

Identification of user requirements for an energy scenario database

Klara Reder^{a1}, Mirjam Stappel^a, Christian Hofmann^b, Hannah Förster^c, Lukas Emele^c, Ludwig Hülk^b and Martin Glauer^d

^a Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Königstor 59, 34119 Kassel, Germany

^b Reiner Lemoine Institut, Rudower Chaussee 12, 12489 Berlin, Germany

^c Öko-Institut e. V., Schicklerstraße 5-7, 10179 Berlin, Germany

^d Otto-von-Guericke-Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

ABSTRACT

Energy scenarios assist decision making regarding the transformation of the energy supply system. A multitude of scenarios exists in various formats. Thus, for scientists and policy stakeholders alike, it remains difficult to distinguish and compare scenario data. Hence, the aim of the project *SzenarienDB* is to establish an energy scenario database containing data in comparable and machine-readable format. *SzenarienDB* will do so by extending the *OpenEnergyPlatform* (OEP). To ensure that the extension fulfils the requirements of the modelling community, we conducted an online survey. We asked the participants about what they expected of an energy scenario database. Along with input from expert meetings and GitHub issues on that topic, we derived user requirements from the answers. In total, we identified 69 requirements. Out of these, around 44% were considered as very urgent. Hence, we conclude that there is a great need for the development of a consistent energy scenario database. To tackle the requirements we grouped these into twelve categories: input and output, data review process, bug-fixes, documentation, factsheets, features, functions to modify data, layout, metadata, ontology, references, and other. Each category is resolved according to its intrinsic properties.

Keywords:

User Requirements;
Energy Scenarios;
Open Source;
Open Data;
OpenEnergyPlatform

URL: <http://doi.org/10.5278/ijsepm.3327>

1. Introduction

The transformation of the energy supply system is complex and the identification of impacts is influenced by the results of scientific reports based on energy scenarios. In general, a scenario is used to express that a future condition or development of a certain aspect is seen as “possible” [1]. Energy scenarios describe possible future developments in the energy supply system and e.g. may include effects on greenhouse gas emissions. They can aid the identification of optimal or appropriate paths of development and serve as a factual basis for political decision-making [2]. There are several kinds of scenarios, from which two types are popular in the field of energy scenarios. These types are called “forecasting”

and “backcasting”. The type of “forecasting” generates exploratory scenarios that take a look from today into the future. In these types of scenarios, no certain goal or plan is predetermined, where a development shall go. Whereas in “backcasting” a target scenario is created with given future conditions, looking for a development that reaches these conditions [1]. Nonetheless, the term scenario is not defined and thus may have different implications depending on the person using it. Hence, this leads to less transparency and comparability when working with multiple scenarios.

Several studies and energy scenarios are published each year, usually by research institutes on behalf of public authorities, companies or civil society

¹Corresponding author – e-mail: klara.reder@iee.fraunhofer.de

organisations [3] [4]. For stakeholders and even the energy modelling community it has become increasingly difficult to compare different scenarios, as methods and objectives usually differ and assumptions may be expressed in different ways [1]. Even the reconstruction of a single scenario can be complex or impossible, since assumptions are often not published in full detail [5], thus lacking transparency. Furthermore, the collection and processing of input data for scenarios has become more time consuming and costly. This lack of transparency fosters distrust, but trust in this research does matter because it contributes to policies and strategic decision making on energy, as [6] explicates. Some approaches were made to meet the need for transparency and comparability in the energy system modelling and scenario community. A transparency checklist was developed by [7], to improve the quality and traceability of scenario studies, for example. Other studies focus on the topic of transparency by open access of data and models [8] [9] [10] and data enrichment of those [11].

In our project SzenarienDB, we focus on transparency and comparability of (complex) energy scenarios. The project SzenarienDB aims to create a database for energy scenarios as an extension of the OpenEnergyPlatform (OEP) [12] [13], an open source platform for energy data. Here, scenario data of several studies will be uploaded to the database, freely and easily accessible under an open license. They can serve as a reference and help to establish more transparency and comparability. In addition, it is part of the project to ensure the maintenance of the database even after the project has ended. We assume that easily accessible data from the database via a user-friendly interface will increase accessibility as well as scientific exchange. This will contribute to reducing the necessary effort for model comparisons and sensitivity analyses. Furthermore, the data platform has potential to facilitate scientific and political decision making due to a generally improved level of transparency and comparability. Finally, in the ideal case, the platform will contain the most recent developments in scenario generation and modelling.

The development of the OEP was started in the research project open_eGo, by the implementation of an open and community driven energy database. The database is based on a PostgreSQL database that is made available via a web-interface on the OEP [12] [14]. The main focus is to exchange and provide open data via an online data portal which could be used by the project

partners and across research projects [15]. Furthermore, the OEP includes the possibility to version-controlled data sets and assign rich meta data to data sets. An application programming interface (API) allows secure and documented interactions and data exchange. Many python-based tools use SQLAlchemy to communicate with existing databases that also allows the usage of different database interfaces by so-called dialects. In order to ease the use of the OEP the oedialect [12] has been developed to enable the use of SQLAlchemy structures to access the data available on the OEP.

In European energy systems research several open source modelling approaches emerged. These include projects like SciGrid [16], oemof [17], GENESYS [18], open_eGo [12], OPSD [19], PyPSA-Eur [20] and others.

In the past, there have been several approaches to distribute open access energy data. In 1991 the project IKARUS [21] set up a free database. Despite a considerable demand the approach failed. This was due to technical and conceptual restrictions such as the distribution of data via hardware and a proprietary database management system. Another open database from the early days is OpenEI [22]. OpenEI is based on the CKAN system of the open knowledge foundation. The CKAN system is also used by the World Bank database that focuses on developing countries. CKAN is in widespread use, but during the initial assessment of possible frameworks it did not use modern web frameworks such as Flask or Django for the web architecture and was still based on python2 and Pylons. The migration of CKAN to a more modern python3- and Flask-based foundation is currently in progress. To address such shortcomings, the OEP was developed as a Django based open-source application [23]. This gives the OEP a flexible foundation which can be extended easily and independently from data specific aspects. Further recent projects include the European Union project OpenENTRANCE which aims to develop, apply and disseminate an open, transparent and integrated modelling platform for target scenarios in 2020, 2030 and 2050. The database itself will be hosted by the *International Institute for Applied Systems Analysis* (IIASA) [24].

The past approaches to distribute open access energy data show that it is important to include the user requirements, in order to ensure the success of such a database. Establishing user requirements is a common method to capture the most important

features, functionality and requirements for a software development project [25]. Stakeholders and users have individual requirements for a particular software. User requirements provide the basis for specification sheets that allow meeting these needs. Considering user requirements during the software development stage requires relatively little effort but the effect on the final result is often significant [26]. The Institute of Electrical and Electronics Engineers [27] defines a requirement as:

- 1) *A condition or capability needed by a user to solve a problem or achieve an objective.*
- 2) *A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents.*
- 3) *A documented representation of a condition or capability as in (1) or (2).*

Therefore, it is necessary to capture the user requirements of the targeted stakeholders in order to develop an energy scenario database that will be accepted and used by the target group.

The objectives of the energy scenario database are:

- provision of access through an API regardless of the system or programming language
- versioning of the data, including old results, correction of errors, addition of scenarios and other
- open licences (CC0) for all uploaded scenario data
- serving as a role model for similar projects of other disciplines and regions
- triggering a broad discussion on standards for the exchange on data, code and description of models and scenarios

The novelties of our database compared to existing databases in the energy sector are:

- helping to reduce the expenses in energy modelling due to easy access of existing energy scenarios
- serving as a central repository of consistent and, as far as possible, complete energy scenario data
- fostering the comparability of scenarios and thereby improving the support of policy decisions
- creating an ontology with open access in the field of energy modelling

In the following are the methods and results described how we generated user requirements for an energy scenario database.

2. Methods to generate user requirements

User requirements can be established via different methods, such as interviews, comparison to other systems, user observation at the point of application, and more [25]. Our approach for developing user requirements for the energy scenario database is based on an online survey, on expert meetings as well as on GitHub issues. The details of this approach are described in the following.

In the course of this study we conducted an online survey among potential users of our database from the energy scenario and modelling community. We chose this method because of its high accessibility to the target group, as well as the relatively modest preconditions regarding time and cost [28] [29]. Our main research question was: 'What are the user requirements for an open-source database containing energy scenarios?'. The online survey consisted of two parts. The first part considered the day-to-day work of the target group. The second part focused on features and criteria a scenario database should ideally fulfil from their point of view. A complete list of all questions is available in Appendix Table A1. We invited the target group to take part in the online survey via several channels that focus on energy modelling and scenario topics:

- E-mail list of *Strommarkttreffen* (www.strommarkttreffen.org)
- E-mail list of *openmod initiative* (www.openmod-initiative.org)
- posting on the platform of *Forschungsnetzwerk Energie – Systemanalyse* (research network energy system analysis) (www.forschungsnetzwerke-energie.de/systemanalyse)
- internal E-mail list of colleagues and target persons at Fraunhofer IEE
- internal E-mail list of colleagues and target persons at Öko Institute
- internal E-mail list of colleagues and target persons at Reiner Lemoine Institute (RLI)
- internal E-mail list of colleagues and target persons at Projektträger Jülich (PtJ)
- internal E-mail list of colleagues and target persons at Federal Ministry of Economic Affairs and Energy (BMWi)
- E-mail list of the *VDI Richtlinien-Gruppe zu Energieszenarien* (VDI guidelines group on energy scenarios)

We derived user requirements from the online survey's multiple-choice answers and free text comments

the participants were able to provide, too. These user requirements were phrased so that they followed the structure of *user stories*, i.e. “As *<a type of user >*, I want *<goal>*, [so that some reason]” (<.>required, [.] optional) [30]. For example: *As a user I would like to use a wizard to upload .csv files in order to use the SzenarienDB without any technical precognition.* All user requirements had to satisfy the criteria in Table 1 [31] [26].

Another common method used to define user requirements is called the INVEST method [32]. The acronym stands for Independent, Negotiable, Valuable to the customers, Estimable, Small and Testable user stories. All of the INVEST criteria but “Negotiable” are included in the criteria listed in Table 1. Negotiability is attempted however, by publishing all gathered requirements in the OEP’s online repository and by writing this paper to hopefully reach more people who can generate feedback and thereby improve the database.

Further user requirements were derived in meetings and web conferences with experts from within the project who discussed one topic at a time. These topics were ‘What metadata should be included?’, ‘How to reference original data?’, ‘How to review uploaded data?’ and ‘Requirements for tutorials of the oedialect’.

Finally, the issues on the OEP GitHub repositories were used as another base for user requirements. Since the repositories are constantly changing, we set the cut-off date for consideration to be the 29th of October 2018. We collected 147 issues from GitHub in total. These issues did not satisfy our user requirement criteria mentioned above. In some cases these issues were opened before we could decide on the method of user

stories and other issues were raised by people who don’t participate in the project. Moreover, several of these issues described similar problems, as well as problems addressed in the survey. Overlapping issues and requirements were therefore merged together. Furthermore, some issues were very specific while others were very broad. Issues were filtered and aggregated into subsets, preserving the initial intentions, but embedding them into a bigger picture; e.g. requested bug fixes were grouped together, as well as calls for documentation, while some time consuming feature requests, such as a global search function, were discarded. This resulted in 27 cumulative requirements condensed from GitHub.

In order to merge these different sources of user requirements, we removed duplicate requirements. We classified each requirement according to the following criteria:

- estimated time for completion
- urgency
- overall estimate
- category

Time and *urgency* were assessed roughly, using the T-shirt size estimation method [33]. We defined the sizes as follows: S = small/one day/not urgent, M = medium/one week/somewhat urgent and L = large/one month/very urgent.

The *overall estimate* rates the importance of a requirement. We jointly rated requirements, following the German school grade system from 1 to 6, with 1 being very important and 6 being insufficient.

Finally, requirements were classified into one of twelve *categories*: input and output, data review process,

Table 1: Criteria applied for user requirements [29] [24]

criteria	description
estimation of priority	The estimate arranges user requirements by priority.
completeness	Completeness requires that all aspects of the requirement are formulated without implicit assumptions.
documentation	Each user requirement shall have a documentation on where the requirement originates from and the evolution to the final formulation.
correctness	User requirements should be correct and accepted by the stakeholders to be necessary.
clarity	The user requirements are written in non-technical understandable text so that everyone involved, from stakeholder to developer, understand what is meant.
consistency	User requirements are consistent among each other. That means that no competing interests between user requirements exist.
verifiability	All user requirements can be tested to ensure that each user requirement is functioning.
uniqueness	User requirements have to be unique in the sense that they must describe only one issue at a time.
action	User requirements should describe an action, hence some type of functionality which can be used by the person who requires it.

bug-fixes, documentation, factsheets, features, functions to modify data, layout, metadata, ontology, references and other (further explanation in section 3.3). Categorisation was implemented in order make sure that all different kinds of requirements are addressed. A categorisation also facilitates the distribution of tasks with different capabilities in the working team. The final requirements with the corresponding *estimated time*, *urgency*, *overall estimate* and *category* served as input for the specification sheet.

3. Results and discussion

The results and discussion are presented together in this chapter, starting with the online survey in section 3.1. It is followed by the evaluation of the specification sheet in section 3.2 and concludes with a description on how the requirements of the specification sheets built the energy scenario extension of the OEP in section 3.3.

3.1. Analysis of the online survey

The online survey was started by 177 participants and fully completed by 101 participants. The following numbers all refer to those participants who completed the questionnaire. We received the first response on 12th of June 2018 and closed the survey on 27th of August 2018. About 90% of the responses were given between 13th of June and 10th of July. Most participants work in research institutes (71%) and are involved in scenario generation as well as in making use of scenarios created by others (69%). Only 6 participants do not work with energy scenarios at all. About 56% frequently use external databases, such as Eurostat, OpenStreetMap and others. Only 11% do not use databases at all.

The survey revealed a large interest in the topic, especially by the scientific energy modelling community. Participants stated that they are willing to use energy scenarios from an energy scenario database like the OEP (96%) and also to publish their own scenarios there (92%). However, a precondition for publishing scenarios for many participants (41%) is financing. Obstacles in contributing to such a database lie in the difficulty to provide open-source licensing of data or in the commercial nature of scenarios. The participants were asked about their willingness to implement an interface between OEP and their models. The majority (53%) was willing to do so under certain conditions. In the free text these conditions included for example: simple and intuitive API and little effort for the implementation of the

API. Out of the participants, 26% were willing to implement a port without any preconditions.

The majority of participants (52%) require highly resolved scenario data, e.g. hourly time series for one year, spatial resolution in scale of kilometres. Only 19% use data with a low level of detail, such as aggregated values for countries or years.

Furthermore, quality of data (52%) is most important to the participants followed by quantity of data (25%) and user friendliness of the platform (23%).

The participants were asked to assign different levels of importance to six features. Figure 1 shows the results in decreasing order: 'filter data', 'description of metadata', 'text search', a 'glossary/ontology', 'preview of data' and 'ad-hoc visualization'. The possibility to 'filter data' was selected most often (70%) as being indispensable. The features 'description of the metadata', 'text search', 'glossary/ontology' and 'preview of the data' are seen as indispensable or quite important by the majority of participants. The feature 'ad-hoc visualization' was considered by most participants merely as nice-to-have (60%). Only very few participants selected that a feature was a waste ($\leq 7\%$) or I don't know ($< 4\%$).

The preferred formats for uploads and downloads on such a database were interrogated. The participants had the possibility to choose multiple formats. They predominantly favored .csv, .xlsx, API and table. We further prompted the participants to prioritise different criteria into ranked classes from 1 to 6 (Figure 2). A 'list of references for all datasets' was most often (56%) selected

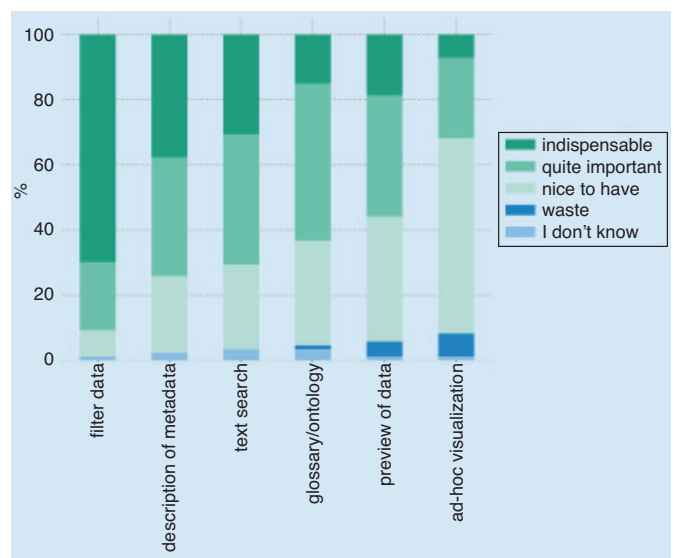


Figure 1: Assignment of different features to the categories indispensable, quite important, nice to have, waste and I don't know

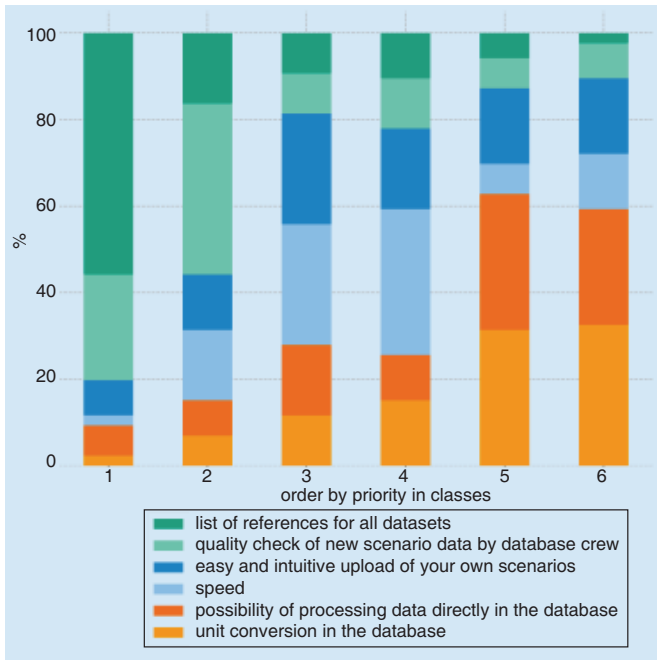


Figure 2: Different criteria for a scenario database arranged by priority in classes 1 to 6

to have the highest priority (class 1). Furthermore, 24% found ‘quality check of new scenario data by database crew’ to have the highest priority, about 40% see it the second highest class 2. The criteria ‘easy and intuitive upload of your own scenarios’ and ‘speed’ have a similar distribution. For these two criteria the participants selected most often a class 3 to 4 (between 19-34%) and less often a class with high or low priority. The criteria of least importance are the ‘possibility of processing data directly in the database’ and ‘unit conversion in the database’ (27% and 33% in class 6 respectively).

The expert meetings revealed that the term ‘scenario’ may be understood quite differently, hence a question was included in the online survey to find out what the participants understood by ‘scenario’. A list of possible scenario elements was suggested, where the participants could choose multiple answers. The possible answers were: ‘general framing parameters and assumptions (e.g. geographical and temporal scope, ...)’, ‘scenario type (e.g. extreme scenario, objective scenario, ...)’, ‘model input data’, ‘justification/explanation on assumption’, ‘modelling parameters’, ‘model output data’ and ‘other [free text field]’. All of the above answers apart from ‘other’ were selected with similar shares (around one sixth each) but the distribution between the different answers varied depending on the participant answering

the online survey. From the participants 36% selected all six possible answers, and 24% and 23% selected five and four answers out of six, respectively. This shows that not a single answer explains the term ‘scenario’ and it is hard to find a consistent definition within the community. Hence, we derived that the energy scenario database has to offer the possibility to include data for all of the six answers above and arbitrary permutations of a subset. This definition is especially helpful for the ontology which ensures that everyone is using the same terminology and hence fosters transparency and comparability.

3.2. Specification sheet evaluation

In total, 69 user requirements were derived from the online survey, expert discussions and OEP GitHub issues. These requirements create the specification sheet. We examined and compared the requirements according to the methods in section [methods]. We found that the requirements do not compete with one another. The only requirement which has an overlap is *Create a discussion space for tables and schemas*. It does not compete with another requirement but with the openmod Wiki [34] and openmod forum [35]. Despite this slight overlap in topic, a discussion forum for tables and schemas is very specific and is not covered by the openmod Wiki and openmod forum, which is why we kept this requirement. However, such a forum may have topics and discussion similar or duplicates to those of the openmod Wiki and openmod forum. Moreover, in our analysis we did not accept fifteen requirements because

- the functionality of the issue is already implemented. E.g. *As a user I want the name of the homepage to be displayed high up on google (1-5), so that I don’t confuse the homepage and don’t have problems finding it.*
- the functionality of the issue was ranked unimportant or requested by only one person of the online survey, and posed huge implementation/conceptual work which was disproportionate to the importance of the functionality. E.g. *As a user I would like to work with multidimensional tables (like Eurostat) to assign complex values.*

The evaluation of the specification sheet showed that 44% of the user requirements were considered very urgent and 26% as not urgent. This implies that there is a great need for a scenario database and its specific requirements. The estimation of urgency is furthermore

helpful in the upcoming project management. Very urgent issues can be worked off first. For the implementation of all user requirements, we roughly estimate 24 months, originating from 16 user requirements with the duration of one month, 27 of one week and 25 of one day. Hence, together with the urgency this gives very fast improvement possibilities: to first work off the issues with short time estimation and high urgency.

Most user requirements (20%) fall into the category input and output, i.e. upload and download of data, and in the category feature (20%) (Figure 3). Third most frequent category is metadata (16%), followed by OEP layout (12%), functions to modify the dataset (9%), documentation wanted (9%) and others which are below 5%.

Interestingly, the user requirements, while explicitly meant to reflect on energy scenario needs, did not end up being very specific for the energy scenario domain. Most requirements would be the same for e.g. a water quality database. Generally, the compiled requirements should hold true for any database that stores modelling input and output data and may contain georeferenced and temporal data. Therefore, an established energy scenario database may be of interest for other disciplines as well. Our chosen approach is thus transferable to other disciplines of research, too.

All user requirements can be accessed at GitHub at <https://github.com/OpenEnergyPlatform> with the tag ‘specification sheet’.

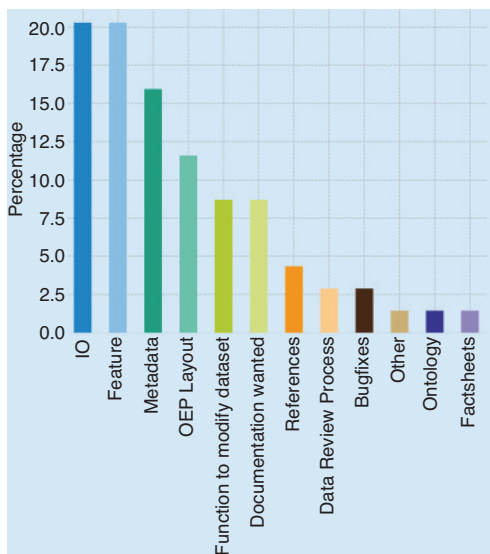


Figure 3: Percentage of user requirements grouped into twelve categories

3.3. Integration and extension of the OEP

The requirements for an open database are very diverse. To take this into account we have twelve categories:

- input and output
- data review process
- bug-fixes
- documentation
- factsheets
- features
- functions to modify data
- layout
- metadata
- ontology
- references
- other

These twelve categories enable a structured integration of the user requirements we identified.

Figure 4 shows a schematic overview on the work flow of users interacting with the OEP. The work flow is as follows: an energy scenario developer or modeller generates e.g. scenario data, which is uploaded into the OEP (tile: data) and correct metadata is supplied (tile: metadata). The developer or modeller also completes the factsheets (tile: factsheets) which are distinguished into model factsheets and scenario factsheets. The model factsheets contain information on how the model works and the scenario factsheets contain information on how the scenario is characterised. The factsheets and the metadata are coupled to the ontology (tile: ontology) which ensures that the same terminology is used throughout the OEP. The uploaded scenario may now be downloaded (category: input/output) by other energy modellers. This enables them to use the data for their own modelling exercises. Furthermore, users may participate in the reviewing process for data, which is designed to allow for peer review.

‘Inputs and output’ are managed via an API which is programmed in python. This allows that users only need to invest in establishing a routine on how to interact with the OEP once and can then easily use this routine repeatedly. Since not all users indicated that they would like to use an API, we identified the need for an up- and download wizard as one of the major requirements in our specification sheet. The use of the wizard shall be intuitive while using the API might be more challenging for first time users. Hence, to fulfil the category documentation written tutorials which are presented in *Jupyter Notebooks* will be provided along with video tutorials on

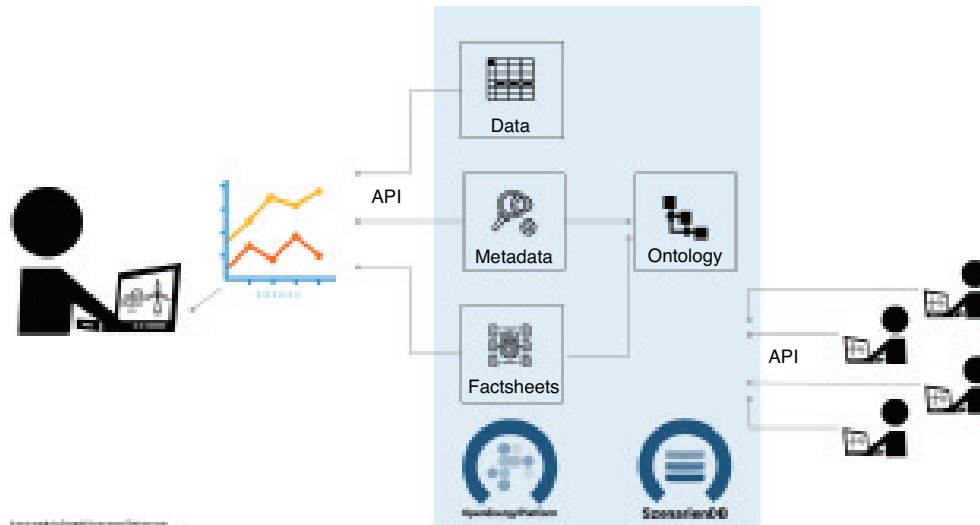


Figure 4: Work flow of users uploading and downloading data to the OEP

the details of the API and also the upload/download wizard. Documentation in form of tutorials will also be provided for all other important features of the OEP.

How the data is displayed in the OEP provides the user with several ‘functions to modify the data’ such as filtering data. These functions are all in separate GitHub issues due to their independence of each other. These function will ensure an easy usability of the data. These changes are often supported by layout changes (category: layout) to enhance usability.

The current ‘metadata’ format implemented in the OEP will be extended by a standardised, energy scenario specific metadata string. This string includes a human readable description, as well as machine readable name, spatial and temporal context, references to sources and licenses, a list of contributors, a detailed description of the data structure, information on conducted data reviews and additional metadata keys that help to evaluate, compare and contextualise any uploaded dataset.

The OEP ‘factsheets’ are a standardised collection and presentation of information about modelling frameworks, models and scenarios used in climate and energy system modelling. The use of interactive fields and pre-defined responses is designed to make it easy to add new factsheets and to filter for existing entries. The goal is to create a full set of linked factsheets (and datasets) to improve transparency. The current focus is on extending the scenario factsheets to the heterogeneous landscape of different energy scenarios and to link the information in the ontology. An ‘ontology’ provides the

possibility to create a standardised language for a domain of interest: it is a system of concepts including the descriptions of how these concepts relate to one another. The ontology created for the OEP harmonises and defines terms and concepts used throughout the OEP, for example in factsheets and the metadata. In the course of the *SzenarienDB* project, the current ontology on the OEP is extended by terminology specific to energy scenarios. This includes information needed for target scenarios, temporal and regional concepts, sector concepts, modelling assumptions and constraints.

The user can also upload input data and in that case set ‘references’ to individual data tables and cells. These references can be used to include the uploaded data in Linked Open Data schemes (LOD) and make them more accessible to potential users and allow the integration of other sources, e.g. by concepts defined in the ontology.

The requested ‘features’ (category: features) for the energy scenario extension of the OEP are of different kinds, but many refer to preview functionality such as the requirement: *As a user, I would like to use the preview function to display data, for example as a table, in order to be able to evaluate the content of the scenarios.*

The ‘data review’ process is planned to include a badge system like bronze, silver and gold. Other users of the OEP, besides the person contributing a dataset, may rank the dataset and comment on missing or questionable entries. This will ensure that on the one hand the datasets are complete (including metadata, references, licences etc.) and on the other hand that the uploaded

scenario data is correct and fulfils a scientific standard. The reviewers will be encouraged to participate by a ranking system of their profile similar to stack overflow. The more reviews they have done the more e.g. stars they get. The review functionality shall also include a commenting function, where comments can be up-voted or down-voted.

The final two categories are 'bug-fixes' and 'other'. Bugs unfortunately always occur in a software development project, and have to be fixed. These can be of very different kind. Either misspelled text on the web-page, links which are not working or features which are broken etc.

The last category is 'other' which contains all requirements which could not be included in the other eleven categories. This includes for example the requirement *As a user, I want to access old versions of data if I accidentally entered something wrong. These requirements will be tended to one by one.*

4. Conclusion

Our main research question was: 'What are the user requirements for an open source database containing energy scenarios?'. We addressed this by an online-survey as well as by expert meetings and GitHub issues. Our main findings were:

- The modelling community has a high interest in an energy scenario database.
- They are willing to upload their energy scenarios and use energy scenarios of others.
- More than 50% of the participants would use an API for upload and download, with .csv being the preferred download format.
- The two most important features were 'filtering of data' and 'description of metadata'.
- The two most important ranked criteria were 'references for all datasets' and 'quality check of uploaded data'.
- Of the requirements, around 40% were rated as very urgent showing the great need for an energy scenario database.

In the further development of the OpenEnergyPlatform these findings are addressed in realising the user requirements. To aggregate the 69 user requirements they have been clustered into twelve categories: input and output, data review process, bug-fixes, documentation, fact-sheets, features, functions to modify data, layout, metadata, ontology, references and other. Hence, these

categories give an overview on the main development areas.

The geographic scope of the OEP is currently Germany. Thus the target group for the survey had to originate from there. Since the German energy system modelling community is relatively small, in turn was the sample size. Once the OEPs focus becomes more international, future surveys can be conducted; based on larger samples sizes. We assume that scientists in this research field will have similar user requirements on such databases, no matter where in the world they conduct their research.

Further limitations are given by the duration of the project. User requirements had to be selected so that they can all be worked of within the duration of the project.

Hence, further Research includes

- complex visualisation of data within maps (e.g. how to display polygons),
- automatic identification of missing data in time series,
- data review which is not only human based but also supported by artificial intelligence,
- including multidimensional tables,
- promoting the OEP including the new energy scenario extension worldwide,
- further promoting of the OEP in general to ensure its use and usefulness.

Acknowledgements

This research has been funded by the Federal Ministry of Economic Affairs and Energy of Germany as part of the project SzenarienDB (03ET4057A-D).

References

- [1] Christian Dieckhoff, Hans-Jürgen Appelrath, Manfred Fishedick, Armin Grunwald, Felix Höffler, Christoph Mayer, and Wolfgang Weimer-Jehle. Zur Interpretation von Energieszenarien. *Schriftenreihe Energiesysteme der Zukunft*, 2014. https://www.akademienunion.de/fileadmin/redaktion/user_upload/Publikationen/Stellungnahmen/141203_Energieszenarien_Web_final.pdf
- [2] Editha Kötter, Ludwig Schneider, Frank Sehnke, Kay Ohnmeiss, and Ramona Schröer. The future electric power system: Impact of power-to-gas by interacting with other renewable energy components. *Journal of Energy Storage*, 5:113-119, 2016. <http://doi.org/10.1016/j.est.2015.11.012>

- [3] Deutsches Zentrum für Luft und Raumfahrt (DLR), Stuttgart Institut für Technische Thermodynamik, Abt. Systemanalyse und Technikbewertung, Fraunhofer Institut für Windenergie und Energiesystemtechnik (IWES), Kassel and Ingenieurbüro für neue Energien (IFNE), Long-term scenarios and strategies for the deployment of renewable energies in Germany in view of European and global developments. URL https://www.dlr.de/dlr/Portaldata/1/Resources/documents/2012_1/leitstudie2011_kurz_bf.pdf
- [4] Öko-Institut, Fraunhofer ISI, and Hans-Joachim Ziesing. Klimaschutzszenario 2050 - 2. Endbericht, 2015. <https://www.oeko.de/oekodoc/2451/2015-608-de.pdf>.
- [5] Fraunhofer UMSICHT and Fraunhofer IWES. Abschlussbericht Metastudie Energiespeicher. <https://www.umsicht.fraunhofer.de/content/dam/umsicht/de/dokumente/pressemitteilungen/2015/Metastudie-Energiespeicher-Kurzfassung-web.pdf>.
- [6] Stefan Pfenninger. Energy scientists must show their workings, *Nature*, (7642), 2017. <http://doi.org/10.1038/542393a>
- [7] Karl-Kiên Cao, Felix Cebulla, Jonatan Gómez Vilchez, Babak Mousavi, and Sigrid Prehofer. Raising awareness in model-based energy scenario studies - a transparency checklist. *Energy, Sustainability and Society*, (6), 2016. <http://doi.org/10.1186/s13705-016-0090-z>
- [8] Stefan Pfenninger, Lion Hirth, Ingmar Schlecht, Eva Schmid, Frauke Wiese, Tom Brown, Chris Davis, Matthew Gidden, Heidi Heinrichs, Clara Heuberger, Simon Hilpert, Uwe Krien, Carsten Matke, Arjuna Nebel, Robbie Morrison, Berit Müller, Guido Pleßmann, Matthias Reeg, Jörn C. Richstein, Abhishek Shivakumar, Ian Staffell, Tim Tröndle, and Clemens Wingenbach. Opening the black box of energy modelling: Strategies and lessons learned. *Energy Strategy Reviews*, (19), 2018. <http://doi.org/10.1016/j.esr.2017.12.002>
- [9] Robbie Morrison. Energy system modeling: Public transparency, scientific reproducibility, and open development, *Energy Strategy Reviews*, (20), pp. 49-63, 2018. <http://doi.org/10.1016/j.esr.2017.12.010>
- [10] Fabian Gotzens, Heidi Heinrichs, Jonas Hörsch, and Fabian Hofmann. Performing energy modelling exercises in a transparent way - the issue of data quality in power plant databases. *Energy Strategy Reviews*, (23), 2018. <http://doi.org/10.1016/j.esr.2018.11.004>
- [11] Francisco Javier Miguel-Herrero, Víctor Iván Serna-González and Gema Hernández-Moral. „Supporting tool for multi-scale energy planning through procedures of data enrichment,“ *Int J Sustain Energy Plan Manag* (24) 2019. <https://doi.org/10.5278/ijsepm.3345>
- [12] Ulf Phillip Müller, Birgit Schachler, Wolf-Dieter Bunke, Julian Bartels, Martin Glauer, Clara Büttner, Stephan Günther, Editha Kötter, Ilka Cußmann, Ludwig Hülk, Malte Scharf, Till Mossakowski, Jochen Wendiggensen . open_eGo.: Netzebenenübergreifendes Planungsinstrument — zur Bestimmung des optimalen Netz- und Speicherausbaus in Deutschland — integriert in einer OpenEnergyPlatform. <https://doi.org/10.1016/j.esr.2018.08.014>.
- [13] Ludwig Hülk, Berit Müller, Martin Glauer, Elisa Förster, and Birgit Schachler. Transparency, reproducibility, and quality of energy system analyses – a process to improve scientific work. *Energy Strategy Reviews*, 22:264-269, 2018. <http://doi.org/10.1016/j.esr.2018.08.014>.
- [14] Martin Glauer, Stephan Günther, Ludwig Hülk, and Wolf-Dieter Bunke. An open database concept for open energy modelling. *From Science to Society: The Bridge provided by Environmental Informatics-Adjunct Proceedings of the 31st EnviroInfo conference*, vol. 5, p. 2018, 2017. ISBN 978-3-8440-5495-8.
- [15] Ludwig Hülk, Lukas Wienholt, Ilka Cußmann, Ulf Philipp Müller, Carsten Matke, Editha Kötter. Allocation of Annual Electricity Consumption and Power Generation Capacities Across Multiple Voltage Levels in a High Spatial Resolution. *Int J Sustain Energy Plan Manag*, (13) 2017. <https://doi.org/10.5278/ijsepm.2017.13.6>
- [16] Wided Medjroubi, Ulf Philipp Müller, Malte Scharf, Carsten Matke, and David Kleinhans. Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models. *Energy Reports*, 3:14-21, 2017. <http://doi.org/10.1016/j.egy.2016.12.001>
- [17] Simon Hilpert, Cord Kaldemeyer, Uwe Krien, Stephan Günther, Clemens Wingenbach, and Guido Plessmann. The Open Energy Modelling Framework (oemof) - a novel approach in energy system modelling. *Energy Strategy Reviews*, 22:16-25, 2018. ISSN 2211467X. <http://doi.org/10.1016/j.esr.2018.07.001>
- [18] Christian Bussar, Melchior Moos, Ricardo Alvarez, Philipp Wolf, Tjark Thien, Hengsi Chen, Zhuang Cai, Matthias Leuthold, Dirk Uwe Sauer, and Albert Moser. Optimal Allocation and Capacity of Energy Storage Systems in a Future European Power System with 100% Renewable Energy Generation, *Energy Procedia*, 46:40-47, 2014. <http://doi.org/10.1016/j.egypro.2014.01.156>
- [19] Frauke Wiese, Ingmar Schlecht, Wolf-Dieter Bunke, Clemens Gerbaulet, Lion Hirth, Martin Jahn, Friedrich Kunz, Casimir Lorenz, Jonathan Mühlentpfordt, Juliane Reimann, and Wolf-Peter Schill. Open Power System Data - frictionless data for electricity system modelling. *Applied Energy*, 236:401-409, 2019. <http://doi.org/10.1016/j.apenergy.2018.11.097>
- [20] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. PyPSA-Eur: An Open Optimisation Model of the European Transmission System. *Energy Strategy Reviews*, 22:207-215, 2018. <http://doi.org/10.1016/j.esr.2018.08.012>

- [21] Peter Markewitz and Gotthard Stein. Das IKARUS-Projekt: energietechnische Perspektiven für Deutschland, Forschungszentrum Jülich, 2003. URL http://juser.fzjuelich.de/record/136093/files/Umwelt_39.pdf
- [22] OpenEI, Open Energy Information. URL https://openei.org/wiki/Main_Page
- [23] Django software foundation. URL <https://www.djangoproject.com/>.
- [24] SINTEF. URL <https://www.sintef.no/en/projects/open-entrance/>.
- [25] Gerhard Versteegen, editor. *Anforderungsmanagement*. Berlin, Heidelberg, New York: Springer, 2004. ISBN 3-540-00963-9.
- [26] Klaus Pohl. *Requirements Engineering: Grundlagen, Prinzipien, Techniken.*, Heidelberg: dpunkt.verlag, 2008. ISBN 978-3-89864-550-8.
- [27] Didar Zowghi, editor. *2015 IEEE 23rd International Requirements Engineering Conference (RE): 24-28 Aug. 2015, Ottawa, ON, Canada*, Piscataway, NJ, 2015. IEEE. ISBN 978-1-4673-6905-3.
- [28] Michael D. Kaplowitz, Timothy D. Hadlock, and Ralph Levine. A Comparison of Web and Mail Survey Response Rates, *Public Opinion Quarterly*, 68(1): 94–101, 2004. <http://doi.org/10.1093/poq/nfh006>.
- [29] Jeremy C. Wyatt. When to Use Web-based Surveys. *Journal of the American Medical Informatics Association*, 7(4):426-430, 2000. <http://doi.org/10.1136/jamia.2000.0070426>
- [30] Mike Cohen. *Succeeding with Agile: Software Development Using Scrum*. Addison-Wesley, Bosten, 2010. ISBN 0-321-57936-4.
- [31] Mike Cohen. *User Stories Applied: For Agile Software Development*. Addison-Wesley, Boston, 2004. ISBN 0-321-20568-5.
- [32] Bill Wake. INVEST in Good Stories, and SMART Tasks. 2003. URL <https://xp123.com/article-s/invest-in-good-stories-and-smart-tasks/>
- [33] Steve McConnell. *Software Estimation: Demystifying the Black Art*. Safari Books Online. Microsoft Press, Redmond, Washington, 2006 ISBN 9780735605350. URL <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10762188>
- [34] Openmod wiki. URL https://wiki.openmod-initiative.org/wiki/Main_Page
- [35] Openmod forum. URL <https://forum.openmod-initiative.org/>

Appendix: List of questions of the online survey

Table A1: Questions of the online survey on the topic: users expect/wish of a scenario database

Nr.	question	answers
1	Are you working as ...?	<ul style="list-style-type: none"> • Programmer • Natural scientist • Engineer • Economist • Social scientist • Other
2	Information about your institute/organization/enterprise (optional).	<ul style="list-style-type: none"> • Research center • Government authorities • NGO • System operator/utility company • Other • Position
3	Are you working with scenarios?	<ul style="list-style-type: none"> • Yes, I am involved in the generation of scenarios and working with scenarios generated by others. • Yes, I am working with scenarios generated by others. • Not yet.
4	Would it be an option for you to provide your own scenarios for "SzenarienDB"?	<ul style="list-style-type: none"> • Yes, I would provide my own scenarios and publish all assumptions, as far as possible. • Yes, in case this is part of my project and will be financed. • No, this is not an option because of the license. • No, this is not an option for me because of other reasons, which are ...
5	Would it be an option for you to include a database like "SzenarienDB" in your workflow by using scenarios from it?	<ul style="list-style-type: none"> • Yes, sounds good. • No, using scenarios from "SzenarienDB" is not an option for me, because ...
6	Would it be an option for you to have a port implemented/ implement a port by yourself between your models and "SzenarienDB", which enables an easy access for further usage?	<ul style="list-style-type: none"> • Definitely. • Yes, in case of... • No. • Explanation:
7	There are several definitions and understandings of what a "scenario" is, in the context of energy system modelling. Which parameters are part of a scenario in your daily work?	<ul style="list-style-type: none"> • General framing parameters and assumptions (e.g. geographical and temporal scope, ...) • Scenario type (e.g. extreme scenario, objective scenario, ...) • Model in-put data • Justification/explanation on assumptions • Modelling parameters • Model output data • Other:
8	Are you using (external) databases in your daily work, such as OpenStreetMap, Eurostat, etc., to download data, e.g. as input to your models?	<ul style="list-style-type: none"> • Yes, in many cases. • Rarely. • Never. • When using a database, it is usually one of these:
9	How important are the following database features for you when looking for data/using a database	<ul style="list-style-type: none"> • Possibility to filter data while searching • Text search • Preview of data, e.g. as table • Ad-hoc visualization of data, e.g. as diagram • Description of metadata as text • Glossary/ontology
10	In addition, the following features are important to me when looking for data in the internet/using external databases:	

- 11 Which type of data provision is the most comfortable for your purposes? (Download interface, formats)
- API
 - table
 - study
 - .xlsx
 - .csv
 - .json
 - .xml
 - .pdf
 - .nc or .cdf
 - datapackage
 - other:
- 12 Please rate the following criteria for using a database like "SzenarienDB" according to your personal opinion and order by priority: No. 1 = "most important" to No. 6 = "least important".
- List of references for all datasets
 - Quality check of new scenario data by database crew
 - Speed
 - Easy and intuitive upload of your own scenarios
 - Possibility of processing data directly in the database
 - Unit conversion in the database
- 13 Please value the criteria "Quality of data", "Quantity of data" and "User friendliness" for "SzenarienDB" by distributing 100 points among the three categories. The more important a category is to you, the more points you assign.
- Quality of data
 - Quantity of data
 - User friendliness
- 14 Which level of detail do you need for your data?
- High (e.g. hourly time series for one year, spatial resolution in scale of kilometers)
 - Medium (e.g. typical days, ...)
 - Low (e.g. aggregated values for countries, years,...)
 - Further explanations:
- 15 Do you know the Open Energy Platform? <http://openenergy-platform.org/>
- Yes, I am quite familiar with it.
 - Yes, but I don't have many experiences with it.
 - No.
- 16 Which properties/features do you appreciate when using OEP?
- 17 Which properties are uncomfortable for you when using OEP?
- 18 News on the project "SzenarienDB" are available here https://www.iee.fraunhofer.de/de/projekte/suche/laufende/SzenarienDB.html?q_ck=1529400732290 or here <http://reiner-lemoine-institut.de/szenariendb/>
Are you interested in getting information via Email about further actions in the project "SzenarienDB"?
- No, thank you.
 - Yes, please. Email:
-

