The Open Energy Ontology

Martin Glauer[4], Meisam Booshehri[4], Lukas Emele[1], Simon Fluegel[4], Hannah Förster[1], Johannes Frey[2], Ulrich Frey[3], Janna Hastings[4], Christian Hofmann[5], Carsten Hoyer-Klick[3], Ludwig Hülk[5], Anna Kleinau[4], Kevin Knosala[6], Leander Kotzur[6], Patrick Kuckertz[6], Till Mossakowski[4], Fabian Neuhaus[4,8], Martin Robinius[6], Mirjam Stappel[7], and Detlef Stolten[6]*

[1] Öko-Institut, Berlin, Germany
[2] InfAI / Leipzig University, Germany
[3] German Aerospace Center (DLR), Germany
[4] Faculty of Computer Science, Otto-von-Guericke-Universität Magdeburg, Germany
[5] Reiner-Lemoine-Institut, Berlin, Germany
[6] Institute of Techno-economic Systems Analysis (IEK-3), Forschungszentrum Jülich GmbH, Jülich, Germany
[7] Fraunhofer Institute for Energy Economics and Energy System Technology, Germany
[8] KRDB Research Centre, Free University of Bozen-Bolzano, Italy

Abstract. We describe and evaluate the Open Energy Ontology (OEO) that we are developing for the energy systems analysis domain. Energy systems analysis uses computational models to create scenarios reflecting possible future developments in distributed networks of energy supply and consumption. To date, the energy systems analysis domain is still fragmented and it is difficult to integrate or compare results across studies. The goal of the OEO is to build a common and shared conceptualisation that will be used in the energy systems analysis community for multiple purposes, including unification of terminology, metadata annotation, semantic search and increased interoperability for the large volumes of data in this field.
We present the OEO ontology structure and content, and discuss important challenges that have arisen during the development – in particular due to the breadth and complexity of the energy systems analysis domain. Finally, we evaluate the ontology for coverage of terms from the Open Energy Platform fact sheets, with respect to a set of competency questions, and for annotation agreement among experts.

Keywords: Energy systems modelling · Ontology evaluation · Linked open data · Metadata annotation · Energy systems analysis.

1 Introduction

Projections about the future consequences implied by energy systems scenarios are made possible by complex mathematical models. Scientific recommendations are often derived from the results of such models and can provide important

*Alphabetical, except for lead author.

information to the society and policy makers. However, different modelling approaches, and different assumptions may lead to different projections. In the debate about how a future, resilient and climate-friendly energy system could appear – considering societal and geopolitical challenges –, it is essential that the information upon which decisions are made are open, transparent and comparable. Only such an approach can ensure that there is full access to all the information needed for decision making.

The Open Energy Ontology (OEO) is an ontology for the domain of energy systems modelling that is being developed to support annotation, sharing, re-use and integration of energy systems models. In this paper we give a short introduction to the domain of energy systems modelling and point out different problems that exist within the community (Section 2). Many of these problems arise from inconsistencies in term usage or interface design. We will present the design and structure of the OEO (Section 3) and notable issues that have been solved during this process (Section 4). In addition, we evaluate the current state of the OEO with the help of a coverage study, competency questions, and an inter-annotator agreement study (Sections 5-7).

2 The Energy Systems Domain

The domain of energy systems modelling focuses on research about complex energy systems, e.g. of a country. Existing energy systems and possible changes to energy systems are studied by creating different scenarios and analysing the behaviour of models of energy systems in these scenarios. In particular, they are used to assess impacts such as costs, greenhouse gas emissions, demand for resources or feasibility of those systems. They provide the decision basis for long term projections either to support (1) policy decision making from a macroeconomic perspective, (2) mid-term investment decision from a single business perspective, or (3) control decisions in real time operation. Therefore, the domain applies mathematical models to simulate or optimise different aspects of the energy systems, especially energy supply and demand, potentially at various levels of detail. The term *energy systems model* has been established for the entirety of tools and applications that perform these computations.

Energy systems models and scenarios are quite heterogeneous in scope, focus, methodology and level of detail. Some studies focus mainly on electrical energy systems whereas others integrate different sectors or commodities simultaneously, e.g., heat, chemicals or transportation. The temporal and spatial resolution levels may also vary significantly. Some research focuses on the behaviour of different actors in the energy supply chain and their influence on the energy markets, whilst others are concerned with the monetary impact of renewable energy sources on the energy prices for government or consumers. External effects of the energy systems on the environment, for example greenhouse gas emissions, can also be part and sometimes the main interest of scenario-driven energy systems analyses. The projections can address far-off targets like climate neutrality, or evaluate the development of the energy system over the next couple of years. This heterogeneity is partially explained by the different research scopes but also by modelling limitations due to the mathematical complexity of the energy systems [14]. Consequently, the coupling of different models is required in order to answer more complex research questions, but this is challenging to implement due to inconsistent data and interface descriptions of the individual models.

A number of assumptions form the foundation for those computations, which – in turn – guide political and business decisions. These assumptions can include future developments of carbon prices, subsidies, climate change impacts, the spatial distribution of energy plants and their technological advances. A critical analysis and comparison of different energy systems models is required to justify these decisions, but is hindered by the inherent heterogeneity of the energy domain [11]. This problem is especially pertinent for that part of the energy modelling community that is focused on open source and open data solutions – the open energy community.

Recently, the awareness of open science principles has risen throughout the energy systems modelling community. The publication of source code and data increases the transparency and reproducibility of energy systems modelling. But it also reveals some of the inconsistencies in how certain terms and concepts are used within the community [7]. Partially, this is due to the different scientific areas that intersect in the energy systems domain: physics, engineering, economics, meteorology and climate research.

These problems may be solved by the introduction of a comprehensive ontology for the energy systems domain. An ontology includes a taxonomy of terms with definitions that clarify the intended meaning of those terms. These terms can be used to annotate data and define interfaces that address existing integration problems between different tools. The taxonomy is further enhanced by logical axioms in the ontology. The logical formalisation enables the inference of certain constraints that prevent inconsistencies and may make knowledge and data that is currently implicit more explicit. Ontologies are successfully being applied to solve similar problems in biomedicine, for example the annotation of gene functions [2] and chemicals [5].

3 Design and Structure of the OEO

3.1 Ontology Background, Context and Outline

The *Open Energy Family* provides an open source toolbox for open data within the field of energy systems analysis research. At the heart of this effort is the *Open Energy Platform* (OEP)[9], a collaborative online platform with an underlying database for energy and climate analysis data. The scope of the OEP includes any kind of data set that is of relevance for energy systems analysis, as long as it is published under an open license. The data covers a wide range of subjects important for modelling of energy systems including the energy grid (e.g., power lines, substations, power plants), energy demand, climate (e.g., wind,

[9] `https://openenergy-platform.org/`

precipitation, temperature, cloud coverage), geography (e.g., administrative and political boundaries) and economy (e.g., GDP and price of commodities). In addition, the OEP provides information about energy scenarios and output data from computations. The OEP serves as a reference and facilitates scientific and political decision making due to an improved level of transparency and comparability. The data sets in the OEP reflect the wide variety of academic interests and backgrounds of the members of the community who provided them. One incentive for the development of the OEO is to provide some level of integration and enable search across the different datasets within the OEP.

The OEO is written in the Web Ontology Language (OWL) and consists of about 900 classes. About 300 of these are specific to the OEO, the rest are reused from other ontologies as described in Section 3.3. There are around 80 object properties; about 50% reused from other ontologies. In total, the OEO contains over 8000 axioms. The ontology can be accessed via GitHub[10] and its official releases are published on the OEP[11].

3.2 Approach and Best Practices

The OEO is built on top of the Basic Formal Ontology (BFO), a foundational ontology that describes basic cross-domain types of entity, such as objects and processes [1]. In addition, the OEO also adopts best practices and design principles in line with the broader community of the OBO Foundry [25].

The ontology has a modular organisation (see Section 3.3). The classes are introduced within a tree-like taxonomic structure; however, multiple inheritance may be inferred from logical axioms. OEO entities are assigned semantics-free alphanumeric identifiers in the namespace *OEO:x* (where x is a unique number). Each entity in the ontology is assigned a unique *label* and a textual *definition*.

When possible, the OEO strives to reuse existing resources. As illustrated in Fig. 1, the OEO imports parts of four other ontologies (BFO, the Information Artifact Ontology (IAO) and its subset the Ontology Metadata Ontology (OMO), the Relationship Ontology (RO), and the Unit Ontology (UO). For this purpose we use the ROBOT library [12] to extract custom sub-modules, which contain exactly the content that is required. Further, some content of the Financial Business Ontology (FIBO) is reused, but not directly imported. The reuse of resources will be discussed in more detail in the next section where we explain the structure of the OEO and its main modules.

One goal of the OEO development team is to achieve transparency. The discussions and decisions about changes to the ontology are documented in Github's issue tracker. Further, we reuse the "term tracker item" annotation from the IAO to link entities in the ontology to GitHub issues and pull requests that defined or changed the entity. This approach enables rapid access to information about the history of a class, as well as the discussions that took place around it.

[10] `https://github.com/OpenEnergyPlatform/ontology/`
[11] `https://openenergy-platform.org/ontology/`

3.3 Structure and Submodules

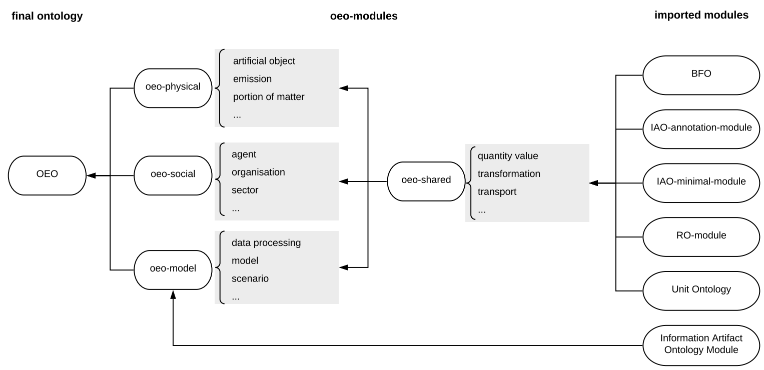

Fig. 1. Submodules and imports of the OEO

Fig. 1 illustrates the structure of the OEO. Its main modules cover three aspects of energy systems analysis: 1) models and data, 2) socio-economic aspects and 3) the physical side of energy systems. Each of these three main modules imports a shared module, which contains content that is used by more than one of them. The three main modules are imported into the main ontology, which adds relations between the separate modules. The modular structure of the OEO reduces the likelihood of clashes during concurrent editing of the ontology. In addition, the modular structure helps to avoid modelling mistakes by enforcing a clear distinction between objects in the real world (e.g., the electrical grid) and information artefacts (e.g., a model of the electrical grid). Since research on energy systems is both concerned with the facts and the results of models and simulations, both need to be represented, but need to be distinguished. In the following we will discuss the four modules in more detail.

The *oeo-model* module comprises all entities related to data and energy systems models. Beyond a taxonomy of models, most entities defined in this module relate to either transformations of data or information entities, e.g. model calculations and the data processing methods used in energy systems models. This module reuses parts of the Information Artifact Ontology (IAO).[12] In particular, the *oeo-model* module reuses the taxonomy of different information content entities (such as documents, data items, symbols and figures). This taxonomy is extended by classes that are specifically relevant for the energy systems analysis domain; e.g., the scenario class as well assumptions and constraints of scenarios.

The *oeo-social* covers socio-economic aspects of energy systems. Included are basic classes such as "population" and "organisation" as well as their social roles;

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

e.g. “author”, “producer” and “user”. The multiple and complex ways of classifying and defining sectors in energy systems modelling are reflected by the “sector” and “sector division” classes. Since economic conditions have a major impact on factors like energy prices and energy consumption, the OEO also needs to include economic terms. To cover these, we used the *Financial Industry Business Ontology* (FIBO[13]) as source. Since FIBO is not aligned with BFO, directly importing parts of FIBO would lead to inconsistencies. Instead, we largely re-used the definitions from FIBO but made minor adjustments to the definitions of the entities in order to fit them into the framework of BFO.

The *oeo-physical* module covers the physical world of energy systems; e.g., power plants, batteries, or energy carriers. Since many of these entities are physical objects, many classes in oeo-physical are subclasses of BFO’s “material entity” class. Power plants are categorised by their inputs, e.g. wind farms or biofuel power plants. One important topic within this module is matter, materials and fuels, such as coal, peat, water and methane. Linguistically, these are mass terms that are notoriously difficult to model in logic-based languages like OWL [20]. To address this issue we represented them as kinds of “portion of matter”, which we introduced as a subclass of BFO’s “object aggregate”. Using axioms that enable automated classification these materials are arranged into different categories based on their properties and capabilities, such as greenhouse gases or fuels. In particular, fuels have been categorised into detailed subtypes such as biofuels, renewable fuels or nuclear fuels.The related entities for greenhouse gas emission and pollution are defined as subclasses of BFO’s “process”. To describe quantitative amounts of physical entities, the Unit Ontology [9] was imported into this module. It defines power units and energy units, thus usefully covering an important part of the energy systems domain.

The *oeo-shared* module covers entities that are used in more than one of the main modules. The majority of these entities are imported from existing resources. We have already mentioned BFO and the Unit Ontology. Another imported module is based on the *Relations Ontology (RO)*, and contains a subset of the object properties defined by the RO [24]. These include some mereological, temporal and spatial relationships, as well as some other domain independent object properties, e.g., ‘has disposition’ and ‘has output’. Further, oeo-shared contains all metadata annotations defined by the Information Artifact Ontology, including the “term tracker item" annotation that we mentioned in Section 3.2. Based on the imported modules, oeo-shared adds some additional entities that are shared across the main modules, some of them are domain independent (e.g., “transformation” and “process attribute of”), and some of which are specific to energy systems (e.g., “primary energy consumption”).

3.4 Outreach, Sustainability and License

Consensus across perspectives is important for acceptance and usability of an ontology, thus it is essential to include a diverse group of contributors during

[13] `https://spec.edmcouncil.org/fibo/`

development. The OEO development process is therefore focused on openness and accessibility. The ontology is developed on GitHub[14] and all developments are discussed in issues that are publicly accessible. Discussions involve experts with different backgrounds and expertise. This ensures that new definitions match with a wide range of experts' views of the domain. Development in such a heterogeneous team raises the need for guidance during the development process. Guidelines[15] have been established that regulate the way new entities and definitions can be proposed, issues and bugs can be reported and ultimately changes to the ontology are applied. A steering committee[16] has been established that includes stakeholders from different institutions and communities. This, on the one hand, raises the awareness of the ontology within the community and, on the other hand, enables these stakeholders to influence the direction of future developments. The steering committee will also ensure the sustainability of the ontology development beyond the scope of specific projects or funding awards. The involvement of different stakeholders enables the development of an ontology that may be used in different applications. This is also reflected in the choice of the ontology license. A CC0 license allows the use of the ontology in a wide range of contexts including proprietary and commercial tools. The ontology can therefore also be used as a mediator between data sets and solutions arising within these different contexts.

3.5 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!, [22]) 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[17] and perform a number of quality checks such as consistency and coherence[18]. These checks control the general quality of the ontology – but are agnostic w.r.t. 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 6). Each contribution to the ontology is automatically checked against the robot profiles, competency questions and for consistency.

[14] https://github.com/OpenEnergyPlatform/ontology
[15] https://github.com/OpenEnergyPlatform/ontology/wiki
[16] https://openenergy-platform.org/ontology/oeo-steering-committee/
[17] http://robot.obolibrary.org/validate-profile
[18] http://robot.obolibrary.org/report

4 Challenges During Ontology Development

One of the challenges encountered during the development of the OEO has been the breadth and complexity of its domain, which covers a broad range of entities of different types. In this respect, it is unlike many of the most widely used ontologies, which cover just a single or a small number of entity types. For example, the Gene Ontology [2] consists of more than 40,000 classes (as of November 2020), but covers only three types of entity: biological processes, molecular functions and cellular locations. In contrast, the OEO covers diverse sorts of entities such as energy, portions of matter, technologies, physical attributes, measures, units, models, parameters, sectors, economic attributes, and others. For each of these different entity types, challenges can arise with determining the best formulation for that entity in terms of its classification, definition and relationship to other entities. Moreover, the nature of the energy systems domain means that even the physical entities that are within scope range over very complex phenomena which can be difficult to define and represent.

The central topic for the energy systems modelling domain is of course energy. Energy is a term that can mean different things in different contexts, and there are many different ways to define energy, as can be witnessed in any dictionary. The OEO has defined *energy* as *'A quality of matter and radiation which is manifest as a capacity to perform work.'* The decision to classify energy as a *quality* reflects that energy is regarded as a *dependent* entity. This was a subject to debate at length; there are of course perplexing issues relating to the nature of matter and energy that arise from fundamental physics, but these must be avoided if one aims to make pragmatic progress. The energy class lists as examples of energy *'causing motion or the interaction of molecules'*. An additional comment is included to elaborate on the mathematical connection with other entities: *'Energy is power integrated over time.'* Different subclasses of energy are included, such as chemical energy, kinetic energy, electrical energy and thermal energy.

The challenge of polysemy pervades the vocabulary associated with energy systems. Another example is *fuel*, which may refer to a type of *substance*, but may also be used to mean a way that a substance can be *used* in a certain context. Thus, there are some substances that are always described as fuels, such as coal or petrol, but there are other substances that are usually never associated with fuel but nevertheless may be used as a fuel in the right context, such as waste. For this reason, we defined fuel as follows: *'A portion of matter that has the disposition to be an energy carrier and which has a fuel role that is realised in processes that release the carried energy by transforming the portion of matter into a different kind of portion of matter in a way that releases heat or does work.'* The important aspect of this definition is the connection between a portion of matter to its *use* in the relevant type of process, which is a *role* in BFO terms.

Another challenging area for the representation of entities in the energy systems modelling domain is that the distinction between entities at different levels of granularity is not clear in the terminology used in the domain. For example, the term *generator* is usually used to mean the specific technological unit that converts motive power into electrical energy, but in some cases the term is used to mean complex aggregations of technological units forming an overarching component in an energy system, for example, a windmill or even a whole power plant consisting of several windmills. In the OEO, the class *generator* (OEO:00000188) refers to the specific component, and there is a separate class *power generating unit* (OEO:00000334) that is defined as having a generator as a part.

As the energy systems modelling domain deals with models that simulate or optimise energy demands, production and usage over periods of time, the representation of time series and time steps is important. The nature of the temporal information in such system models does not, however, always map exactly onto the nature of the temporal information in other domains. For example, model time steps may have variable durations depending on the data or on the progress of the steps in a sequential simulation model. For time series, values may be aggregated (integrated or averaged over the duration of the time step) or represent momentary values. These reasons made it difficult to reuse existing resources for temporal information such as OWL-Time[19]. For example, OWL-Time requires the specification of temporal information with regard to a reference that maps onto actual dates and times. Thus, it is possible to specify temporal durations (such as days or weeks) with OWL-Time, but if a temporal duration of a day is specified, then that day should start and end at midnight. In contrast, in the energy systems domain, a day can mean *any* 24-hour period. Hence, we could not use OWL-Time but had to develop our own formalisation.

The temporal representation example illustrates the different representational layers in this domain: entities in the world, information content entities that are about entities in the world, and information content entities that are models, simulations and approximations. The modelling branch of the ontology that explicitly includes information about models and their associated inputs and outputs, adds an additional layer of complexity to the ontology, because an energy systems model may used to study situations and conditions that deviate from reality. The distinction between the entity as it is, needed for annotation, and the entity as it appears in models and simulations, relates also to the formalisation of axioms in the ontology. There has to be a clear distinction between – for example – a real-world power plant and its representation in a specific energy systems model. The latter may *refers to* the existing power plant, but is in fact an information content entity.

For energy systems modelling, a quantitative description of certain entities, especially from the *oeo-physical* branch, is necessary. E.g., we may need to indicate how many gigawatts of wind power plants are assumed to be installed in a region for a certain scenario. To represent such quantitative information, the class *quantity value* was created as an information content entity that contains a numeral together with a unit of measurement. Quantity values relate to gaugeable properties of physical entities, like *height*, *coefficient of performance* or *declared net capacity*.

[19] https://www.w3.org/TR/owl-time/

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%

5 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. Macro-economic data such as population, gross domestic product and energy prices is 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 [8]. 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 which 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 sub-classes (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" is not yet included in the OEO, whereas "(portion of) coal" is. *Good match* indicates a full match.

The evaluation results of the coverage study are shown in Table 1 and have been made publicly accessible[20]. In total, the annotation of 153 fact sheet fields was tested, as depicted in the first table row (*"ALL"*). More than half of the

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

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 3, 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 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.

6 Evaluation II: Competency questions

Competency questions provide a methodology for capturing and evaluating semantic requirements for an ontology[10]. 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 [19]. 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 3.5.

For example, as mentioned in Section 4 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; of these currently 41 are answered successfully. The answer of the remaining nine competency questions is 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 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 [26]. 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 5 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.

For this study, we only included data from participants who fully completed the study. Of the 20 such participants, 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.[21]

[21] `https://doi.org/10.5281/zenodo.3870654`

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 [15]. Using this metric, the inter-annotator agreement for our study was $\kappa = 0.668$. According to the classification in [17] 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 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 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 Related Work

The Open Energy Ontology is the first publicly available and open ontology for the energy systems domain. An ontology had been developed within the *EnArgus project* [21] as part of an information system[22] that collects data on energy science related projects and funding. The publicly available wiki for this project shows a number of entities that are useful for the domain of energy science. However, the EnArgus ontology is proprietary, which limits its usability in an open science and open data context. Moreover, EnArgus is mainly a simple taxonomy, with only very few object properties and essentially only hierarchical axioms.

Technical objects are a fundamental part of the energy domain, e.g., power generators and energy grids. Some of these have been modelled as parts of the Internet of Things (IoT), which includes ontologies covering the topics of

[22] `https://www.enargus.de/`

urban planning and energy management [18], energy efficient buildings [13] and distributed energy sources[23].

The domain of energy modelling is comprised of several sub-domains for different areas including social sciences, economics, ecology, climate science. An ontology for wind energy [16] bridges one of the gaps between the physical energy domain and the climate domain. Yet, numerous subdomains are still not integrated into the the ontological context of energy systems models. The FIBO ontology [3] - as discussed earlier - 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. An ontology for electricity markets [23] attempted to bridge the gap between the economical domain of markets and the technological domain of energy systems. Yet, these ontologies include definitions that are tailored too tightly to electricity markets and do not cover other energies, e.g. heat. Most of the above ontologies consider a limited fraction of different energy systems. The OntoPowSys [6] ontology follows a more integrated approach. It as developed as part of the larger *J-Park Simulator*[24] which supports efforts towards more sustainable, carbon-neutral industry parks. The main focus of this project is the development of a knowledge management system for these parks that increases the interoperability between different industrial park components. The resulting ontology is focused on generation, transport and consumption of electricity as well as monetary aspects. Recent developments within the open energy systems community show the need for a more holistic view on energy systems, which includes the integration of other kinds of energy. The OEO aims to cover these different energy systems as well as additional social and political aspects.

This shows that there are existing ontologies for some specific facets of the energy systems domain but the Open Energy Ontology is the first to attempt to cover it as a whole.

9 Conclusion and Future Work

We reported on the development-in-progress as well as the syntactic and semantic evaluation of an open community-driven ontology for the energy systems analysis domain. While ontologies are not completely novel to this domain, 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. Following the advice of the 'ten simple rules' [4], we conducted an evaluation in the spirit of 'evaluate early, evaluate often'. Our findings show that we are moving in the right direction. However, work still remains to be done to enhance both the coverage of the ontology and the quality of its documentation. Upcoming releases will expand the socioeconomic aspects of the energy systems domain, harnessing content from already existing ontologies where possible to comprehensively describe socioeconomic entities. Energy models themselves, including assumptions,

[23] https://innoweb.mondragon.edu/ontologies/dabgeo/common-domain/ders/1.0/index-en.html

[24] http://www.theworldavatar.com/

algorithms and modelling apporaches will also be in the focus of future releases. In addition to these extensions we will formalise different fact sheets that have been developed by domain experts in an RDF knowledge graph that is based on the OEO. Generally, the Open Energy Ontology will be used for metadata annotation of energy systems modelling data (cf. the Open Energy Platform).

The Open Energy Ontology has been developed as a first holistic, open domain ontology for the energy systems domain. The inclusion of the open energy community in the development team and the steering committee guarantees a good consensus for specific definitions and design decisions as well as a good foundation for future applications. Open solutions are playing an increasing role in energy science with wide-ranging benefits over and above the practical objectives of re-use and reproducibility. An open ontology can be the mediating factor between different sub-communities and aid in the integration of results and data sets[25], whereby the latter is of increasing importance in the wake of the rising number of open data platforms, like the Open Energy Platform[26], Open Power System Data[27], and Renewable Ninja[28]. These benefits will help to increase the transparency of data from energy systems analysis and, therefore, improve trust in the respective results. Trust and openness is especially important in the context of the ongoing climate crisis and the challenges of diminishing public trust in scientific results. There is a need for robust, reproducible evidence that can be amalgamated and compared across different modelling approaches and stakeholder groups. As discussed, the Open Energy Ontology is an essential tool to address that need.

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

References

1. Arp, R., Smith, B., Spear, A.D.: Building Ontologies with Basic Formal Ontology. MIT Press (2015)
2. Ashburner, M., et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics **25**(1), 25–29 (May 2000)
3. Bennett, M.: The financial industry business ontology: Best practice for big data. Journal of Banking Regulation **14**(3-4), 255–268 (2013)
4. Courtot, M., Malone, J., Mungall, C.: Ten simple rules for biomedical ontology development. In: Proceedings of ICBO (2016)
5. Degtyarenko, K., et al.: ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Research **36**(Database issue), D344–D350 (Jan 2008). https://doi.org/10.1093/nar/gkm791
6. Devanand, A., Karmakar, G., Krdzavac, N., Rigo-Mariani, R., Eddy, Y.F., Karimi, I.A., Kraft, M.: Ontopowsys: A power system ontology for cross domain interactions in an eco industrial park. Energy and AI p. 100008 (2020)

[25] An ontology-based data bus is currently developed as part of the project LOD-GEOSS. See: https://reiner-lemoine-institut.de/en/lod-geoss/

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

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

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

7. Emele, L., et al.: Conversion Necessities in Climate and Energy System Modelling (Jul 2020). https://doi.org/10.5281/zenodo.3949692
8. Euzenat, J., Shvaiko, P.: Ontology Matching, Second Edition. Springer (2013)
9. Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The Units Ontology: a tool for integrating units of measurement in science. Database (Oxford) (2012)
10. Grüninger, M., Fox, M.S.: Methodology for the design and evaluation of ontologies. In: Proc. of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95 (1995)
11. 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 **22**, 264–269 (2018)
12. Jackson, R., Balhoff, J., Douglass, E., Harris, N.L., Mungall, C.J., Overton, J.A.: ROBOT: A tool for automating ontology workflows. BMC Bioinformatics **20**, 407 (2019). https://doi.org/10.1186/s12859-019-3002-3
13. Kofler, M.J., Reinisch, C., Kastner, W.: A semantic representation of energy-related information in future smart homes. Energy and Buildings **47**, 169–179 (2012)
14. Kotzur, L., et al.: A modeler's guide to handle complexity in energy system optimization (2020)
15. Kraemer, H.C.: Extension of the kappa coefficient. Biometrics pp. 207–216 (1980)
16. Küçük, D., Küçük, D.: OntoWind: An improved and extended wind energy ontology. arXiv preprint arXiv:1803.02808 (2018)
17. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. biometrics pp. 159–174 (1977)
18. Madrazo, L., Sicilia, A., Gamboa, G.: Semanco: Semantic tools for carbon reduction in urban planning. In: Proc. 9th Europ. Conf. Product & Process Modelling (2012)
19. Neuhaus, F., et al.: Towards ontology evaluation across the life cycle – the ontology summit 2013. Applied Ontology **8**(3), 179–194 (2013)
20. Nicolas, D.: The Logic of Mass Expressions. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Stanford University, winter 2018 edn. (2018)
21. Oppermann, L., Hinrichs, E., Schade, U., Koch, T., Rettweiler, M., Ohrem, F., Plötz, P., Beier, C., Prinz, W.: EnArgus: Zentrales Informationssystem Energieforschungsförderung. INFORMATIK 2015 (2015)
22. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. International Journal on Semantic Web and Information Systems (IJSWIS) **10**(2), 7–34 (2014)
23. Santos, G., et al.: Electricity markets ontology to support mascem's simulations. In: Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection. pp. 393–404. Springer International Publishing, Cham (2016)
24. Smith, B., Ceusters, W., Klagges, B., et al.: Relations in biomedical ontologies. Genome Biol. (2005). https://doi.org/10.1186/gb-2005-6-5-r46
25. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology **25**(11), 1251–1255 (2007)
26. Stevens, R., Lord, P., Malone, J., Matentzoglu, N.: Measuring expert performance at manually classifying domain entities under upper ontology classes. Journal of Web Semantics **57**, 100469 (2019)