109/SUPERGEN.2009.5430854>, ISSN 2156-969X, 2009.
- Papakonstantinou, K. and Shinozuka, M.: *Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory*, *Reliab. Eng. Syst. Safe.*, 130, 202–213, <https://doi.org/10.1016/j.ress.2014.04.005>, 2014.
- Partridge, C., Mitchell, A., Cook, A., Sullivan, J., and West, M.: *A Survey of Top-Level Ontologies – to inform the ontological choices for a Foundation Data Model*, CDBB, <https://doi.org/10.17863/CAM.58311>, 2020.
- Pease, S. G., Sharpe, R., van Lopik, K., Tsalapati, E., Goodall, P., Young, B., Conway, P., and West, A.: *An interoperable semantic service toolset with domain ontology for automated decision support in the end-of-life domain*, *Future Gener. Comp. Sy.*, 112, 848–858, <https://doi.org/10.1016/j.future.2020.06.008>, 2020.

- Premkumar, V., Krishnamurthy, S., Wileden, J. C., and Grosse, I. R.: A semantic knowledge management system for laminated composites, *Adv. Eng. Inform.*, 28, 91–101, <https://doi.org/10.1016/j.aei.2013.12.004>, 2014.
- Pronost, G., Mayer, F., Marche, B., Camargo, M., and Dupont, L.: Towards a Framework for the Classification of Digital Twins and their Applications, in: 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, United Kingdom, 21–23 June 2021, IEEE, 1–7, <https://doi.org/10.1109/ICE/ITMC52061.2021.9570114>, 2021.
- Qi, Q., Terkaj, W., Urgo, M., Jiang, X., and Scott, P. J.: A mathematical foundation to support bidirectional mappings between digital models: an application of multi-scale modelling in manufacturing, *P. Roy. Soc. A-Math. Phys.*, 478, 20220156, <https://doi.org/10.1098/rspa.2022.0156>, 2022.
- Quaeghebeur, E., Sanchez Perez-Moreno, S., and Zaaier, M. B.: WESgraph: a graph database for the wind farm domain, *Wind Energ. Sci.*, 5, 259–284, <https://doi.org/10.5194/wes-5-259-2020>, 2020.
- Reder, M. D., Gonzalez, E., and Melero, J. J.: Wind Turbine Failures – Tackling current Problems in Failure Data Analysis, *J. Phys. Conf. Ser.*, 753, 072027, <https://doi.org/10.1088/1742-6596/753/7/072027>, 2016.
- Rogushina, J. V. and Gladun, A. Y.: Semantic Processing of Metadata for Big Data: Standards, Ontologies and Typical Information Objects, in: *Information Technologies and Security (ITS)*, Kyiv, Ukraine, 10 December 2020, 114–128, ISSN 1613-0073, <https://ceur-ws.org/Vol-2859/paper10.pdf> (last access: 25 March 2024), 2020.
- Rosenberg, S.: PSEG and the promise of wind power, *The CASE Journal*, 16, 51–74, <https://doi.org/10.1108/tcj-03-2019-0024>, 2019.
- Sanchez-Fernandez, A. J., González-Sánchez, J.-L., Luna Rodríguez, I., Rodríguez, F. R., and Sanchez-Rivero, J.: Reliability of onshore wind turbines based on linking power curves to failure and maintenance records: A case study in central Spain, *Wind Energy*, 26, 349–364, <https://doi.org/10.1002/we.2793>, 2023.
- Schreiber, G. and Raimond, Y.: RDF 1.1 Primer, <http://www.w3.org/TR/rdf11-primer/> (last access: 25 March 2024), 2014.
- Sempreviva, A. M., Vesth, A., Bak, C., Verelst, D. R., Giebel, G., Danielsen, H. K., Mikkelsen, L. P., Andersson, M., Vasiljevic, N., Barth, S., Sanz, R. J., Gancarski, P., Reigstad, T. I., Bolstad, H. C., Wagenaar, J. W., and Hermans Koen, W.: Taxonomy And Metadata For Wind Energy Research & Development, Zenodo, <https://doi.org/10.5281/zenodo.1199489>, 2017.
- Seyr, H. and Muskulus, M.: Decision Support Models for Operations and Maintenance for Offshore Wind Farms: A Review, *Applied Sciences*, 9, 278, <https://doi.org/10.3390/app9020278>, 2019.
- Shapiro, B. T. and Spivak, D. I.: Dynamic Operads, Dynamic Categories: From Deep Learning to Prediction Markets, *Electronic Proceedings in Theoretical Computer Science*, 380, 183–202, <https://doi.org/10.4204/eptcs.380.11>, 2023.
- Soussi, N. and Bahaj, M.: Exploiting NoSQL Document Oriented Data Using Semantic Web Tools, in: *Advanced Intelligent Systems for Sustainable Development (AI2SD'2018)*, edited by: Ezziyyani, M., Springer International Publishing, Cham, 110–117, ISBN 978-3-030-11928-7, 2019.
- Spivak, D. I.: Functorial data migration, *Inform. Comput.*, 217, 31–51, <https://doi.org/10.1016/j.ic.2012.05.001>, 2012.
- Spivak, D. I.: Poly: An abundant categorical setting for mode-dependent dynamics, arXiv [preprint], <https://doi.org/10.48550/arXiv.2005.01894>, 11 June 2020.
- Spivak, D. I. and Kent, R. E.: Ologs: A Categorical Framework for Knowledge Representation, *PLoS ONE*, 7, e24274, <https://doi.org/10.1371/journal.pone.0024274>, 2012.
- Spyns, P., Meersman, R., and Jarrar, M.: Data Modelling versus Ontology Engineering, *SIGMOD Rec.*, 31, 12–17, <https://doi.org/10.1145/637411.637413>, 2002.
- St. Clere Smithe, T.: Open Dynamical Systems as Coalgebras for Polynomial Functors, with Application to Predictive Processing, *Electronic Proceedings in Theoretical Computer Science*, 380, 307–330, <https://doi.org/10.4204/eptcs.380.18>, 2023.
- Strack, B., Lenart, M., Frank, J., and Kramer, N.: Ontology for maintenance of onshore wind turbines, *Forsch. Ingenieurwes.*, 85, 265–272, <https://doi.org/10.1007/s10010-021-00466-x>, 2021.
- Studer, R., Benjamins, V. R., and Fensel, D.: Knowledge engineering: Principles and methods, *Data Knowl. Eng.*, 25, 161–197, 1998.
- The Editors of Encyclopaedia Britannica: Formal system 1 Logic, Symbols & Axioms, <https://www.britannica.com/topic/formal-system> (last access: 1 October 2023), 2012.
- Tsalapati, E., Jackson, T., Johnson, W., Jackson, L. M., Vasilyev, A., West, A. A., Mao, L., and Davies, B.: The Role of Semantic Technologies in Diagnostic and Decision Support for Service Systems, in: *Hawaii International Conference on System Sciences*, Waikoloa Village, Hawaii, 3–6 January 2018, <https://doi.org/10.24251/HICSS.2018.196>, 2018.
- Tsiliamanis, G. P., Wagg, D. J., Antoniadou, I., and Worden, K.: An Ontological Approach to Structural Health Monitoring, in: *Topics in Modal Analysis & Testing*, edited by: Dilworth, B. and Mains, M., Springer International Publishing, Cham, 8, 51–59, ISBN 978-3-030-47717-2, 2021.
- Uhlenkamp, J.-F., Hauge, J. B., Broda, E., Lütjen, M., Freitag, M., and Thoben, K.-D.: Digital Twins: A Maturity Model for Their Classification and Evaluation, *IEEE Access*, 10, 69605–69635, <https://doi.org/10.1109/ACCESS.2022.3186353>, 2022.
- van der Valk, H., Haße, H., Möller, F., and Otto, B.: Archetypes of Digital Twins, *Bus. Inf. Syst. Eng.*, 64, 375–391, <https://doi.org/10.1007/s12599-021-00727-7>, 2021.
- Veers, P., Dykes, K., Lantz, E., Barth, S., Bottasso, C. L., Carlson, O., Clifton, A., Green, J., Green, P., Holttinen, H., Laird, D., Lehtomäki, V., Lundquist, J. K., Manwell, J., Marquis, M., Meneveau, C., Moriarty, P., Munduate, X., Muskulus, M., Naughton, J., Pao, L., Paquette, J., Peinke, J., Robertson, A., Sanz Rodrigo, J., Sempreviva, A. M., Smith, J. C., Tuohy, A., and Wiser, R.: Grand challenges in the science of wind energy, *Science*, 366, eaau2027, <https://doi.org/10.1126/science.aau2>, 2019.
- Vicuña, D. M. and Pérez, J. E.: Wind energy policy and its effects on local development. An analysis from the stakeholder system (Loja, Ecuador), *Anales de Geografía de la Universidad Complutense*, 40, 73–95, <https://doi.org/10.5209/aguc.69333>, 2020.
- Villegas, M., Melero, M., and Bel, N.: Metadata as Linked Open Data: mapping disparate XML metadata registries into one RDF/OWL registry, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation*

- (LREC'14), Reykjavik, Iceland, 26–31 May 2014, European Language Resources Association (ELRA), 393–400, ISBN 978-2-9517408-8-4, http://www.lrec-conf.org/proceedings/lrec2014/pdf/664_Paper.pdf (last access: 25 March 2024), 2014.
- Vrandečić, D.: *Ontology Evaluation*, Springer, Berlin, Heidelberg, 293–313, https://doi.org/10.1007/978-3-540-92673-3_13, ISBN 978-3-540-92673-3, 2009.
- W3C: *Linked Data*, <https://www.w3.org/standards/semanticweb/data>, last access: 10 March 2023.
- Wagg, D. J., Worden, K., Barthorpe, R. J., and Gardner, P.: *Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications*, ASME J. Risk Uncertainty Part B, 6, 030901, <https://doi.org/10.1115/1.4046739>, 030901, 2020.
- Weber, J. and Köppel, J.: Can MCDA Serve Ex-Post to Indicate ‘Winners and Losers’ in Sustainability Dilemmas? A Case Study of Marine Spatial Planning in Germany, *Energies*, 15, 7654, <https://doi.org/10.3390/en15207654>, 2022.
- Wever, L., Krause, G., and Buck, B. H.: Lessons from stakeholder dialogues on marine aquaculture in offshore wind farms: Perceived potentials, constraints and research gaps, *Mar. Policy*, 51, 251–259, <https://doi.org/10.1016/j.marpol.2014.08.015>, 2015.
- Wiens, M., Meyer, T., and Thomas, P.: The Potential of FMI for the Development of Digital Twins for Large Modular Multi-Domain Systems, in: *Proceedings of 14th Modelica Conference 2021, Modelica 2021*, Linköping, Sweden, 20–24 September 2021, Linköping University Electronic Press, 235–240, <https://doi.org/10.3384/ecp21181235>, ISSN 1650-3686, 2021.
- Wierling, A., Schwanitz, V. J., Altinci, S., Bałazińska, M., Barber, M. J., Biresselioglu, M. E., Burger-Scheidlin, C., Celino, M., Demir, M. H., Dennis, R., Dintzner, N., el Gammal, A., Fernández-Peruchena, C. M., Gilcrease, W., Gładysz, P., Hoyer-Klick, C., Joshi, K., Kruczek, M., Lacroix, D., Markowska, M., Mayo-García, R., Morrison, R., Paier, M., Peronato, G., Ramakrishnan, M., Reid, J., Sciullo, A., Solak, B., Suna, D., Süß, W., Unger, A., Fernandez Vanoni, M. L., and Vasiljevic, N.: FAIR Metadata Standards for Low Carbon Energy Research – A Review of Practices and How to Advance, *Energies*, 14, 6692, <https://doi.org/10.3390/en14206692>, 2021.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>, 2016.
- Wu, M., Juty, N., WG, R. M. S., Collins, J., Duerr, R., Ridsdale, C., Shepherd, A., Verhey, C., and Castro, L. J.: *Guidelines for publishing structured metadata on the web*, Version 3.1, Zenodo, <https://doi.org/10.15497/RDA00066>, 2021.
- Zardini, G., Milojevic, D., Censi, A., and Frazzoli, E.: Co-design of Embodied Intelligence: A Structured Approach, in: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 27 September 2021–1 October 2021, 7536–7543, <https://doi.org/10.1109/IROS51168.2021.9636513>, 2021.
- Zheng, X., Lu, J., and Kiritsis, D.: The emergence of cognitive digital twin: vision, challenges and opportunities, *Int. J. Prod. Res.*, 60, 1–23, <https://doi.org/10.1080/00207543.2021.2014591>, 2021.
- Zhou, A., Yu, D., and Zhang, W.: A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA, *Adv. Eng. Inform.*, 29, 115–125, <https://doi.org/10.1016/j.aei.2014.10.001>, 2015.
- Zhu, Y.-L., Wang, X.-Y., and Cheng, D.-L.: *Ontology-Based Research on Wind Power Plant Information Interaction*, in: *First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008)*, Adelaide, SA, Australia, 23–24 January 2008, IEEE, 169–172, <https://doi.org/10.1109/WKDD.2008.88>, 2008.