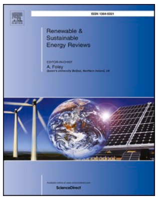
Renewable and Sustainable Energy Reviews 162 (2022) 112472

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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



Data harmonisation for energy system analysis – Example of multi-model experiments

H. Gardian [a,*], J.-P. Beck [b], M. Koch [c], R. Kunze [d], C. Muschner [e], L. Hülk [e], M. Bucksteeg [f]

[a] *Department of Energy Systems Analysis, Institute of Networked Energy Systems, German Aerospace Center (DLR), Curiestr. 4, 70563, Stuttgart, Germany*
[b] *Helmut Schmidt University / University of the Federal Armed Forces Hamburg, Institute of Automation Technology, Holstenhofweg 85, Hamburg, 22043, Germany*
[c] *Oeko-Institute e.V., Merzhauser Str. 173, 79100, Freiburg, Germany*
[d] *Energy Systems Analysis Associates (ESA^2), Bernhardstr. 92, 01187, Dresden, Germany*
[e] *Reiner Lemoine Institut (RLI), Rudower Chaussee 12, 12489, Berlin, Germany*
[f] *House of Energy Markets and Finance, Essen, Germany, University of Duisburg-Essen, Universitätsstr. 12, 45117, Essen, Germany*

ARTICLE INFO

Keywords:
Data harmonisation
Model comparison
Metadata
Energy system modelling
Energy system analysis

ABSTRACT

A variety of models have emerged in the field of energy system analysis to answer a wide range of research questions centred around a sustainable future for the energy sector. Even models designed to address similar issues often have a different focus or modelling approach. Thus, model experiments are a vital tool to provide an overview of the range of models and enable decision-makers to make meaningful model choices. Such comparisons are executed based on a harmonised data set to ensure a high degree of comparability. In the MODEX project cluster, six model experiments, including 40 energy system models, were conducted, and efforts were made to harmonise the input data within the individual comparison and beyond them in the consortium. The experiences and findings of the consortium on how data harmonisation could be performed are presented in this paper. In particular, the focus lies on data transparency to ensure a high degree of reproducibility. A key finding is that while model heterogeneity complicates harmonisation, an early focus on data research and scenario design promotes the creation of a common data set. The metadata collection can provide a significant advantage for the use of model experiment results by external scientists and the data acquisition process itself because of the predefined machine-readable and standardised format.

1. Introduction

1.1. Background and motivation

In recent decades, due to the pressing issue of climate change and the resulting energy transition, a large number of models for energy system analysis have become established, which differ from each other in their methodological approach and focus [1]. These energy modelling activities influence the development of energy systems being therefore a matter of public interest and thus require accountability and transparency. With this ever-increasing demand for more transparency, the topic of open input data, open software and accessible result data is coming more into focus. While modelling activities have remained rather opaque in the past due to a lack of transparency regarding the modelling assumptions [2], leading to a large variance in the modelling results [3] and limited stakeholder involvement [4], the energy research community is expanding to facilitate reproducibility in methods and to share research processes more openly [5]. Efforts have already been undertaken to make the models available to a broad community and thus promote the exchange between modellers [6]. However, the field of system analysis remains convoluted for decision makers as to which model choice to make for a specific research question. One way to gain more insight into the increasingly complex models is to conduct methodological model comparisons. On the one hand, this aims to increase transparency in the model landscape for outsiders, such as project planners or policy makers. On the other hand, it also allows scientists to improve their models or validate their model results. Modelling hurdles can be overcome, as the methodology of model experiments enables to increase the understanding of one's own model and to identify the strengths and weaknesses of the respective modelling approach.

Beyond the mere provision of data, another key to transparency in energy system analysis is the precise description and origin of these using metadata according to the FAIR (Findable, Accessible,

---

* Corresponding author.
E-mail address: hedda.gardian@dlr.de (H. Gardian).

https://doi.org/10.1016/j.rser.2022.112472
Received 31 August 2021; Received in revised form 8 April 2022; Accepted 12 April 2022

List of abbreviations

BDI	Federation of German Industries
CRF	Common Reporting Format
csv	comma-separated values
DWH	Data Warehouse
EMF	Energy Modeling Forum
EMS	Energy Models System
ENTSO-E	European Network of Transmission System Operators for Electricity
FAIR	Findable, Accessible, Interoperable and Reusable
GDP	gross domestic product
IAM	integrated assessment model
NTC	net transfer capacity
NUTS	Nomenclature of Territorial Units for Statistics
OEO	Open Energy Ontology
OEP	Open Energy Platform
RE	renewable energy
SQL	Structured Query Language
TYNDP	Ten-Year Network Development Plan
UNFCCC	United Nations Framework Convention on Climate Change

Interoperable and Reusable) principle [7,8]. If all data used in model experiments are publicly available and sufficiently described, as well as the procedure for comparisons documented, a model experiment can be a great enrichment for the modelling community, because only in this way traceability is guaranteed.

One approach to these principles is the MODEX model experiment cluster. It consists of six research projects (FlexMex, MEO, MODEX-EnSAVes, MODEX-NET, MODEX-POLINS, and open_MODEX) in which in total 39 partners with 40 models (see Table 3 in the appendix) perform model comparisons on current issues in system analysis [9]. Each project has an individual thematic focus within the broad spectrum of energy system analysis topics. Accordingly, they differ in the set-up of the involved models and the spatial, temporal, and technical resolution of the underlying scenario (see Fig. 1).

Although both projects, MODEX-EnSAVes, and FlexMex, deal with sector coupling technologies, they differ significantly in their objectives. While FlexMex has a strong technological focus with a comparison of various load balancing options, MODEX-EnSAVes analyses the influences of the development of electromobility and power-to-heat technologies on the security of supply. In MEO, there is a methodological emphasis on comparing the system operation, whereas in MODEX-NET, models with a similar technical focus on network modelling are compared. The open_MODEX project takes a broader approach, focusing on frameworks with open licenses. The model comparison in MODEX-POLINS includes models for evaluating policy instruments, such as carbon pricing or coal phase-out plans. While the various sectors (electricity, heat, gas, and transport) are considered in different detail in the projects, all model comparisons have the consideration and focus on the power sector in common.

Energy system model results are of particular importance for the assessment of expansion paths and provide the basis for political decisions. They represent a key planning tool for integrating the continuously growing renewable energy production and sector coupling technologies. Incorrect model results and poor traceability could lead to high economic, social, and environmental costs if climate change policy relies on ineffective incentives and steering mechanisms based on false estimates. Therefore, the reproducibility of the results is of crucial importance so that all stakeholders can understand and verify the model

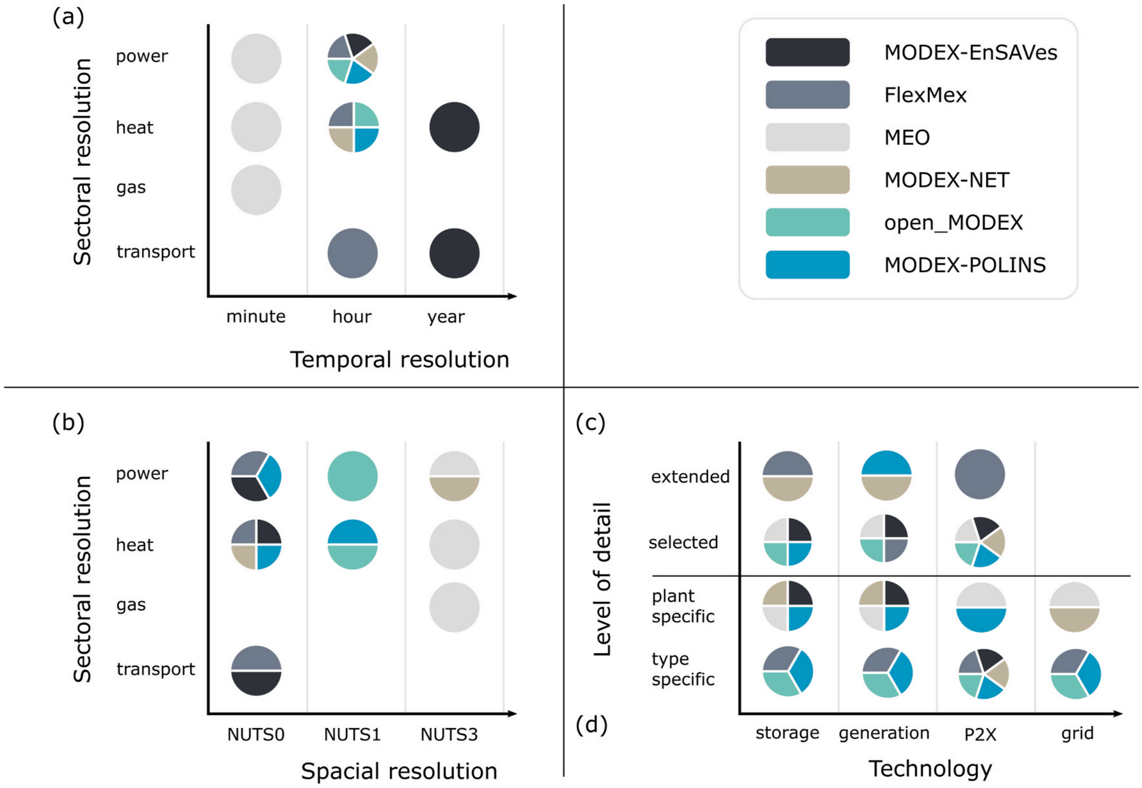


Fig. 1. Classification of MODEX projects. Figure (a) shows the temporal resolution of the individual sectors, while figure (b) depicts the regional resolution. The two-part figure regarding the level of detail in the lower right quadrant shows at the top (c) whether a large range of technologies of a class or only selected ones were analysed and in the lower area (d) whether plants were modelled aggregated or individually. If circle segments of the same colour appear twice in a graphic, the individual models within a project fall into different categories.

outputs. Only in this way do the results provide a real contribution to the energy transition. Also, both the selection of a suitable modelling methodology and the selection of adequate input data greatly influence on the quality and significance of the results of a model-based system analysis. In the MODEX cluster, methodical model comparisons are in the foreground to identify model differences. To increase comparability across the modelling experiments, the input data of the models are harmonised as far as possible.

1.2. State of research

The history of model comparisons in energy system analysis goes back several decades. Already in 1976, for example, the Energy Modeling Forum (EMF) was founded at Stanford University with currently 37 ongoing or completed model intercomparison studies [10].

When classifying past model comparisons, a distinction must be made between theoretical (as carried out in Refs. [11,12]) and practical studies, i.e. those that were conducted based on a literature review and those in which the models considered were applied to identical study cases and the deviations in results subsequently compared. In this paper focused on input data harmonisation, the latter type of comparisons is of importance, some examples of which explicitly mention data harmonisation are listed in chronological order in Table 1 below.

The emphasis of previous model comparisons in energy system analysis lay less on open science or transparent methods and traceability but rather aimed at investigating specific content-related research questions. These thematic focuses play only a subordinate role in the MODEX comparison cluster. Instead, special attention is paid to methodological comparisons of the models, open data sets, and a high degree of transparency and reproducibility.

Furthermore, the topic of data preparation and data harmonisation that is addressed in this paper has so far been of secondary importance in the research landscape but is increasingly coming into focus. Not only in model comparisons is comprehensive input data harmonisation described as relevant for the quality of results [24], but also in studies on model coupling such as [25] database harmonisation is considered essential to obtain reliable results. Only [26] analysed in detail the data harmonisation process for model comparisons and proposed a structured methodology for a comprehensive input data harmonisation that includes a stepwise approach. However, the methodology only applies to the input data harmonisation of IAMs and the model volume was limited to six plus one additional reference model.

But not only the harmonisation process itself represents a challenge and leaves room for further research. There are also still no uniform standards for the semantics, structuring, and storage of energy system analysis data [6]. One of the main reasons for this might be the fact that the methodological development of models and the answering of research questions are often the main focus of system analysis, while a transparent input data basis was hardly an issue. Since the models are usually developed as stand-alone applications, their data management is often also organised individually and tailored specifically to each model. This complicates the transferability and thus the reusability of elaborately created databases in the research landscape of energy system analyses.

Also, the research for reliable data sources with unrestricted access e.g. technical-economic parameters proves challenging, as the few open databases are mostly limited in their technological scope [27]. Even when relevant data for energy systems analysis is made publicly available, there is often a lack of appropriate licensing, lack of indexing to find the data, or missing contextual information to evaluate the data factually [5].

Evolving metadata strings to describe published data attempt to address these shortcomings, but are themselves very heterogeneous, due to the wide variety of energy data to be described [28,29]. To ensure a minimum of openness and reusability through metadata, FAIR principles and the principles of open science have been integrated into the

Table 1
Literature review of conducted model experiments with regard to the degree of harmonisation of the input data: ✓ complete harmonisation, o partial harmonisation and x non-harmonisation.

Model experiment	Reference	Harmonisation effort	Focus
CASCADE MINTS	[13,14]	o	In the European project “CAse Study Comparisons And Development of Energy Models for INtegrated Technology Systems” (CASCADE-MINTS), electricity coupled with hydrogen scenarios were investigated upon key technologies and policy options. A harmonisation of quantitative assumptions was conducted.
AMPERE	[15]	o	In the international “Assessment of climate change Mitigation Pathways and Evaluation of the Robustness of mitigation cost Estimates“ (AMPERE) project, 17 energy-economy and integrated assessment models (IAM) are participating to map the possible pathways of meeting the climate targets, while taking climate policy into account. The input data concerning the baseline and reference policy assumptions were harmonised.
ADVANCE	[16,17]	✓	In another European project called ADVANCE (“Advanced Model Development and Validation for the Improved Analysis of Costs and Impacts of Mitigation Policies”), IAMs were compared and used to investigate the implications of the Paris Agreement.
Multi-model assessment in Japan	[18]	x	The study applies six energy-economic and IAMs to examine the decarbonisation challenges in the energy system of Japan.
RegMex	[19,20]	✓	A model experiment with the aim of publishing and documenting detailed data sets was RegMex, which aimed to derive robust conclusions on the transformation of the energy system.
EMF 34	[21]	x	The EMF 34 study on energy integration and energy trading in North America compares the results of 17 energy system models and discusses their policy motivation.
Multi-model study of the North American energy system	[22]	o	This North American-focused comparison of eight energy economic models assesses the impact of renewable energy policy instruments on economic, environmental and

(*continued on next page*)

Table 1 (*continued*)

Model experiment	Reference	Harmonisation effort	Focus
			planning outcomes in the power sector. Harmonisation is described as incomplete for parameters related to the techno-economic, cost and policy framework assumptions.
4NEMO	[23]	✓	The 4NEMO project aims at integrating economic as well as social dynamics and their related uncertainties into energy system models. The results obtained can indicate which are the strengths, weaknesses and advantages compared to other models.

development of metadata, which has led to some standardisation [7,30]. In this work, we use a metadata string that follows these standards to describe the data and increase the transparency from the different MODEX projects.

1.3. Contribution of this paper

This paper focuses on methods for data harmonisation - data harmonisation in this context refers to the alignment of model inputs of multiple models and not, as in other research fields, the unification of data points from different data sources - and data documentation as well as on the topic of open data. Especially in the application case of model comparisons, data harmonisation is closely linked to the quality of the results, since a discrepancy in the data basis can distort the results and lead to misinterpretations. For this reason, this study focuses on this topic, while the literature sources cited mention data harmonisation, no direct conclusions can be drawn about the underlying methodology and the quality of the implementation.

In addition to methodological model comparisons, the MODEX cluster pays special emphasis to these issues. The aim is to stimulate the exchange of experience between modellers on these topics and to identify barriers and basic approaches to handling them. With this in mind, not only the necessary data matching for the implementation of the methodical model comparisons within the individual MODEX projects will be carried out, but also an effort will be made to achieve a high degree of harmonisation across the projects. This is intended to raise modellers' awareness of these important issues and to provide the community with the opportunity to benchmark their model using detailed documentation of used data and methods.

With this paper, we extend previous studies in this area, as data harmonisation in a consortium of the size and heterogeneity of models has not been investigated before. The methods are shown and their boundaries increase their relevance as they have been tested based on this large-scale project consortium. In addition, the expertise of a large number of renowned research institutes, which are all active in the field of energy system analysis but always with a different perspective, could be incorporated, as well as the experience from six independent research projects.

The cross-project data coordination was the responsibility of the group of data managers, which consisted of one partner from each of the MODEX projects. Sections 2.1 to 2.3 describe the approach and results of the data managers group in the MODEX cluster. The focus was on the most important framework data for modelling (esp. wholesale fuel prices, emission factors, etc.), which are usually required by a large part of the models used in the entire network and at the same time represent essential model drivers.

The cross-project data management formed the basis for the necessary further data harmonisation within the individual MODEX projects for the model comparisons. Depending on the analysis focus and the models used, different approaches were followed. Section 2.4 presents the methodological approach of MODEX-EnSAVes as a representative example, since in this project not only models are compared, but also coupled with each other through a common database. The latter places special demands on data harmonisation, which are illuminated against the background of this still rather young trend in energy system analyses [25]. Section 2.5 is then dedicated to the topic of open data and the documentation and reusability of model data. Finally, chapter 3 collects, clusters, and describes the hurdles and challenges encountered during the harmonisation process. Approaches to solutions and resulting consequences are explained accordingly.

2. Materials and methods

In the following, we will describe our data harmonisation procedure in more detail, focusing on the harmonisation of input data. For these, several harmonised modelling aspects could be identified. Because of the different project foci and the correspondingly divergent results data as well as the different solutions for formatting and data transfer, it was as a consequence infeasible to harmonise the output data. Therefore, these will not be considered further in the following.

Also, this study does not mainly focus on model validation. The models used have been tested for suitability in numerous other publications and projects. Likewise, extensions of the models are not in the focus, but the comparison on the current development status. The models were compared in their typical environment and therefore the input data were oriented as close as possible to the data used so far. In this context, many models have high requirements for realistic data. Therefore, the utilisation of synthetic inputs was only useful to a limited extent.

2.1. General modelling aspects

First of all, general modelling aspects were examined. These were divided into two groups. Under the spatial system boundaries group, the regions & countries, Nomenclature of Territorial Units for Statistics (NUTS) levels, and network levels of the MODEX projects were compared. Under the analysis period group, the start year, analysis horizon, base year, temporal resolution, weather year, and real price reference year were collected. For the analysis period, several aspects could be harmonised, while others remained unequal due to the different focal points of the studies.

Regarding the considered regions & countries, MODEX-EnSAVes, open_MODEX, and MEO perform their investigations only for Germany or single subregions. MODEX-POLINS, FlexMex, and MODEX-NET, however, take the neighbouring and other European countries up to all European Network of Transmission System Operators for Electricity (ENTSO-E) countries into account. In terms of regional resolution, some projects consider details at the NUTS 3 level, while the majority distinguishes only NUTS 0 regions. Different grid levels are considered in MEO only, where the low-voltage grid is modelled as well.

For the analysis period, the start year was harmonised for all six MODEX projects such that all use 2016. The analysis horizon varies significantly from project to project. While MODEX-POLINS, MODEX-EnSAVes, and MODEX-NET consider an analysis period up to the year 2030, FlexMex and open_MODEX chose the year 2050. The MEO project does not perform a classical system analysis and thus has no analysis horizon, but relates its investigations exclusively to the start year. The base years vary, of course, due to the different analysis horizons. While MODEX-POLINS chose 2025 and 2030 as base years, FlexMex uses only 2050 and open_MODEX considers base years in intervals of 10 years. Without an analysis horizon, MEO strictly speaking has no base years, but forecasts for 2024 and 2034 are used. MODEX-EnSAVes performs calculations for 2020, 2025, and 2030 and the MODEX-NET project chose 2016 and 2030 as base years. The temporal resolution is very similar in almost all projects, as calculations in the electricity sector are carried out on an hourly basis. Only MEO has a higher temporal resolution of 15 min, for its detailed network calculations. For the weather year, all projects except FlexMex (2012) and MEO (2011) chose the start year 2016. MODEX-EnSAVes and MODEX-NET use additionally other weather years, e.g. to simulate an extreme weather year. For the price reference, all projects uniformly use 2016.

2.2. Gathering of the data requirements

While the focus was previously placed on the harmonisation of general modelling aspects, the harmonisation of input data will now be considered in the following.

Depending on the type and focus of the individual models, there are different requirements for the input data needed, for example about their spatial and temporal resolution, but also generally for individual categories of input data. For instance, power grid models as applied in MODEX-NET and MEO inherently rely on input data on the topology and the degree of expansion of the power grid and also require all data on power generation and demand in a corresponding spatial resolution at the level of the grid nodes.

To get an overview of the required input data and the resulting outputs of the 40 models used in the individual MODEX projects, these were systematically recorded in an input-output table based on various categories. The eight overarching data categories are:

- Macro-economic and statistical data, including subcategories for the gross domestic product (GDP), population, employees, households, buildings, freight volume, policy objectives, etc.
- Environmental data, including subcategories for weather data, fuel type, technology specific emission factors, and land usage, etc.
- Demand data, including subcategories for demand volumes and profiles for electricity, heat, fuel types and other energy carrier as well as for CO_2 needed in methanation, etc.
- Installed infrastructure data, including subcategories for power plant portfolios, storages for electricity, heat or gas, electricity, heat and gas grids, vehicle fleets, load points for electric vehicles, etc.
- Technology- and plant-specific parameters, including subcategories for efficiencies, load gradients, specific investments, lifetimes, specific costs, availabilities, etc.
- Prices and costs, including subcategories for fuel, electricity, heat, CO_2 for methanation, CO_2 certificates, and redispatch costs, etc.
- Actor behaviour and acceptance, including subcategories for self-consumption maximisation, driving profiles, remediation activity, etc.
- Deployment/utilisation of infrastructure, including subcategories for deployment profiles for generation and storage of electricity and heat, utilisation of electricity grids, congestion management, expansion or deconstruction of infrastructure, etc.

As expected, the input-output table result shows a wide range of input data used across the more than 150 subcategories. In addition, some data are input data for some of the models, while they are part of the outputs for others. For example, this is the case for dispatch-only models compared to investment models for installed generation capacity. The input-output table is online available at openenergy-platform.org.[1]

Thus, only a small subset of the possible input data can be considered for data harmonisation, which enters all models as input data and for which data harmonisation also appears to make sense. The following input data were identified as the lowest common denominator across all models and projects: Prices for fuels and CO_2 certificates, fuel-specific CO_2 emission factors, country-specific load profiles for electricity and district heating, and net transfer capacities (NTC) at the interconnection points in the European electricity grid.

To ensure the highest possible transparency regarding the input data used in the individual MODEX projects in addition to the data harmonisation, their metadata were collected and published (see 2.5).

2.3. Description of model input data

As part of the data harmonisation process, a literature search was carried out for the individual input data and a default value was formed for each. For the development of fuel and CO_2 prices, the “Current policies” scenario from the World Energy Outlook 2018 was selected as the base source [31]. In addition to the prices for CO_2 certificates, this affects the energy carrier hard coal, natural gas, and oil. For lignite, the Grid Development Plan for Electricity 2030 (version 2019) was used [32], for nuclear energy the Ten-Year Network Development Plan (TYNDP) 2018 [33] and biomass products the assumptions of the reference scenario from the Federation of German Industries (BDI) study “Climate Paths for Germany“ [34]. All prices were converted according to the price reference year 2016.

The fuel-specific CO_2 emission factors were determined by the United Nations Framework Convention on Climate Change (UNFCCC) reporting and based on the national inventory report on the German greenhouse inventory [35]. The methodology for compiling the UNFCCC greenhouse gas inventories is based on the so-called Common Reporting Format (CRF), a standardised format for all sectors and sub-categories. Greenhouse gas emissions are usually calculated as the product of activity rate and emission factor.[2] The activity rate often corresponds to the fuel input in the energy sector. The emission factors are stored in country-specific tables for each greenhouse gas and energy source.[3] For lignite, the average value between East German and West German coalfields is calculated and for biomass, the CO_2 emission factors are set at zero.

The load and renewable energy (RE) profiles for 2016 are based on the ENTSO-E Transparency Platform and have been normalised to values between 0 and 1 in terms of installed capacity or maximum electricity demand [36]. Based on these normalised profiles, the absolute profiles can then be formed as a function of the annual energy quantity for the individual scenarios.

The harmonised input data was made available to all MODEX projects as default values via comma-separated values (csv) files. If possible, these default values should be used in project-specific scenarios. If this was not possible, other data could also be used. This was then briefly justified in the respective scenario documentation.

Table 2 shows an overview of the proposed data sets and to what extent a harmonisation could be conducted, where individual projects deviate from the suggested data, or even whether the proposed data set is not used in any project.

Due to the very different research questions of the individual MODEX projects, a complete harmonisation was not possible and this would also not have been expedient. However, Table 2 shows that some parameters could be harmonised across the projects. The first column describes which data is involved in each case, e.g. the electricity load profile. The proposed data set or the source is mentioned in the second column. The

---

[1] https://openenergy-platform.org/dataedit/view/reference/modex_data_publication_i_o_table Please note that classification of the frameworks only refers to the requirements of the scenarios in the specific projects. The frameworks are more versatile.

[2] https://unfccc.int/process-and-meetings/transparency-and-reporting/reporting-and-review-under-the-convention/greenhouse-gas-inventories-annex-i-parties/reporting-requirements.

[3] https://www.umweltbundesamt.de/sites/default/files/medien/361/dokumente/co2_ef_liste_2022_brennstoffe_und_industrie.xlsx.

Table 2
Data harmonisation – model inputs of scenario dependent parameters.

Parameter	Suggested source	MODEX-POLINS	FlexMex	open_MODEX	MEO	MODEX-EnSAVes	MODEX-NET
Electric load profile	Eurostat 2018a	X	X + ENTSO-E 2016	Own Data	SimBench	X + me	X
Heat load profile	Bründlinger et al 2018, IEA 2018a, Renewables.ninja 2020	X	X	Own Data	Own data	X	Own Data
RE generation/ RE feed-in	OPSD 2020, Entso-E 2018b	X	Own data from EnDat-Modell, Scholz2012	me	me	X	X
Emission factors/ energy generation	UBA 2016 [35]	(X)	X	X	X	(X)	X
Installed generation capacity	Bundesnetzagentur 2019, Entso-E 2018b	X	me	MaStR2021	me	X	X
NTCs	Entso-E 2018b, Rippel et al 2019	X	X	DLR SciGrid	X	X	X
Fuel prices	IEA 2018b, BMWi Langfristszenarien, da Szenariojahr 2050	X	X	X	X	X	X
CO_2 price	IEA 2018b, Hirst and Keep 2018, BMWi Langfristszenarien, da Szenariojahr 2050	X	X	Own data	X	WEO2018 (Scenario "Sustainable devel.")	X

X	harmonises with the proposed MODEX values
(X)	almost harmonised with the proposed MODEX values, only slight deviation
me	model endogen
▨	other data set / own data set

following columns list for each MODEX project whether the proposed source has been used or if there are deviations from it. These deviations can be minor, in which case it is considered as harmonised. However, data values can also be generated endogenously in the models and thus are not included in the models as a data set at all. In addition, the table indicates as well the utilisation of other data sources, which are then labelled with the respective source.

In summary, it can be said that the emission factors, the prices for fuels and CO_2, and the NTCs have been harmonised well. For the electricity and heat profiles, it is clear that harmonisation was more difficult, as can be seen from the use of various data. The same applies to the RE generation and the installed electrical capacity.

2.4. Need and background of data harmonisation for model coupling – example of MODEX-EnSAVes

The challenge of harmonising models, which all perform the same task, becomes apparent in the described methodology and the limited number of harmonised data sets. An additional component of complexity arises when these models do not compute in parallel but in cascade.

Many model developments take a comprehensive system view and depict the relevant stages of the energy supply chains in a highly aggregated or simplified manner, but as completely as possible. In recent years, however, there has been a trend for many model developments to focus only on individual aspects or sub-areas of the energy system (e.g. on the development of energy demand in the transport sector or on the design of regional energy markets), but to model these in much greater detail [25].

This is essentially due to two factors. Firstly, during the energy transition, an enormous variety of new stakeholders and options for action in system design are emerging. Secondly, the requirements from politics and industry for the findings of system analysis have increased significantly. In particular, more robust assessments of economic viability, market success or investor behaviour as well as the consideration of acceptance in the economy and the society, are becoming more important. At the same time, the holistic assessment of the impacts of alternatives to system design in terms of life cycle assessment or social aspects is becoming increasingly important.

This development has led to more detailed and thus more complex research questions, which usually cannot be modelled and analysed in the required depth of detail with a single comprehensive system model. Nevertheless, the cross-sectoral system view still plays an important role in order not to neglect the existing interdependencies between the individual parts of the energy system. To adequately take these interactions into account, the specific models for individual system components must be coupled with each other and, if necessary, also with models with a comprehensive aggregated system view and to form an Energy Models System (EMS) [37]. The coupling primarily concerns the data exchange between the models, in which the results of one model serve as input data for other models. The model application within the EMS usually takes the form of an iterative process. This process may require several iterations to converge the results of the models involved.

Due to the numerous degrees of freedom in modelling as well as the lack of standards for structuring and holding the necessary data for the energy system analyses, this point represents the greatest challenge in model coupling. The creation of an EMS is usually done in the following three steps:

Step 1: Selection of the models to be used and definition of the interfaces for data exchange

The selection of the models to be used essentially depends on the problem to be investigated, the boundaries of the system under consideration, and the required spatial, temporal and technical resolution of the modelling. The model selection must be made in such a way that all relevant system aspects and their interactions are covered with the necessary level of detail.

To define the interfaces for data exchange, the defined input data are first compared with the output data of all models involved and the overlaps of input and output data sets for the possible data exchange are identified. It is recommended to focus on the most important model drivers. On this basis, the needed interaction of the individual models within the EMS is designed and agreed upon. It must be taken into account that the models used may well have overlaps in the modelling of individual system aspects. For this reason, in addition to determining the data sets to be exchanged, the modelling of the individual system aspects must also be delineated and the model responsible for each must be determined. At the same time, the handling of necessary data transformations during data exchange between the models is methodically determined. This can concern, for example, the breakdown of spatially aggregated results to local type regions or the conversion of wholesale prices into end-user prices.

Furthermore, a detailed model application plan is created in this step. This contains the sequence and the schedule of the individual model applications and schedules the respective data exchange. The model application plan is particularly important if models from different institutions are to interact smoothly, as delays in one model usually affect all models involved due to the iterative model application process.

Step 2: Generation of a common harmonised database

The second step essentially corresponds to the procedure of the data harmonisation in the MODEX cluster already described in Section 2.2. From the matching of the data sets in step 1, the input data sets of all participating models with the same content are identified, which are not provided as an output by one of the participating models within the framework of the data exchange, i.e. which do not represent endogenous variables within the EMS. These are essential model drivers that usually represent exogenous influencing factors for the development of an energy system (e.g. the development of the population or industrial production) or are to be defined at the system boundaries (e.g. world market prices for energy carriers). This information forms the common framework data that must be harmonised between the models to ensure a consistent analysis within the EMS. Special attention should be paid to how a specific data set is used in the individual models. For example, in some models interest rates are used to model specific stakeholder preferences, while in others they represent the usual discount rate in an economy or industry.

Step 3: Configuration of the model interfaces for data exchange

The third step involves defining the required data mapping. Since the models to be coupled were usually developed independently of each other as stand-alone applications with separate data storage, they usually have significant differences in the data set structures. As a further harmonisation aspect, the mapping thus forms the basis for achieving data compatibility between the models and is therefore also a decisive prerequisite for their consistent interaction. Challenges arising here are of various types:

- The models use different levels of detail or aggregation of the data (e.g. regarding countries, sectors, technologies, energy carriers, etc.).
- Different identifiers for the same datasets exist (count and labelling) and there are different restrictions for the labelling of identifiers in the models (allowed characters and length).
- The time structures of the same datasets can differ (yearly, seasonal, hourly, etc.).
- The provided or required data have different file formats (e.g. csv, structured query language (SQL), etc.).

Another major challenge arises from the size of the data sets to be exchanged, which results from the increasing requirements for high spatial and temporal resolutions in system modelling. A small example to illustrate the dimensions that typically occur: A single data set on German electricity demand, broken down to NUTS 3 level for 8760 h per year, already reaches a size of over 10.5 million data points with an analysis horizon of 30 years with only 3 modelled interpolation years. If the electricity demand was further differentiated into 50 individual energy application processes, e.g. for adequate modelling of load management measures, more than 500 million data points would be reached. Applied to EU-wide modelling, this data set would comprise over 1.7 billion data points. These dimensions place enormous demands on the common model data management and the tools for data exchange. Specified models are necessary to handle these challenges efficiently.

These data management tools include functionalities such as the integration of harmonised data (step 2) into a common database, the provision of an interface for data exchange (step 3), and linking input data to result data for publication, ideally combined with the corresponding metadata. Some tools already support these or similar functionalities, such as the data warehouse (DWH) of the project partner ESA[2] used in MODEX-EnSAVes [38], but also pyam [39], Spine [40], and the Open Energy Platform (OEP) [41] in combination with the oedatamodel [42] can automate the coupling process. The utilisation of those tools makes it possible to meet the increasing requirements for transparency of system analysis as well as documentation and reusability of data. These essential aspects are considered in more detail below.

2.5. Required metadata management

As a consequence of the methods and experiences described, sound data and model management is necessary for modelling exercises, but they also must be transparent to ensure good scientific practice. Metadata, as a flexible vehicle to document information of information, supports the development towards more transparency along the modelling process from scenario and model definition, selection of input data, pre-and post-processing of data to results communication. Of these steps, the documentation of input data is the most cost-effective and promising to increase transparency due to its low complexity and evolving documentation standards within subdomains of energy modelling [28,29]. The other process steps' documentation and comparability across different models and modelling exercises remain more complex or less uniform (in the domain of energy modelling), despite standardisation efforts by various research groups. These efforts rely largely on metadata and focus on scenario documentation and interoperability; the linkage of existing databases and provenance documentation on modified data sets; the reproducibility and reliability of scenario processing; and the development of an open energy ontology as an accelerator for improved interoperability [8,43–46]. These projects follow principles of open science as guidelines to promote transparency and reproducibility. Some MODEX modelling experiments publish their input data, allowing reproducibility, and some are also under an open license for free re-use.

This reflects the fact that transparency and reproducibility in energy system research is still not a matter of course, despite the recent positive developments [6]. Although documenting and publishing the data is the simplest form of contributing to transparency, it is not yet a complete contribution to its reusability. Data sources must be licensed and documented in the metadata accordingly (the same applies to published methods) [47]. Enough meta-information must be given in the metadata to ensure clear data interpretation. Data sets in energy modelling, however, are very large and the types of data are very heterogeneous, stemming from fields of geography, meteorology, economics, and engineering, posing challenges to metadata documentation. In particular, the requirements for the level of detail of documentation differ between domain experts, who tend to need a higher density of information, and the non-domain experts, whose information needs are less detailed to ensure subsequent use of the data. Metadata documentation must be flexible enough to meet these needs to form a standard in the domain of energy system modelling.

Despite the incompleteness of data publications in the MODEX cluster, a complete set of metadata on all modelling parameters is compiled.[4] Challenges remain in publishing open datasets, often due to a lack of open and properly licensed data. When republishing data openly, the rights of the copyright holders of each data point must be known to choose an appropriate open license. Furthermore, the chosen license should be as unrestricted as possible to be compatible with other open licenses and to avoid data silos. Funding institutions such as the European Union have recognised this and increasingly demand the use of open data in tenders. The minimum target of publishing metadata for the input data has been defined in the MODEX cluster to increase transparency in research and to take a step towards reproducibility. Ultimately, the collected metadata will facilitate understanding and comparability of the modelling activities, increase the transparency of scientific practice, promote reproducibility of research and thus lead to better interpretation of the research output. The metadata evaluations of the MODEX cluster show though, that the input data across the projects is largely under copyright protection.

Among the MODEX model comparisons, the greatest possible data harmonisation among selected input parameters has been targeted. However, as seen above, this has only been achieved partly due to the diversity of participating energy models or energy modelling frameworks and their largely heterogenous need for data input and scenario frames and model assumptions. Its documentation across all projects has been facilitated with the *oemetadata (v.1.4.1)*, as both human-and machine-readable standard of data sources, licenses, and scenario assumptions [28]. The oemetadata string has been used in the energy system modelling community for a few years. In addition to other types of documentation, it helps to find and process data sources and scenarios more easily. It allows to document multiple resources for the parameterisation of input parameters and provides detailed documentation possibilities of temporal and geographic information, licenses, and more in-depth information. Simultaneously, the oemetadata follow the FAIR Guiding Principles for scientific data management and stewardship [48]. The FAIR principles promote guidelines for data stewardship established to facilitate the process of discovery, evaluation, and reuse of data and other digital assets in publicly funded projects. The principles explicitly include “computational stakeholders” who, in addition to human reuse of resources, play an increasing role in exploring and processing data in data-driven research projects. The principles in the short state, that a digital asset should be findable and described with rich metadata; it should be accessible and metadata accessible, even when the digital asset is no longer available; it should use interoperable formal and accessible metadata written in a broadly applicable language; it should ensure reusability by indicating a clear and accessible license. The *oemetadata (v.1.4.1)* implements those principles and remain flexible in its structure to meet the future challenges of data interoperability.

3. Results and discussion

The difficulty of the targeted undertaking is the harmonisation beyond the individual model comparison projects in the MODEX cluster to establish comparability here as well. The different model and project foci reduce the possible extent of the unified input data to specific parameters of the electricity sector. Nevertheless, using the methodology described above, we were able to standardise the part of the data that was harmonised in the MODEX cluster. However, experience has been gathered on the challenges that have arisen in the course of harmonising model inputs for model comparisons. A survey among the participating modellers was able to capture the different aspects of the hurdles related to a harmonisation of input data and to collect the resulting issues (see Fig. 2). They were asked what model-independent and model-dependent hurdles they encountered during the harmonisation process, both within and across projects. We identified four categories of problem areas: issues due to model differences, challenges related to the scenario definition, difficulties in adjusting data in the models, and in the data collection itself.

3.1. Model differences

Due to the different levels of detail concerning spatial, temporal, and technological resolution, a reduction of the scenario definition and the required data respectively to the lowest common denominator is inevitable. The models also differ in the exogenously required and the endogenously calculated parameters. Even though data requirements of the models within the projects show a significantly larger intersection than in the overall MODEX consortium, nevertheless, the consensus on a uniform data basis reduces the model scope.

3.2. Scenario definition

Reducing the model scope to the smallest common denominator has led to scenarios of lower complexity that deviate from common plausibility in some MODEX projects. While other projects analysed extensive case studies and were able to provide recommendations for action on their basis, this goal was secondary, since the focus was placed on the comparison of the models and the scenarios were merely seen as a test case.

3.3. Data adaptation

In addition, transferring common input data into the model’s parameterisation can be error-prone and introduces hurdles. Due to the vast number of parameters that models must adjust, multiple iterations are required to correct errors in the transfer or to add forgotten data. This iterative process can be lengthy for model comparisons involving a cascade, that is, using model coupling as described in section 2.4. Also, adjusting parameters may be more time-consuming for models that use databases in the background. It has been shown that an early comparison of the results and the input data that the models ideally provide when they report their results, simplifies the outlined process, shortens it, and makes it more manageable. The precise description of the parameters and the assignment of units ensure a uniform understanding. In addition, the use of the Open Energy Ontology (OEO) [49] can provide a remedy as it promotes a uniform understanding. Current further developments of the ontology make it possible to describe a large part of the input data of energy system models [46]. Following the subsequent translation of the parameter names into the model parameter names, it may be necessary to convert the values into the correct unit to bring them into the proper form of the model inputs.

3.4. Data acquisition

Another issue that can complicate the transparent comparison of models is the acquisition of the data itself. Since this project aims at maximum transparency and traceability, the use of open and licensed data sets is desired. The thus limited data pool often does not include scenario data and for the non-electricity sectors, the data situation is scarce. As a result, the projection of current data and the use of expert guesses are often resorted to. Current funding of various data-focused projects such as LOD-GEOSS [50] or SEDOS [51] as well as the criticism of duplicated work and difficulties in data research in Ref. [6] also point to the need for well-documented, findable and open input data for energy system analysis, to which the consideration of the FAIR principles in the MODEX projects contributes. By publishing the metadata of the data used in addition to uploading the actual data to the OEP and other platforms, the MODEX project cluster has been able to provide maximum transparency to deliver another source of a well-described data collection to be used in system analysis. Even if not all data

---

[4] https://openenergy-platform.org/dataedit/view/reference?query=MODEX.

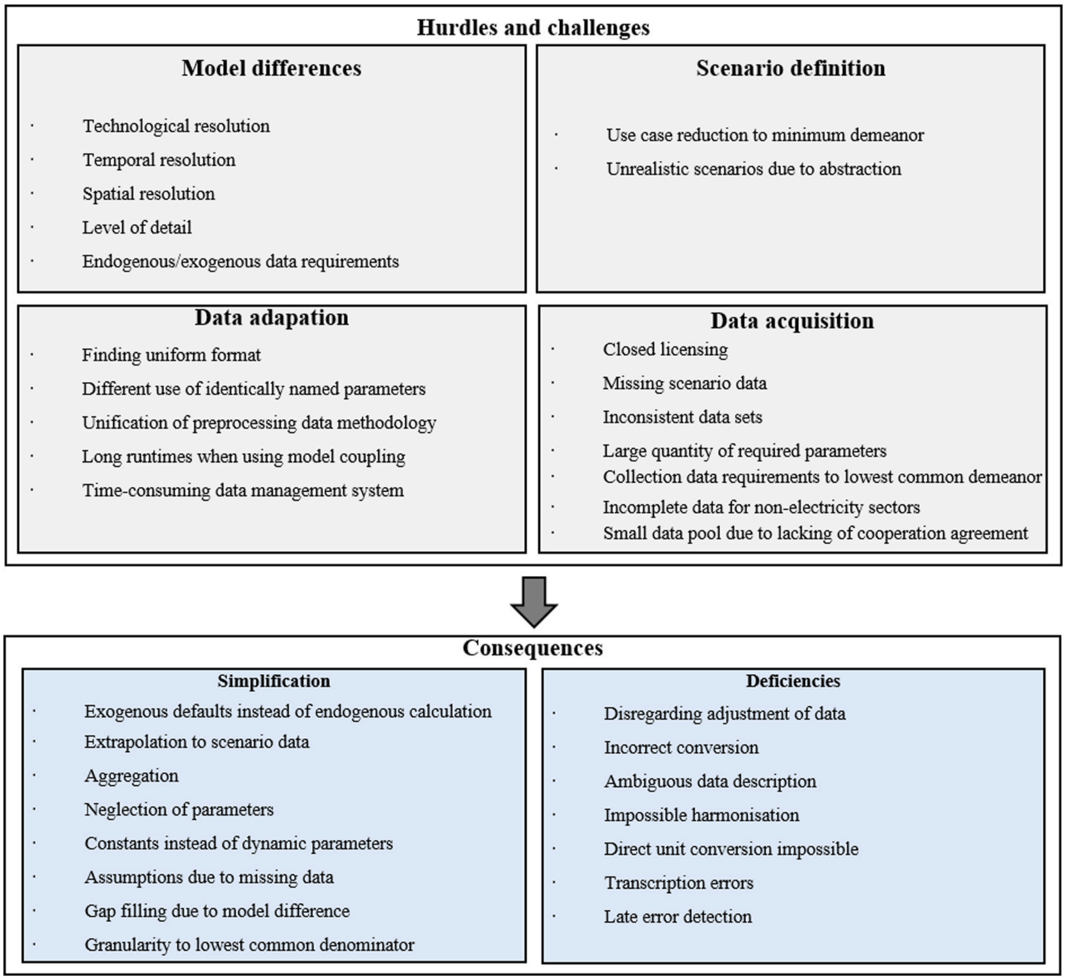


Fig. 2. Collection of hurdles and challenges encountered during the harmonisation process and the resulting simplifications and deficiencies.

sources are publicly accessible, the data provenance is documented. The metadata created[5] in the MODEX projects based on the oemetadata v1.4.1 meet all the requirements of the FAIR criteria for findability (F1–F4) and accessibility (A1-A2) but have shortcomings regarding the second interoperability criterion (I2). While the metadata vocabulary is openly discoverable, documented, and defined on GitHub, it is not associated with globally unique and persistent identifiers. This shortcoming stems from the development status of the metadata standard and is addressed in oemetadata v.1.5.1, which allows the metadata vocabulary to be mapped to a community-developed ontology (OEO) that uses globally unique and persistent identifiers alongside vocabulary definition. About the reusability criteria (R1), we are confident that this is met in full. We provide a licence for the metadata and a licence for the data sources whenever indicated. In cases where no license is given, we acknowledge the authors' full copyright in the source. We show provenance when necessary, and with our documentation, we fulfil the community standard to comprehend the modelling exercise with the documented data. Using oemetadata v.1.4.1, we also employ a standard metadata format from the energy system modelling community. Table 4 in the appendix shows the evaluation of the FAIR criteria in more detail.

To quantify the impact of problem areas discussed above, the survey also asked about deficiencies in the preparation and harmonisation of data and necessary simplifications. The modellers have pointed out that a full harmonisation was infeasible. It was noted that errors in the parameterisation could not be excluded entirely due to the possibility of a different understanding and use of parameters as well as the large quantity of input data used. In some places, a direct conversion into the required unit was not possible (e.g., specification of transport service demand in person- or ton kilometers vs. specification of number and demand of individual passenger cars). Also, far-reaching simplifications have to be accepted to create a common data basis, such as the aggregation to the same level of detail or the omission of parameters that would allow a more detailed description of technologies.

4. Conclusions

The quality of model comparison results is directly dependent on the degree of harmonisation of the input data. A distortion of the results due to deviating input data can be excluded only through a maximally harmonised data basis. However, in addition to a harmonised database, which is decisive for the outcome of a model comparison, its documentation and licensing is essential to make the results useable for scientists beyond the model experiment. The experiences during the implementation of this process within the MODEX project cluster were described in detail and the applied metadata string was introduced. Nevertheless, the work has shown that the procedure of data harmonisation challenges modellers with a considerable task. To point out the additional challenges of data harmonisation in a model coupling context, this topic has been taken up in a short excursus.

It could be observed that the described hurdles and issues concerning the harmonisation of input data for model comparisons are amplified according to how much the models under consideration differ from each

---

[5] https://openenergy-platform.org/dataedit/view/reference?query=MODEX.

other. Moreover, in the case of the MODEX cluster, the variety of project-related research questions increased harmonisation challenges. Although the data used could be clustered into six thematic groups (costs, environmental, installed capacity, load, supply, techno-economic), which depict the main data needs of each energy system modelling project, individual MODEX project foci allowed only for a fraction of data to be harmonised. Solely the unification of basic parameters, such as emission factors and fuel costs, was feasible to realise. However, for the harmonised data a duplication of research work could be avoided.

Although overarching harmonisation was not possible, the efforts of the group of data managers had several positive outcomes. Using the overarching work, a rough framework could be created and consistency within the project cluster could be established. MODEX was a unique opportunity to develop a sound methodology for harmonisation in model comparisons, as it is rare to find such a large number of models in one project cluster, representing a broad range of energy system analysis. With the help of the knowledge gained, it was possible to show how complex and time-consuming data harmonisation processes could be performed. In the process, several hurdles could be identified and solutions could be proposed. However, the limitations of the harmonisation of input data were also revealed.

The steps to be taken are first the mutual description of the models involved to get a basic understanding of the model characteristics. Then a collection and description of all models' required input and generated output data has to be created to check for overlaps. Within a common data search, the identified parameters are assigned values and these are provided with units. The resulting data set is annotated with the described metadata string and uploaded in the standardised format of the OEP. Each model is now responsible for a correct transformation of the provided data into the required form. The resulting publicly accessible dataset can henceforth be used without restrictions. However, it is difficult to quantify the benefits of this approach, as there are no counters to prove downloads of the data and the use of the metadata by other scientists.

Furthermore, several general ways could be identified to improve a harmonisation process of input data for future projects. Among them is a detailed assessment of data requirements and the avoidance of possible comprehension related errors by using the OEO to describe the parameters accurately. Moreover, errors in parameterisation can be identified at an early stage by comparing the model result data and the input parameters used. During the project planning phase, the time-consuming process of data harmonisation should be sufficiently considered. A focus should be placed on a uniform data format and data acquisition to create a sound basis for model comparisons. The use of open and licensed data increases the transparency and reproducibility of model comparisons. In addition, the provision of metadata for model inputs has a favourable cost-benefit ratio. Even without a full harmonisation of input data, sound documentation can provide a high degree of reproducibility and avoid parameterisation errors.

Besides the uniform data itself, a uniform format for the input and output data also has an advantage for performing model couplings and comparisons. Similar data processing increases interoperability and minimises the susceptibility to errors in parameterisation not only for model comparisons but also model couplings. A uniform data format would thus be desirable in energy system analysis, but would involve considerable effort and expense due to the historically evolved models and different technical solutions.

The aim of the work of the data managers in the overarching MODEX cluster was not only to harmonise data across projects but also to raise awareness amongst scientists in the field of energy system analysis of the relevance of open and well-documented data, which is an essential basis for robust results. The recommendations proposed in this paper are therefore intended to be simple in their application to keep the barriers to widespread use low but to have a major impact on improving transparency and traceability.

Author contribution

Hedda Gardian: Conceptualisation, Methodology, Investigation, Data curation, Writing - Original draft preparation, Supervision, Writing - Review \& Editing, Visualization **Jan-Philip Beck:**Conceptualisation, Methodology, Investigation, Data curation, Writing - Original draft preparation, Supervision, Writing - Review \& Editing **Matthias Koch:** Conceptualisation, Methodology, Investigation, Data curation, Writing - Original draft preparation, Writing - Review \& Editing **Robert Kunze:** Conceptualisation, Methodology, Investigation, Data curation, Writing - Original draft preparation, Writing - Review \& Editing. **Christoph Muschner:**Conceptualisation, Methodology, Investigation, Data curation, Writing - Original draft preparation, Writing - Review \& Editing. **Ludwig Hülk:**Conceptualisation, Methodology, Investigation, Data curation**Michael Bucksteeg:**Conceptualisation, Methodology, Investigation, Data curation, Writing - Review \& Editing.

Declaration of competing interest

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

Acknowledgements

The research for this paper was performed within the projects 'MODEX' supported by the German Federal Ministry for Economic Affairs and Climate Action under the grant numbers 03ET407 plus the individual project number 4–9. The authors thank all project partners for their assistance and participation in the harmonisation process.

Annex.

The models involved in the MODEX thematic network are documented in a model specific factsheet on the open energy platform.[6] The models can be found using the tag "MODEX".

Table 3
Models participating in the MODEX model comparison cluster.

Model name	Abbreviation	Institute/Company	MODEX project	Open Source
	μGRiDS	Offenburg University of Applied Science	MEO	true

(*continued on next page*)

---

[6] https://openenergy-platform.org/factsheets/models/, https://openenergy-platform.org/factsheets/frameworks/.

Table 3 (*continued*)

Model name	Abbreviation	Institute/Company	MODEX project	Open Source
Microscale Grid Reactive Decentralised Energy Systems				
ALternative Automobiles and Diffusion and INfrastructure	ALADIN	Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe	MODEX-EnSAVes	false
ASsessment of TRAansport Strategies	ASTRA	M-Five GmbH, Karlsruhe	MODEX-EnSAVes	false
Balmorel Model	Balmorel	Technical University of Denmark (DTU), Copenhagen	open_MODEX	true
Building Simulation Model	BSM	ESA^2, Dresden	MODEX-EnSAVes	false
Dispatch and Investment Evaluation Tool with Endogenous Renewables	DIETER	German Institute for Economic Research (DIW), Berlin	FlexMex	true
DIMENSION	DIMENSION	Institute of Energy Economics (EWI), University of Cologne	MODEX-POLINS	false
European Electricity Market Model	E2M2	Institute of Energy Economics and Rational Energy Use (IER), University of Stuttgart	FlexMex	false
European Electricity Market Model	E2M2s	Chair for Data Management Systems and Knowledge Representation, University of Duisburg-Essen	MODEX-EnSAVes	false
ELectricity MODel	ELMOD	Chair of Energy Economics, Technical University Dresden (TUD)	MODEX-NET	true
electricity LOad curve ADjustment	eLOAD	Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe	MODEX-EnSAVes	false
ELectricity TRAnsshipment MODel	ELTRAMOD	Chair of Energy Economics, Technical University Dresden	MODEX-EnSAVes	false
Electricity Market Model	EMMA	Neon Neue Energieökonomik GmbH, Berlin	MODEX-POLINS	false
Energy Agent	Energy Agent	Chair for Data Management Systems and Knowledge Representation, University of Duisburg-Essen	MEO	true/ false
electricity Transmission Grid optimization	eTraGo	Institute of Networked Energy Systems, German Aerospace Center (DLR), Oldenburg	MODEX-NET	true
EuroPower	EuroPower	Forschungszentrum Jülich (FZJ IEK-3)	MODEX-NET	false
FORecasting Energy Consumption Analysis and Simulation Tool	FORECAST	Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe	MODEX-EnSAVes	false
Global ENergy System MODel	GENeSYS-MOD	German Institute for Economic Research (DIW), Berlin	open_MODEX	true
GENetic optimization of a European Energy Supply SYStem	GENESYS-2	Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University	FlexMex, open_MODEX	true
Distribution Grid Energy Model	GridSim	Research Center for Energy Economics (FfE), Munich	MEO	false
HeatSim	HeatSim	Chair for Data Management Systems and Knowledge Representation, University of Duisburg-Essen	MODEX-EnSAVes	false
IDILES	IDILES	University of Duisburg-Essen	MODEX-EnSAVes	false
IntegraNet/TransiEnt Library	IntegraNet/ TransiEnt	Gas- und Wärme-Institut Essen e.V.	MEO	true
Integriertes Simulationsmodell zur Anlageneinsatz- und ausbauplanung mit Regionalisierung	ISAaR	Research Center for Energy Economics (FfE), Munich	FlexMex, MODEX-NET	false
Joint Market Model	JMM	Chair for Management Science and Energy Economics (EWL), University of Duisburg-Essen	FlexMex, MODEX-POLINS, MODEX-EnSAVes	false
Market Simulation	MarS	Institute of Power Systems and Power Economics (IAEW) at RWTH Aachen University	FlexMex, MODEX-NET	false
Model of InternationaL Energy Systems	MILES	TU Dortmund University, Institute of Energy Systems, Energy Efficiency and Energy Economics	MODEX-NET	false
MOSAIK	MOSAIK	Institut für Informatik (OFFIS), Oldenburg	MEO	true
open energy system modelling framework	oemof	Reiner Lemoine Institute (RLI), Berlin	FlexMex, open_MODEX	true
Test- and Simulation Environment	OpSim	Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Kassel	MEO	false
Pandapower Pro	Pandapower Pro	Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Kassel	MEO	false
Program Package for Emission Reduction Strategies in Energy Use and Supply	PERSEUS	Institute for Industrial Production, Chair of Energy Economics, Karlsruhe Institute of Technology (KIT), Karlsruhe	MODEX-NET	false
PowerACE	PowerACE	Karlsruhe Institute of Technology (KIT), Karlsruhe	MODEX-EnSAVes	true
PowerFlex	PowerFlex	Oeko-Institute e.V., Freiburg	MODEX-NET, MODEX-POLINS	false
Renewable Energy Mix	REMix	Institute of Networked Energy Systems, German Aerospace Center (DLR), Stuttgart	FlexMex	false
Renewable Electricity Supply and STORage in Europe	RESTORE	Wuppertal Institut (WI), Wuppertal	FlexMex	false
SCOPE - cross sectoral invest	SCOPE - cross sector	Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Kassel	MODEX-POLINS	false
SCOPE - electricity market	SCOPE - EM	Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Kassel	MODEX-POLINS	false
Transport energy, economics and environment model	TE3 model	Karlsruhe Institute of Technology (KIT), Karlsruhe	MODEX-EnSAVes	true
urbs	urbs	Chair of Renewable and Sustainable Energy Systems (ENS), Technical University Munich (TUM)	open_MODEX	true
ZuBer	ZuBer	University of Wuppertal	MEO	false

Table 4
Evaluation of FAIR principles in MODEX.

	FAIR Principle	Evaluation in MODEX	Example
Findable	F1. (Meta)data are assigned a globally unique and persistent identifier	Applies.	https://openenergy-platform.org/data edit/view/reference/modex_flexmex_techno_economic
	F2. Data are described with rich metadata (defined by R1 below)	Applies with 68 available keys in the metadata string.	https://github.com/OpenEnergyPlatform/oemetadata/blob/develop/metadata/v141/metadata_key_description.md
	F3. Metadata clearly and explicitly include the identifier of the data they describe	Applies with key 11.3 in oemetadata v.1.4.1.	11.3 path - URL to original source
	F4. (Meta)data are registered or indexed in a searchable resource	Applies. They are registered on the OEP. The OEP will be connected to the Databus.	
Accessible	A1. (Meta)data are retrievable by their identifier using a standardised communications protocol	Applies. They can be accessed with the REST-API.	
	A1.1 The protocol is open, free, and universally implementable	Applies. The metadata string is licenced under a CC0-1.0 and uses open standards.	
	A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	Applies. With the given license on metadata and data exact conditions under which the data are accessible are specified.	
	A2. Metadata are accessible, even when the data are no longer available	Applies. The Open Energy Platform is funded for the next ten years.	
Interoperable	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Applies. Oemetadata v1.4.1 are based on JSON format. Oemetadata v1.5 and later versions are based on JSON LD.	
	I2. (Meta)data use vocabularies that follow FAIR principles	Applies to a limited extent. The metadata documentation is easily findable and accessible by anyone who uses the dataset, but the documented, controlled vocabulary used to describe datasets is not resolvable using globally unique and persistent identifiers. Oemetadata v1.5.1 and later versions are using the OpenEnergyOntology to resolves the controlled vocabulary with URIs.	https://github.com/OpenEnergyPlatform/oemetadata/blob/develop/metadata/v141/metadata_key_description.md
	I3. (Meta)data include qualified references to other (meta)data	Applies. Cross references to other datasets given, where available. All datasets are properly cited.	
Reusable	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	Applies. Each source is specified with various attributes, tables are summarized with a scope description and keywords.	
	R1.1. (Meta)data are released with a clear and accessible data usage license	Applies. Where licences are available they are indicated, otherwise authors' full copyright is indicated.	
	R1.2. (Meta)data are associated with detailed provenance	Applies. Where provenance information was given they were documented.	
	R1.3. (Meta)data meet domain-relevant community standards	Applies. By using the oemetadata we apply a domain typical standard.	

References

[1] Bhattacharyya SC, Timilsina GR. A review of energy system models. Int J Energy Sect Manag 2010;4(4).

[2] Doukas H, Nikas A, González-Eguino M, Arto I, Anger-Kraavi A. From integrated to integrative: delivering on the paris agreement. Sustainability 03 July 2018;10(7).

[3] Fuss S, Canadell JG, Peters GP, Tavoni M, Andrew RM, Ciais P, Jackson RB, Jones CD, Kraxner F, Nakicenovic N, Le Quéré C, Raupach MR, Sharifi A, Smith P, Yamagata Y. Betting on negative emissions. Nat Clim Change 21 September 2014: 850–3.

[4] Doukas H, Nikas A. Decision support models in climate policy. Eur J Oper Res 1 January 2020;280(1):1–24.

[5] Pfenninger S, et al. Opening the black box of energy modelling: strategies and lessons learned. Energy Strategy Rev 2018;19:63–71.

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

[7] Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data; 2016.

[8] Huppmann D, et al. pyam: analysis and visualisation of integrated assessment and macro-energy scenarios. Open Res Eur June 2021;1:74.

[9] Jülich Projektträger, Jülich GmbH Forschungszentrum. Modell experimente für die Energiewende [Online]. Available: https://www.energiesystem-forschung.de/foerdern/modex. [Accessed 15 June 2021]. Accessed.

[10] Stanford University. Energy modeling Forum [Online]. Available: https://emf.stanford.edu/. [Accessed 7 June 2021]. Accessed.

[11] Fattahi A, Sijm J, Faaij A. A systemic approach to analyze integrated energy system modeling tools: a review of national models. Renew Sustain Energy Rev 2020;133.

[12] Jebaraj S, Iniyan S. A review of energy models. Renew Sustain Energy Rev 2006; 10:281–311.

[13] Uyterlinde M, et al. Energy trends for Europe in a global perspective: Baseline projections by twelve E3-models in the CASCADE MINTS project. Petten: ECN; 2004.

[14] Capros P, Panos V, Mantzos L, Zeka-Paschou M, Papandreou V, Argiri E, Uyterlinde M, Rösler H, Smekens K, Martinus G, Thuijl Ev, Zwaan Bv d, Coninck Hd, Criqui P, Mima S, Menanteau P, Bellevrat E, Schrattenholzer L. CASCADE mints - CAse Study Comparisons And Development of Energy Models for INtegrated Technology Systems. 2004.

[15] Kriegler E, et al. Assessing Pathways toward Ambitious Climate Targets at the Global and European levels: A Synthesis of Results from the AMPERE Project. IIASA Policy Report; 2014.

[16] Luderer G, Kriegler E, Delsa L, Arvesen A, Edelenbosch O, Emmerling J, Krey V. Deep carbonisation towards 1,5°C-2°C stabilisation: A synthesis of results from the ADVANCE project. 2016.

[17] Luderer G, Pietzcker RC, Carrarab S, Boerd HSd, Fujimori S, Johnson N, Mima S, Arent D. Assessment of wind and solar power in global low-carbon energy scenarios: an introduction. Energy Econ May 2017;64:542–51.

[18] Sugiyama M, Fujimori S, Wada K, Endo S, Fujii Y, Komiyama R, Kato E, Kurosawa A, Matsuo Y, Oshiro K, Sano F, Shiraki H. Japan's long-term climate mitigation policy: multi-model assessment. Energy 9 November 2018;167: 1120–31.

[19] Gils HC, Sterchele P, Kost C, Brucker L, Janßen T, Krüger C, Schüwer D, Luhmann H-J. RegMex - Modellexperimente und -vergleiche zur Simulation von Wegen zu einer vollständig regenerativen Energieversorgung. 2018.

[20] Gils HC, Pregger T, Flachsbarth F, Jentsch M, Dierstein C. Comparison of spatially and temporally resolved energy system models with a focus on Germany's future power supply. Appl Energy 2019;225.

[21] Huntington HG, Bhargava A, Daniels D, Weyant JP, Avraam C, Bistline J, Edmonds JA, Giarola S, Hawkes A, Hansen M, Johnston P, Molar-Cruz A, Nadew M, Siddiqui S, Vaillancourt K, Victor N. Key findings from the core North American scenarios in the EMF34. Energy Pol September 2020;144.

[22] Bistline JE, Brown M, Vaillancourt K, Siddiqui SA. Electric sector impacts of renewable policy coordination: a multi-model study of the North American energy system. Energy Pol October 2020;145.

[23] Siala K, Mier M, Schmidt L, Torralba-Díaz L, Sheykhha S, Savvidis G. Which model features matter? An experimental approach to evaluate power market modeling choices. Energy April 2022;245:15.

[24] O'Neill BC, Carter TR, et al. Achievements and needs for the climate change scenario framework. Nat Clim Change 25 November 2020;10:1074–84.

[25] Kunze R, Schreiber S. Model coupling approach for the analysis of the future European energy system. In: Möst D, Schreiber S, Herbst A, Jakob M, Martino A, Poganietz W, editors. The future European energy system, vol. 1. Cham: Springer; 2021. p. 27–51.

[26] Giarola S, Mittal S, Vielle M, Perdana S, Campagnolo L, Delpiazzo E, Bui H, Annela AK, Kolpakov A, Sognnaes I, Peters G, Hawkes A, Köberle AC, Grant N, Gambhir A, Nikas A, Doukas H, Moreno J, van de Ven D-J. Challenges in the harmonisation of global integrated assessment models: a comprehensive methodology to reduce model response heterogeneity. Sci Total Environ April 2021;783.

[27] Shiraki H, Sugiyama M. Back to the basic: toward improvement of technoeconomic representation in integrated assessment models. Climatic Change 12 May 2020; 162:13–24.

[28] OpenEnergyPlatform. Open Energy Metadata - Open Energy Metadata Description. 2021 [Online]. Available: https://github.com/OpenEnergyPlatform/oemetadata/blob/a37da7c736fb65f3b556608bdc97ad6fd537ac54/metadata/v141/metadata_key_description.md. [Accessed 23 August 2021]. Accessed.

[29] Pritoni M. Metadata schemas and ontologies for building energy applications: a critical review and use case analysis. Energies 2021;7(14).

[30] Barbosa L, et al. Practical guide on Open Science for researchers. 2021 [Online]. Available: 10.5281/ZENODO.3968115.

[31] IEA. World Energy Outlook 2018. 2018. Paris.

[32] 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, TransnetBW GmbH, Netzentwicklungsplan Strom, Version 2019, zweiter Entwurf der Übertragungsnetzbetreiber, vol. zweiter Entwurf der Übertragungsnetzbetreiber, Berlin, Dortmund, Bayreuth, Stuttgart. 2019.

[33] ENTSO-E, ENTSOG, TYNDP 2018, Scenario Report. Brussels: Main Report; 2019.

[34] Boston Consulting Group. Prognos, Klimapfade für Deutschland. Hamburg, Basel, Berlin: München; 2018.

[35] Umweltbundesamt. Berichterstattung unter der Klimarahmenkonvention der Vereinten Nationen und dem Kyoto-Protokoll 2020, Nationaler Inventarbericht zum Deutschen Treibhausgasinventar 1990 - 2018. Dessau-Roßlau; 2020.

[36] ENTSO-E. Entsoe Transparency Platform. 2019 [Online]. Available: https://transparency.entsoe.eu/. [Accessed 15 June 2021]. Accessed.

[37] ESA2. Shaping our energy system – combining European modelling expertise: Case studies of the European energy system in 2050. 2013.

[38] ESA2. Database of the REFLEX project in the ESA2 data warehouse [Online]. Available: https://data.esa2.eu/tree/reflex. [Accessed 19 August 2021]. Accessed.

[39] IIASA and the pyam developer team. pyam: analysis and visualization of integrated-assessment & macro-energy scenarios. IIASA and the pyam developer team; 2021 [Online]. Available: https://pyam-iamc.readthedocs.io/en/stable/. [Accessed 31 August 2021]. Accessed.

[40] VTT Technical Research Centre of Finland Ltd. Spine. 2021 [Online]. Available: http://www.spine-model.org/contacts.htm. [Accessed 31 August 2021]. Accessed.

[41] Otto von Guericke Universität Magdeburg. RLI, Europa-Universität Flensburg. et al.. 2021 [Online]. Available: https://openenergy-platform.org/about/. [Accessed 31 August 2021]. https://openenergy-platform.org. Accessed

[42] RLI. oedatamodel. 2020 [Online]. Available: https://github.com/OpenEnergyPlatform/oedatamodel. [Accessed 31 August 2021]. Accessed.

[43] OpenEnergyPlatform. Overview - Factsheets. 2021 [Online]. Available: https://openenergy-platform.org/factsheets/overview/. [Accessed 23 August 2021]. Accessed.

[44] German Aerospace Center - Institute of Networked Energy Systems. Research Project LOD-GEOSS. 2021 [Online]. Available: https://www.dlr.de/ve/en/desktopdefault.aspx/tabid-14162/24548_read-59660/. [Accessed 23 August 2021]. Accessed.

[45] Reiner Lemoine Institut. Automated comparison of energy scenarios – sirop. 2021 [Online]. Available: https://reiner-lemoine-institut.de/en/automated-comparison-energy-scenarios-sirop/. [Accessed 23 August 2021]. Accessed.

[46] Booshehri M, Emele L, Flügel S, Förster H, Frey J, Frey U, Glauer M, Hastings J, Hofmann C, Hoyer-Klick C, Hülk L, Kleinau A, Knosala K, Kotzur L, Kuckertz P, Mossakowski T, Muschner C, Neuhaus F, Pehl M, Robinius M, Sehn V, Stappel M. Introducing the Open Energy Ontology: enhancing data interpretation and interfacing in energy systems analysis. Energy AI 2021;5. September.

[47] Morrison R. Energy system modeling: public transparency, scientific reproducibility, and open development. Energy Strategy Rev 2018;20:49–63.

[48] Wilkinson MD, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3.

[49] OpenEnergyPlatform. Open Energy Ontology. 2022 [Online]. Available: https://openenergy-platform.org/ontology/. [Accessed 27 February 2022]. Accessed.

[50] DLR, RLI, PIK, FZJ, Leipzig University and University of Stuttgart, "LOD-GEOSS. github.io," [Online]. Available: https://github.com/LOD-GEOSS/LOD-GEOSS.github.io. [Accessed 22 02 2022].

[51] DLR. Research Project SEDOS - the importance of sector integration within the energy transition in Germany - modeling with a national open source reference energy system [Online]. Available: https://www.dlr.de/ve/en/desktopdefault.aspx/tabid-18291/29094_read-76584/. [Accessed 22 February 2022]. Accessed.