The Open Energy Ontology

Lukas Emele[1], Hannah Förster[1], Johannes Frey[2], Ulrich Frey[3], Martin Glauer[4], Janna Hastings[4], Christian Hofmann[5], Carsten Hoyer-Klick[3], Ludwig Hülk[5], Anna Kleinau[4], Patrick Kuckertz[6], Till Mossakowski[4], Fabian Neuhaus[4], Martin Robinius[6], and Mirjam Stappel[7] (alphabetical)

[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, Kassel, Germany

Abstract. We introduce and evaluate the Open Energy Ontology (OEO) which we develop for the energy system analysis domain. Energy system analysis uses computational models to create scenarios reflecting possible future developments in distributed networks of energy supply and consumption. To date, the energy system analysis domain is still fragmented and it is difficult to integrate results across studies.
The goal of the OEO is to build a common and shared conceptualisation that will be used in the energy system analysis community for multiple purposes, including annotation of the large amounts of data that result from various research projects assembled – for example – on the Open Energy Platform (OEP). Adding annotations will make this data semantically searchable, exchangeable, re-usable and interoperable.
We present the OEO ontology structure and content, and evaluate it for coverage of terms from the OEP fact sheets, and for annotation agreement among experts. We also describe how the ontology will be used for Linked Open Data.

Keywords: Collaborative ontology development · Ontology evaluation · Linked open data · Metadata annotation · Energy system analysis.

1 Challenges within a Heterogeneous Domain

The transition to sustainable energy systems and achieving the COP21 climate goals[8] are global societal challenges. Science has the task of supporting this ongoing transformation through objective assessment (for example detailed assessment of progress towards climate and energy targets as in [5]), new findings and innovative methods and strategies. Against this background, energy system

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

analysis is faced with complex research questions about scenarios concerning the state and development of the energy system and its transition to more and more (up to 100%) deployment of renewable energy. These research questions cannot be answered by individual scientists, institutions or even countries, thus the energy system analysis community is dependent on networking and cooperation, and needs to ensure an extensive and frictionless scientific exchange. However, this is not an easy task since scientists working in this area come from a wide range of diverse disciplines, including engineering, natural and social sciences, physics, mathematics, computer science, economics, meteorology and geography.

Energy scenarios are one of the main research outputs of energy system analysis. They contain statements about possible future developments of the energy system, based on a coherent and internally consistent set of assumptions, including their motivations. Energy system models map aspects of the (future) energy system, and accompany scenarios in many cases. Both are used to conduct studies and experiments and have distinctive characteristics in which they often differ greatly. Regardless of their differences, the great variety of scenarios and models depict the range of possible developments for the future energy system and their manifold aspects. A single scenario or model by itself is not able to map all relevant aspects of energy systems with sufficient accuracy. The regional scope of energy scenarios and models can vary from small municipalities to whole countries or continents. Although their specific focus is often on electricity supply, the boundaries of the investigated system may be extended to also account for e.g. heating, cooling or gas supply. Thereby, they can span different sectors such as industry, residential or mobility, and depict technological, economical, ecological and social aspects. The scenarios also differ in their respective levels of detail, e.g. their temporal and spatial resolutions. Even if two scenarios have the same characteristics and basic definitions, they still often differ in their underlying approaches.

The data used as input for scenarios and models originates from a large variety of data sources belonging to many different domains. Data is provided by public agencies, gathered from scientific papers and commercial or public databases, stem from crowd sourcing initiatives or are measured by researchers themselves or by remote sensors. The respective data formats range from single values or time series to multidimensional fields, thereby representing information in various spatial and temporal resolutions, e.g. hourly wind speeds at various sites and various heights above ground. In addition to this extensive data basis, the energy system 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 promotes 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.

As of yet, there is no ontology tailored to energy system 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 [9,10,11]. The *Open Energy Ontology* (OEO) has been developed with the goal of easing cooperation and exchange of information across the energy system analysis domain. The OEO is designed to map the complexity of the research area and to organise the ambiguous terminology of this domain. Its steady growth increasingly enables the precise, unequivocal and comprehensible annotation and interpretation of research data. Serving as a basis for an international and frictionless scientific exchange, the OEO enables the consolidation and reuse of distributed data inventories across domains, thereby harnessing synergies within the global and interdisciplinary energy system analysis community and supporting the robust transition to sustainable energy systems.

The remainder of this article is organised as follows: Section 2 provides an overview of open and collaborative approaches for ontology development and ontologies connected to the energy domain. In Section 3, the OEO is introduced as a domain ontology for energy system modelling and analysis, and the chosen design patterns and content structures are described. Section 4 elaborates on the OEO's open collaborative development processes and on how they are embedded in the energy system analysis community to ensure its sustainable development. Thereafter, Section 5 describes two use cases which are currently part of the third-party funded LOD-GEOSS research project. In Section 6 the setup of two evaluation experiments is explained. While the first is concerned with the ontology's coverage of domain terminology, the second addresses its current usability for domain experts. The results of the evaluation are then discussed in Section 7. Section 8 closes this article with general conclusions and an outlook of future work.

2 Related Work

2.1 Open and Collaborative Ontology Development

Ontologies such as the OEO that are developed to serve a scientific community as a whole rely on workflows, standards and technologies that enable collaborative development. Many ontology development methodologies have been proposed (e.g. [7, 9, 20, 22]). 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 [4], one of the most

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

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

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

important aspects of good ontology development is to re-use existing ontology content as far as is possible. This allows cumulative extension of the available knowledge resources without duplication of effort. Hence, the OEO has been designed to import relevant content where possible. Having a clear approach to ontology versioning and using an 'open' license are also key to many of the methodological recommendations, as is using a version control system for keeping track of the ontology's evolution. Furthermore, such methodologies typically include recommendations for setting the scope of the ontology, and for evaluation, which 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 [19] is an initiative in the biological and biomedical domain which has brought together ontology authors to create a set of design principles and standards for ontologies which can be semi-automatically verified. These design principles and standards have also allowed the implementation of tools such as the ontology library ROBOT [10] which automates many common ontology development tasks. While the OEO addresses a different domain to the biomedical, many of the standards which we have adopted in its development are based on those developed for the Foundry. For example, we re-use Foundry metadata standards and common relationships.

2.2 Ontologies in the Energy Domain

The only well-known terminological resource for energy system analysis is the *EnArgus Ontology* [15]. 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 has been collected in a semi-automatic fashion. The related wiki promises a rich resource of useful terms and definitions. However, the EnArgus Ontology is proprietary and has not been made available to the community. Therefore, it cannot be reused by other energy system analysis projects, and in particular not by *open* energy system analysis projects. Based on the publicly available information in the wiki, the EnArgus ontology consists mainly of a subclass hierarchy and is only lightly axiomatised.

Energy markets and price developments are a central part of many energy system models. Electricity markets are the subject of the *Electricity Markets Ontology* [17], and financial markets of the *Financial Industry Business Ontology* [2]. Recent developments in energy system 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 this is added, pre-existing ontologies may be harnessed where possible, supplemented by additional content according to the 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 [14] and the Energy Resource Ontology [11] cover energy-related aspects of the housing sector. Other physical systems, their relations and properties are modelled in the SEAS ontology [13], which was developed as a generalisation of the semantic sensor ontology (SSN) [3]. Many sources of renewable energy depend on some kind of meteorological phenomena and most energy simulations involve assumptions regarding weather and climate to predict 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[12].

In summary, no publicly available ontology exists covering the full energy system analysis domain. The OEO addresses this gap.

3 General Design Choices of the OEO

3.1 Ontology Background, Context and Outline

The OEO was created as a part of the *Open Energy Family*, an open source toolbox and database for open data within the field of energy system analysis research. This toolbox is built around the *Open Energy Platform* (OEP)[12], a collaborative online platform with an underlying database for energy and climate analysis data. A wide range of data types, from single energy data sets to complete energy scenarios, can be uploaded to the database. All data sets are published under an open license and become freely and easily accessible to others. The OEP serves as a reference and facilitates scientific and political decision making due to an improved level of transparency and comparability.

The OEO has been developed within the project *SzenarienDB*, augmented by the project *LOD-GEOSS*. The project *SzenarienDB* extends the functionality of the *Open Energy Platform* to a transparent and user friendly database for energy scenarios [16]. As an essential part of the energy system analysis domain, scenarios are complex and heterogeneous and thus are in need of an ontology as a common understanding. The aim of the project *LOD-GEOSS* is to create a network of heterogeneous databases for input and output data from energy system 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 meta data catalogue which makes the data findable and accessible. Both projects implement the FAIR principles[13] of open data to energy system analysis data.

The OEO is developed using the the Web Ontology Language (OWL). It contains around 900 classes. About 300 of these are OEO-owned classes, while the remainder is imported from one of the external ontologies as described in Section 3.3. There are around 80 object properties. About 50% are imported and 50% are created internally. In total, the OEO contains over 8000 axioms.

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

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

The first official release 1.0 of the OEO is scheduled for June 2020. The ontology can be accessed via GitHub[14] and its official releases are published on the OEP[15].

3.2 BFO, Design Patterns and Best Practices

As is common practice for many scientific ontologies, the OEO is structured based on a shared 'upper level' or foundational ontology that describes basic cross-domain types of entity, such as objects and processes. 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', such as qualities and other attributes.

As mentioned above, the OEO also adopts ontology design patterns and best practices, in line with the broader community of the OBO Foundry [19]. Best practice principles concerning terminology, definitions and taxonomy were also derived from the book [1]. The ontology has a modular organisation (described in Section 3.3) and maintains as far as possible a single asserted superclass taxonomic structure, although in some cases additional superclasses can be inferred from logical axioms. 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 in order to prevent clashes during concurrent editing. Each entity in the ontology is assigned a unique *label* and a text *definition*, while additional synonyms, comments and logical axioms may be included if needed.

3.3 Structure and Submodules

The OEO consists of three main modules (Fig. 1) covering these aspects of the energy system analysis domain: 1) models and data, 2) social and economic aspects and 3) the physical side of energy systems. All modules are imported into the main ontology, which adds relations between the separate modules.

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, e.g. model calculations and the data processing methods used in energy system models. Information-related entities included in this module are largely an imported subset of the Information Artifact Ontology[16]. This imported module includes

[14] `https://github.com/OpenEnergyPlatform/ontology/`

[15] `https://openenergy-platform.org/ontology/`

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

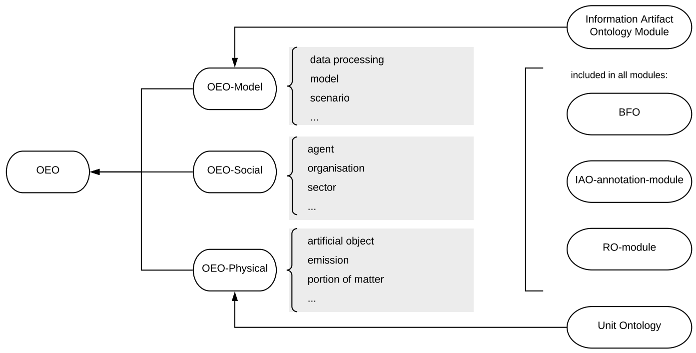

Fig. 1. Submodules and imports of the OEO

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, for example, the scenario class and different types of data descriptors, as well as assumptions and constraints.

The *oeo-social* module depicts social roles and entities to describe the social and economic aspects of energy systems. Included are basic classes such as "person" and "organisation" as well as sectors. Sectors are implemented as a combination of a "sector" class alongside overarching "sector divisions" that delineate which sectors are relevant within a particular context. Different kinds of roles for people or organisations in the domain are defined beneath "agent", including "author", "producer" and "user". An important kind of organisation for the energy systems domain are energy producers, implemented though the "organisational energy producer" class and its subclasses.

The *oeo-physical* module includes all entities related to the physical world of energy systems. Generators, batteries, different materials and technologies are covered. Most entities described are physical objects and therefore subclasses of BFO's "material entity" class. One important topic within this module is matter, materials and fuels. These are represented beneath a "portion of matter" class. Here different materials such as coal, peat, water and methane are defined. Using axioms that enable automated classification based on logical equivalences, 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". The ontology also includes artificial objects such as batteries and power plants. Power plants are categorised by their inputs, e.g. wind farms or biofuel power plants. To describe quantitative amounts of physical entities, the Unit Ontology [8] was imported into this module. It defines

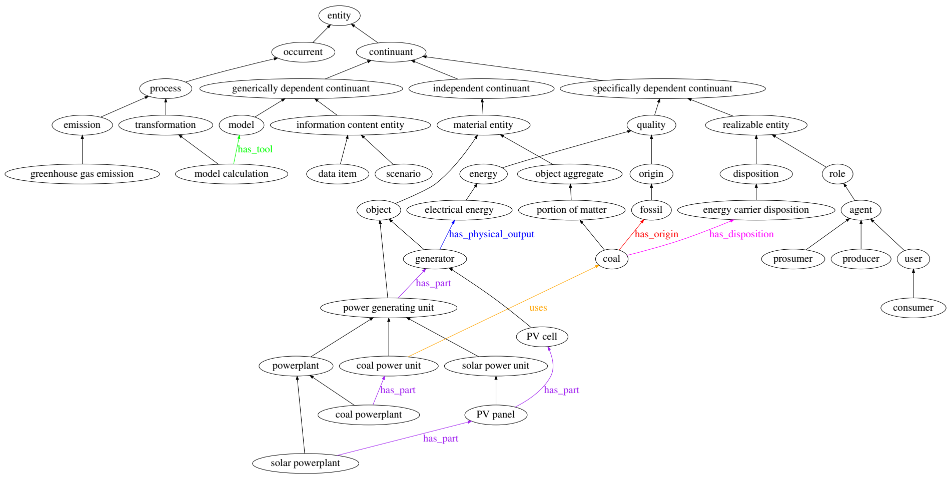

Fig. 2. Overview of a subset of classes and properties of the OEO to illustrate how they are organised inside the OEO. A black arrow denotes "is a", i.e. a subclass relation.

power units and energy units, thus usefully covering a part of the energy systems domain.

As mentioned above and shown in Fig. 1, the OEO imports parts of other ontologies to avoid "re-inventing the wheel". Reuse is facilitated by using the ROBOT library [10] to extract just the needed content as sub-modules. Aside from BFO, two other ontologies are also reused. First, the *Relations Ontology (RO) module* contains a subset of the object properties defined by the Relations Ontology [18]. We chose to only include a subset of RO, as many of the relations are not relevant for energy system analysis. Examples of object properties imported 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, the OEO contains all metadata annotations defined by the Information Artifact Ontology. This module includes standardised annotations such as the "term tracker item" annotation that 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.

Fig. 2 shows some of the classes and properties of the final structure of the OEO, indicating that beyond a mere taxonomy, there is a rich set of properties (relations) linking classes. If a relation just affects classes of one specific module it is defined in that module, while relations that link classes of different modules together are defined in the parent OEO file.

4 Open Collaborative Development

As discussed in 3.2, the OEO follows the OBO principles[17] and thus, is developed as an interdisciplinary, collaborative, public and open source[18] project. The chosen workflow reflects these characteristics, with a special focus on openness. All technical discussions and developer meetings are held publicly on the project's GitHub page[19] and anyone is invited to contribute. Furthermore, a steering committee with experts from different related disciplines was created to guide the development of the ontology.

4.1 Git Workflow

The development of the OEO takes place mainly on GitHub. Detailed manuals for usage[20] and contribution[21] allow new collaborators and users a facilitated entry to the ontology, and describe the workflow, which ensures quality and traceability of decisions. The workflow requires that every suggested change to the ontology has to be discussed in an issue before making the actual change. Issues are 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 project, larger changes at least three. These members should include one domain expert and one ontology expert. In order to reflect the diverse background of the members and facilitate rapid group formation when tackling an issue, developers can join GitHub teams in their fields of expertise. There are teams for the fields of economy, energy modelling, linked open data, meteorology and ontology. If an agreement is hard to reach within the issue discussions, it is then added to the agenda of the next ontology developer meeting / telephone conference.

The procedure is designed to be slow, but thorough. After an issue's solution is agreed upon, technical implementation of the change can follow a quick protocol and can be carried out by any member.

4.2 Community Embedding

The workflow on GitHub is supplemented with online developer meetings in which progress is reviewed and challenging topics are discussed. These meetings are held approximately every other month and are organised by the members of the research projects SzenarienDB and LOD-GEOSS. In such cases where no agreement on an issue can be found and where there are several possible solutions, the issue, along with the different options and their ramifications are passed to the OEO-Steering Committee (OEO-SC) which will debate and decide on it. Aside from helping with directional decisions, the OEO-SC helps to raise

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

awareness of the ontology and its adoption in currently running and planned projects. It convenes approximately every 3 months. To ensure a widespread acceptance of the committee and the OEO, a diverse group of experts with several years of experience in the domain and from different organisations was selected. The OEO has been introduced to several hundred scientists in the field: It was presented to the openmod community and to the Forschungsnetzwerke-Energie (FNE), which has more than 250 participants. Furthermore as it is developed as part of the Open Energy Family toolbox, a community of over 350 registered Open Energy Platform users is exposed to it.

5 Initial Use Cases

The vision of the LOD-GEOSS project is to build an interoperable, distributed database architecture (guided by Linked (Open) Data principles) for establishing a platform economy for data life cycles in the energy domain. Alongside the broad range of potential applications of the ontology, such as data set summarisation, user categorisation or tagging, and semantic search, it is currently being employed for two use cases. First, the OEO has been used for data representation. With the help of a mapping to the OEO, the German *Core energy market data register* is transformed from CSV to RDF[22], such that it is queryable using SPARQL[23].

Second, the LOD-GEOSS project aims to directly connect the distributed database architecture to energy system models. Using the OEO, the annotation of data inventories and the functional parameters expected or provided by model interfaces should be homogenised in such a way that clear assignments can be made. At the same time, the heterogeneity of interface descriptions is to be reduced and thus the effort of programmers and users to produce or understand them minimised. Within the project, the interfaces of several well established energy system models of different type are analysed to ensure a broad integration of the most important data categories. The FINE Framework, for example, is an open source Python package[24] that provides functionalities for modelling, optimisation and analysis of high-resolution energy system models in terms of time, space and technology [23]. 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[25] in which the individual function parameters are named and defined, it is currently being investigated to what extent there is already coverage with the terminology of the OEO, and at which points the interface or the ontology must be adapted. Using these specific model applications, the project aims to develop

[22] https://databus.dbpedia.org/jj-author/mastr/bnetza-mastr/

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

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

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

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.

6 Evaluation Setup

6.1 Coverage Study

The ontology should cover use cases such as annotation of various fact sheets and databases. An ontology coverage study was based on scenario fact sheets that are being developed within the project SzenarienDB. Scenario fact sheets collect information in a structured form about energy system scenarios.

The information collected in the scenario fact sheets covers a variety of topics including general information such as title and authors, publication format and license, and 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. The scenario fact sheets are used to describe energy scenarios and filled by their authors when providing the corresponding scenario data to the OEP.

We used the field names of the fact sheet form as input for a semi-automated concept annotation task. In the first stage, five concept 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 [6]. 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. *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.

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	72 / 47%	36 / 24%	45 / 29%	108 / 71%
ESE counts / ratio	109	71 / 65%	15 / 14%	23 / 21%	86 / 79%

6.2 Inter-annotator Agreement Study

The classes and definitions included in an ontology should be comprehensible and unambiguous. When annotating resources with ontology concepts for improved findability and query functionality, it is crucial that different annotators are able to use the ontology 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 [21].

We selected five text fragments from model fact sheets to study if energy domain experts can annotate them consistently. We selected such text fragments where annotation with ontology terms was not obvious, i.e. such that there was no perfect match between portions of the text fragment and labels of ontology terms, but rather several only roughly matching ontology terms could be relevant. For every text fragment, using the same string similarity technique and manual refinement by ontology developers as in Section 6.1 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 system 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.

7 Evaluation Results and Discussion

7.1 Coverage

The evaluation results of the coverage study are shown in Table 1 and have been made publicly accessible[26]. In total, the annotation of 153 fact sheet fields was tested, as depicted in the first table row (*"ALL"*). About half of the fields (47%) have a good match, whereas 29% have no match at all and cannot be described by the OEO yet.

About 30% of the fact sheet fields (44) 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 now, the main focus of the OEO development for the initial release has been on the *oeo-physical* module, with the other modules scheduled for becoming the focus area in subsequent

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

releases. Thus, the other modules have not yet been comprehensively developed, and especially the *oeo-social* module is still in an early state of development.

To mitigate for this, the second row of the table (*"ESE"*) just considers those fields (109) that are not related to socioeconomic aspects. Here, about 65% of the concepts have a good match and 21% have no match at all. Comparing the total counts of both results (*"ALL"* and *"ESE"*), it can be seen that there is only one field within the socioeconomic part that has a good match.

7.2 Inter-annotator Agreement

For the inter-annotator agreement 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. It turned out that for 71% of the candidate ontology entities, 70% or more of the participants agreed in their annotation of a text fragment. However, an inter-annotator agreement of 75% or more was only achieved for 59% of the candidate ontology entities. This result shows a medium level of inter-annotator agreement with significant room for improvement.[27]

According to our analysis, several factors contributed to the relatively low inter-annotator agreement:

1. 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 a redundant annotation does not cause problems, but in the context of this evaluation it reduced the measured inter-annotator agreement.
2. In some cases the choices provided to the participants did not contain an entity that would describe a text fragment optimally, and there was no obvious second-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.
3. It is likely that the documentation of some of the entities in the OEO was not sufficient to enable the participants to use them consistently.

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 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.

[27] Instead of one or more ontology concepts, it was also possible to select "None of the above". This has been chosen only very few times, which suggests that our set of candidate entities had sufficient coverage for annotating the given text fragments.

8 Conclusion and Future Work

We reported on the development-in-progress and initial evaluation of an open community-driven ontology for the energy system 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. In energy system analysis, aside from the practical benefits for re-use and reproducibility, openness has important consequences for transparency and the building of trust and accountability. Increasingly, open data platforms such as Renewable Ninja [28] and Open Power System Data [29] are working towards transparently allowing the community to share data, align models and work together, which will be even further facilitated by the ontology. Transparency and trust are ever more important in the context of the advancing climate crisis, as the outputs 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.

At its first release the ontology is still in an early stage of development. 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, much work will still remains to be done to enhance both the coverage of the ontology and the quality of its documentation. The second release, planned for later in the summer of 2020, will focus in particular on the missing socioeconomic aspects of the energy systems domain, harnessing content from already existing ontologies where possible to comprehensively describe socioeconomic entities. Subsequent releases will expand the coverage more generally and add more detail to the modelling module, including algorithms, constraints and assumptions.

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. Bennett, M.: The financial industry business ontology: Best practice for big data. Journal of Banking Regulation **14**(3-4), 255–268 (2013)
3. 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 **17**, 25–32 (2012)
4. Courtot, M., Malone, J., Mungall, C.: Ten simple rules for biomedical ontology development. In: Proceedings of ICBO (2016)

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

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

5. European Environment Agency: Trends and projections in Europe 2019: Tracking progress towards Europe's climate and energy targets. https://doi.org/10.2800/51114
6. Euzenat, J., Shvaiko, P.: Ontology Matching, Second Edition. Springer (2013)
7. Fernandez-Lopez, M., Gomez-Perez, A., Juristo, N.: Methontology: from ontological art towards ontological engineering. In: Proceedings of the AAAI97 Spring Symposium. pp. 33–40. Stanford, USA (March 1997)
8. Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The Units Ontology: a tool for integrating units of measurement in science. Database (Oxford) (2012)
9. Grüninger, M., Fox, M.: Methodology for the design and evaluation of ontologies. In: IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995 (07 1995)
10. 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
11. 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)
12. Küçük, D., Küçük, D.: OntoWind: An improved and extended wind energy ontology. arXiv preprint arXiv:1803.02808 (2018)
13. 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, (July 2017)
14. Madrazo, L., Sicilia, A., Gamboa, G.: Semanco: Semantic tools for carbon reduction in urban planning. In: Proc. 9th Europ. Conf. Product & Process Modelling (2012)
15. 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)
16. 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 **25**(3), 95–108 (2020). https://doi.org/10.5278/ijsepm.3327
17. 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. pp. 393–404. Springer (2016)
18. 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
19. 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)
20. Spyns, P., Tang, Y., Meersman, R.: An ontology engineering methodology for DOGMA. Applied Ontology **2**(3), 13–39 (2008)
21. 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)
22. Suárez-Figueroa, M., Gómez-Pérez, A., Fernández-López, M.: The NeOn Methodology for Ontology Engineering, pp. 9–34. Springer, Berlin, Heidelberg (2012)
23. Welder, L., Ryberg, S.D., Kotzur, L., Grube, T., Robinius, M., Stolten, M.: Spatio-temporal optimization of a future energy system for power-to-hydrogen applications in Germany. Energy **158**, 1130–1149 (2018)