Energy and AI 5 (2021) 100074

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Perspective

Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis

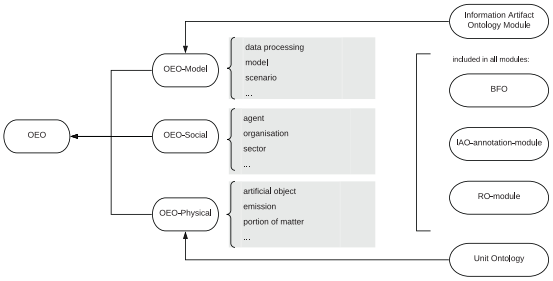
Meisam Booshehri[d], Lukas Emele[a], Simon Flügel[d], Hannah Förster[a], Johannes Frey[b], Ulrich Frey[c], Martin Glauer[d], Janna Hastings[d], Christian Hofmann[e], Carsten Hoyer-Klick[c,*], Ludwig Hülk[e], Anna Kleinau[d], Kevin Knosala[f], Leander Kotzur[f], Patrick Kuckertz[f], Till Mossakowski[d], Christoph Muschner[e], Fabian Neuhaus[d], Michaja Pehl[h], Martin Robinius[f], Vera Sehn[g], Mirjam Stappel[i]

[a] *Institute for Applied Ecology, Berlin, Germany*
[b] *Institute for Applied Informatics (InfAI) / Leipzig University, Germany*
[c] *Department of Energy Systems Analysis, Institute of Networked Energy Systems, German Aerospace Center (DLR), Stuttgart, Germany*
[d] *Faculty of Computer Science, Otto-von-Guericke-Universität Magdeburg, Germany*
[e] *Reiner Lemoine Institut, Berlin, Germany*
[f] *Institute of Techno-economic Systems Analysis (IEK-3), Forschungszentrum Jülich, Jülich, Germany*
[g] *Institute of Energy Economics and Rational Energy Use, University of Stuttgart, Germany*
[h] *Potsdam Institute for Climate Impact Research, Potsdam, Germany*
[i] *Fraunhofer Institute for Energy Economics and Energy System Technology, Kassel, Germany*

HIGHLIGHTS

- The field of energy systems analysis suffers from heterogeneous data, incompatible definitions and irreproducible models. Ontologies, particularly the presented Open Energy Ontology helps to solve these problems.
- Ontologies are a precondition for model coupling, semantic analyses of data, and data re-use.
- The Open Energy Ontology offers a common description of knowledge and vocabulary which is used across domains and different modelling approaches.
- The Open Energy Ontology is embedded within a broad community process to ensure the broadest coverage possible.
- Use cases demonstrate the added value of an ontology in the energy domain.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 January 2021
Received in revised form 4 April 2021
Accepted 5 April 2021
Available online 27 April 2021

ABSTRACT

Heterogeneous data, different definitions and incompatible models are a huge problem in many domains, with no exception for the field of energy systems analysis. Hence, it is hard to re-use results, compare model results or couple models at all. Ontologies provide a precisely defined vocabulary to build a common and shared conceptu-

* Corresponding author.
E-mail address: carsten.hoyer-klick@dlr.de (C. Hoyer-Klick).

https://doi.org/10.1016/j.egyai.2021.100074

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

alisation of the energy domain. Here, we present the Open Energy Ontology (OEO) developed for the domain of energy systems analysis. Using the OEO provides several benefits for the community. First, it enables consistent annotation of large amounts of data from various research projects. One example is the Open Energy Platform (OEP). Adding such annotations makes data semantically searchable, exchangeable, re-usable and interoperable. Second, computational model coupling becomes much easier. The advantages of using an ontology such as the OEO are demonstrated with three use cases: data representation, data annotation and interface homogenisation. We also describe how the ontology can be used for linked open data (LOD).

1. Introduction

The objective assessment of current and future energy system design and operation is a global and highly multidisciplinary research question in the domain of energy systems analysis. Experts in engineering, natural and social sciences, physics, mathematics, computer science, economics, meteorology and geography often work together: Important examples constitute analyses of pathways towards a sustainable energy system in line with the Paris Agreement[1]. Countries, institutions and researchers depend on networking and cooperation within the energy systems community. Extensive scientific exchange between all relevant actors is needed to solve one of our most urgent societal challenges. But different communities amongst these actors have developed different nomenclatures and conceptualisations, that are also reflected in the respective documentation of research data and results. This heterogeneous structure of this research domain entails a number of problems regarding data and knowledge management that hinder friction-less collaboration between scientists. In the following, we will introduce the Open Energy Ontology (OEO) as a means to address these problems. First, we dicuss the domain of energy systems modelling and detail specific challenges that the OEO can address. In Section 3 we provide an introduction to ontologies and their benefits for energy system modelling. In Section 4 we present existing approaches in the energy domain. In Section 5, we introduce the OEO as a domain ontology for energy system modelling and analysis, and we describe its design choices, patterns and content structures. Section 6 elaborates on the OEO's open and collaborative development process and on how we embed these in the energy systems analysis community to ensure the OEO's sustainable development. Thereafter, in Section 7 we describe the evaluation of several aspects of the OEO. Section 8 is about use cases, which are part of the third-party funded SzenarienDB, LOD-GEOSS and SIROP research projects. In Section 9 we close this article with general conclusions and an outlook of future work.

2. Challenges within a heterogeneous domain

Research in this domain is often accomplished by using computational models (energy system modelling) which describe the behaviour and possible evolution of energy and related systems. A great variety of energy models and scenarios, based on a coherent and internally consistent set of assumptions and motivations, depict possible pathways for future energy systems. However, a single scenario or a model by itself can never map all relevant aspects with sufficient accuracy. Thus, researchers usually build a set of scenarios to address a certain problem, focusing on special questions built on individual narratives. A scenario may focus on technological, economical, ecological or social aspects, or a combination of these. Energy models differ in their regional and sector scope (e.g. industry, residential or mobility), their level of detail (e.g. their temporal and spatial resolutions), and their initial assumptions. The core of knowledge generation is often the comparison and interpretation of scenario and model outputs based on variations of the input data and assumptions to understand the system behaviour. The facilitation of inter-model data transfer enables better analysis by combining data from well proven and tested model frameworks with other approaches and domains.

[1] https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

2.1. Evolution of energy systems modelling

Pfenninger et al [49] provide an overview of the evolution of energy system modelling facing the various upcoming challenges throughout history. In this paragraph, we present a short wrap-up on how energy systems analysis evolved from technological models to modelling approaches covering various domains of science with a steadily increasing complexity in data handling. Energy systems analysis began in the early 1970s as a reaction to the oil crisis, e.g. with founding of the Energy Technology Systems Analysis Programme (ETSAP) of the International Energy Agency (IEA) in 1974 and the International Institute for Applied Systems Analysis (IIASA) in 1972.

The first energy system models were based on linear programming [11]. Widely used examples are the MARKAL/TIMES model family [18] or the MESSAGE model [56]. These models focused on the technological evolution of the energy systems optimising towards the least-cost solution. The next innovation was the development of hybrid models [27] which extended the modelling to the economic domain by coupling the bottom-up technology energy system models with economic general equilibrium models, for example with the model MACRO. The MARKAL-MACRO linking was obtained by hard-linking two models and directly solving the coupled system models[44]. The MESSAGE-MACRO linking used a soft approach [46] by defining interfaces between the MESSAGE and the MACRO model, feeding the output into the other model and solving them in an iterative approach.

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

These modelling approaches already combined data from the energy technology domain as well as the economy domain. In the beginning, these models were mainly dealing with conventional energy systems based on fossil and nuclear power sources. Since these fuels are storable and usable energy can be produced on demand, these models did not need to deal with temporal and spatial variability on the production side. These models only covered a few time slices for the different seasons, day and night, and peak demand, summing up to 12 time slices per year accounting for differences on the demand side. New challenges then arose with the increasing importance of renewable power sources in climate-neutral energy systems. Specifically, solar and wind power are highly variable in space and time. Thus, energy system modelling that included larger shares of these sources needed to deal with their spatial and temporal availability patterns. Therefore, a new class of energy system models emerged after the turn of the century. These typically used 8760 hourly time steps per year and required climate and weather data as an important, and new, input. Since then, energy system modelling also included the domains of climate and weather. Typical representatives of these are models such as REMIx [55], PyPSA [6] or SCOPE [19]. Newer approaches focus on the increased inclusion of social and societal aspects. Examples are socio-technical scenario development [63] and agent-based modelling [12,33]. In addition to these more classical modelling approaches, new artificial intelligence, machine learning and big data methods receive increasing recognition in the domain, e.g. in the use of machine learning in model parametrisation [21]. Donghan et al. [30] show a broad range of applications which use artificial intelligence (AI) methods in energy research. Within energy systems analysis, the first AI applications focus on building analysis and

management systems [2,15,31,39,59,65], local integrated energy systems [5,14,41], smart charging [20], demand prediction [4,7,8,40,51], big data analysis including data mining [42,45] and investment decisions [21]. Algorithms therefore need to be enabled to understand and interpret this large variety of data sources from heterogeneous domains.

With increasing resolution in technologies, time, space and the inclusion of more thematic domains, data handling becomes increasingly complex. The correct interpretation of data from different domains is thus key for successful analysis within single models and even more if models are coupled through data interfaces. The Open Energy Ontology (OEO) is an approach for an exact definition of data and how they should be interpreted by the models used in energy systems analysis. Since both fields – machine learning and big data – are rapidly growing, this becomes an urgent need for future research. Examples include the problem of ”data silos” - big data that is, in principle, available, but that cannot be reused by other researchers, because its curation and formats are not reproducible, or too expensive in terms of time to reformat and hence not usable. If the same data were to be annotated with an ontology, it would be immediately reusable as its meaning and mapping to other data sources would be unambiguous. Ontologies also make it possible in machine learning pipelines to use data from various sources without rewriting the complete work-flow, since different data can be treated the same way if annotated consistently with an ontology. The Open Energy Ontology is therefore an important enabler for the application of these methods.

2.3. Research driven by heterogeneous data from heterogeneous sources

Research in the domain of energy systems analysis is driven by data to a very large extent. Results are also highly dependent on the quality of input data, since scenarios vary considerably depending on input variations. Von Scheidt et al. did an extensive review of data analytics in the electricity sector [54] and found a large variety of data analysis approaches along the whole value chain. Input data for scenarios and models usually originate from a large variety of data sources belonging to many different domains. Harmonising and interpreting the data from heterogeneous domains remains a major challenge at the beginning of each research task: data is provided by public agencies, gathered from scientific papers and commercial or public databases, stems from crowd-sourcing initiatives or is measured by researchers or via remote sensors. The respective data formats range from single values or time series to multidimensional fields. The data represent information in various spatial and temporal resolutions, e.g. hourly wind speeds at various sites and various heights above ground. In addition to the extensive data basis, the energy systems analysis community is – as a result of its modelling efforts – itself generating data at a large scale. Without the means of permanently and consistently annotating data with contextual information and documentation, databases are at risk of becoming “data graveyards” in which it is difficult to find, link, retrieve and update existing and relevant data. This situation furthers the emergence of isolated and quickly outdated data silos. Such silos lead to cycles of assembling data inventories again and again, resulting in poor data handling efficiency across the community. A positive counter-example comes from biology: the Gene Ontology[2], founded in 1998, is at the very center of biomedical knowledge about gene functions. It is a shared, distributed and ubiquitously used collection of over 6 million functional annotations of more than 4400 species in a machine-readable format. It includes findings from over 150,000 papers and has itself been used in tens of thousands of scientific studies. It powers databases, is widely used for any kind of annotation task and is thus arguably the most successful resource in computational biology. As an example from a different domain, the terms and relations defined in ontologies form the foundation for many internet of things (IoT) applications and play a fundamental role in the development of digital twins that are consistently usable across different domains. But the benefits of the OEO exceed the mere annotation of existing knowledge. Each element of this ontology is part of a large logical theory that is based on the expressive OWL2 semantics [32] and can be used to infer implicit knowledge. This enables the development of information systems that not only integrate the data from different scientific contexts, such as between chemistry and biology [26], but also fill existing gaps automatically in accordance with the ontology’s logical theory. The knowledge represented in ontologies can also be used as a foundation for novel AI approaches. For example, the ontology class structure of the CHEBI ontology has recently been used as the foundation for a deep-learning approach for the classification of chemical entities[25].

2.4. Research gap - the road to the open energy ontology

As has been described in the previous subsections 2.1 through 2.3, dealing with increasing complex data structures in energy systems analysis has historically been a neglected topic. However, there are now a number of initiatives of open data platforms and forums which have begun to discuss these challenges, as described in this subsection. In addition, we describe related ontology development in Section 4. We will link these to our work.

Nevertheless, none of the existing approaches covers the broad range of terminologies we need for our domain, nor has a suitable structure for our requirements. As of now, there is no ontology tailored to energy systems analysis that describes the relevant data and modelling approaches with all their characteristics. Thus, the management, exchange, comparison and interpretation of scientific data, approaches and results represent difficult challenges continuously addressed by third-party funded projects and community initiatives[3,4,5]. The openmod glossary[6] was an initial effort to develop a community-managed knowledge store in the energy modelling domain and served as a basis for the OEO. The glossary[7] included 323 terms centred around the modelling of photovoltaic modules gathered by the community and from a series of lectures at the HTW[8]. Its web application has enabled the allocation of synonyms and acroynms, sub- and generic terms and the creation of discussion threads for each term. The glossary’s usefulness for shared comprehension has become clear alongside its technical and structural limits as (machine-)readable and a structured storage of knowledge. The terminology of the third-party funded project openENTRANCE[9] only tackles terms relevant to project-specific models and is missing relational links between its terms. The lack of semantic linkage between the terms in the two projects mentioned above hinders their application to AI, which however, is being addressed in the OEO.

One notable ontology has been released, based on use-cases involving industry parks [13]. This ontology contains important entities relating to energy grid structures and demand-supply chains. We aim to align a variety of entities in both ontologies. Yet, many aspects that are important for a conceptualisation of the energy systems analysis domain are not covered. This specifically includes environmental factors and the description of data and scientific processes, both major elements of the OEO’s domain.

We developed the OEO with the objective of easing cooperation and exchange of information across the energy systems analysis domain. We also designed the OEO to map the complexity of the research area and

[2] geneontology.org

[3] https://www.energieforschung.de/forschung-und-innovation/systemanalyse/modex

[4] https://www.forschungsnetzwerke-energie.de/systemanalyse

[5] https://openmod-initiative.org/manifesto.html

[6] https://wiki.openmod-initiative.org/wiki/Category:Glossary

[7] an alphabetical directory of terms of a domain with definitions and their sources

[8] https://re-master.htw-berlin.de/

[9] https://github.com/openENTRANCE/nomenclature

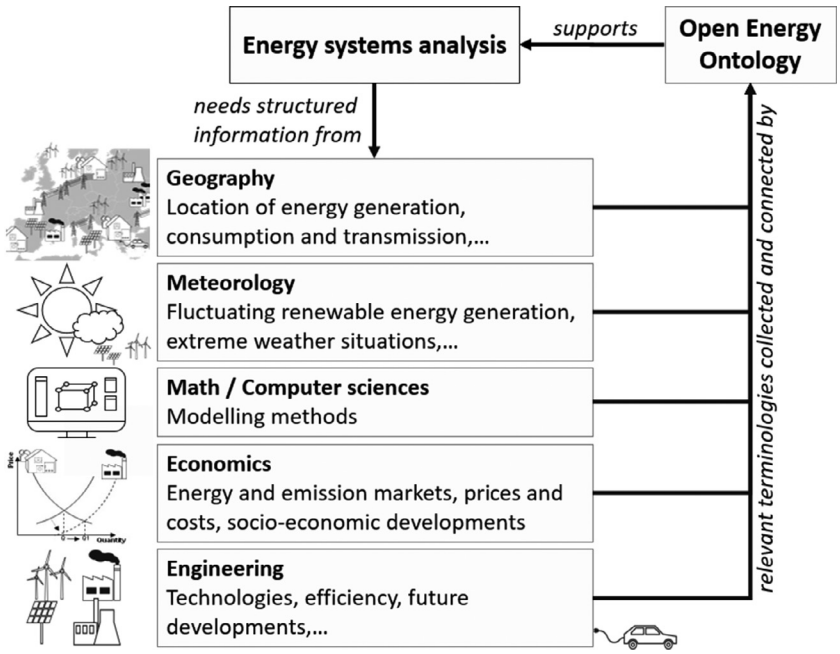


Fig. 1. The OEO collects, connects and structures parts of domain terminologies relevant for energy systems analysis.

to collect, connect and structure the ambiguous terminology of adjacent domains, the energy systems analysis domain needs information from (Fig. 1). Its steady growth increasingly enables the precise, unequivocal and comprehensible annotation and interpretation of research data. Serving as a basis for international and friction-less scientific exchange, the OEO enables consolidation and re-use of distributed data inventories across domains, thereby harnessing synergies within the global and interdisciplinary energy systems analysis community and supporting the robust transition to sustainable energy systems.

2.5. Objectives

Our objectives are tailored to the increased sophistication and interdependence of energy system modelling as described above. New needs arise compared to the past: Models and modellers increasingly interact; more and heterogeneous domain data and knowledge becomes available; computational capacities grow. Previously suitable routines – such as exchanging data as files, adding data to models by hand, coupling models by pre-defined, static interfaces – become less and less feasible. Increased automation of these interfaces by machines requires a semantic understanding of the data.

Further, when interacting, different experts may be expressing the same thing, but using different terms – those that are common within their discipline. This poses challenges, not only in investing the time to understand one another before one can work efficiently together, but also in terms of investing time to find such a common understanding again and again as these challenges occur repeatedly and in different project contexts. An example of a common misunderstanding regards final energy consumption of the industry sector. While models which are calibrated to the European energy statistic (Eurostat) define final energy consumption of the industry sector excluding the fuel uses for non-energetic fuel consumption, models which are calibrated to international statistics (IPCC) define it including non-energy uses. If a clear definition of this result variable is missing, these different approaches are not easily traceable and lead to confusion.

An ontology can help to ease these challenges. Our goals with this paper are to demonstrate the value added by ontologies and to describe how, with the Open Energy Ontology – in the energy system modelling domain – we have taken some steps towards a common vocabulary for

- data understanding across domains (see Section 6),
- data representation (see Section 8.3),
- data annotation for data to be machine and AI interpretable (see Section 8.2),
- interface-homogenisation for coupling of models using clear model-interface descriptions (see Section 8.4),
- automated data validation (see Section 9).

While the Open Energy Ontology is growing, it is by no means the only ontology in this field. Existing ontologies, how they relate and how they are integrated is described in Section 4. This section also describes the novelty of our approach and how it enhances scientific knowledge in the domain of energy systems analysis.

3. What is an ontology and what is it good for?

Ontologies, as the term is used here, are formal descriptions of entities in a certain domain and their relationships to one another. This is different from Ontology as a sub-field of philosophy, which is about the study of being and the fundamental categories of existence. In contrast to taxonomies (like the familiar taxonomy of animals and plants), or thesauri or vocabulary lists, ontologies also define the relations between entities in a formal way. This means that, typically, an ontology consists of different kinds of generic classes (e.g. buildings, house, roof, colour, or tilt), which can be related to one another, e.g. *a house has a roof as part and is located in a village* and *a roof has a colour and a tilt* (“has part”, “located in”, “has colour” and “has tilt” are relations). Specific instances of classes can be defined as well, e.g. *the Eiffel tower is a building and has a grey colour*. Here, the *Eiffel tower* is an instance of the class of buildings and *grey* is an instance of colour.

Ontologies, as formal specification of entities within a domain, usually include definitions and provides several advantages. Ontologies

- provide a *common vocabulary* within a field. This facilitates *sharing of information* and avoids ambiguities – even for software agents. Hence they ease *cooperation*.
- enable researchers to *better navigate the complexity of a domain*, since they provide a well-thought-out structure of definitions and relationships. Ontologies make it easy to check for *consistency*.
- enable *re-using domain knowledge*. Existing work does not have to be repeated and can be combined with own efforts.

- *separate domain knowledge from operational knowledge*. Processes may be independent of the involved components. For example a robot turns screws (process, i.e. operational knowledge) and the screws for that process can come in different sizes (process components, i.e. domain knowledge). Separating the two conceptually allows for easier reuse when describing conceptually similar things. Domain knowledge can be reused without the need for knowledge of the operational details, while operational knowledge can still be represented.
- *allow increasing knowledge by automatic inference*. Axioms are logical expressions in the underlying logical language in which the ontology is written. Axioms are associated with classes in an ontology (e.g. all trees have trunks) and apply to all instances of the class. In practice, this means that if, for example, a data set of trees is added as instances to an ontology, a so-called reasoner is able to infer that the new tree instances also have trunks, thus creating new knowledge.
- *map between isolated data*. Typically, institutions have their own data formats, work-flows and terms, sometimes called data silos. If the same ontology is used, then the data can be easily transferred, exchanged, and updated.

We can distinguish between two types of ontologies: upper-level ontologies and domain ontologies. Domain ontologies focus on a certain part of reality, a domain, such as energy systems. Upper-level ontologies provide classifications and relations of very generic sorts of things, such as “object” or “process”, which are used across domains. Examples of upper-level ontologies include the Suggested Upper Merged Ontology (SUMO) and the Basic Formal Ontology (BFO). Domain ontologies usually use an external upper-level ontology for their basic structure and extend these in a domain-specific way. BFO is the upper-level ontology used by the OEO and is further described in Section 5.2.

The most widely used family of knowledge representation languages for authoring ontologies is the Web Ontology Language (OWL). It builds upon the Resource Description Framework (RDF), which is able to represent information about entities and their relationships. The Protégé[10] software is a popular tool for implementation and exploration of OWL ontologies in a graphical user interface.

4. Ontologies in the energy domain

To date, the only well-known terminological resource for energy systems analysis is the *EnArgus Ontology* [48]. The German state and its federal governments use this ontology to support decision-makers with energy science related findings. It includes a wide range of terminology that was collected in a semi-automatic fashion. The related wiki offers a rich resource containing useful terms and definitions. However, the EnArgus Ontology is, to this date, proprietary and thus currently not available to the community for reuse. Based on the publicly available information in its wiki, we infer that the EnArgus ontology mainly consists of a subclass hierarchy and is only lightly axiomatised (i.e. has only simple logical expressions, see Section 3).

Energy markets and price developments are a central part of many energy system models. Electricity markets form the subject of the *Electricity Markets Ontology* [53], and financial markets of the *Financial Industry Business Ontology* [3]. Recent developments in energy systems analysis necessitate a more holistic approach to the representation of markets, including for heat, gas and other energy carriers as well as the transitions between those. The OEO does not yet include a comprehensive treatment of markets, but when we add the respective terms, we will harness pre-existing ontologies where possible, supplemented by additional content according to our needed scope. Semantic technologies have been applied in many smart home applications for data management and data integration. Therefore, the domains of houses and urban development have been covered by ontologies. For example, the SEMANCO Ontology [43] and the Energy Resource Ontology [34] cover energy-related aspects of the housing sector. Other physical systems, their relations and properties are modelled in the SEAS ontology [38], which was developed as a generalisation of the semantic sensor ontology (SSN) [9]. Many sources of renewable energy depend on some kind of meteorological phenomena and most analyses in the climate and energy field involve assumptions regarding weather and climate to project the behaviour of those energy sources. The annotation of meteorological and climate data and the involved technologies was the main use case for the development of the OntoWind ontology [36]. In summary, as of today, no publicly available ontology covers the full domain of energy systems analysis. With the OEO, we have begun to address this gap.

5. The OEO and its general design choices

5.1. Ontology background, context and outline

We created the OEO as a part of the *Open Energy Family*, an open source toolbox and database for open data within the field of energy systems analysis research. We built this toolbox around the *Open Energy Platform* (OEP)[11]. The OEP is a collaborative online platform with an underlying database for energy and climate analysis data. To this database, users can upload a wide range of data types, for example time-series, geographic data and lookup-tables. Single energy data sets and complete energy scenarios can be uploaded to the database. Our users publish all data sets under an open license and thus data becomes freely and easily accessible to others. The OEP therefore serves as a reference and facilitates scientific and political decision-making by fostering an improved level of transparency and comparability.

Currently, we develop the OEO within the project *SzenarienDB*, augmented by the projects *LOD-GEOSS* and *SIROP*. In *SzenarienDB* we extended the functionality of the *Open Energy Platform* to become a transparent and user friendly database for energy scenarios [52]. Scenarios are an essential part of the domain of energy systems analysis and at the same time they are complex and heterogeneous in nature. To make scenarios transparent and comprehensible, an ontology is needed to generate a common understanding across research areas.

The aim of the project *LOD-GEOSS* is to create a network of heterogeneous databases for input and output data from energy systems analysis. The idea is to share the data in decentralised databases which stay with the data owners, so they can take care of data updates and maintenance. The databases are connected through a metadata catalogue which makes the data findable and accessible.

In the project *SIROP* we strive towards a better interoperability of energy scenarios. The comparison of scenario data sets is a laborious process which is usually done manually. By using and extending the OEO, a semi-automated comparison of energy scenarios becomes possible.

SzenarienDB, LOD-GEOSS and *SIROP* implement the FAIR principles[12] of open data to energy systems analysis data.

We develop the OEO using the Web Ontology Language (OWL). Currently, the OEO contains around 870 classes. About 350 of these are OEO-owned classes. The remainder is imported from one of the external ontologies as described in Section 5.3. Furthermore, the OEO uses 80 object properties. About half of these are created internally for domain-specific purposes, while the other half are imported. To date, the OEO contains over 8500 axioms (logical assertions).

We made the first official release – 1.0.0 – of the OEO on June 11, 2020, and we released version 1.4.0 on March 02, 2021. The OEO can

[10] https://protege.stanford.edu/

[11] https://openenergy-platform.org/

[12] https://www.go-fair.org/fair-principles/

be accessed via GitHub[13] and its official releases are published on the OEP[14].

5.2. BFO, Design patterns and best practices

We structure the OEO based on a shared "upper level" or foundational ontology that describes basic types of entity, such as "object" and "process", which are not domain specific and serve as a basic framework. The energy specific entities are integrated as subclasses of that basic framework. This is common practice for many scientific ontologies. The OEO has adopted the widely used Basic Formal Ontology (BFO) for this purpose [1]. BFO distinguishes between "occurrent" entities that unfold in time and have temporal parts (e.g. processes, transformations, flows), and "continuant" entities that continue to exist as the same individual over time (e.g. objects, organisms, devices). Among continuant entities, BFO further distinguishes between those that are "independent" and those that are "dependent" on other entities, such as qualities and other attributes.

The OEO also adopts ontology design patterns and best practices, in line with those of the broader scientific ontology community as represented by the OBO Foundry [58]. We derived best practice principles concerning taxonomy, terminology and definitions from [1]. The ontology has a modular organisation (described in Section 5.3). It follows – as far as possible – a single asserted superclass taxonomic structure, which means that every class in the OEO is allowed to have exactly one parent class (monohierarchy). Additional superclasses are inferred from logical axioms where needed, using automated reasoning. For example, water has the parent class "portion of matter", but because its axioms state that it is renewable and can be used as an energy carrier, automated reasoning infers a second parent class "renewable energy carrier".

Each entity in the ontology is assigned a unique *label* and a text *definition*, while additional synonyms, comments, examples of usage and relations to other entities may be included if needed. We label classes with commonly used domain terminology, although, especially with ambiguous terms, this is not always possible. To prevent confusion, each OEO-owned class and relation comes with a distinct definition. For classes, we choose the Aristotelian definition format, that consists of a reference to the superclass (the *genus*) and a clear specification of what distinguishes the members of the subclass from other members of the superclass (the *differentia*).[15]

Each entity in the ontology is assigned an alphanumeric primary identifier in the namespace *OEO:x* (where *x* is a unique number). The numbers are sequential and semantics-free, however, specific sub-ranges are assigned to different ontology curators. We do this to prevent clashes during concurrent editing.

5.3. Structure and submodules

The OEO consists of three main domain-specific modules (Fig. 2) covering the following aspects of the energy systems analysis domain:

1. models and data (oeo-model),
2. social and economic aspects (oeo-social),
3. the physical side of energy systems (oeo-physical)

Furthermore, there is an additional module for classes and relations that are needed in multiple modules (oeo-shared).

All modules are imported into the main ontology, which adds relations between the separate modules. We chose this modular approach because it makes maintenance easier: different groups can work on different files without risking clashes from concurrent changes. This approach also helps organise the content into logical sub-divisions within the overall domain.

The *oeo-model* module comprises all entities related to data and models. Apart from the different types of models, most entities defined in this module relate to either transformations of data or information entities. This includes for example model calculations and the data processing methods used in energy system models. Information-related entities that we include in this module are largely an imported subset of the Information Artifact Ontology[16]. This imported module includes the class "information content entity", with subclasses to define types of information content entity, such as data items, documents, symbols and figures. The OEO's own information content entities are classified as subclasses of these more general information entities. Examples include the scenario class, different types of data descriptors, as well as assumptions and constraints.

The *oeo-social* module depicts social, economic and political entities to describe the socio-economic aspects of energy systems: Included are basic classes such as "population" and "organisation". Sectors are implemented as a combination of a "sector" class alongside overarching "sector divisions". The "sector divisions" delineate which sectors are relevant within a particular context. Different kinds of roles are defined, such as "agent", "author", "producer" and "user". An important kind of organisation for the domain are energy producers, implemented via the class "organisational energy producer" and its subclasses. Economic entities are also relevant for the domain of energy systems modelling. To cover these, we decided to re-use the well-established existing *Financial Industry Business Ontology* (FIBO[17]) [3]. It provides a rich resource of entities and relations pertaining to the domain of economics and financial markets that are important for many energy systems models. Since FIBO does not use the BFO, we have adjusted the FIBO classes and definitions to add a fitting BFO classification. Thus, FIBO content is not imported as-is but used as a source and annotated as cross-references to OEO-owned classes. The selected economic terms include, for example, "price", "gross domestic product" and "exchange".

The *oeo-physical* module includes all entities related to the physical world of energy systems. Basic concepts like energy, power and matter are classified as well as technical objects like power plants or batteries. Many entities describe physical objects and are therefore subclasses of BFO's "material entity" class. Matter, materials and fuels are represented beneath the "portion of matter" class, like coal, peat, water and methane. We use axioms that enable automated classification based on logical equivalences. Thus, these materials are arranged into different categories based on their properties and capabilities, such as greenhouse gases or fuels. In particular, we categorised fuels into detailed subtypes such as biofuels, renewable fuels or nuclear fuels. We defined the related entities for greenhouse gas emission and pollution as subclasses of BFO's "process". The class "artificial object" contains technical devices such as batteries and generators. We categorise power plants by their inputs, e.g. "wind farms" or "biofuel power plants". Further, different kinds and usages of energies and transformation processes are part of this module, "primary energy production" or "final energy consumption". To describe quantitative amounts of physical entities, the OEO imports the Unit Ontology [22] into this module. The Unit Ontology defines power units and energy units, thus usefully covering a part of the energy systems domain.

The *oeo-shared* module includes those entities and relations that are needed across multiple different sub-modules. For example, we define here classes such as quantity values.

[13] https://github.com/OpenEnergyPlatform/ontology/

[14] https://openenergy-platform.org/ontology/

[15] E.g., in the definition *A square is a rectangle in which all sides have the same length* the superclass (genus) of "square" is "rectangle". The differentia is the property of having sides of equal length.

[16] https://github.com/information-artifact-ontology/IAO/

[17] https://spec.edmcouncil.org/fibo/

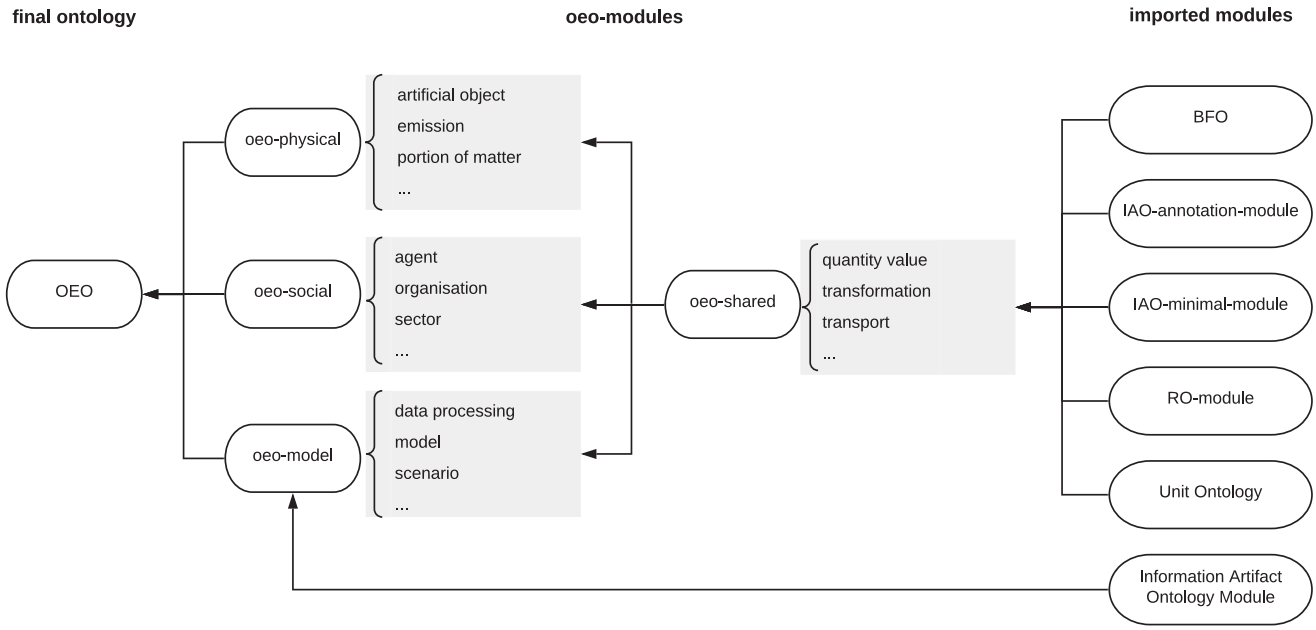


Fig. 2. Submodules and imports of the OEO.

As we mentioned above and show in Fig. 2, the OEO imports parts of other ontologies to avoid "re-inventing the wheel". Aside from BFO, we reuse significant parts of two other ontologies:

First, we import the *Relations Ontology (RO) module* which contains a subset of the object properties defined by the Relations Ontology [57]. We chose to only include a subset of RO, as many of the relations are not relevant for energy systems analysis. Examples of object properties we import through this module are properties such as "has quality" and "has disposition", some basic properties such as "part of", and properties to define temporal and spatial relations including "starts with" and "located in". Second, we reuse all metadata annotation classes defined by the Information Artifact Ontology (IAO), e.g. "document" or "reference", in the oeo-model module. The IAO also contains unseful standardised annotations, such as the "term tracker item" annotation, which are reused. It is used to reference a GitHub issue and pull request that defined or changed the entity, creating transparency by allowing rapid access to further information and the history of a class, as well as the discussions that took place around it. The annotations, along with some important IAO classes that are useful for all modules, are imported via the oeo-shared module.

We facilitate the reuse by utilising the ROBOT library [29] to extract just the content we need, as further described in Section 6.1.

Fig. 3 illustrates some of the classes and properties of the OEO: beyond a mere taxonomy, there is a rich set of properties (relations) that link classes. If a relation just affects classes of one specific module we defined it in that particular module. Relations that link classes from different modules are defined in the parent OEO file.

6. Open collaborative development

We discussed in 5.2, that the OEO follows the OBO principles[18]. We thus develop the OEO in an interdisciplinary, collaborative, public and open source[19] way. Our chosen workflow reflects these characteristics, and our specific focus is on openness: All our technical discussions and developer meetings are held publicly on the project's GitHub page[20]. Anyone is invited to contribute. Furthermore, we established a steering committee comprising of experts from different related disciplines. The steering committee guides the development of the OEO.

6.1. Ontology development

There are, broadly speaking, two different approaches to building a domain ontology. One is that the ontology can be generated by means of an automated approach using AI to analyse text corpora (e.g., scientific publications or resources such as Wikipedia). This assembles relevant information about a certain domain and converts it into an ontology. The second approach to creating an ontology is that human domain experts collect and develop relevant entities manually, defining and interrelating them in the ontology. The latter approach is the one used for the OEO development. Clearly, this is a slower process.

However, automatic approaches struggle to resolve noise and varying levels of quality in the source material, terminological ambiguities, diverging terminologies and different points of view that are represented in scientific texts. For this reason, no automatically generated ontology has so far been successfully adopted as a scientific reference ontology. In contrast human developers are able to identify these issues during the ontology development process, and, thus, are able to develop a consistent representation of the domain and a well-defined vocabulary. Another advantage of the manual approach is avoiding an unintentional bias that might exist in specific data sources. Thus, the domain ontology can be harnessed by other AI applications without reinforcing an unwanted bias; see chapter 8.

Ontologies such as the OEO are developed to serve a scientific community. Their creation processes rely on workflows, standards and technologies which enable collaborative development. Many ontology development methodologies have been proposed (e.g. [17,24,60,62]). In many ways these are similar to the workflows and methodologies associated with *open source software*: they aim to make the ontology development process reliable and repeatable, while focusing on quality throughout the development. As exemplified by the recommendations in a recent short article offering "ten simple rules" for ontology development [10], one of the most important aspects of good ontology development is to re-use existing ontology content as much as possible. This allows for cumulative extension of available knowledge resources and

[18] http://obofoundry.org/principles/fp-000-summary.html
[19] https://github.com/OpenEnergyPlatform/ontology/blob/dev/LICENSE
[20] https://github.com/OpenEnergyPlatform/ontology/issues

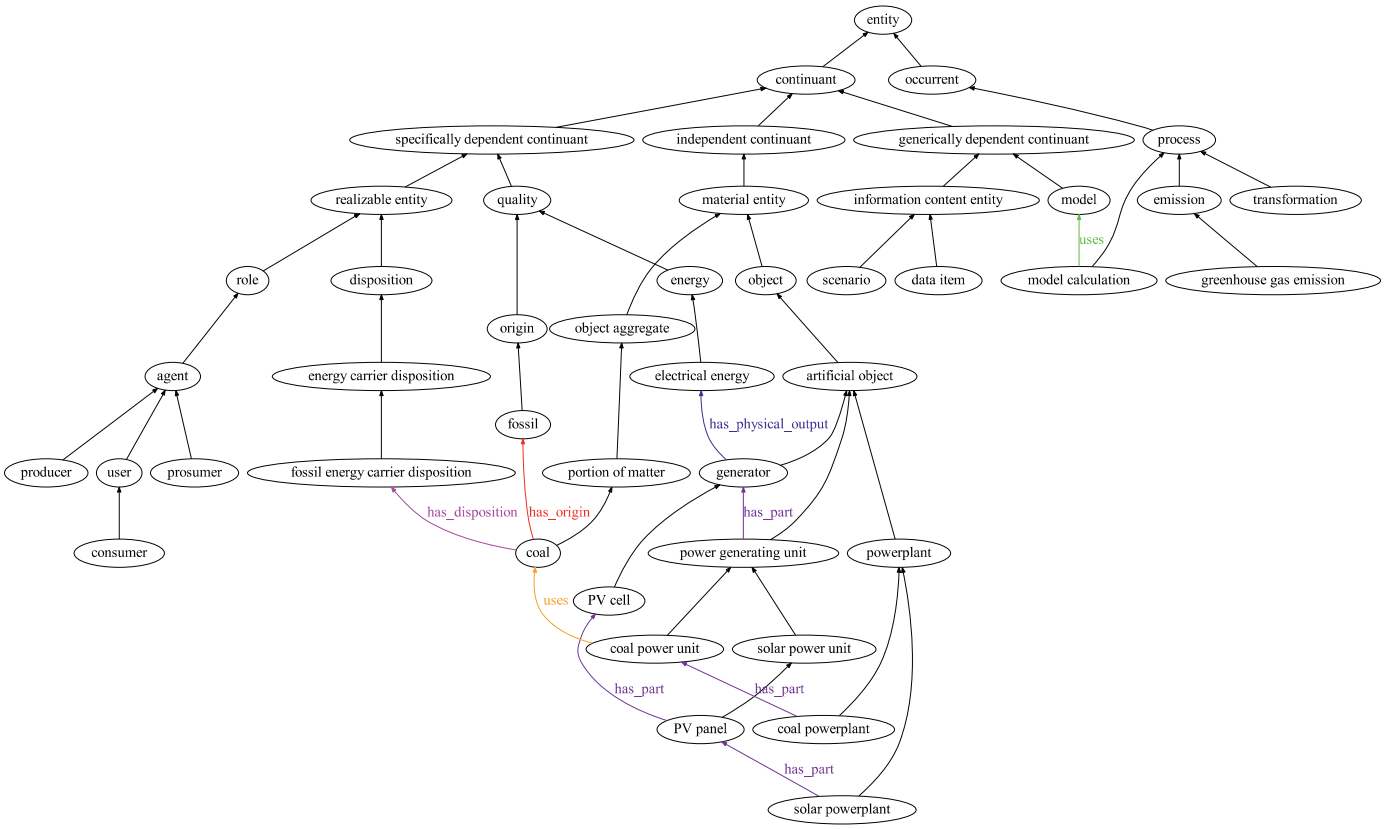


Fig. 3. Overview of a subset of classes and properties of the OEO to illustrate how they are organised inside the OEO. A black arrow denotes "is a", i.e. a subclass relation.

prevents duplication of effort. Hence, we designed the OEO to import relevant content where possible.

A clear approach to ontology versioning control and an 'open' license are considered key elements of methodological recommendations (e.g. in the OBO Foundry Principles [58]). Furthermore, these methodologies typically include recommendations for setting the scope of the ontology, and for its evaluation. The latter should be performed early, frequently and openly. Finally, they recommend community engagement and documentation of design patterns.

To facilitate re-use and collaborative exchange of ontology content between different communities and different domain areas, it is particularly important that common standards are adhered to. To help facilitate the development of such common standards, the OBO Foundry [58] is an initiative in the biological and biomedical domain. It has brought together ontology authors to create a set of ontology design principles and standards which can be semi-automatically verified. These design principles and standards have also allowed the implementation of tools such as the ontology library ROBOT [29] which automates many common ontology development tasks. While with the OEO we address a different domain, many of the standards which we have adopted in its development are based on those developed by the OBO Foundry. For example, we re-use Foundry metadata standards and common relationships.

6.2. Git workflow

We develop the OEO publicly: its code and all discussions are available on GitHub. Our detailed manuals for usage[21] and contribution[22] allow new collaborators and users a facilitated entry to the ontology. The description of the workflow ensures quality and traceability of decisions. Our workflow requires that every suggested change to the ontology has to be discussed in a GitHub issue before proceeding with an actual change. We characterise issues categorised into one of four categories:

- adding new entity
- restructuring existing parts
- updating definitions of existing entities and
- other

Small changes need the agreement of at least two members of the developers, larger changes at least three. These members should include one domain expert and one ontology expert. To reflect the diverse background of the OEOs developers and to facilitate rapid group formation when tackling an issue, OEO developers join GitHub teams in their fields of expertise. Currently we have teams for these domains: economy, modelling, linked open data, meteorology and formal ontology. If agreement is challenging to reach by discussing in a GitHub issue, we add it to the agenda of the next ontology developer meeting. These meetings generally take place as online conferences. In addition to the teams of domain experts there is also a team that carries out new releases.

Our development procedure is slow, but thorough by design. After we agree on an issue's solution, technical implementation of the change can follow quickly along a specific protocol and can be carried out by any member.

This development workflow is enhanced by several automated tests, that ensure a certain level of quality standards. These checks include syntactic constrains and an automated reasoner is used to check for log-

[21] https://github.com/OpenEnergyPlatform/ontology/blob/dev/README.md

[22] https://github.com/OpenEnergyPlatform/ontology/blob/dev/CONTRIBUTING.md

ical consistency. Additionally, the pitfall checker OOPS[23] and the OBO-foundry tool ROBOT[24] are used for quality assurance.

6.3. Community embedding

We supplement the workflow on GitHub with online developer meetings. In these, we review progress and discuss challenging issues. We schedule these meetings every month as jour-fixes, so we maintain a reliable schedule. Currently, these meetings are organised and prepared by members of the research projects *SzenarienDB, LOD-GEOSS* and *SIROP*. In cases where we cannot find agreement regarding an issue – neither in GitHub nor in the developer meeting – we pass this issue and possible solutions to the OEO-Steering Committee (OEO-SC). The OEO-SC discusses and provides a decision. Thus, the OEO-SC helps with directional decisions. While this is one focus, the other focus of the OEO-SC is to raise awareness of the ontology and its adoption in active and planned projects. The steering committee convenes approximately every 3 months. To ensure a widespread acceptance of the committee and the OEO, the OEO-SC members are experts from various domain-related backgrounds and organisational contexts and with several years of experience in their respective domain.

To ensure an appreciative interaction between all OEO-developers, we follow a self-chosen code of conduct. This code of conduct is based on the principles of non-violent communication and is thus in line with GitHub community guidelines. While the subject of the OEO is, in principal, a neutral matter – having such a code of conduct in place helps to concentrate on the issue and avoid heated discussions that may be hurtful to some or all participants.

To date, we introduced the OEO to several hundred scientists in the field: We presented it to the international openmod community which has approximately 550 registered users and to the Forschungsnetzwerke-Energie (FNE). The latter has has more than 250 participants in Germany. Currently, a community of over 350 registered Open Energy Platform users is exposed to the OEO's development.

6.4. Testing and continuous integration

The large number and diversity of contributors makes regular checks of coherence and consistency necessary. A number of automated and semi-automated tests have thus been implemented. Protégé is used as the default development tool for the OEO, and the OWL reasoners that are supported by Protégé are used to ensure consistency. The OntOlOgy Pitfall Scanner (OOPS!, [50]) defines a set of common pitfalls that occur during ontology development processes, such as missing naming conventions or missing annotations. OOPS! is used manually to ensure that releases of the ontology do not violate these rules. As discussed earlier, the ROBOT library is used in different parts of the ontology development process, e.g. module extraction. It is also a central part of our automated testing and continuous integration process as it is used to validate the ontology against different OWL profiles[25] and perform a number of quality checks such as consistency and coherence[26]. These checks control the general quality of the ontology – but are agnostic with respect to the specific domain. Therefore, we designed a number of competency questions to ensure that the entities in the ontology match their intended semantics (see Section 7.2). Each contribution to the ontology is automatically checked against the ROBOT profiles, competency questions and for consistency.

[23] http://oops.linkeddata.es/
[24] http://robot.obolibrary.org/
[25] http://robot.obolibrary.org/validate-profile
[26] http://robot.obolibrary.org/report

7. Evaluation

In this section we evaluate three different aspects of the OEO. Firstly, we evaluate its coverage of the domain. Secondly, we evaluate the quality of its axiomatisation with the help of competency questions. Lastly, we evaluate the quality of the natural language definitions of the terms in the OEO with the help of an inter-annotator agreement study. The evaluation studies were influenced by the requirements of our use cases, which are discussed in Section 8.

7.1. Evaluation I: Coverage study

Our first evaluation concerns whether the OEO contains the terms that are needed for a typical use case. One intended use case of the ontology is the annotation of various fact sheets and databases. Our ontology coverage study was based on scenario fact sheets that are being developed within the project SzenarienDB. These fact sheets are used to describe energy scenarios when the corresponding scenario data is provided to the OEP. The fact sheets include general information, such as title and authors, publication format and license, as well as the temporal and spatial analysis space of the energy models. Information on the performed modelling are covered in detail by different fields for energy and demand sectors, fuels, energy flows and environmental effects. Macroeconomic data such as population, gross domestic product and energy prices are also covered.

We used the field names of the fact sheet form as input for a semi-automated entity annotation task. In the first stage, five entity candidates from the OEO were automatically retrieved for each field label from the fact sheet form, based on label string similarity, more specifically, a combination of word tokenisation, soft Jaccard index on the token sets, and Levenshtein distance for softening the Jaccard index [16]. In the second stage, a group of ontology developers selected the correct entities or combination of entities from the candidates. Furthermore, they identified relevant entities from the ontology that were not discovered by the automatic approach. We excluded fact sheet fields that served as broad fallback descriptions (e.g. *Other Fuels*) from the evaluation, as these are deliberately not included in the ontology. Introducing such fallbacks in an ontology is considered to be bad design; for annotation purposes the same expression can be formally achieved through use of the parent class (e.g. *Fuel*) intersected with complements of subclasses (e.g. not *Fossil Fuel*). Further, ontology properties were excluded.

For the evaluation, a three-stage rating was applied to measure how well a fact sheet entity was covered by one or a combination of OEO entities: *No match* indicates that the OEO does not contain any matching entities (yet) to annotate a given fact sheet field. *Partial match* indicates that a fact sheet entity can be annotated in part by one or a combination of OEO entities. For example: "costs of coal" can only be expressed partially, because "costs" was at the time of the study not yet included in the OEO, whereas "(portion of) coal" was. *Good match* indicates a full match.

The evaluation results of the coverage study are shown in Table 1 and have been made publicly accessible[27]. In total, the annotation of 153 fact sheet fields was tested, as depicted in the first table row (*"ALL"*). More than half of the fields (52%) have a good match, whereas 20% have no match at all and cannot be described by the OEO yet.

About 30% of the fact sheet fields (46) relate to socioeconomic aspects of the domain. These refer to e.g. costs of fuels or prices for CO_2 emissions, as well as populations or gross domestic products (GDP). As described in Section 5, the OEO is structured into three modules. Until recently, the main focus of the OEO development has been on the *oeo-physical* module, with the other modules scheduled for becoming the focus area during subsequent releases. Thus, the other modules have

[27] https://doi.org/10.5281/zenodo.3870654

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*).

	# 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%

not yet been comprehensively developed, and especially the *oeo-social* module is still in a relatively early state of development.

To mitigate for this, the second row of the table (*"ESE"*) just considers those fields (107) that are not related to socioeconomic aspects. Here, about 60% of the concepts have a good match and 20% have no match at all. Comparing the total counts of both results (*"ALL"* and *"ESE"*), it can be seen that there are only 14 fields (30%) within the socioeconomic part that have a good match. Since we will focus next on the development of *oeo-social*, we expect significant improvements of this coverage in the near future.

7.2. Evaluation II: Competency questions

Competency questions provide a methodology for capturing and evaluating semantic requirements for an ontology [23]. As a first step, ontology developers work together with domain experts to develop usage scenarios and document which kind of questions the ontology is expected to answer in a given scenario. The combination of a scenario, a question and its intended answer constitute a kind of proof obligation: the formal representation of the scenario together with the axioms of the ontology is supposed to logically entail the formal representation of the intended answer. These proof obligations may be validated automatically with the help of an automated theorem prover.

Competency questions are particularly useful for the development of ontologies that have a well-specified role within the context of a larger information system, because in these circumstances the usage scenarios are restricted and well-defined and, thus, the development of these scenarios and the associated competency questions may drive the whole ontology engineering process [47]. In particular, for these kinds of ontology development projects the competency questions may be used as a measurement of a kind of completeness: if the ontology is able to answer all competency questions, all of the documented requirements are met, and, thus, the ontology development process has succeeded.

Reference ontologies such as the OEO are used to provide a shared terminology for a large community. Hence, there is no specific application context and no specific set of requirements for which the OEO is built. Thus, there is no notion of "completeness" that could be evaluated with the help of competency questions. Nevertheless, we found competency questions quite useful for the semantic evaluation of our ontology, since they allow us to evaluate whether the axioms of the ontology match the semantics that is *intended* by the domain experts. Some of our competency questions reflect the consensus position on particularly ambiguous or contentious terms. These competency questions enable us to detect changes to the ontology that are in conflict with the result of previous agreements. This kind of domain-specific semantic evaluation complements the checks for consistency and coherence mentioned in Section 6.4.

For example, the appropriate representation of *fuel* provided a challenge for the OEO. The design decisions that arose from this debate have been transformed into competency questions and formalised in OWL. An example of one of those questions is *"Is charcoal an energy carrier that is solid under normal conditions?"*. The HermiT reasoner is then used to check the entailment relation between the ontology and the questions. This process has been integrated into the continuous integration strategy in order to assure that future developments within the ontology preserve these inferences. Currently, there are 50 competency questions[28]; of these currently 41 are answered successfully. The answers of the remaining nine competency questions are currently not entailed, because of missing entities and axioms. In the future we will extend the ontology in a way that will enable these inferences.

7.3. Evaluation III: Inter-annotator agreement study

The classes and definitions included in an ontology should be comprehensible and unambiguous. When annotating resources with terms from an ontology for improved findability and query functionality, it is crucial that different annotators are able to use these terms consistently. Thus, one way to evaluate ontologies is to ask users to annotate texts with terms from the ontology and measure the agreement of their answers [61]. Thus, we used five text fragments from model fact sheets to study whether energy domain experts can annotate them consistently. We selected only text fragments where the annotation with an ontology term was not obvious, i.e. there was no perfect match between portions of the text fragment and labels of ontology terms, but rather several only roughly matching ontology terms. Hence, the domain experts had to read and understand the definitions of the terms to perform the annotation task.

For every text fragment, using the same string similarity technique and manual refinement by ontology developers as in Section 7.1 above, six ontology entities were selected. Together with the respective text fragment, annotators were given a multiple choice among those six entity definitions, plus a seventh field "None of the above". Researchers at institutes with energy systems analysis focus were identified as potential participants of this study and were invited by email. Participants in the study had no previous experience using the OEO.

Out of 34 participants, 20 completed the full survey. For this study, we only include data from these 20 participants. Among these, two had previous experience with ontologies, and 17 had at least one year of experience with energy systems modelling. The questions and responses have been made publicly accessible.[29]

As a measure of inter-annotator agreement, we use an extension of the kappa coefficient for multiple annotators with a multiple-choice setup, developed by Kraemer [35].

Using this metric, the inter-annotator agreement for our study was $\kappa = 0.668$. According to the classification in [37] this indicates a 'substantial' level of inter-annotator agreement. Moreover, while it was also possible to select "None of the above", this was chosen only very few times, which suggests that our set of candidate entities had reasonable coverage for annotating the given text fragments.

However, there is still room for improvement in the agreement. Notably, participants did not follow our guidance to only select the best match, and also picked broader matches. For example, if "greenhouse gas emission" was chosen as a match, the participants were not supposed to also choose "greenhouse gas". The second annotation is redundant, since the ontology already contains an axiom that states: "Greenhouse gas emissions involve the emission of some greenhouse gas". In practice, adding this redundant annotation does not usually cause problems, but

[28] https://github.com/OpenEnergyPlatform/ontology/tree/dev/tests/competency_questions

[29] https://doi.org/10.5281/zenodo.3870654

in the context of this evaluation it made it more difficult to evaluate the true agreement on the best match.

Participants also noted that in some cases the choices provided did not contain an entity that would describe a text fragment optimally, and for that reason there was no obvious 'best' match. Hence, the gaps in the coverage of our domain that were detected in the first evaluation had negative impacts on the inter-annotator agreement study.

We are in the process of revising the OEO according to the insights from this evaluation. One major task is to increase the coverage of the OEO in order to ensure that it provides all the terminology that is necessary to describe energy scenarios and models. Equally important is to improve the documentation of the entities in the ontology. Thus far the main focus was on providing ontologically sound and logically correct definitions. But to achieve better inter-annotator agreement we need to add more explanations, examples and synonyms.

8. Use cases

Alongside the broad range of potential applications for the OEO, such as summarising data sets, user categorisation or tagging, and semantic search, we want to present four use cases for which the OEO is currently being employed. All current use cases arise from the projects *SzenarienDB, LOD-GEOSS* and *SIROP*.

8.1. Implementation: Scenario description

The coverage study was conducted on a dataset extracted from existing scenario factsheets. This is inspired by the inherently heterogeneous structure of energy research results in energy scenarios. Published results, the datasets they used, and the implicit and explicit assumptions that were involved are often only loosely connected, which hinders the transparency and reproducibility of scientific results in the domain of energy systems modelling [28]. A framework is needed that allows the annotation of these scenarios and studies as well as their related elements such as the used and produced datasets or models that underlie the results. The OEO specifies the general structures that can be used to build such a system. A number of gaps that were identified in this analysis were addressed, which allows for a more exhaustive annotation of scenarios and their related components. A collection of properties of energy scenarios and energy studies that covers the most important information to adequately describe their properties and relations. These properties were compiled into spreadsheets and filled with data for a collection of studies and scenarios. An RDF knowledge graph that uses the classes and relations defined in the OEO and other prominent vocabularies (e.g. Dublin Core, FOAF) was created manually based on these spreadsheets and the Open Energy Platform was extended by an additional section than enables users to view and edit the information stored in the knowledge graph. This allows researchers to make their research more publicly available and facilitates a more structured landscape of datasets, energy models and scientific results.

8.2. Data representation for the core energy market data register

As of 2019 the German Federal Network Agency publishes the core energy market data register (Marktstammdatenregister, German abbreviation MaStR). It is a complete list of all registered power stations in Germany that connect to the grid. It includes all power stations, regardless of size: from large coal, lignite and water plants to small wind and private solar power modules. The data are published under an open license[30]. MaStR data is made available via an application programming interface (API). This can be seen as improved accessibility, compared to the more traditional way of downloading files. However, the API documentation[31] is complex. The so called SOAP protocol that is used lacks a standardised interaction model, so any access to the data needs has to be set up manually and will only work for the MaStR API. This constitutes a hurdle in terms of accessibility, which is amplified by limitations on the number of data points that can be requested at once. This limit is set to 10,000 requests per user per day[32]. Consequently, for a large number of users, data access is not barrier-free. Putting technical and conceptual challenges aside, the provision of open and accessible data is important and valuable, as it facilitates reproducible research. Therefore, we decided to enhance this existing infrastructure: we developed an open-source tool[33] to extract data from the MaStR interface. We enhanced the extracted data with metadata and made it available on the Open Energy Platform[34]. There, users can download the data without the need to register. We aim to make regular updates using the developed scripts. With the OEO we cover many entities that are important to describe the domain of energy market data contained in MaStR. Such entities include, for example, power plants and spatial regions. With these entities we can annotate the MaStR dataset. The annotations allow conceptual queries that are closer to natural language and do not depend on the actual representation of data. A user who wants to collect the data of all power plants in the dataset has to query each of the individual endpoints of the MaStR-API (*GetAnlageEegWind, GetAnlageEegSolar*, ...). Such complex queries can be simplified by mapping the API to an endpoint for the SPARQL Protocol[35] that uses the terms defined in the ontology. This allows for much simpler queries, based on terms that are commonly used and have agreed-upon definitions. Data on power plants can be accessed via a simple SPARQL query `SELECT ?answer WHERE {?answer a oeo:powerplant}`. The logical foundation of the OEO allows not only the definition of those queries, but also the enrichment of the dataset with logical dependencies. It is – for example – possible to limit the above query to geothermal power plants by using the class *geothermal power plant* or one might query all power plants that have a part that is a *geothermal power unit*. This flexibility allows the definition of a versatile data interface that allows for complex data aggregation but that is still easy to use and to understand.

8.3. Data annotation of an energy meteorological time series data set

Time series of different kinds serve as input data for, or result as output data from energy system models. A common use-case is energy meteorological time series data: weather times series are fed into a model which calculates times series of weather dependent energy generation data, e.g. from wind farms or solar collectors. However, their consistent and complete annotation is not trivial. Missing or ambiguous information can lead to blunders when re-using or interpreting the data.

There are different – but nevertheless equivalent – ways of describing the content of a time series, and much work is spent on discovering and adapting the definition of a specific energy meteorological time series, as habits in annotation vary across different domains. Sometimes there are conventions in certain domains, but even those are not always followed. A time series is defined as a set of data points (measured or modelled values) referencing to a set of points or intervals in time (time steps). These time steps in turn are defined by either a) a start and ending time, or b) a time stamp and the length of the time step, as well as the alignment of the time stamp within the time step (i.e. time stamp indicates either the start, middle, or ending of the time step). Energy system models use many different kinds of time series from several dis-

[30] Creative Commons Attribution-NoDerivs 3.0 Germany (CC-BY-ND-3.0-DE)

[31] https://www.marktstammdatenregister.de/MaStRHilfe/files/webdienst/2019-08_15%20Dokumentation%20MaStR%20Webdienste%20V1.2.18.pdf

[32] https://www.marktstammdatenregister.de/MaStRHilfe/subpages/webdienst.html

[33] https://github.com/OpenEnergyPlatform/open-MaStR

[34] https://openenergy-platform.org/dataedit/view/supply/bnetza_eeg_anlagenstammdaten

[35] https://api.triplydb.com/s/U9p6sbrkg

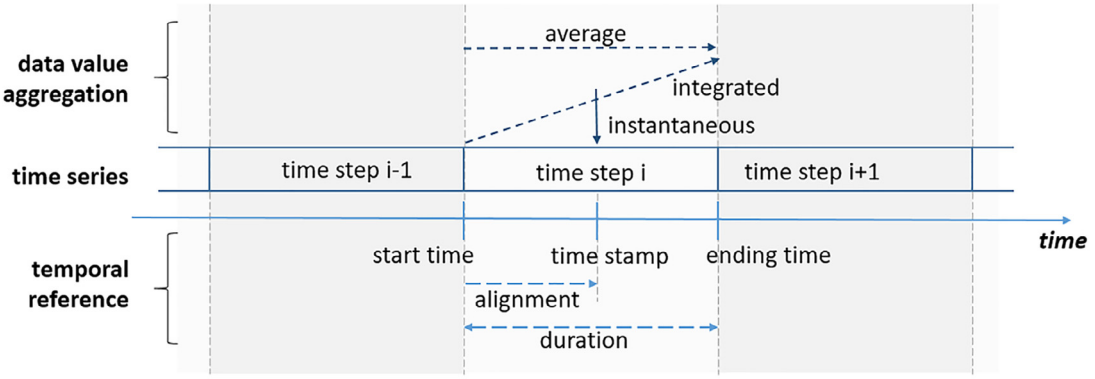


Fig. 4. Overview of time series annotation structure.

ciplines, e.g. meteorological data, time series of power generation and consumption or economic time series, like energy prices. Thus, there is no common way of defining time series. E.g. in meteorology it is a habit to use the ending of an interval as a time stamp, as data loggers often mark the time when the recording of the data is finished. However, this is not a fixed requirement and using meteorological times series often entails guessing whether the data provider followed this habit. For solar data this can be solved by comparing it to the solar geometry, but for other types of time series a verification is difficult. In climate science there are at least the climate conventions[36].

Furthermore, the type of aggregation done on the measured or modelled value (e.g. instantaneous, averaged or integrated values) for the time steps has to be annotated. Within the climate conventions this is done via the attributes `bounds` and `cell_method`. For example, wind speeds and temperatures are usually recorded as an average over a certain time interval. Solar irradiation is recorded as integrated energy in kWh/m^2 and rain as integrated amount of water on a defined area (l/m^2). Wind gusts may be maximum values within a time interval. In climate data sets, solar radiation is also often represented as an averaged rate, namely irradiance in W/m^2. As a rule of thumb, values that describe a rate (e.g. power) are averages, while values that describe an amount (e.g. energy) are integrated values. For a correct interpretation this needs to be made explicit. Especially for solar radiation the values in Wh/m^2 or W/m^2 are the same for typical time steps of one hour. However, they are conceptually different and need to be treated differently in further processing of the data.

Another common source of misinterpretation in time series data are time zones and daylight saving in the temporal information. Omission of time zone information sometimes means local standard time, sometimes UTC. These also need be made explicit in the documentation of the data. Fig. 4 illustrates the annotation structure for such time series.

When time series are exchanged in the energy system modelling community, it currently takes a substantial effort to identify and record these time series specifications. The OEO facilitates this process by providing concise and unambiguous definitions for the different annotation concepts. It eventually allows a complete description and effortless interdisciplinary identification of the structure and content of the energy meteorological and other time series.

8.4. Interface homogenisation of the FINE energy system model framework

Within this use case we aim to directly connect the distributed database architecture mentioned in Sect. 5.1 to an energy system model. Using the OEO, we want to homogenise the annotation of data inventories and the functional parameters expected or provided by model interfaces in such a way that clear assignments can be made. At the same time, we reduce the heterogeneity of interface descriptions and thus minimise the effort of programmers and users to produce and understand them. Currently the interfaces of several well-established energy system models of different types are analysed to ensure a broad integration of the most important data categories.

The FINE Framework, for example, is an open source Python package[37] that provides functionalities for modelling, optimisation and analysis of high-resolution energy system models in terms of time, space and technology [64]. 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. Based on the currently existing interface description[38] in which the individual function parameters are named and defined, we currently explore to what extent there is already coverage with the terminology of the OEO, and at which points we have to adapt the interface or the ontology. Using these specific model applications, we aim to develop best practices that can be used to homogenise the connection of data to models and ultimately the exchange of data between the models themselves and to promote scientific exchange within the international energy system community.

9. Conclusion and future work

We reported on the development-in-progress of an open and community-driven ontology for the energy systems analysis domain: the Open Energy Ontology (OEO). While ontologies are not completely novel to this domain, many pre-existing efforts were focused either on a specific sub-area of the overall domain, or were developed as proprietary resources without general open accessibility. In energy systems analysis, aside from the practical benefits for reuse and reproducibility, openness has important consequences for transparency and the building of trust and accountability. Increasingly, the teams that build open data platforms such as Renewables Ninja[39], Open Power System Data[40] and the Open Energy Platform[41] work towards transparently allowing the community to share data, align models and work together. This will be further facilitated by the OEO. Transparency and trust are even more important in the context of the advancing climate crisis, as the outputs

[36] https://cfconventions.org

[37] https://github.com/FZJ-IEK3-VSA/FINE

[38] https://vsa-fine.readthedocs.io/en/master/componentsDoc.html

[39] https://www.renewables.ninja/

[40] https://open-power-system-data.org/

[41] https://github.com/OpenEnergyPlatform/ontology/tree/dev/tests/competency_questions

of modelling efforts may be used in decision-making processes where there are strong feelings about particular possibilities. There is a need for robust, reproducible evidence that can be amalgamated and compared across different modelling approaches and stakeholder groups.

We have seen that the amount and complexity of data is rapidly increasing in the domain of energy systems analysis. As data driven methods such as machine learning, big data and AI are gaining increasing importance, machine interpretable data annotation is a key enabler for the increased use of these methods. The use and increased coverage of the Open Energy Ontology will help to make use of the growing amount of public data in the energy system.

A first evaluation in Section 7 shows that the OEO is an adequate solution for a better annotation of data in the domain of energy systems analysis.

The OEO is still under development, but already shows benefits in use for some promising applications which we described in Section 8. The development of an ontology for a specific domain is a consensus-building process within the domain, not only extending and deepening a shared comprehension of interrelated concepts but also promoting a common understanding of what constitutes valid data. If a consensus is formed on what conditions datasets have to meet to be considered correct, this knowledge can be added to the OEO and be used for automated data validation. Therefore, further ontology development should be based on broad participation within the domain. We hope that our presentation of the ontology in this article will serve as an invitation to others to join this development process and to start using the ontology for the annotation of data sets, which then can be shared and used more easily.

The development process is organised on GitHub, as described in Section 6. It started within the *SzenarienDB* project which came to an end by March 2021. It will be continued by the projects *LOD-GEOSS* and *SIROP*. Further contributing projects are already planned and some of the partners will use their institutional base to ensure the sustainability and continuation of this effort.

We see the OEO as a basis for an enhanced collaboration of various models and methods to take modelling of future energy systems to the next level by enabling flexible coupling of models through well defined data interfaces, as a part of the development of a data eco-system. By networking different modelling approaches through defined data interfaces we can grasp some more of the complexity of the energy system transformation process. The further we travel the road of transforming our energy systems to sustainable ones, the more we need to consider details and dependencies of the transformation process, which needs the collaboration of models and methods within the domain of energy systems analysis.

Author contributions:

This paper is a joint effort of all authors. The paper is based on an extensive collaborative discussion process of the content as well as a collective writing process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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).

References

[1] Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. MIT Press; 2015.
[2] Azuatalam D, Lee W-L, de Nijs F, Liebman A. Reinforcement learning for whole-building hvac control and demand response. Energy and AI 2020;2:100020. doi:10.1016/j.egyai.2020.100020. http://www.sciencedirect.com/science/article/pii/S2666546820300203
[3] Bennett M. The financial industry business ontology: best practice for big data. Journal of Banking Regulation 2013;14(3–4):255–68.
[4] Biswas MR, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. Energy 2016;117:84–92. doi:10.1016/j.energy.2016.10.066. http://www.sciencedirect.com/science/article/pii/S0360544216315006
[5] Boretti A. Integration of solar thermal and photovoltaic, wind, and battery energy storage through ai in neom city. Energy and AI 2021;3:100038. doi:10.1016/j.egyai.2020.100038. http://www.sciencedirect.com/science/article/pii/S2666546820300380
[6] Brown T, Hörsch J, Schlachtberger D. Pypsa: python for power system analysis. J Open Res Softw 2018;6(4). doi:10.5334/jors.188. eprint 1707.09913.
[7] 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 and AI 2020;2:100028. doi:10.1016/j.egyai.2020.100028. http://www.sciencedirect.com/science/article/pii/S2666546820300288
[8] Chou J-S, Bui D-K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build 2014;82:437–46. doi:10.1016/j.enbuild.2014.07.036. http://www.sciencedirect.com/science/article/pii/S037877881400574X
[9] Compton M, Barnaghi P, Bermudez L, García-Castro R, Corcho O, Cox S, Graybeal J, Hauswirth M, et al. The SSN ontology of the W3C semantic sensor network incubator group. Journal of Web Semantics 2012;17:25–32.
[10] Courtot M, Malone J, Mungall C. Ten simple rules for biomedical ontology development. In: Proceedings of ICBO; 2016.
[11] Dantzig G. Linear programming and extensions. Pinceton, NJ: Prinction University Press; 1965.
[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] Devanand A, Karmakar G, Krdzavac N, Rigo-Mariani R, Eddy YF, Karimi IA, et al. Ontopowsys: a power system ontology for cross domain interactions in an eco industrial park. Energy and AI 2020:100008.
[14] Devanand A, Karmakar G, Krdzavac N, Rigo-Mariani R, Foo Eddy Y, Karimi IA, Kraft M. Ontopowsys: a power system ontology for cross domain interactions in an eco industrial park. Energy and AI 2020;1:100008. doi:10.1016/j.egyai.2020.100008. http://www.sciencedirect.com/science/article/pii/S2666546820300082
[15] Dorokhova M, Ballif C, Wyrsch N. Rule-based scheduling of air conditioning using occupancy forecasting. Energy and AI 2020;2:100022. doi:10.1016/j.egyai.2020.100022. http://www.sciencedirect.com/science/article/pii/S2666546820300227
[16] Euzenat J, Shvaiko P. Ontology matching, second edition. Springer; 2013.
[17] 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. p. 33–40.
[18] Fischbone L, Abilock H. MARKAL, A linear programming model for energy systems analysis: technical description of the bnl version. Intnational Journal of Energy Research 1981;5:353–75.
[19] 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.
[20] Frendo O, Graf J, Gaertner N, Stuckenschmidt H. Data-driven smart charging for heterogeneous electric vehicle fleets. Energy and AI 2020;1:100007. doi:10.1016/j.egyai.2020.100007. http://www.sciencedirect.com/science/article/pii/S2666546820300070
[21] Frey UJ, Klein M, Deissenroth M. Modelling complex investment decisions in germany for renewables with different machine learning algorithms. Environmental Modelling & Software 2019;118:61–75. https://www.sciencedirect.com/science/article/abs/pii/S1364815218304353
[22] Gkoutos GV, Schofield PN, Hoehndorf R. The units ontology: a tool for integrating units of measurement in science. Database (Oxford) 2012.
[23] 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.
[24] Grüninger M, Fox M. Methodology for the design and evaluation of ontologies. IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995; 1995.
[25] 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.
[26] 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. doi:10.1186/1471-2164-14-513.
[27] Hourcade J, Jaccard M, Bataille C, Ghersi F. Hybrid modelling: new answers to old challenges. Energy 2006;2:1–12.

[28] 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 reviews 2018;22:264–9.

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

[30] Jin D, Ocone R, Jiao K, Xuan J. Energy and ai. Energy and AI 2020;1:100002. doi:10.1016/j.egyai.2020.100002. http://www.sciencedirect.com/science/article/pii/S2666546820300021

[31] Kaur R, Schaye C, Thompson K, Yee DC, Zilz R, Sreenivas R, Sowers RB. Machine learning and price-based load scheduling for an optimal iot control in the smart and frugal home. Energy and AI 2021;3:100042. doi:10.1016/j.egyai.2020.100042. http://www.sciencedirect.com/science/article/pii/S2666546820300422

[32] 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/

[33] Klein M, Frey UJ, Reeg M. Models within models - agent-based modelling and simulation in energy systems analysis. Journal of Artificial Societies and Social Simulation 2019;22:6. doi:10.18564/jasss.412. http://jasss.soc.surrey.ac.uk/22/4/6.html

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

[35] Kraemer HC. Extension of the kappa coefficient. Biometrics 1980:207–16.

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

[37] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977:159–74.

[38] 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.

[39] Lissa P, Deane C, Schukat M, Seri F, Keane M, Barrett E. Deep reinforcement learning for home energy management system control. Energy and AI 2021;3:100043. doi:10.1016/j.egyai.2020.100043. http://www.sciencedirect.com/science/article/pii/S2666546820300434

[40] Luo X, Oyedele LO, Ajayi AO, Akinade OO, Delgado JMD, Owolabi HA, Ahmed A. Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings. Energy and AI 2020;2:100015. doi:10.1016/j.egyai.2020.100015. http://www.sciencedirect.com/science/article/pii/S266654682030015X

[41] Luo X, Oyedele LO, Akinade OO, Ajayi AO. Two-stage capacity optimization approach of multi-energy system considering its optimal operation. Energy and AI 2020;1:100005. doi:10.1016/j.egyai.2020.100005. http://www.sciencedirect.com/science/article/pii/S2666546820300057

[42] 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. doi:10.1016/j.apenergy.2016.08.079. http://www.sciencedirect.com/science/article/pii/S0306261916311679

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

[44] Manne A, Wene C. Markal-macro: a linked model for energy-economy analysis.. Technical Report BNL-47161. Upton, NY (United States): Brookhaven National Lab; 1992. http://www.osti.gov/energycitations/product.biblio.jsp?osti-id_10131857

[45] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: lessons from assembling a very large national database of building energy use. Appl Energy 2015;140:85–93. doi:10.1016/j.apenergy.2014.11.042. http://www.sciencedirect.com/science/article/pii/S0306261914012112

[46] Messner S, Schrattenholzer L. MESSAGE-MACRO: Linking an energy supply model with a macroeconomic modules and solving it iteratively. New Horizons in Environmental Economics 2000;25:267–82.

[47] Neuhaus F, et al. Towards ontology evaluation across the life cycle – the ontology summit 2013. Appl Ontol 2013;8(3):179–94.

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

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

[50] Poveda-Villalón M, Gómez-Pérez A, Suárez-Figueroa MC. OOPS! (Ontology pitfall scanner!): an on-line tool for ontology evaluation. International Journal on Semantic Web and Information Systems (IJSWIS) 2014;10(2):7–34.

[51] 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. doi:10.1016/j.rser.2015.04.065. http://www.sciencedirect.com/science/article/pii/S1364032115003354

[52] Reder K, Stappel M, Hofmann C, Förster H, Emele L, Hülk L, Glauer M. Identification of user requirements for an energy scenario database. International Journal of Sustainable Energy Planning and Management 2020;25(3):95–108. doi:10.5278/ijsepm.3327.

[53] 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. p. 393–404.

[54] vom Scheidt F, Medinová H, Ludwig N, Richter B, Staudt P, Weinhardt C. Data analytics in the electricity sector - a quantitative and qualitative literature review. Energy and AI 2020;1:100009. doi:10.1016/j.egyai.2020.100009. http://www.sciencedirect.com/science/article/pii/S2666546820300094

[55] 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.

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

[57] Smith B, Ceusters W, Klagges B, et al. Relations in biomedical ontologies. Genome Biol 2005. doi:10.1186/gb-2005-6-5-r46.

[58] 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.

[59] 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. doi:10.1016/j.apenergy.2017.08.188. http://www.sciencedirect.com/science/article/pii/S0306261917312151

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

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

[62] Suárez-Figueroa M, Gómez-Pérez A, Fernández-López M. The NeOn Methodology for Ontology Engineering. Springer, Berlin, Heidelberg; 2012. p. 9–34.

[63] Weimer-Jehle W, Buchgeister J, Hauser W, Kosow H, Naegler T, Poganietz W-R, Pregger T, Prehofer S, von Recklinghausen A, Schippl J, Vögele S. 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.

[64] 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.

[65] Westermann P, Evins R. Using bayesian deep learning approaches for uncertainty-aware building energy surrogate models. Energy and AI 2021;3:100039. doi:10.1016/j.egyai.2020.100039. http://www.sciencedirect.com/science/article/pii/S2666546820300392