ORIGINAL RESEARCH **Open access**

A Community-Driven Open Energy Ontology for Linked Open Data and Energy Model Integration

Sara Ben Youssef[1*], Amal Trabelsi[1], Karim Boudiaf[2]

Abstract

Diverse data structures, varying terminologies, and disjointed models present significant challenges across numerous disciplines, including energy systems analysis. Consequently, reusing findings, comparing model outputs, and integrating different models become highly difficult. Ontologies supply a rigorously established lexicon that establishes a unified, shared conceptual framework for the energy sector. In this paper, we introduce the Open Energy Ontology (OEO), specifically engineered for energy systems analysis. Adopting the OEO yields multiple advantages for researchers. Primarily, it facilitates the standardized annotation of massive datasets derived from diverse scientific investigations, such as those found on the Open Energy Platform (OEP). Incorporating these annotations renders data semantically discoverable, transferable, reusable, and cross-compatible. Furthermore, it significantly simplifies the computational linking of distinct models. The benefits of applying an ontology like the OEO are illustrated through three specific applications: data representation, metadata annotation, and interface standardization. Additionally, we outline the ontology's utility in the context of linked open data (LOD).

Keywords Collaborative ontology development, Linked open data, Metadata annotation, Energy systems analysis

*Correspondence:
Sara Ben Youssef
sara.benyoussef@utunis.tn
[1] Department of Artificial Intelligence in Energy Engineering, Faculty of Engineering, University of Tunis, Tunis, Tunisia
[2] Department of Smart Grid Intelligence, Faculty of Science and Technology, University of Tunis El Manar, Tunis, Tunisia

Introduction

The impartial evaluation of existing and prospective energy system architectures and functions represents a worldwide, highly interdisciplinary research challenge within energy systems analysis. Specialists across engineering, the natural and social sciences, physics, mathematics, computer science, economics, meteorology, and geography frequently collaborate; prominent examples include examining trajectories leading to sustainable energy frameworks that align with the Paris Agreement1. Nations, organizations, and academics rely heavily on collaboration and connectivity within the energy systems network. Comprehensive scholarly dialogue among all involved stakeholders is essential to address one of our most critical societal issues. However, various subgroups among these participants have established distinct terminologies and conceptual models, which consequently appear in their corresponding research documentation and findings. The resulting diversity in this research field creates several data and knowledge management complications that obstruct seamless cooperation among scientists. In the subsequent sections, we present the Open Energy Ontology (OEO) as a mechanism to resolve these issues. First, we examine the field of energy systems modeling and outline the specific obstacles that the OEO can overcome.We offer an overview of ontologies and their advantages concerning energy system modeling. We review current methodologies utilized within the energy sector. We introduce the OEO as a specialized domain ontology tailored for energy system modeling and analysis, detailing its structural choices, design frameworks, and content organization. We explain the transparent, cooperative formulation process of the OEO and its

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

integration into the energy systems analysis community to ensure its long-term viability. We outline the assessment of multiple dimensions of the OEO. We explore practical applications derived from the externally funded SzenarienDB, LOD-GEOSS, and SIROP research initiatives. We conclude the paper with overarching deductions and perspectives on future research directions.

Challenges Within a Heterogeneous Domain

Investigations in this field frequently rely on computational frameworks (e.g., energy system modeling) to delineate the dynamics and potential for advancement of energy and its interconnected networks. A wide array of energy models and scenarios, grounded in logical and unified premises and objectives, illustrate potential trajectories for tomorrow's energy landscapes. Nevertheless, an isolated scenario or an individual model cannot accurately capture all pertinent variables. Consequently, investigators typically construct a collection of scenarios to address a specific issue, focusing on distinct questions derived from unique underlying narratives. A given scenario might emphasize technical, financial, environmental, or societal dimensions, or an amalgamation thereof. Energy models vary according to their geographical and sectoral focus (e.g., manufacturing, housing, or transportation), their granularity (e.g., time-based and geographic resolution), and their foundational hypotheses. The primary engine for generating knowledge often lies in contrasting and analyzing scenario and model outcomes, while varying the initial data and premises to understand how the system operates. Smoothing the pathway for inter-model data exchange enhances analytical depth by integrating information from rigorously validated modeling structures with those from alternative methodologies and disciplines.

Evolution of energy systems modelling

Pfenninger *et al*. [1] provide a comprehensive review of the historical progression of energy system modeling as it adapted to a variety of emerging challenges. This paragraph offers a summary of the transition of energy systems analysis from purely technical paradigms to comprehensive modeling techniques that encompass multiple scientific disciplines, accompanied by an escalating complexity in data management. The discipline of energy systems analysis originated in the early 1970s in response to the global oil crisis, marked by the establishment of the Energy Technology Systems Analysis Programme (ETSAP) by the International Energy Agency (IEA) in 1974 and the International Institute for Applied Systems Analysis (IIASA) in 1972. The initial energy system models utilized linear programming techniques [2]. Prominent examples include the MARKAL/TIMES suite of models [3] and the MESSAGE framework [4]. These frameworks focused on the technical evolution of energy networks, aiming to optimize for the most economically efficient outcomes. The subsequent breakthrough involved creating hybrid models [5], which broadened the analytical scope to include the economic domain by linking granular, bottom-up technological energy models with macroeconomic general equilibrium models, such as the MACRO model. The integration of MARKAL and MACRO was achieved through hard-linking the two architectures, allowing the combined system model to be solved concurrently [6]. Conversely, the connection between MESSAGE and MACRO employed a soft-linking strategy [7], establishing specific interfaces between the MESSAGE and MACRO models in which the output of one served as the input to the other, solving them through iterative cycles.

Rising challenges in complex data handling and emerging big data and artificial intelligence approaches

These earlier modeling strategies successfully integrated information from both the energy technology and economic sectors. Initially, these frameworks predominantly addressed traditional energy networks reliant on nuclear and fossil fuel resources. Because such fuels can be stockpiled and energy generated on demand, these models did not need to account for geographic or temporal fluctuations in production. To reflect demand-side variations, these simulations incorporated only a limited number of time segments—typically covering different seasons, diurnal cycles, and peak

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

consumption periods —totaling roughly 12 time slices annually. Subsequent challenges emerged as renewable energy resources gained prominence within carbon-neutral energy paradigms. In particular, wind and solar energy exhibit significant temporal and spatial volatility. Consequently, energy system models incorporating substantial proportions of these renewables had to accommodate their complex geographical and chronological availability patterns. As a result, a novel category of energy system models was developed post-2000. These paradigms typically employed 8,760 hourly increments annually and required meteorological and climatological records as a crucial, novel data input. From that point forward, energy systems analysis integrated the disciplines of climate and weather. Standard representatives of these models include REMIx [8], PyPSA [9], and SCOPE [10]. More recent methodologies prioritize the broader incorporation of societal and sociological factors. Prominent examples involve sociotechnical scenario development [11] and agent-based modeling [12, 13]. In addition to these traditional modeling techniques, novel artificial intelligence, machine learning, and big data strategies are gaining substantial recognition within the field, such as the application of machine learning for model parameterization [14]. Jin *et al*. [15] illustrate a wide range of implementations of artificial intelligence (AI) techniques in energy studies. In the realm of energy systems analysis, early AI implementations have targeted building analysis and management systems [16-21], localized integrated energy networks [22-24], intelligent charging infrastructure [25], load forecasting [26-30], big data evaluation such as data mining [31, 32], and capital allocation strategies [14]. Consequently, algorithms must be equipped to comprehend and translate this extensive array of data streams originating from disparate disciplines. As the resolution of technologies, time, and space escalates alongside the integration of broader thematic areas, managing this information grows exponentially complicated. Accurate translation of cross-disciplinary data is therefore crucial for effective assessments within standalone models, and this necessity is even more pronounced when models are interconnected via data linkages. The Open Energy Ontology (OEO) provides precise data definitions and specifies how models operating within energy systems analysis should interpret them. Because both the big data and machine learning fields are expanding rapidly, addressing this issue has become a pressing necessity for future research. Instances of this include the “data silo” dilemma—massive datasets that are theoretically accessible but remain unusable by other researchers because their formatting and curation lack reproducibility, or because reformatting them demands prohibitive time investments. If the same data were annotated using an ontology, it would become instantly reusable, given that its semantic meaning and alignment with alternative data sources would be unequivocal. Furthermore, ontologies enable machine learning pipelines to ingest information from diverse sources without requiring a complete overhaul of the workflow, since heterogeneous data can be treated uniformly when consistently annotated with an ontology. Thus, the Open Energy Ontology acts as a vital catalyst for applying these advanced methods.

Research driven by heterogeneous data from heterogeneous sources

Data serves as the primary engine for research in energy systems analysis. Because scenarios fluctuate significantly with different inputs, the outcomes depend heavily on the quality of the underlying data. In a comprehensive assessment of data analytics within the electricity industry, Von Scheidt *et al*. [33] identified a wide array of analytical methodologies applied across the entire value chain. Information utilized for models and scenarios typically stems from diverse data sources spanning numerous disciplines. At the outset of any research endeavor, aligning and translating this cross-disciplinary data poses a significant hurdle: data may be supplied by government bodies, extracted from academic literature, sourced from public or commercial repositories, generated through crowd-sourcing efforts, or directly recorded by investigators and remote sensing technologies. These data formats span from isolated figures and chronological sequences to complex multidimensional arrays. Additionally, the information spans differing geographic and temporal resolutions, including site-specific hourly wind velocities measured at multiple elevations.

Beyond relying on a massive foundational dataset, the energy systems analysis sector concurrently produces immense volumes of its own data through ongoing modeling activities. Without robust mechanisms to continuously and consistently tag data with necessary documentation and context, repositories risk becoming “data graveyards,” where locating, connecting to, extracting, and refreshing pertinent information becomes highly problematic. This predicament accelerates the creation of disconnected, rapidly obsolete data silos. Consequently, researchers are forced into repetitive cycles of recompiling data inventories, drastically degrading data management efficiency throughout the discipline.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Biology offers an encouraging counterexample: established in 1998, the Gene Ontology2 serves as the core of biomedical understanding of genetic functions. This collaborative, decentralized, and universally applicable framework contains upwards of 6 million machine-readable functional annotations across roughly 4,400 species. It integrates discoveries from over 150,000 publications and has been heavily cited in tens of thousands of academic investigations. Driving numerous databases and facilitating extensive annotation processes, it stands as arguably the most effective tool in computational biology. Looking at another sector, the relationships and terminology established by ontologies serve as the bedrock for numerous Internet of Things (IoT) platforms. They are crucial for creating digital twins that function reliably across distinct fields.

However, the advantages of the OEO extend far beyond simply tagging available information. Every component in this ontology functions within an expansive logical framework grounded in the sophisticated OWL2 semantics [34], which enables the deduction of implicit concepts. This capability enables data systems to seamlessly blend information from diverse scientific arenas—such as merging biological and chemical data [35]—while simultaneously bridging structural voids in an automated fashion, guided by the ontology's logical foundations. The structured knowledge inherent in ontologies can additionally provide a base for innovative artificial intelligence methodologies. For instance, the categorical architecture of the CHEBI ontology was recently used to underpin a deep learning model for classifying chemical compounds [36].

Research gap - the road to the open energy ontology

The management of increasingly intricate data architectures within energy systems analysis has traditionally been overlooked. Nonetheless, various open data platforms and discussion groups have recently initiated dialogues regarding these obstacles, which we outline in this subsection. Furthermore, parallel ontology construction efforts are detailed and subsequently connected to our current research.

Even so, no existing methodology adequately encompasses the expansive lexicon of our field, nor does it provide a structural framework that meets our specific demands. Presently, no dedicated ontology exists for energy systems analysis that delineates pertinent data and modeling techniques, along with their full array of attributes. Consequently, handling, transferring, contrasting, and evaluating scientific findings, methods, and data remain formidable obstacles that are constantly addressed by externally funded programs and collaborative community efforts345.

The OpenMOD glossary6 represented an early attempt to construct a community-governed knowledge repository for energy modeling, functioning as a foundational blueprint for the OEO. This glossary7 contained 323 distinct terms focused on photovoltaic module modeling, aggregated by community members and derived from a lecture series at the HTW8. The associated web interface facilitated the assignment of acronyms, synonyms, broader categories, and specialized sub-terms, alongside enabling dedicated discussion threads for individual entries. While the glossary successfully fostered mutual understanding, its inherent technical and structural constraints as a structured, machine-readable knowledge archive simultaneously became apparent. The lexicon developed by the externally funded openENTRANCE9 initiative strictly addresses vocabulary pertinent to its specific models and lacks interconnections among its concepts. The absence of semantic connections among the entries in these two previously mentioned projects impedes their utility for artificial intelligence applications—an issue that the OEO specifically resolves.

A prominent ontology was previously introduced based on industrial park use cases [37]. It incorporates critical entities associated with supply-demand networks and energy grid architectures. We intend to harmonize numerous components across both ontologies. Still, multiple vital facets required for a comprehensive conceptualization of the energy systems analysis sector remain omitted. This gap notably involves environmental variables as well as descriptions of scientific methodologies and data frameworks—all of which are fundamental pillars of the OEO's scope.

We engineered the OEO to facilitate smoother collaboration and information transfer throughout the energy systems analysis discipline. Additionally, the OEO is structured to accurately reflect the intricacy of this scientific field by gathering,

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

linking, and organizing the often ambiguous lexicons of neighboring disciplines upon which energy systems analysis relies **(Figure 1)**. The continuous expansion of this ontology progressively facilitates the accurate, unambiguous, and easily understandable tagging and translation of scientific data. As a foundational tool for seamless global academic communication, the OEO enables the integration and repurposing of scattered cross-disciplinary data repositories. This mobilizes cooperative synergies across the worldwide, multidisciplinary energy systems analysis network, ultimately bolstering the resilient shift toward sustainable energy infrastructures.

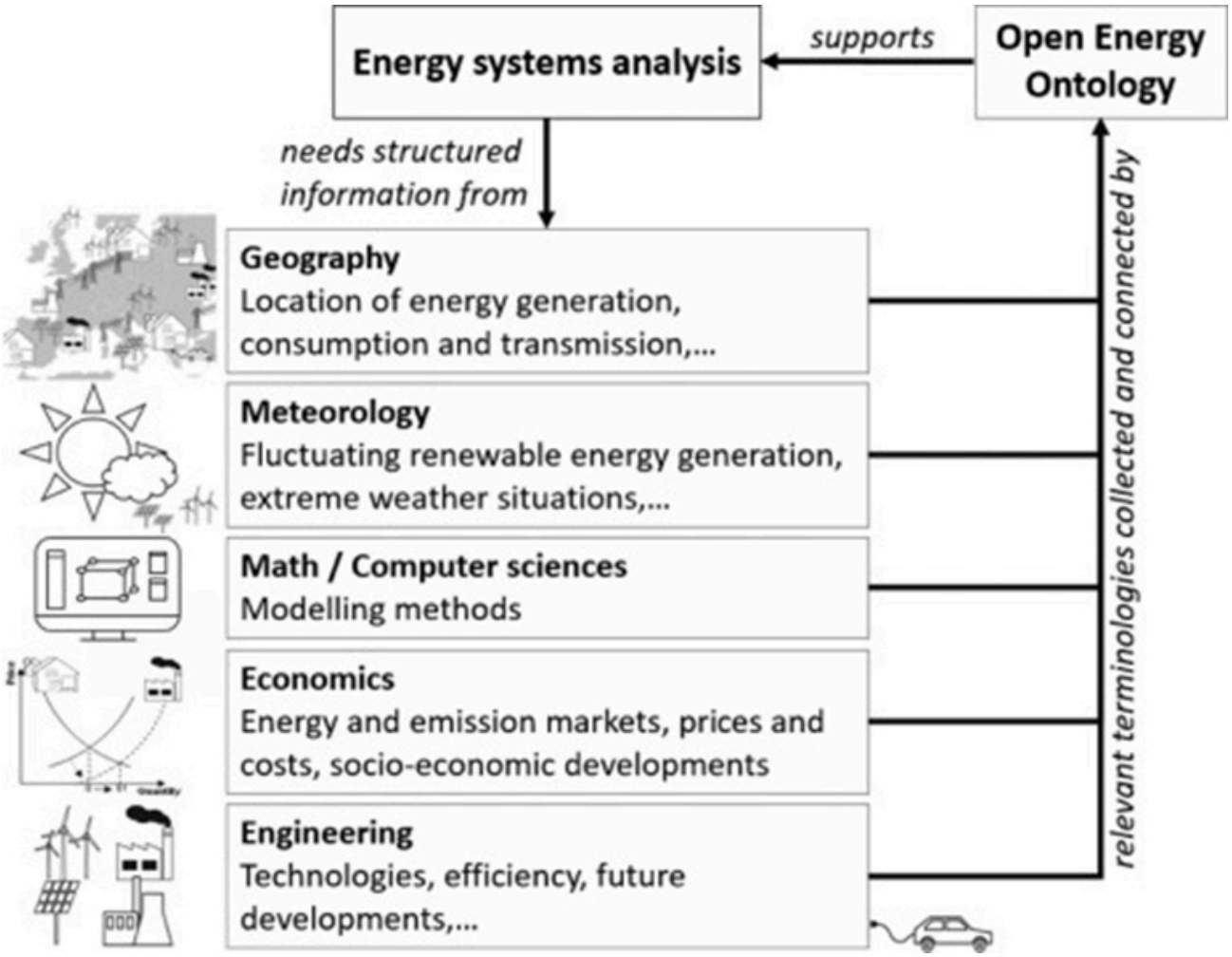

Figure 1. The OEO collects, connects and structures parts of domain terminologies relevant for energy systems analysis.

Objectives

Our goals directly address the escalating complexity and interconnectedness inherent in modern energy system modeling, as detailed previously. Contemporary requirements differ significantly from historical needs: interactions between models and researchers are intensifying; an expanding volume of varied domain knowledge and data is emerging; and computing power continues to surge. Formerly acceptable practices—like transferring files manually, hand-coding data into simulations, and linking systems through rigid, pre-established interfaces—are becoming increasingly impractical. To automate these linkages efficiently, machines must have semantic understanding of the information being processed.

Moreover, during collaborative efforts, specialists might describe identical concepts using distinct terminology unique to their respective fields. This creates significant hurdles, requiring substantial time to achieve mutual understanding before

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

productive work can commence, and forcing teams to repeatedly establish this shared understanding across successive, varied project settings. An illustrative example of such miscommunication concerns the industrial sector's final energy consumption. Models aligned with European energy statistics (Eurostat) characterize industrial final energy consumption by omitting fuels utilized for non-energetic purposes. In contrast, frameworks aligned with global metrics (e.g., IPCC) incorporate these non-energy applications into their definitions. Without an unequivocal definition of this outcome variable, tracing the divergent methodologies becomes arduous, inevitably leading to analytical confusion.

Applying an ontology can alleviate these complications. The primary aims of this article are to illustrate the substantial benefits provided by ontologies and to outline how the Open Energy Ontology has advanced the establishment of a standardized lexicon within energy system modeling, specifically targeting:

· Cross-disciplinary data comprehension

· Information representation

· Metadata tagging to render information interpretable by AI and machine systems

· Interface standardization to facilitate model integration via unambiguous interface definitions

· Automated data validation

Although the Open Energy Ontology continues to expand, it is certainly not the sole ontology operating within this sector. The relationships, integrations, and landscape of currently available ontologies are detailed in the following discussion. This discussion also highlights the innovative nature of our methodology and illustrates how it enhances scientific understanding in energy systems analysis.

What is an Ontology and What is it Good for?

In this context, ontologies are structured delineations of entities within a specific field, along with their mutual connections. This definition diverges from the philosophical sub-discipline of Ontology, which explores the nature of existence and fundamental categories of being. Unlike standard taxonomies (such as the well-known classifications of flora and fauna), vocabulary lists, or thesauri, ontologies rigorously define the associations linking different entities. This indicates that an ontology generally incorporates various overarching categories (such as buildings, houses, roofs, colors, or tilt) that interconnect. For instance, a house is a component of a roof and is situated within a village.

In contrast, a roof possesses a specific colour and tilt (here, "has part", "located in", "has colour", and "has tilt" function as relations). Furthermore, distinct instances of these classes can be established; for example, the Eiffel Tower is categorized as a building and exhibits a gray color. In this scenario, the Eiffel Tower represents an instance of the building class, while the gray represents an instance of the color class.

Operating as formalized representations of domain-specific entities, ontologies typically contain precise definitions and supply multiple benefits. Ontologies:

· Establish a unified lexicon within a specific area. This promotes seamless information exchange and eliminates ambiguity—even for automated software agents. Consequently, they simplify collaborative efforts.

· Allow investigators to navigate a domain's intricacies by providing a rigorously constructed framework of relationships and definitions. Ontologies facilitate straightforward consistency verifications.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

· Permit the recycling of domain-specific knowledge. Previous efforts need not be duplicated; they can be smoothly integrated into new initiatives.

· Decouple operational logic from domain knowledge. Procedures can function independently of the specific components they manipulate. To illustrate, a robot driving screws represents a process (operational knowledge), while the screws themselves may vary in size (process components, representing domain knowledge). Conceptually isolating these two aspects simplifies reuse when characterizing conceptually related items. Domain knowledge remains reusable without requiring insight into operational specifics, while the operational knowledge can still be distinctly modeled.

· Facilitate knowledge expansion via automated inference. Axioms act as logical statements formulated in the ontology's foundational logical syntax. Within an ontology, axioms are linked to specific classes (for instance, asserting that all trees possess trunks) and apply universally to every instance within that class. Practically, this implies that if a new dataset of trees is incorporated as instances into the ontology, a computational tool known as a reasoner can deduce that these novel tree instances must also feature trunks, thereby generating new knowledge.

· Create linkages between compartmentalized data. Institutions frequently employ proprietary terms, workflows, and data structures, resulting in what are often termed data silos. By adopting a shared ontology, information becomes straightforward to update, exchange, and transfer.

Ontologies can be categorized into two primary classifications: domain ontologies and upper-level ontologies. Domain ontologies focus on a specific segment of reality, such as energy systems. Conversely, upper-level ontologies provide classifications and relationships for highly generalized concepts, such as "process" or "object," that apply universally across domains. Prominent instances of upper-level ontologies are the Basic Formal Ontology (BFO) and the Suggested Upper Merged Ontology (SUMO). Generally, domain ontologies adopt an external upper-level ontology to form their foundational architecture, which they then expand in a domain-tailored manner. The BFO serves as the upper-level framework for the OEO .

The Web Ontology Language (OWL) is the predominant family of knowledge representation languages for constructing ontologies. OWL is built on the Resource Description Framework (RDF), a standard for defining entities and their relationships. The Protégé10 application remains a widely utilized graphical user interface for the exploration and implementation of OWL ontologies.

Ontologies in the Energy Domain

Presently, the EnArgus Ontology represents the only widely recognized terminological asset for energy systems analysis [38]. The German federal government and states utilize this ontology to equip policymakers with insights derived from energy science. It encompasses a broad spectrum of terminology gathered through semi-automated processes. The associated wiki provides a comprehensive repository of valuable definitions and terms. Nevertheless, the EnArgus Ontology currently remains proprietary, restricting the community from reusing it. Based on publicly accessible details in its wiki, we deduce that the EnArgus ontology is primarily structured as a subclass hierarchy and features minimal axiomatization (i.e., it uses only rudimentary logical statements).

Price fluctuations and energy markets are core components of numerous energy system models. The Electricity Markets Ontology centers on electricity markets [39], whereas the Financial Industry Business Ontology addresses financial markets [40].

Contemporary advancements in energy systems analysis demand a more comprehensive methodology for modeling markets that encompasses gas, heat, and alternative energy carriers, alongside the transitions linking them. Although the OEO does not currently provide an exhaustive treatment of market dynamics, future additions of relevant terminology will leverage existing ontologies where feasible, augmented by supplementary content tailored to our specific operational

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

scope. Semantic technologies have been widely used in numerous smart home systems for data integration and management. Consequently, ontologies well represent the urban development and housing sectors. For instance, the Energy Resource Ontology [41] and the SEMANCO Ontology [42] address energy-centric elements within the residential sector.

Furthermore, the SEAS ontology models various physical systems, including their properties and relationships [43], serving as a broadened adaptation of the semantic sensor network (SSN) ontology [44]. A significant portion of renewable energy resources rely on meteorological events, and the majority of climate and energy analyses require weather and climate projections to forecast the performance of these energy sources. The primary application driving the creation of the OntoWind ontology was the annotation of climate and meteorological data, along with the associated technologies [45]. Ultimately, no publicly accessible ontology currently encompasses the entire scope of energy systems analysis. The development of the OEO represents our initial step in bridging this divide.

The OEO and its General Design Choices

Ontology background, context and outline

We developed the OEO as part of the Open Energy Family, an open-source database and toolkit for managing open data in energy systems analysis research. We centered this toolkit around the Open Energy Platform (OEP). The OEP functions as a collaborative web platform supported by an underlying database focused on climate and energy analysis data. Users can upload a diverse array of data formats to this repository, including lookup tables, geographic information, and time series. The database accommodates both comprehensive energy scenarios and standalone energy datasets. Because our users release all datasets under an open license, the data becomes freely and readily available to the public. Consequently, the OEP serves as a central reference point, enhancing comparability and transparency to support political and scientific decision-making.

At present, the OEO is being developed under the SzenarienDB initiative, with support from the SIROP and LOD-GEOSS projects.

Through the SzenarienDB project, we expanded the capabilities of the Open Energy Platform, transforming it into an accessible and transparent repository for energy scenarios [46]. While scenarios form a critical element of energy systems analysis, they are inherently heterogeneous and intricate. To render these scenarios understandable and transparent, an ontology is necessary to foster shared understanding across various research disciplines.

The objective of the LOD-GEOSS project is to establish a network linking disparate databases containing both input and output data for energy systems analysis. This concept involves sharing data across decentralized databases managed directly by the data owners, allowing them to independently handle maintenance and updates. These distinct databases are linked via a metadata catalog, ensuring the data remains accessible and discoverable.

Within the SIROP project, our goal is to improve the interoperability of energy scenarios. Contrasting scenario datasets is a demanding task typically executed manually. By applying and expanding the OEO, it becomes feasible to conduct semi-automated comparisons of these energy scenarios.

Collectively, SIROP, LOD-GEOSS, and SzenarienDB apply the FAIR principles12 of open data to the realm of energy systems analysis information.

We are constructing the OEO using the Web Ontology Language (OWL). At this time, the OEO comprises approximately 870 classes. Roughly 350 of these are exclusive to the OEO. The remaining classes are imported from external ontologies. Additionally, the OEO employs 80 object properties. Approximately half were formulated internally to serve domain-specific needs, while the rest are imported. Currently, the OEO houses over 8500 axioms (logical assertions).

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

We launched version 1.0.0—the inaugural official release of the OEO—on June 11, 2020, followed by version 1.4.0 on March 02, 2021. The OEO is accessible via GitHub13, and official versions are hosted on the OEP14.

BFO, Design patterns and best practices

We organize the OEO using a common, foundational or "upper level" ontology that characterizes fundamental entity types, such as "process" and "object," which operate independently of specific domains to provide a structural scaffold. Entities specific to the energy sector are then incorporated as subclasses within this foundational framework. This methodology is standard practice across numerous scientific ontologies. To achieve this, the OEO has integrated the extensively utilized Basic Formal Ontology (BFO) [47]. The BFO distinguishes between "occurrent" entities that possess temporal segments and unfold over time (such as flows, transformations, and processes) and "continuant" entities that persist as identical individuals across time (such as devices, organisms, and objects). Within continuant entities, the BFO makes a further distinction between entities that are "independent" and those that are "dependent" upon other entities, including various attributes and qualities.

Additionally, the OEO incorporates established ontology design patterns and best practices that align with the wider scientific ontology community, specifically those endorsed by the OBO Foundry [48].

We drew our core best-practice guidelines on definitions, terminology, and taxonomy from Ref. [47].

The ontology utilizes a modular architecture. Whenever feasible, it adheres to a single asserted superclass taxonomic organization, requiring that each class within the OEO have only one parent class (monohierarchy). Where necessary, automated reasoning deduces additional superclasses via logical axioms. For instance, water is assigned the parent class "portion of matter"; however, because its axioms establish that it is renewable and functions as an energy carrier, automated reasoning successfully derives an additional parent class: "renewable energy carrier".

Every entity within the ontology receives a text definition and a distinct label. In contrast, supplementary elements such as relations to other entities, usage examples, comments, and synonyms are incorporated when necessary. We assign commonly utilized domain terminology as class labels; however, this proves challenging with particularly ambiguous terms. To avert misunderstandings, every internally generated OEO class and relation includes an explicit definition. In our classes, we employ the Aristotelian definitional structure, which involves referencing the superclass (the genus) and providing a precise articulation of the characteristics that separate the subclass members from all other members of that superclass (the differentia).

Furthermore, each entity in the ontology is assigned an alphanumeric primary identifier under the namespace OEO:x (where x is a unique integer). These identifiers are sequential and devoid of inherent semantic meaning; however, specific numerical ranges are dedicated to different ontology curators. This strategy is implemented to avoid conflicts during simultaneous editing sessions.

Structure and submodules

The OEO is structured into three primary domain-specific modules **(Figure 2)**, which address distinct facets of energy systems analysis:

a. Models and data (oeo-model),

b. Social and economic aspects (oeo-social),

c. The physical side of energy systems (oeo-physical)

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

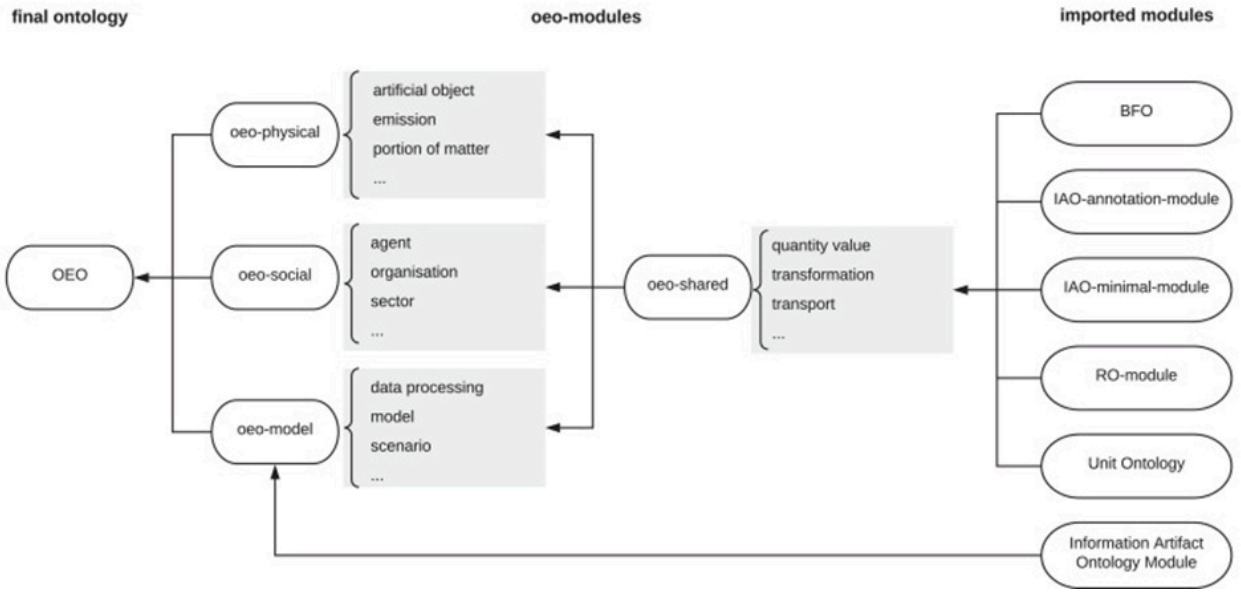

Figure 2. Submodules and imports of the OEO.

Additionally, a supplementary module (oeo-shared) exists to host classes and relations shared across multiple modules. These components are universally imported into the core ontology, which establishes connections among the distinct modules. This modular architecture was selected to facilitate maintenance; it allows separate teams to modify different files simultaneously without the risk of conflict. Consequently, this method effectively categorizes the material into coherent subdivisions across the broader domain.

The oeo-model module encompasses all elements about data and models. Beyond various model classifications, the majority of the concepts outlined here pertain either to data transformations or information entities. Examples include computational procedures and data-processing techniques used in energy system modelling. The information-centric concepts incorporated within this module primarily consist of a selected subset imported from the Information Artifact Ontology. This incorporated subset includes the “information content entity” class, along with subclasses that specify distinct types of information, including data items, documents, symbols, and figures. The unique information content entities native to the OEO are categorized as subclasses beneath these broader information frameworks. Notable examples include the scenario class, various data descriptors, and fundamental assumptions and constraints.

The oeo-social module captures social, economic, and political concepts to articulate the socio-economic dimensions of energy systems. This includes fundamental classes like “population” and “organisation”. Sectors are structured by pairing a general “sector” class with broader “sector divisions”. These “sector divisions” specify the relevant sectors for a given context. Various roles are also characterized, including “agent”, “author”, “producer”, and “user”. A critical organisational type within this field is the energy producer, represented by the “organisational energy producer” class alongside its subclasses.

Furthermore, economic concepts play a crucial role in energy systems modelling. To encompass these, we integrated the recognized Financial Industry Business Ontology (FIBO17) [40]. This framework supplies an extensive collection of entities and relationships related to economics and financial markets, which are essential for numerous energy models. Because FIBO operates independently of the BFO, we modified FIBO’s definitions and classes to integrate an appropriate BFO categorization. Consequently, FIBO material is not directly imported; rather, it serves as a foundational resource and is annotated via cross-references to indigenous OEO classes. Examples of incorporated economic terminology include “price”, “gross domestic product”, and “exchange”.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

The oeo-physical module encompasses concepts linked to the tangible components of energy systems. It categorizes foundational concepts such as energy, power, and matter, alongside technical infrastructure like power plants and batteries. Numerous entities denote tangible objects, positioning them as subclasses within the BFO's "material entity" category. Substances, materials, and fuels—such as coal, peat, water, and methane—are nested under the "portion of matter" class. We employ axioms to facilitate automated categorization grounded in logical equivalences. Consequently, these substances are grouped into distinct categories, such as greenhouse gases and fuels, based on their intrinsic properties and functions. Specifically, fuels are systematically categorized into subcategories, including biofuels, renewable fuels, and nuclear fuels. Concepts associated with greenhouse gas emissions and environmental pollution are designated as subclasses of the BFO's "process".

Additionally, the "artificial object" class includes technical equipment such as generators and batteries. Power plants are classified by their energy inputs, yielding categories such as "wind farms" and "biofuel power plants". Moreover, various forms and applications of energy, along with conversion processes such as "primary energy production" and "final energy consumption", are integrated into this module. To quantify physical entities, the OEO integrates the Unit Ontology [49]. By defining units for power and energy, the Unit Ontology effectively addresses a crucial segment of the energy systems field.

The oeo-shared module contains the concepts and relationships required across various submodules. For instance, foundational classes such as quantity values are established here.

As previously noted and illustrated in **Figure 2,** the OEO integrates segments from external ontologies to prevent redundant efforts. Beyond the BFO, we heavily leverage portions of two additional ontologies. Initially, we incorporate the Relations Ontology (RO) module, which encompasses a specific subset of object properties detailed in the Relations Ontology [50]. We opted for a partial inclusion of the RO because numerous relations do not pertain to energy systems analysis. Instances of object properties integrated via this module include "has quality" and "has disposition", alongside fundamental properties like "part of", and spatiotemporal descriptors such as "starts with" and "located in". Secondly, we incorporate all metadata annotation categories established by the Information Artifact Ontology (IAO)—such as "document" or "reference"—into the oeo-model module. The IAO also provides valuable standardized annotations, such as the "term tracker item", which we implement. This specific annotation links to GitHub issues and pull requests documenting entity modifications or definitions, fostering transparency by providing swift access to supplementary details, historical context, and the surrounding discourse for a given class. These annotations, alongside pivotal IAO classes that are beneficial across all modules, are assimilated through the oeo-shared module **(Figure 3)**. We streamline this integration by employing the ROBOT library [51] to selectively isolate necessary content.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press



Figure 3. Overview of a subset of classes and properties of the OEO to illustrate how they are organised inside the OEO. A black arrow denotes “is a”, i.e. a subclass relation.

Open Collaborative Development

The OEO adheres to the OBO principles18. Consequently, the development of the OEO is conducted through an interdisciplinary, collaborative, transparent, and open source19 methodology. The selected workflow mirrors these traits, placing a pronounced emphasis on transparency: all technical deliberations and developer assemblies occur publicly via the project’s GitHub page20. Contributions from anyone are welcome. Additionally, a steering committee composed of specialists from various allied disciplines has been formed. This committee provides overarching guidance for the OEO’s progression.

Ontology development

Generally, the construction of a domain ontology follows one of two distinct methodologies. The first entails an automated generation process that uses AI to parse text corpora (such as academic papers or Wikipedia). This method aggregates pertinent domain information and translates it into an ontological structure. The alternative methodology requires human domain specialists to manually curate and construct relevant entities, explicitly defining and connecting them within the framework. This manual method was selected for the development of the OEO. This is unavoidably a more time-consuming procedure.

Nevertheless, automated strategies often fail to accurately filter noise, assess fluctuating source quality, or resolve terminological conflicts and divergent perspectives prevalent in academic literature. Consequently, no automatically produced ontology has yet achieved widespread acceptance as a rigorous scientific reference. Conversely, human curators can detect and rectify these discrepancies throughout the development phase, enabling the creation of a coherent domain representation and a meticulously articulated vocabulary. An additional benefit of this manual curation is

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

the mitigation of inadvertent biases inherent in certain data repositories. By doing so, the resulting domain ontology can be safely used by subsequent AI applications without perpetuating undesirable biases.

Frameworks like the OEO are engineered to benefit the scientific community. Their construction relies heavily on specific workflows, technological tools, and standards that foster collaborative growth. Numerous developmental paradigms for ontologies have been introduced (e.g., Refs. [52-55]). These paradigms often mirror the protocols used in open-source software engineering: they aim to make the development process both dependable and reproducible, maintaining a stringent focus on quality assurance throughout development. As highlighted by the guidelines presented in a brief publication detailing “ten simple rules” for constructing ontologies [56], a paramount principle of robust ontology design is maximizing the reuse of pre-existing content. This strategy facilitates the progressive expansion of current knowledge bases and circumvents redundant labor. Therefore, the OEO was structurally designed to integrate applicable external content wherever feasible.

Establishing explicit version-control protocols and adopting an ‘open’ license are recognized as fundamental components of these methodological guidelines (e.g., the OBO Foundry Principles [48]). Moreover, these strategies routinely include guidance on defining the ontology’s boundaries and conducting evaluations. Such assessments must be conducted transparently, early, and regularly. Ultimately, these guidelines advocate for active community participation and the thorough documentation of structural design patterns.

Adherence to universal standards is exceptionally critical for promoting the reuse and cooperative sharing of ontological materials across disparate domains and communities. To aid in the formulation of these shared standards, the OBO Foundry [48] was established as a pioneering effort within the biological and biomedical sectors. This initiative has unified ontology creators to construct a framework of design principles and standards capable of semi-automated validation. Furthermore, these standardized protocols have enabled the creation of utilities like the ROBOT ontology library [51], which automates numerous routine developmental procedures. Although the OEO targets an entirely different field, a significant portion of the developmental standards we integrated are rooted in the OBO Foundry’s specifications. For instance, we leverage the Foundry’s established metadata criteria and standardized relationships.

Git workflow

The OEO is developed in a public setting: all source code and related discussions are hosted openly on GitHub. To facilitate the onboarding of new users and collaborators, we provide comprehensive manuals detailing usage 21 and contribution guidelines 22. Documenting this workflow guarantees accountability and maintains the quality of our decision-making. Specifically, our process requires that any proposed modification to the ontology be deliberated in a GitHub issue before implementation. We classify these issues into four distinct categories:

- Adding a new entity
- Restructuring existing parts
- Updating definitions of existing entities
- Other

Minor revisions require the consensus of at least two development team members, while major alterations require approval from at least three. These endorsing members must comprise at least one domain specialist and one formal ontology expert. To accommodate the varied expertise of the OEO development team and to promote swift collaboration when addressing issues, developers are organized into GitHub teams based on their specific specialties. At present, these teams cover the following areas: economy, modelling, linked open data, meteorology, and formal ontology. If a consensus proves difficult to achieve through GitHub issue threads, the topic is escalated to the agenda of the

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

subsequent developer assembly. These assemblies typically occur as virtual meetings. Alongside the domain-specific teams, a dedicated team handles new software releases.

By design, our developmental approach is gradual but rigorous. Once an issue resolution is agreed upon, any team member can rapidly implement the technical modifications following a standardized protocol. This procedural workflow is supported by a suite of automated tests that meet specific quality benchmarks. These evaluations encompass syntactic validation, along with the deployment of an automated reasoner to verify logical consistency. Furthermore, the ROBOT24 OBO-foundry tool and the OOPS23 pitfall scanner are utilized to ensure comprehensive quality assurance.

Community embedding

We complement our GitHub-based workflow with virtual developer assemblies. During these sessions, we evaluate ongoing progress and address complex technical issues. By scheduling these gatherings as monthly *jour-fixes*, we establish a consistent developmental rhythm. At present, members from the LOD-GEOSS, SzenarienDB, and SIROP research initiatives coordinate and prepare these meetings.

In scenarios where an agreement regarding a specific issue cannot be reached—neither on GitHub nor during the developer assemblies—the matter, along with prospective solutions, is forwarded to the OEO-Steering Committee (OEO-SC). The OEO-SC reviews the situation and issues a final ruling. Consequently, the OEO-SC helps guide the project's strategic trajectory. In addition to conflict resolution, the OEO-SC is tasked with promoting awareness of the ontology and encouraging its integration into ongoing and future projects. The committee gathers roughly every three months. To guarantee broad recognition of both the committee and the OEO itself, the OEO-SC is composed of seasoned professionals from diverse organizational and domain-specific backgrounds, each possessing extensive experience in their respective fields.

To foster a respectful and collaborative environment among all OEO contributors, we adhere to a self-imposed code of conduct. Grounded in the principles of non-violent communication, this code aligns seamlessly with established GitHub community guidelines. Although the technical content of the OEO is inherently neutral, implementing this code of conduct helps maintain focus on the underlying issues and prevents contentious debates that could alienate or harm participants.

To date, the OEO has been presented to hundreds of researchers within the domain. Specifically, we introduced the framework to the international OpenMod network, which boasts roughly 550 registered members, as well as the Forschungsnetzwerke-Energie (FNE), which includes over 250 participants across Germany. Currently, more than 350 registered users on the Open Energy Platform are actively engaged in the OEO's ongoing development.

Testing and continuous integration

Given the expansive and varied nature of our contributor base, routine verifications for consistency and coherence are essential. Consequently, we have established multiple automated and semi-automated testing protocols. Protégé serves as the primary development environment for the OEO, and we use the OWL reasoners integrated into Protégé to ensure structural consistency. The Ontology Pitfall Scanner [57] outlines a series of common errors encountered during ontology construction, such as omitted annotations or a lack of naming conventions. We manually apply OOPS! To ensure that official ontology releases comply with these structural rules. As noted previously, the ROBOT library plays a critical role across the ontology's lifecycle, including module extraction. It also functions as a core component of our continuous integration and automated testing pipeline, verifying the ontology against distinct OWL profiles25 and executing numerous quality assessments, including coherence and consistency evaluations26. While these validations ensure the ontology's overall structural integrity, they remain domain-agnostic.

To address this, we formulated several competency questions to verify that the ontology's entities align with their designated semantic meanings. Every new contribution to the OEO is systematically evaluated against consistency metrics, competency questions, and the defined ROBOT profiles.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Evaluation

In this segment, we assess three distinct dimensions of the OEO. First, we examine its thematic coverage within the target domain. Second, we appraise the caliber of its axiomatisation utilizing competency questions. Finally, we analyze the quality of the natural-language definitions for OEO terms using an inter-annotator agreement study. These evaluative measures were directly informed by the specific demands of our practical use cases.

Evaluation I: Coverage study

Our initial assessment investigates whether the OEO encompasses the necessary terminology for a standard application scenario. One primary intended application is the annotation of various databases and fact sheets. To conduct this coverage analysis, we utilized scenario fact sheets currently under development in the SzenarienDB project. These sheets serve to characterize energy scenarios when the corresponding data is uploaded to the OEP. The fact sheets capture metadata (such as authors, title, license, and publication format) alongside the spatial and temporal scope of the energy models. Detailed modeling parameters are recorded via distinct fields representing energy flows, fuels, demand sectors, energy sectors, and environmental impacts. Additionally, macroeconomic indicators like gross domestic product, energy prices, and population statistics are included.

We leveraged the field labels from these fact sheet forms to conduct a semi-automated entity annotation process. During the initial phase, an automated system retrieved five potential OEO entity candidates for each fact sheet label. This selection was based on string similarity, utilizing a methodology that combined word tokenisation, a soft Jaccard index applied to the token sets, and a Levenshtein distance metric to moderate the Jaccard index [58]. In the subsequent phase, a team of ontology developers manually identified the appropriate entity, or combination thereof, from the generated candidates. They also pinpointed relevant OEO entities that the automated system failed to detect. For this evaluation, we excluded broad, fallback fact sheet fields (e.g., Other Fuels), as the ontology intentionally omits them. Integrating such vague fallback terms is widely regarded as poor ontology design; for annotation contexts, identical semantic expressions can be formally constructed by combining a parent class (e.g., Fuel) with the complement of specific subclasses (e.g., not Fossil Fuel). Furthermore, ontology properties were omitted from this analysis.

To measure the extent to which OEO concepts encompassed a fact sheet entity, we employed a three-tier rating system: *No match* denotes that the OEO currently lacks the requisite entities to annotate the specific field. *A partial match indicates that the field can be incompletely described using one or more* OEO entities. For instance, “costs of coal” could only achieve a partial match because, at the time of the evaluation, “(portion of) coal” was present in the OEO, whereas “costs” was absent. Finally, *a good match* indicates a comprehensive and exact alignment.

The findings of this coverage assessment are detailed in **Table 1** and have been published for public access27. As indicated in the initial row (“ALL”), the evaluation tested the annotation of 153 total fact sheet fields. Over half of these fields (52%) demonstrated a good match, while 20% yielded no match and remain undescribed by the current OEO iteration.

Table 1. OEO coverage for scenario fact sheet field names measured for ALL evaluated field names, and for a subset excluding socio-economic related fields (ESE).

Empty cell	# Fields	Good match	Partial match	No match	Matches combined
ALL counts/ratio	153	79/52%	43/28%	31/20%	122/80%
ESE counts/ratio	107	65/60%	21/20%	21/20%	86/80%

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Approximately 30% of the evaluated fact sheet fields (46 fields) correspond to the domain's socio-economic dimensions. These include metrics such as GDP, population statistics, or the financial costs of emissions and fuels. The OEO is divided into three distinct modules. Historically, developmental efforts have concentrated predominantly on the oeo-physical module, with the remaining modules slated for expansion in upcoming iterations. Consequently, the remaining sections, particularly the oeo-social module, are still in premature stages of development and lack comprehensive coverage.

To account for this discrepancy, the second row of the table ("ESE") isolates the 107 fields that do not pertain to socio-economic factors. Within this subset, roughly 60% of the concepts had a good match, while 20% lacked a match entirely. By comparing the broader "ALL" dataset with the focused "ESE" subset, it becomes apparent that only 14 socio-economic fields (30%) achieved a good match. Because our immediate developmental roadmap prioritizes expanding the socio-economic module, we anticipate substantial enhancements in socio-economic coverage in the near future.

Evaluation II: Competency questions

Competency questions offer a structured approach for defining and validating the semantic prerequisites of an ontology [59]. Initially, domain experts collaborate with ontology developers to draft practical use cases, explicitly detailing the types of inquiries the ontology must be able to resolve in those contexts. A specific scenario, paired with a related question and its expected answer, establishes a proof obligation: the formal representation of the given scenario, when combined with the ontology's underlying axioms, must logically deduce the formal representation of the correct answer. Automated theorem provers can be utilized to mechanically validate these proof obligations.

Competency questions prove exceptionally beneficial when constructing ontologies intended for specific roles within broader information architectures. In such instances, the operational scenarios are strictly defined and limited; therefore, formulating these questions and scenarios can actively guide the entire engineering workflow [60]. For these targeted projects, competency questions serve as a metric for completeness: if the system successfully processes every competency question, all documented requirements have been fulfilled, signifying a successful developmental process.

Conversely, foundational frameworks like the OEO function as shared reference terminologies for extensive communities. Because the OEO is not tailored to a singular application context or a rigid set of operational requirements, traditional metrics of "completeness" assessed via competency questions do not apply. Nevertheless, we found that competency questions are highly effective for conducting semantic evaluations, as they verify that the ontology's mathematical axioms accurately reflect domain experts' intentions. Certain competency questions are designed to capture the established consensus regarding highly debated or ambiguous terminology. By codifying these agreements, competency questions allow us to immediately flag any future ontological modifications that contradict previously established consensus. This targeted, domain-specific semantic review effectively supplements the broader coherence and consistency checks.

For instance, accurately mapping the concept of fuel posed a significant challenge during the OEO's development. The resulting design resolutions were translated into formal OWL competency questions. One such query asks, "Is charcoal an energy carrier that is solid under normal conditions?" We then deploy the HermiT reasoner to verify the logical entailment between the proposed questions and the ontology's structure. To ensure that subsequent modifications do not disrupt these critical logical pathways, this validation procedure has been integrated directly into our continuous integration pipeline.

At present, the system features 50 established competency questions28; of these, 41 are resolved successfully. The remaining nine queries currently fail to achieve logical entailment due to an absence of specific axioms and entities. We plan to systematically expand the ontology in upcoming iterations to support these pending inferences.

Evaluation III: Inter-annotator agreement study

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

The definitions and classes within an ontology must be unambiguous and comprehensible. When annotating resources with ontological terms to enhance query functionality and findability, these terms must be applied consistently across different annotators. Consequently, one method for evaluating an ontology involves having users annotate text segments with ontological terms and subsequently quantifying the degree of consensus among their choices [61]. To this end, we utilized five text fragments derived from model fact sheets to examine the consistency of energy domain experts. We specifically selected fragments where the appropriate ontological term was not immediately apparent—meaning there was no precise correspondence between the text and a single label, but rather several terms that were only loosely applicable. This required the experts to analyze and interpret the term definitions to complete the task.

For each fragment, we selected six potential ontological entities using the same manual refinement and string similarity methodology. Along with the text fragment, annotators were presented with these six entity definitions and a seventh option: “None of the above.” Invitations were distributed via email to researchers at various institutes focused on energy systems analysis. Participants in this study possessed no prior experience with the OEO.

Of the 34 invited participants, 20 completed the full survey, and only their data is included in this analysis. Among these 20, 17 had at least 1 year of experience in energy systems modelling, while 2 had prior familiarity with ontologies. The questions and responses are publicly accessible.

To quantify inter-annotator agreement, we utilized an extension of the kappa coefficient specifically designed for multiple-choice setups with multiple annotators, as formulated by Kraemer [62].

Using this metric, the agreement calculated for our study was $\kappa = 0.668$. Based on the classification criteria provided in Ref. [63], this represents a “substantial” degree of inter-annotator agreement. Furthermore, the “None of the above” option was rarely selected, indicating that our candidate entity pool provided reasonable coverage for annotating the provided text samples.

Nevertheless, there remains room for improvement in agreement. Notably, participants frequently disregarded our instructions to select only the best-fit match, opting instead to include broader matches. For instance, if a participant selected “greenhouse gas emission,” they often additionally selected “greenhouse gas.” This redundant secondary annotation is unnecessary, as the ontology already includes an axiom stating: “Greenhouse gas emissions involve the emission of some greenhouse gas.” While such redundancy typically does not hinder practical applications, it complicated the assessment of true agreement regarding the optimal match in this evaluation.

Participants also noted that for certain fragments, the provided options lacked a term that optimally described the content, making it difficult to determine a definitive “best” match. Thus, the domain coverage gaps previously identified in the first evaluation negatively impacted the results of this inter-annotator agreement study.

We are currently revising the OEO in light of these findings. A primary objective is to expand the ontology’s coverage to ensure it encompasses all terminology needed to describe energy scenarios and models. Improving the documentation of these entities is equally critical. While our initial efforts focused on creating logically correct and ontologically sound definitions, we must now incorporate additional examples, synonyms, and explanations to improve inter-annotator agreement.

Use Cases

Beyond the OEO’s broad range of potential applications—including tagging, user categorization, data set summarization, and semantic search—we present four specific use cases where the OEO is currently being applied. All current applications originate from the SIROP, LOD-GEOSS, and SzenarienDB projects.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Implementation: Scenario description

The coverage assessment was conducted using a dataset derived from existing scenario fact sheets, motivated by the inherently heterogeneous structure of energy research data. Published results, the underlying datasets, and the explicit or implicit assumptions involved are frequently only loosely connected, which undermines the transparency and reproducibility of scientific results within the energy systems modelling domain [64]. There is a clear need for a framework that enables the annotation of these studies and scenarios, along with their related elements, such as the models or datasets that support the results. The OEO defines the foundational structures necessary to construct such a system. Several gaps identified during this analysis have been addressed, allowing for more comprehensive annotation of components and their scenarios. A collection of properties for energy studies and energy scenarios—covering the essential information required to adequately describe their relations and characteristics—was compiled into spreadsheets and populated with data. An RDF knowledge graph, utilizing the relations and classes defined in the OEO alongside other prominent vocabularies (such as FOAF or Dublin Core), was manually created using these spreadsheets. Furthermore, the Open Energy Platform was enhanced with a new section that enables users to edit and view data stored in the knowledge graph. This empowers researchers to increase the public availability of their work and promotes a more structured landscape for energy models, datasets, and scientific findings.

Data representation for the core energy market data register

As of 2019, the German Federal Network Agency (Bundesnetzagentur) publishes the core energy market data register, known in German as MaStR (Marktstammdatenregister). This is a comprehensive list of every power station in Germany that is connected to the grid, spanning from small private solar modules and water plants to large lignite and coal facilities. These data are released under an open license30 and are accessible via an application programming interface (API). While this represents improved accessibility compared to traditional file downloads, the API documentation31 is notably complex. The SOAP protocol used lacks a standardized interaction model, meaning any data access must be configured manually and is functional only with the MaStR API. This creates an accessibility barrier, which is further exacerbated by limitations on the number of data points that can be requested per session. This threshold is restricted to 10,000 requests per user per day32. Consequently, for many users, data access is not truly barrier-free. Setting aside conceptual and technical challenges, providing accessible, open data is valuable because it facilitates reproducible research. Therefore, we developed an open-source tool 33 to extract data from the MaStR interface, enhanced the output with metadata, and made it available on the Open Energy Platform 34. Here, users can download the data without registration requirements. We intend to maintain regular updates using these scripts. The OEO covers numerous entities essential for describing the domain of energy market data present in MaStR, such as spatial regions and power plants. By annotating the MaStR dataset with these entities, we enable conceptual queries that are independent of the data's specific representation and closer to natural language.

For instance, a user wishing to aggregate data for every power station would typically have to query individual endpoints of the MaStR-API (e.g., GetAnlageEegSolar, GetAnlageEegWind, ...). Such complex queries can be simplified by mapping the API to an endpoint for the SPARQL Protocol35 using the terms defined within the ontology. This allows for significantly simpler queries based on commonly used, consensus-defined terminology. Data regarding power plants can be retrieved via a simple SPARQL query such as SELECT? answer WHERE {?answer a oeo: powerplant}. The logical foundations of the OEO facilitate not only these definitions but also the enrichment of the dataset with logical dependencies. For example, one might limit the previous query to geothermal power plants by utilizing the geothermal power plant class, or query all power plants that possess a part defined as a geothermal power unit. This flexibility enables the creation of a versatile data interface that supports complex data aggregation while remaining intuitive and accessible.

Data annotation of an energy meteorological time series data set

Time series of various types serve both as input and output data for energy system models. A common application involves energy meteorological time series: weather time series are ingested into a model, which then calculates time

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

series for weather-dependent generation, such as solar collectors or wind farms. However, achieving consistent and complete annotation is challenging; ambiguous or missing information can result in errors when interpreting or reusing the data.

There exist multiple equivalent ways to describe the contents of a time series, and substantial effort is often spent adapting or discovering the definition of a specific energy meteorological time series, as annotation habits differ across domains. While some disciplines maintain conventions, they are not always strictly followed. A time series is defined as a sequence of data points—representing modelled or measured values—referenced against a set of time intervals or points (time steps). These time steps are defined by either: a) a starting and ending time, or b) a time stamp plus the time step length, along with the alignment of the time stamp within that step (indicating whether the time stamp refers to the beginning, middle, or end of the interval). Energy system models utilize time series from multiple disciplines, including economic time series such as energy prices, meteorological data, and consumption or power generation time series. Consequently, there is no unified method for defining time series. For example, in meteorology, it is common practice to use the end of an interval as a timestamp, since data loggers frequently record the time when data collection concludes. However, this is not mandatory and using meteorological time series often requires guessing whether the provider adhered to this convention. While solar data can sometimes be verified by comparing it to solar geometry, such verification is difficult for other types of time series. Climate science does provide the climate conventions36 to assist in this area.

Additionally, the type of aggregation performed on the modelled or measured value (e.g., integrated, averaged, or instantaneous values) for each time step must be annotated. Within climate conventions, this is handled through the cell_method and bounds attributes. For instance, temperatures and wind speeds are typically recorded as averages over an interval. Rain is recorded as an integrated water amount on a specific area (l / m^2), and solar irradiation is recorded as integrated energy in kWh / m^2. Wind gusts may represent maximum values within an interval. In climate datasets, solar radiation is also frequently represented as an averaged rate—namely, irradiance in W/m^2. As a general rule, values describing a rate (like power) are averages, whereas values describing an amount (like energy) are integrated. Explicit annotation is required for correct interpretation. While values in W/m^2 and Wh/m^2 for solar radiation might be identical for typical one-hour time steps, they are conceptually distinct and require different treatments during data processing.

Another common source of misinterpretation in time series data involves daylight saving and time zones within temporal metadata. Omitting time zone information can sometimes imply UTC, and other times local standard time; these must be explicitly stated in the data documentation. **Figure 4** illustrates the annotation structure for such time series.

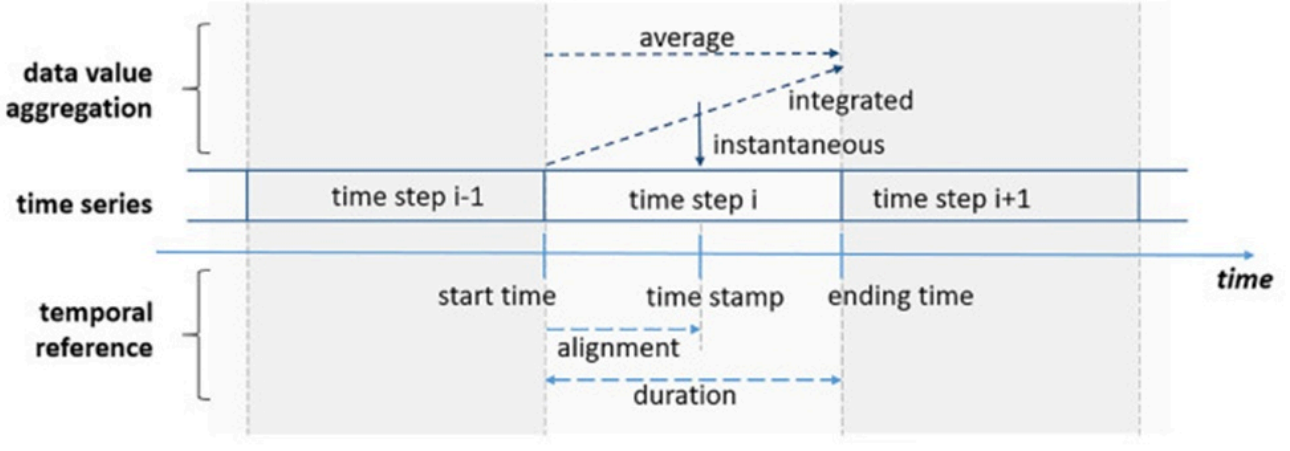

Figure 4. Overview of time series annotation structure.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

When time series are shared within the energy systems modelling community, identifying and recording these specifications currently requires significant effort. The OEO streamlines this process by providing unambiguous, concise definitions for various annotation concepts. This ultimately enables a comprehensive description and effortless interdisciplinary identification of the content and structure of energy, meteorological and other time series.

Interface homogenisation of the FINE energy system model framework

This use case targets the direct linkage of the distributed database architecture introduced with an energy system model. Our goal is to utilise the OEO to standardise the description of data inventories and the functional parameters anticipated or delivered by model interfaces, so that distinct correspondences can be established. At the same time, we mitigate variability in interface specifications, thereby reducing the workload for both developers and users when creating and interpreting them. We are currently examining the interfaces of multiple recognised energy system models across different categories to ensure comprehensive inclusion of the most critical data classes.

The FINE Framework, for example, is an open-source Python package37 that provides functionality for modelling, optimisation and analysis of high-resolution energy system models across time, space and technology [65]. Its four most important component classes, which model source/sink, conversion, transmission and storage technologies, are characterised by approx. 40 different attributes each. All of these attributes must be initialised using static parameters or multidimensional data series before model calculations can be carried out. Building on the currently available interface description38, which labels and specifies the individual function parameters, we are investigating the extent to which alignment with the OEO terminology already exists and identifying where the interface or the ontology requires modification. Through these concrete model applications, we seek to establish best practices for homogenising the linkage of data to models and, eventually, the exchange of data among models, while advancing scientific dialogue within the international energy system community.

Conclusion

We have presented the ongoing development of an open, community-driven ontology for energy systems analysis: the Open Energy Ontology (OEO). Although ontologies are not entirely new to this field, many earlier initiatives focused either on a narrow sub-domain or were built as closed, proprietary resources with limited open availability. Within energy systems analysis, beyond the practical gains in reusability and reproducibility, openness carries significant implications for transparency, trustworthiness, and accountability. Teams behind open data platforms such as Renewables Ninja39, Open Power System Data40 and the Open Energy Platform41 increasingly strive to enable the community to openly exchange data, harmonise models, and collaborate. The OEO will further support these efforts. Transparency and trust become even more crucial against the backdrop of the escalating climate crisis, since modelling outputs may inform decision-making processes where strong sentiments surround certain options. Robust, reproducible evidence that can be synthesised and contrasted across various modelling paradigms and stakeholder constituencies is essential.

We have observed rapid growth in both the volume and complexity of data in the energy systems analysis domain. As data-driven techniques like machine learning, big data, and AI become increasingly significant, machine-actionable data annotation emerges as a fundamental enabler for their broader adoption. The application and expanding scope of the Open Energy Ontology will help harness the increasing body of publicly available energy system data.

An initial assessment demonstrates that the OEO offers a suitable approach for improving data annotation in energy systems analysis.

The OEO remains a work in progress, yet it already delivers value in several promising applications. Creating an ontology for a particular domain constitutes a consensus-finding exercise within that community, which not only broadens and refines a collective understanding of connected concepts but also cultivates a shared notion of what qualifies as valid data. Once a consensus emerges regarding the criteria that datasets must satisfy to be deemed correct, this knowledge

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

can be incorporated into the OEO and leveraged for automated data verification. Consequently, continued ontology development ought to rely on extensive engagement from across the domain. We trust that our introduction of the ontology in this article will encourage others to participate in this developmental undertaking and to begin employing the ontology for dataset annotation, thereby enabling simpler sharing and reuse.

The development workflow is coordinated on GitHub. It commenced under the SzenarienDB project, which concluded in March 2021, and will be sustained through the LOD-GEOSS and SIROP projects. Additional contributing projects are already in preparation, and several partners will draw on their institutional foundations to secure the long-term viability and persistence of this initiative.

We regard the OEO as a foundation for strengthening the interplay of diverse models and methods, elevating the modelling of future energy systems to a higher plane by facilitating flexible model coupling via precisely specified data interfaces, as an element of a burgeoning data ecosystem. By interconnecting distinct modelling approaches via defined data interfaces, we can better capture the complexity of the energy system transition. The further we advance along the path toward sustainable energy systems, the greater the need to account for the nuances and interdependencies of the transformation, which demands cooperation among models and methods within the energy systems analysis domain.

Acknowledgements

None

Conflict of interest

None

Financial support

This work was supported by grants from the Federal Ministry for Economic Affairs and Energy of Germany (BMWi) for the projects SzenarienDB (03ET4057A–D), LOD-GEOSS (03EI1005A–G) and SIROP (03EI1035A–D).

Ethics statement

None

Received: 05 Aug 2024 Revised: 10 Oct 2024 Accepted: 27 Oct 2024
Published online: 15 June 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

1. Pfenniger S, Hawkes A, Keirstead J. Energy systems modelling for twenty-first centy energy challenges. Renewable Sustainable Energy Rev. 2014;33:74-86.

2. Dantzig G. Linear programming and extensions. Princeton: Princeton University Press; 1965.

3. Fischbone L, Abilock H. MARKAL, A linear programming model for energy systems analysis: technical description of the bnl version. Int J Energy Res. 1981;5:353-75.

4. Schrattenholzer L. The energy supply model MESSAGE. 1981. http://webarchive.iiasa.ac.at/Publications/Documents/RR-81-031.pdf.

5. Hourcade J, Jaccard M, Bataille C, Ghersi F. Hybrid modelling: new answers to old challenges. Energy. 2006;2:1-12.

6. Manne A, Wene C. Markal-macro: a linked model for energy-economy analysis. Technical Report BNL-47161, Brookhaven National Lab, Upton, NY (United States); 1992.

7. Messner S, Schrattenholzer L. MESSAGE-MACRO: Linking an energy supply model with a macroeconomic modules and solving it iteratively. New Horiz Environ Econ. 2000;25:267-82.

8. Scholz Y, Gils HC, Pietzcker RC. Application of a high-detail energy system model to derive power sector characteristics at high wind and solar shares. Energy Econ. 2017;64:568-82. https://www.sciencedirect.com/science/article/pii/S0140988316301682.

9. Brown T, Hörsch J, Schlachtberger D. Pypsa: python for power system analysis. J Open Res Softw. 2018;6(4). https://doi.org/10.5334/jors.188.

10. Fraunhofer IWES. SCOPE: Sektorenübergreifende Einsatz- und Ausbauoptimierung für Analysen des zukünftigen Energieversorgungssystems. 2016. http://www.energiesystemtechnik.iwes.fraunhofer.de/content/dam/iwes-neu/energiesystemtechnik/de/Dokumente/-Broschueren/2015_F_SCOPE_web.pdf.

11. Weimer-Jehle W, Buchgeister J, Hauser W, Kosow H, Naegler T, Poganietz WR, et al. Context scenarios and their usage for the construction of sociotechnical energy scenarios. Energy. 2016;111:956-70. https://www.sciencedirect.com/science/article/abs/pii/S0360544216306909.

12. Deissenroth M, Klein M, Nienhaus K, Reeg M. Assessing the plurality of actors and policy interactions: agent-based modelling of renewable energy market integration. Complexity. 2017;2017. https://www.hindawi.com/journals/complexity/2017/7494313/.

13. Klein M, Frey UJ, Reeg M. Models within models - agent-based modelling and simulation in energy systems analysis. J Artif Soc Soc Simul. 2019;22:6. https://doi.org/10.18564/jasss.412.

14. Frey UJ, Klein M, Deissenroth M. Modelling complex investment decisions in germany for renewables with different machine learning algorithms. Environ Modell Softw. 2019;118:61-75. https://www.sciencedirect.com/science/article/abs/pii/S1364815218304353.

15. Jin D, Ocone R, Jiao K, Xuan J. Energy and AI. Energy AI. 2020;1:100002. https://doi.org/10.1016/j.egyai.2020.100002.

16. Azuatalam D, Lee WL, de Nijs F, Liebman A. Reinforcement learning for whole-building hvac control and demand response. Energy AI. 2020;2:100020. https://doi.org/10.1016/j.egyai.2020.100020.

17. Dorokhova M, Ballif C, Wyrsch N. Rule-based scheduling of air conditioning using occupancy forecasting. Energy AI. 2020;2:100022. https://doi.org/10.1016/j.egyai.2020.100022.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

18. Kaur R, Schaye C, Thompson K, Yee DC, Zilz R, Sreenivas R, et al. Machine learning and price-based load scheduling for an optimal iot control in the smart and frugal home. Energy AI. 2021;3:100042. https://doi.org/10.1016/j.egyai.2020.100042.

19. Lissa P, Deane C, Schukat M, Seri F, Keane M, Barrett E, et al. Deep reinforcement learning for home energy management system control. Energy AI. 2021;3:100043. https://doi.org/10.1016/j.egyai.2020.100043.

20. Solano J, Olivieri L, Caamaño-Martín E. Assessing the potential of pv hybrid systems to cover hvac loads in a grid-connected residential building through intelligent control. Appl Energy. 2017;206:249-66. https://doi.org/10.1016/j.apenergy.2017.08.188.

21. Westermann P, Evins R. Using bayesian deep learning approaches for uncertainty-aware building energy surrogate models. Energy AI. 2021;3:100039. https://doi.org/10.1016/j.egyai.2020.100039.

22. Boretti A. Integration of solar thermal and photovoltaic, wind, and battery energy storage through ai in neom city. Energy AI. 2021;3:100038. https://doi.org/10.1016/j.egyai.2020.100038.

23. Devanand A, Karmakar G, Krdzavac N, Rigo-Mariani R, Foo Eddy YS, Karimi IA, et al. Ontopowsys: a power system ontology for cross domain interactions in an eco industrial park. Energy AI. 2020;1:100008. https://doi.org/10.1016/j.egyai.2020.100008.

24. Luo X, Oyedele LO, Akinade OO, Ajayi AO. Two-stage capacity optimization approach of multi-energy system considering its optimal operation. Energy AI. 2020;1:100005. https://doi.org/10.1016/j.egyai.2020.100005.

25. Frendo O, Graf J, Gaertner N, Stuckenschmidt H. Data-driven smart charging for heterogeneous electric vehicle fleets. Energy AI. 2020;1:100007. https://doi.org/10.1016/j.egyai.2020.100007.

26. Biswas MR, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. Energy. 2016;117:84-92. https://doi.org/10.1016/j.energy.2016.10.066.

27. Chen S, Ren Y, Friedrich D, Yu Z, Yu J. Sensitivity analysis to reduce duplicated features in ann training for district heat demand prediction. Energy AI. 2020;2:100028. https://doi.org/10.1016/j.egyai.2020.100028.

28. Chou JS, Bui DK. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build. 2014;82:437-46. https://doi.org/10.1016/j.enbuild.2014.07.036.

29. Luo X, Oyedele LO, Ajayi AO, Akinade OO, Delgado JMD, Owolabi HA, et al. Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings. Energy AI. 2020;2:100015. https://doi.org/10.1016/j.egyai.2020.100015.

30. Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable Sustainable Energy Rev. 2015;50:1352-72. https://doi.org/10.1016/j.rser.2015.04.065.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

31. Ma J, Cheng JC. Estimation of the building energy use intensity in the urban scale by integrating gis and big data technology. Appl Energy. 2016;183:182-92. https://doi.org/10.1016/j.apenergy.2016.08.079.

32. Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T, et al. Big-data for building energy performance: lessons from assembling a very large national database of building energy use. Appl Energy. 2015;140:85-93. https://doi.org/10.1016/j.apenergy.2014.11.042.

33. vom Scheidt F, Medinová H, Ludwig N, Richter B, Staudt P, Weinhardt C, et al. Data analytics in the electricity sector - a quantitative and qualitative literature review. Energy AI. 2020;1:100009. https://doi.org/10.1016/j.egyai.2020.100009.

34. Kendall E, Patel-Schneider P, McGuinness D, Bao J. OWL 2 web ontology language quick reference guide (second edition). Tech. Rep., W3C; 2012. https://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/.

35. Hill DP, Adams N, Bada M, Batchelor C, Berardini TZ, Dietze H, et al. Dovetailing biology and chemistry: integrating the gene ontology with the chebi chemical ontology. BMC Genomics. 2013;14:513. https://doi.org/10.1186/1471-2164-14-513.

36. Hastings J, Glauer M, Memariani A, Neuhaus F, Mossakowski T. Learning chemistry: exploring the suitability of machine learning for the task of structure-based chemical ontology classification. J Cheminform. 2021;13(1):1-20.

37. Devanand A, Karmakar G, Krdzavac N, Rigo-Mariani R, Foo Eddy YS, Karimi IA, et al. Ontopowsys: a power system ontology for cross domain interactions in an eco industrial park. Energy AI. 2020;100008.

38. Oppermann L, Hirzel S, Güldner A, Heiwolt K, Krassowski J, Schade U, et al. Finding and analysing energy research funding data: The EnArgus system. Energy AI. 2021;5:100070. https://doi.org/10.1016/j.egyai.2021.100070.

39. Santos G, Pinto T, Vale Z, Praça I, Morais H. Electricity markets ontology to support MASCEM's simulations. In: International Conference on Practical Applications of Agents and Multi-Agent Systems. Springer; 2016:393-404.

40. Bennett M. The financial industry business ontology: best practice for big data. J Bank Regul. 2013;14(3–4):255-68.

41. Kofler MJ, Reinisch C, Kastner W. A semantic representation of energy-related information in future smart homes. Energy Build. 2012;47:169-79.

42. Madrazo L, Sicilia A, Gamboa G. Semanco: Semantic tools for carbon reduction in urban planning. In: Proc. 9th Europ. Conf. Product & Process Modelling; 2012.

43. Lefrançois M. Planned ETSI SAREF Extensions based on the W3C&OGC SOSA/SSN-compatible SEAS Ontology Patterns. In: Proceedings of Workshop on Semantic Interoperability and Standardization in the IoT, SIS-IoT; 2017.

44. Compton M, Barnaghi P, Bermudez L, García-Castro R, Corcho O, Cox S, et al. The SSN ontology of the W3C semantic sensor network incubator group. J Web Semant. 2012;17:25-32.

45. Küçük D, Küçük D. Ontowind: an improved and extended wind energy ontology. arXiv preprint arXiv:180302808. 2018.

46. Reder K, Stappel M, Hofmann C, Förster H, Emele L, Hülk L, et al. Identification of user requirements for an energy scenario database. Int J Sustainable Energy Plann Manage. 2020;25(3):95-108. https://doi.org/10.5278/ijsepm.3327.

47. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. MIT Press; 2015.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

48. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25(11):1251-5.

49. Gkoutos GV, Schofield PN, Hoehndorf R. The units ontology: a tool for integrating units of measurement in science. Database (Oxford). 2012.

50. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biol. 2005;6(5):R46. PMID: 15892874.
https://doi.org/10.1186/gb-2005-6-5-r46.

51. Jackson R, Balhoff J, Douglass E, Harris NL, Mungall CJ, Overton JA, et al. ROBOT: A tool for automating ontology workflows. BMC Bioinformatics. 2019;20:407.
https://doi.org/10.1186/s12859-019-3002-3.

52. Fernandez-Lopez M, Gomez-Perez A, Juristo N. Methontology: from ontological art towards ontological engineering. In: Proceedings of the AAAI97 Spring Symposium. Stanford, USA; 1997:33-40.

53. Grüninger M, Fox M. Methodology for the design and evaluation of ontologies. In: IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing; 1995.

54. Spyns P, Tang Y, Meersman R. An ontology engineering methodology for DOGMA. Appl Ontol. 2008;2(3):13-39.

55. Suárez-Figueroa M, Gómez-Pérez A, Fernández-López M. The NeOn Methodology for Ontology Engineering. Berlin, Heidelberg: Springer; 2012:9-34.

56. Courtot M, Malone J, Mungall C. Ten simple rules for biomedical ontology development. In: Proceedings of ICBO; 2016.

57. Poveda-Villalón M, Gómez-Pérez A, Suárez-Figueroa MC. OOPS! (Ontology pitfall scanner!): an on-line tool for ontology evaluation. Int J Semantic Web Inf Syst. 2014;10(2):7-34.

58. Euzenat J, Shvaiko P. Ontology matching. 2nd ed. Springer; 2013.

59. Grüninger M, Fox MS. Methodology for the design and evaluation of ontologies. In: Proc. of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95; 1995.

60. Neuhaus F, Vizedom A, Baclawski K, Bennett M, Dean M, Denny M, et al. Towards ontology evaluation across the life cycle – the ontology summit 2013. Appl Ontol. 2013;8(3):179-94.

61. Stevens R, Lord P, Malone J, Matentzoglu N. Measuring expert performance at manually classifying domain entities under upper ontology classes. J Web Semant. 2019;57:100469.

62. Kraemer HC. Extension of the kappa coefficient. Biometrics. 1980;207-16.

63. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;159-74.

64. Hülk L, Müller B, Glauer M, Förster E, Schachler B. Transparency, reproducibility, and quality of energy system analyses–a process to improve scientific work. Energy Strategy Rev. 2018;22:264-9.

65. Welder L, Ryberg SD, Kotzur L, Grube T, Robinius M, Stolten M. Spatio-temporal optimization of a future energy system for power-to-hydrogen applications in germany. Energy. 2018;158:1130-49.
https://www.sciencedirect.com/science/article/abs/pii/S0360544216306909.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press