neon neue energieökonomik

WHITE PAPER

Open Data for Electricity Modeling

An assessment of input data for modeling the European electricity system regarding legal and technical usability

6 November 2018

Authored by
Lion Hirth, Ingmar Schlecht, Jonathan Mühlenpfordt (Neon Neue Energieökonomik GmbH)

A report for the
Federal Ministry for Economic Affairs and Energy, Germany

Open Data for Electricity Modeling

An assessment of input data for modeling the European electricity system regarding legal and technical usability

6 November 2018

Lion Hirth, Ingmar Schlecht, Jonathan Mühlenpfordt (Neon Neue Energieökonomik GmbH)

This report was written for the German Federal Ministry of Economic Affairs and Energy within the project "Open Source Modellierung und Open-Data für quantitative Analysen des Stromsystems im Rahmen der Energiewende" (short "Open Source Energiewende"), project no. 060/17.

We thank Robbie Morrison, Tom Brown, Juliane Reimann, Ludwig Hülk, Matteo De Felice, Frauke Braun, Ralf Uttich, Stefan Pfenninger, Wladimir Tiderko, Nils Börnsen, Felix Arnold, and the participants of the project workshop for helpful comments and inspiring discussions.

Contact:

Neon Neue Energieökonomik GmbH
Karl-Marx-Platz 12
12043 Berlin

Prof. Dr. Lion Hirth
hirth@neon-energie.de
+49 157-55 199 715

neon neue energieökonomik

Table of contents

1 Context: Open modeling

Open electricity system modeling. This report is the first of a series of reports on open electricity system modeling which are part of the project "Open Source Energiewende"[1]. Modeling is running a model (computer code with equations that represent the electricity system) using input data (observed or estimated values) in order to produce output data (results) that can help in answering specific questions (interpretation). At all these stages, modeling can be open or closed (Figure 1).

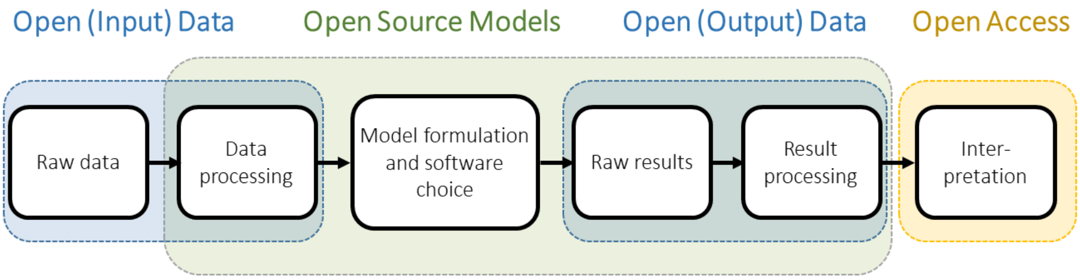

Figure 1. Open data, open source, and open access in relation to the energy modeling process. The focus of this report lies on the first black box: raw data as input to models. Figure updated from Pfenninger et al. (2018) and licensed under CC BY 4.0.

Why open modeling? Open electricity system modeling – using open data in open-source models to produce open results – comes with a range of potential benefits: it can increase the quality of research and policy advice through improved reproducibility and greater scrutiny, increase productivity by allowing reuse and collaborative development, increase credibility and legitimacy in the policy discourse through greater transparency, and make high-quality data and planning tools accessible to researchers and institutions without the funds for commercial alternatives. Finally, ethical considerations suggest that research financed with public money should be public.

This report. This report discusses open input data, i.e. data which is freely accessible and usable. The next report will focus on open source models, i.e. models for which the computer source code is made public for re-use. Open output data and open access are beyond the scope of this project.

Further reading. Recent articles on open energy system modeling include Pfenninger et al. (2017), Pfenninger et al. (2018) and Morrison (2018).

[1] "Open Source Modellierung und Open-Data für quantitative Analysen des Stromsystems im Rahmen der Energiewende" (Open Source Energiewende, project no. 060/17).

2 Data for electricity system modeling

Electricity system modeling requires large amounts of different kinds of input data. Efficient and robust modeling requires data to be of high quality and to be accessible without legal or technical barriers. This introduction defines the scope and the objective of this report.

2.1 Scope of this report

Data for electricity system models. This report is focused on data that are used as inputs for electricity system modeling, such as hour-by-hour load and generation data, information on existing power plants, fuel cost, and electricity price data. Such data are often provided by official or quasi-official institutions such as governments, public authorities, trade associations, system operators, and power exchanges. The data are usually fundamental, well-defined information about the power system, measured or estimated, and concern the present or the past. This report does not cover crowd sourced data or research data. Data are usually published for multiple purposes – not only electricity system modeling – and along with a variety of other data not used for modeling. Modeling an electricity system is an activity that starts and ends with data: data is used as input for the model which, in turn, generates data as output[2].

Open Data. For this report, we refer to "open data" as data that can be easily used both in a legal and technical sense. This report therefore assesses the legal and technical usability of electricity data as inputs for electricity modeling. Legal usability refers to barriers and limitations stemming from intellectual property rights. Technical usability refers to issues such as access options, file formats, documentation and version control.

Geographical scope. This report deals solely with the European electricity system. While the discussion on data quality and usability are more universal, the legal aspects of open data are shaped by European Union law and are therefore specific to the EU.

Cost of data provision. Recording, providing and maintaining data comes with a cost for the data provider who has to maintain technical infrastructure (measuring devices, databases, servers) as well as personnel (data experts, service desk operators, backend developers, web

[2] Model *output* data, in contrast, are usually published by researchers, often along with academic publications – if these are published at all. The focus of this report is on *input* data as characterized above. This is not because publishing results as well-structured and open data is of less importance, but because there is a significant literature on the topic of "open research data", see e.g. Pfenninger et al. (2018) on energy and Ball (2014) on the legal issues in general. The boundary between input and output is to some extent fluid, however, as output data from one model can also serve as input data to another. Examples might include projected fuel costs or CO_2 prices.

designers). At the same time, data creates value by enabling users to conduct analyses contributing to better research, business decisions and policy recommendations. While, when writing this report, we are aware of the costs associated with data provision, we are not in a position to provide an informed cost-benefit assessment. Instead this reports presents the needs and requirements of data users.

Downstream data ecosystem. Electricity system modeling is one of several applications for electricity system data. Other such applications include electricity trading or directly informing the public debate with descriptive statistics and visualizations. For some of these applications, specialized "downstream" data platforms have emerged that collect and republish data from primary sources, catering to the specific needs of different user groups. For electricity system modeling, the Open Power System Data Platform aims to provide a set of ready-to-use input data. For traders and other commercial users, Genscape provides real-time data of power plant generation and transmission needs. For journalists and the interested public, the SMARD Platform provides interactive access to data and visualizations. When considering technical and legal usability of electricity data, one should keep in mind this downstream data ecosystem.

Audience. This report addresses two distinct audiences: first, institutions providing data, such as statistical offices, public agencies, network system operators and electricity market actors. Second, policy makers at the national and European level that determine intellectual property rights and publication obligations.

2.2 Types of data

Modeling is data-intensive. Electricity system models include a range of model types such as power market models, grid expansion planning models, and grid models. The exact data requirements differ by model, but what all these models have in common is that the amount of input data required is substantial.

Basic input data. The following data are needed by practically all types of electricity system models:

- Hourly (or higher temporal resolution) electricity consumption
- Hourly generation potential of variable generation sources (wind power, solar power, run-of-river hydroelectricity)
- Import and export constraints of bidding zones
- Information on existing power plants or generation technologies, including installed capacity and heat rate
- Fuel costs and CO_2 certificate prices

Spatial resolution. Depending on the model type and the specific application, these data need to come in low or high spatial granularity, i.e. specified on a country-level or the level of transmission nodes, major consumers, and municipalities. Electricity spot price modeling, for example, requires data to be specific to each bidding zone while load flow modeling requires detailed geoinformation.

Additional input data. Some models and applications require even more data:

- Uncertainty analyses and Monte Carlo simulations – used for example for generation adequacy assessments – require historical time series data (in particular load, wind, and solar) going back years or decades.
- Unit commitment models require detailed information on individual power stations, such as start-up times, ramping constraints and associated costs.
- Load flow modeling requires engineering data on the electricity grid including lines, substations, transformers and other network elements.
- Investment modeling, e.g. for generation expansion models, requires information about investment costs and resource constraints that can limit the potential for certain technologies.
- Models that cover regions with substantial hydroelectricity, such as Scandinavia or the Alps, require additional information on hydro-power dams including the capacity of their storage, the dam height and details about inflows.
- Models with storage require data on storage efficiencies, capacities and expansion potentials (e.g. underground caverns for hydrogen or compressed air energy storage).
- Models with coupling to other energy sectors required large amounts of additional data: for the heating sector (heat demand time series, heat distribution, data on combined heat and power plants, coefficients of performance for heat pumps, time series for solar thermal collectors, ground heat pump land availability, etc.); for the transport sector (transport demand time series, electric vehicle charging strategies, etc.)

Classes of data. Sometime it makes sense to group the above data into three classes of data:

- Time series data, e.g. load, spot price, generation or outage time series
- Tabular data, e.g. asset inventories or technology cost assumptions
- Geospatial data, e.g. geographic patterns of networks

Electricity data. For brevity, we refer to the above data collectively as "electricity data" in the following.

2.3 Data sources

A multitude of sources. Electricity system modeling regularly requires obtaining input data from multiple sources. Input data for energy models is spread widely across many sources, including governmental agencies, grid operators, market actors, power exchanges, and trade associations. Table 1 lists a selection of important sources[3]. It is meant to give a flavor of the data items and sources that are relevant for electricity system modeling, not a comprehensive list of source or data they provide. Moreover, we restrict the list to examples, e.g. discuss only

[3] The Open Power system Data project provides a more extensive list of data sources at https://open-power-system-data.org/data-sources.

Germany's Statistisches Bundesamt where of course all national statistical offices publish electricity-related data. We will discuss these sources further in the following sections.

Publication mandates. As evidenced by Table 1, it is not only governmental agencies that make energy data available for modeling, but also a number of commercial entities. So why do commercial firms provide data free of charge? The reason is that publication obligations from different kinds of transparency regulation require them to disclose (see 3.2.4). The last column of Table 1 indicates whether the source is a public authority or whether the information is released voluntarily or under mandate. Public sector information and information released under such statutory provisions comprise the bulk of input data required for energy system modeling.

Table 1. European data sources for electricity system modeling (examples).

Source	Type of data provided (selection)	Free of cost	Open license	Why?
Eurostat	• Monthly/yearly electricity consumption, trade and generation per fuel type • Yearly installed capacity per fuel type • Energy balances	Yes	Yes	Public
National statistical offices (e.g. destatis)	• Monthly/Yearly net/gross electricity and heat generation and fuel consumption by fuel type	Yes	Yes	Public
ENTSO-E Transparency Platform	• (Quarter-) hourly load, cross-border trade, dispatch, spot prices and generation per fuel power plant/type • Yearly installed capacity per power plant/fuel type. • Unavailability of generation, consumption and transmission units • Prices and volumes of activated balancing reserves • Weekly filling rate of water reservoirs	Yes	No	Mandated
ENTSO-E Power Statistics	• Hourly load • Monthly electricity consumption, trade and generation per fuel type. • Yearly physical energy flows between countries • Yearly installed capacity per power plant/fuel type. • Yearly installed capacity per fuel type.	Yes	No	Voluntary
REMIT Inside Information Platforms (e.g. EEX Transparency)	• Unavailability of generation, consumption and transmission units • Hourly generation per power plant • Yearly installed capacity per power plant/fuel type. • Filling rate of water reservoirs	View: Yes Download: No	No	Mandatory, some voluntary
Power Exchanges (e.g. EPEX Spot)	• Day-ahead and intraday spot prices	No	No	Voluntary
National regulators (e.g. Bundesnetzagentur)	• List of power plants > 10 MW	Yes	Yes	Public
Bundesnetzagentur SMARD platform	• Quarter-hourly consumption, trade, prices and generation per fuel type • Yearly installed capacity per power plant/fuel type. • Power plant availabilities • Prices and volumes of activated balancing reserves	Yes	Yes	Public
Environmental agencies (e.g. Umweltbundesamt)	• Location and installed electrical-, CHP capacity and turbine type of power plants > 100 MW (>10 MW on request)	Yes	No	Public

Source	Type of data provided (selection)	Free of cost	Open license	Why?
Commercial data providers (e.g. Platts World electric power plants database) or Genscape Power RT	• Detailed data on location, economic and technical characteristics of power plants • Real time generation and transmission flows	No	No	Voluntary
TSOs	• (Quarter-) hourly electricity generation for wind and solar energy, load • Installed power plant capacity • Network data	Yes	No	Voluntary
Common TSO data platforms (e.g., regelleistung.net and netztransparenz.de)	• Quarter-hourly activated balancing reserves • Quarter-hourly electricity generation for wind and solar energy • Location and installed electrical capacity of renewable power plants of all sizes	Yes	No	Mandated
Power exchanges (e.g. EPEX Spot)	• Hourly day-ahead/quarter-hourly intraday prices	No	No	Voluntary
Trade associations (e.g. BDEW)	• List of planned power plant installations	Partly	No	Voluntary
AG Energiebilanzen	• Yearly electricity consumption by sector, gross generation by fuel type	Yes	No	Public
BMWi Energiedaten	• Yearly electricity consumption by sector, gross generation by fuel type	Yes	No	Public
European Environmental Agency	• Protected areas	Yes	Yes	Public
Reanalysis models (e.g. NASA MERRA-2, ECMWF's ERA5)	• Historical wind speeds, solar radiation, temperature at high temporal and spatial resolution	Yes	Yes	Voluntary

2.4 Open data

Defining "open data". One of the most cited definitions of open data comes from Open Knowledge International (OKI)'s Open Definition (OKI 2015b): "Open data and content can be freely used, modified, and shared by anyone for any purpose", or alternatively "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)". In sharp contrast, OECD (2015) defines open data as being "machine-readable". This definition is purely technical without any references to legal aspects.

Our definition. We define open data loosely as "data that is both legally and technically readily usable." This definition has two aspects, a legal side and a technical side. On the legal side we follow the Open Definition, requiring that data can be "freely used, modified, and shared by anyone for any purpose" to classify as open. The technical side is sometimes described as "frictionless data". The purpose of this section is not to provide a clear-cut definition of open data, but rather to delimit the scope of this report. The legal and the technical side of open data are discussed further in the following.

Further reading. The Open Data Handbook (OKI, 2015a) provides a good overview of open data, covering many practical aspects.

2.4.1 Legal aspects and barriers

Under what circumstances is one able to use the data legally? Which restrictions apply? These are aspects of "legal openness".

"Public" is distinct from "open". Unless often believed, "public" in the sense of "available free of charge" is not identical to open. Publication is a precondition for open data, but openness also includes the right to copy, use, modify and share.

Legal barriers. In Europe, structured data can be protected as intellectual property. For electricity data, the so-called "*sui generis* database right" is the most relevant intellectual property right.[4] As other property rights, it gives the data owner – within certain limitations – an exclusive right to use it. Just publishing data online without restricting access does *not* mean these rights have been waived.[5] Therefore, one can only use data legally if the rightsholder grants permission to do so. Such permission is called a "license". Open data means that the rightsholder has granted a license that allows the data to be freely used, modified and shared by anyone. Both the right to modify and the right to publish are essential to open data. Many crucial energy data sources do not provide such an open license.

Modeling practice. It is common practice for energy modelers to use freely-available data anyway without checking the license. In fact, both data users and providers are often unaware of the fact that data use is legally restricted by intellectual property rights. While we are not aware of any court case or other litigation, it remains an unsettling fact that so much electricity system modeling is likely to constitute a violation of intellectual property rights, in particular because the consequences can be severe: in Germany, infringing intellectual property rights can be a criminal offense (§ 108 (1) Nr. 8 UrhG).

What "using data" includes. For effective and high-quality modeling, it is not sufficient to have the right to simply inspect data. Researchers need to be legally able to do the following:

- Retrieve data, i.e. make a local copy. This is obviously necessary for any kind of quantitative analysis.
- Machine-process that data for the purposes listed next as well as for the actual numerical modeling.
- Clean, repair, modify, combine or otherwise amend data. This is important as many datasets are, despite mandatory publication, of poor quality or inconsistent with

[4] A rigorous discussion of legal barriers requires the introduction of terminology and concepts. This is done in section 0.

[5] This has parallels to tangible property: just parking your car on a public street, even if unlocked, does not mean anyone can use it legally – It still remains your property, and you have the exclusive right to use it.

other sources (see Hirth et al. 2018). Often, it is combining databases into larger collections where the real value lies.

- Pass on such amended data to other users.
- Publish original and amended data online for scientific reproducibility. This is particularly important as original sources are sometimes taken offline or are updated.

Use and "reuse". In general, modelers require the right to copy, use, modify, distribute, or publish the data. Such activities are often called "reuse" in legal documents.

Legal aspects of using electricity data for energy system modeling are discussed in detail in section 0.

2.4.2 Technical aspects and barriers

How easy is it to use the data? Which technical barriers exist and how can they be overcome? These are aspects of "technical openness".

Technical barriers. Technical usability of data and publication platforms refers to issues such as download options (including application programming interfaces), machine readability of data (file formats, structure of data), metadata (content, structure, encodings), documentation (data definitions, descriptions of measurement and estimation techniques, technical documentation covering download options and interfaces). Data are "open" if the specifications used enable seamless and frictionless access, processing and analysis.

Technical aspects of using electricity data for energy system modeling are discussed in detail in section 4.

2.4.3 Related aspects beyond "openness"

Data quality. This report focusses on technical and legal usability. A rigorous discussion of data quality in terms of completeness, accuracy, consistency and timeliness of data is beyond the scope of this report.[6]

Data semantics. Similarly, domain-specific glossaries and ontologies (systematized knowledge) covering the energy sector remain outside the scope of this report.

FAIR principles and findability. Apart from legal and technical openness of data, their findability is also important in determining effective use of data. This is reflected in the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, designed and endorsed by a group of stakeholders in research data (Wilkinson et al. 2016). This report discusses these aspects without explicit further reference to the FAIR principles but the suggestions stated here are in line with the FAIR principles.

[6] For an assessment of the data quality of the ENTSO-E Transparency Platform see Hirth et al. (2018) and https://neon-energie.de/transparency-platform/.

3 Legal usability of electricity data

This section assesses the legal status quo of open data for electricity system modeling in Germany and Europe. It identifies legal barriers that prevent the use of data for modeling, and suggests how they can be overcome.

Legal barriers for use. The legal regulations of data ownership and use rights matter, because much of what can and cannot be done with electricity data is governed by intellectual property law. In particular, the use of publicly available data might be restricted by copyright and related rights, which grant the rightsholder exclusive rights and thereby limit the use by others ("reuse").

Non-personal data. Note that all data within the scope of this study is non-personal data, i.e. it does not relate to individuals (recall section 2.2). This is different in other parts of energy, for instance when household smart meters produce information which can relate to individuals. Within the scope of this study, privacy is therefore not an issue and will not be touched upon. Similarly, we will not discuss data as a trade secret that can be used for inference of business decisions.

Further reading. Ball (2014) provides an accessible introduction to open data. Helpful discussion of legal aspects of energy data can be found in a submission to an EU Commission public consultation process by Morrison et al. (2018) and the legal opinions by Jaeger (2017, 2018). Comprehensive legal background is provided by Davison (2008).

3.1 The legal framework

Intellectual property rights. Intellectual property refers to creations of the mind, such as inventions, artistic works or designs. Intellectual property rights (IPRs) grant the creator a time-limited exclusive right over the use of their intellectual output. Specific IPRs apply to different types of creations, such as patents (for inventions), trademarks (for brands) and copyright[7] (for creative works). Structured data can be a creative work and thereby attract copyright. In addition, in Europe a specific intellectual property right for certain structured data exists, the so called *sui generis* database right. Copyright and the *sui generis* database right will be discussed in the following, as well as specific regulation governing data published by public authorities. Readers familiar with intellectual property rights might continue with section 3.2.

[7] Matters are complicated by the fact that the U.S. legal doctrine has a quite different approach to intellectual property than the German doctrine (e.g., the term "copyright" does not exist in German law and does not directly correspond to "Urheberrecht"). For the protection of data, however, European law, which uses the term "copyright", has major relevance, as discussed below.

3.1.1 Intellectual property rights on databases

Statutory framework. The EU Database Directive (Directive 96/9/EC) defines the statutory framework for the legal protection of structured data in Europe.[8]

Data. Data can be thought of as various individual field values (a single observation or fact, a bit of information, an individual data point, a datum). An individual data point cannot be protected as intellectual property. (However, the individual entry might itself be a works that attracts copyright – think of a database of photographs.)

Structured data: database. Databases are defined by the Directive as data that is "arranged in a systematic or methodical way and individually accessible".[9] Note that "database" refers to *the structured collection of bits of information*, rather than the software or hardware used to store that information.[10] In other words, in legal terms a "database" can best be translated as "structured data". It is the *structure* that attracts intellectual property rights, not the raw data.

Protection of databases. The legal protection of databases in Europe is twofold: "classical" copyright and the *sui generis* database right. Both regimes grant the rightsholder certain exclusive rights. The former, which predates the Database Directive, is ratified in Article 3 of the Directive. The latter, which was newly introduced with the Directive, is defined in Article 7. They both emerge automatically when making a database – unlike patents, which have to be applied for explicitly. They have different requirements (creativity versus investment) and different implications (e.g., the duration of protection and the strength of protection), see Table 2. Any given database can be protected under both regimes, either one, or none. Both regimes are discussed in turn.

Table 2. Copyright vs. *sui generis* database right.

	"Classical" copyright	*Sui generis* database right
Articles in Database Directive	Chapter II (Articles 3-6)	Chapter III (Articles 7-11)
Requirement	Creativity • If selection or arrangement of contents are an intellectual creation • = "original" databases (as opposed to "non-original" ones that do not met this criteria)	Investment • If a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents has been made
Exclusive rights ("restricted acts")	• Reproduction, incl. parts • Modification / alteration	Whole or substantial part: • Extraction (copying)

[8] Implemented in Germany as § 4 (2) and § 87a - § 87e Urheberrechtsgesetz (UrhG) (German Act on Copyright and Related Rights). There is no equivalent to the database directive in the United States for instance, despite several such bills being introduced in Congress during the 1990s.

[9] The content of a database does not need to be data. The content can also be "works" "or other materials", which might attract copyright themselves, e.g. a database of books. This case is not relevant for electricity data and hence not further discussed.

[10] Legal and technical terminology diverge at this most. Non-lawyers would probably call a collection of data points a "dataset". However, that terms is neither used nor defined in the relevant law.

	"Classical" copyright	*Sui generis* database right
	• Making available to the public	• Re-utilization (making available to the public)
Duration	70 years after death of author	15 years

Rationale behind the Directive. The legal historical context helps understanding the rationale behind the Database Directive. Prior to 1996, databases were protected in Member States according to their respective legal doctrine. Common law countries that followed the "sweat of the brow" doctrine placed a lower requirement (investment) than those with *droit d'auteur* tradition (individual creativity). The Directive harmonized the requirement by applying the high standard (creativity) for copyright protection while creating a new right with somewhat reduced protection with lower requirements (investment). European Commission (2005), Davidson (2008) and Fisher et al. (2018) recall the rationale for introducing the Database Directive.

3.1.2 Copyright on databases

Requirements. A database attracts copyright if its selection and arrangement is sufficiently creative; it is then called an "original" database. Such "classical" copyright on databases stems from the long-established protection of collections such as encyclopedias. It predates the Database Directive, but is clarified and ratified therein: "Databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright" (Article 3). In other words, if the structuring of the data is sufficiently creative to classify as an intellectual creation, copyright applies. In turn, a database that does not reach a threshold of originality (a so-called non-original database) does not attract copyright. As most electricity data lacks originality (see 3.2.1 below), copyright will not be discussed further.

3.1.3 *Sui generis* database right

Requirements. Any compiler or "maker" of a database (the term used by the Database Directive), who made a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents is granted a so-called "*sui generis* right".[11] This right was newly introduced in the Database Directive. The *sui generis* database

[11] The Database Directive clearly separates "copyright" (Chapter II) from "*sui generis* right" (Chapter III). Parts of the literature are less rigorous and use the term copyright to encompass the *sui generis* right (and call a breach of the *sui generis* database right a "copyright infringement"). The Directive speaks of the "*sui generis* right", some authors prefer "database right"; for clarity we call it more specifically the "*sui generis* database right".

right is similar to copyright but is not granted for creativity but for the financial and professional investment made in obtaining, verifying, and presenting the contents. Hence a database that is structured without creativity but was expensive to make can be protected under the *sui generis* database right, while not being copyrightable.

Scope of the right. The *sui generis* database right grants the database maker the exclusive right of "extraction and/or re-utilization of the whole or of a substantial part". Extraction is defined as "the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form". Re-utilization is defined as "making available to the public" (all quotes from Directive 96/9/EC, Article 7), which refers primarily to online publication. In other words, re-utilization here means "redistribution and online transmission".[12]

Statutory limitations. The Database Directive allows, but does not oblige, member states to establish statutory exemptions[13] from the exclusive rights of the rightsholder. Germany's copyright law (§§ 60a ff UrhG), for example, permits creating a copy of substantial parts (but not entire) protected databases for

- non-commercial personal scientific research,
- non-commercial teaching, and
- personal use, but only for databases that cannot be accessed electronically.

There is no statutory exception that permits making a database available to the public. As all electricity databases can be accessed by electronic means, the last exception is of no relevance for electricity modelers. We will discuss "non-commercial" use below in section 3.1.6 and conclude that in energy system modeling in most cases the use is likely to be commercial, such that the statutory exemptions do not apply.

Case law. A European Court of Justice ruling in 2004 considerably restricted the scope of the *sui generis* database right.[14] The court held that only the investment in obtaining and verifying the data, not the investment in creating the data should be considered when evaluating the investment requirement. As the European Commission (2005, p. 13) acknowledges, the ruling shows "the serious difficulties raised by attempting to harmonize national laws by recourse to untested and ambiguous legal concepts ('qualitatively or quantitatively substantial investments in either the obtaining, verification or presentation of contents')".

---

[12] This is implemented in Germany as §87b Abs. 1 UrhG, which is somewhat clearer in its language: "Der Datenbankhersteller hat das ausschließliche Recht, die Datenbank insgesamt oder einen nach Art oder Umfang wesentlichen Teil der Datenbank zu vervielfältigen, zu verbreiten und öffentlich wiederzugeben.", i.e. „The database owner has the exclusive right to reproduce, distribute and make publicly available the full database or a substantial part (by type or scope) of the database".

[13] "Limitations", "exceptions" and "exemptions" are used interchangeably in the literature.

[14] C-203/02, The British Horseracing Board Ltd and Others v William Hill Organization Ltd.

3.1.4 Specific regulations for the public sector

A significant share of electricity data is published by public sector bodies, such as federal or state-level departments, authorities and agencies (recall section 2.3). On a political level, open data strategies have been adopted such as the G8 Open Data Charta. As a consequence, some public bodies can be obliged to publish certain data as open government data. Im some cases this commitment has resulted in legal obligations.

Public sector information. At the European level, the Public Sector Information (PSI) Directive (Directive 2003/98/EC, amended by Directive 2013/37/EU) regulates the re-use of data published by public sector bodies with the goal to enhance the quantity of open government data. The directive is currently being updated. The European Commission proposes that public sector bodies shall not exercise *sui generis* database rights in order to prevent or restrict the re-use of data, a stipulation not in place in the current version of the PSI directive. In Germany, the PSI Directive has been transposed by the Informationsweiterverwendungsgesetz (IWG). Separately from the PSI directive, the EU Commission Guidelines on recommended standard licenses (2014/C 240/01) recommends a Creative Commons Zero license for public sector information.

TSOs are not public bodies. Article 2 of the PSI directive provides a definition of "public bodies" as the State, regional or local authorities, bodies governed by public law and associations formed by one or several such authorities or one or several such bodies governed by public law. System operators, despite being regulated entities and subject to publication obligations, do not qualify as public sector bodies. The same is true for the European Network of Transmission System Operators for Electricity (ENTSO-E). An extension to public enterprises is currently under discussion in the context of the update of the PSI Directive.

Germany's open data law. For federal public authorities, the German e-government law (§12 EGovG) requires that certain data have to be published ("open-by-default" rule). It also specifies that such data must be published in machine-readable form, amended with metadata which have to be published through the data portal GovData, and be re-usable by anyone without restrictions. Personal data or data protected by copyright are excluded, however.

Official works in Germany. According to the German copyright law (§5 UrhG), official works do not enjoy copyright protection. This is likely to also apply to databases: the would-be landmark case "*Sächsischer Ausschreibungsdienst*" at the Federal Court of Justice (BGH) was cut short by a negotiated settlement before being adjudicated, but the court indicated that the copyright exception for official works would also apply to *sui generis* database rights on databases that are of official character. That is the case where a database has been published for an official purpose and for the general public to take note of. Under this doctrine, no licenses are required for the use of many databases that are made by or for authorities and other public institutions.

3.1.5 Licensing data

The need for guidance. With all the complexity and ambiguity surrounding the legal protection of databases, users need clear guidance from compilers on what they are allowed to do with the data. The usual way to communicate permissions to potential users of data are licenses.

Licenses. Licenses are legal instruments for a rightsholder to permit someone to do things that would otherwise infringe on the rights held. Under German doctrine, a license is considered a contract by which the licensor (rightsholder) allows the licensee (data user) to use otherwise protected material. Licenses for intellectual property are usually written specifically for certain types of property, such as software code, creative works (literature, music or photography) or databases. The license typically includes specifications as to what kind of use is allowed (e.g., making it available to the public) and which obligations apply (e.g., acknowledging the author). This can include the right to sublicense, i.e. the right for the licensee to grant a license themselves. License agreements do not need to be labeled as such. In practice, they are often found as "Terms of Use" in the "Disclaimer" section of an electricity data website.

The default: all rights reserved. Making a database available online by itself does *not* imply that it can be used freely. Without a license agreement, the default is that intellectual property rights apply, even when no claim is expressly made. (Stating "all rights reserved" merely emphasizes that no license has been granted, but is otherwise redundant.) If a database is published without a license, the rightsholder reserves, or holds for her own use, all the rights provided by copyright and the *sui generis* database right. Should those rights apply, using such data will constitute an infringement.

3.1.6 Commercial versus non-commercial use

Relevance. As mentioned above (section 3.1.3), only non-commercial research and teaching are granted relevant statutory exemptions to use otherwise protected databases. As we will discuss below (section 3.2.3), most electricity data providers provide licenses that permit non-commercial usage only. This makes a precise definition of "non-commercial" imperative for energy system modeling.

Defining "non-commercial". "Commercial use" aims at generating income or other economic benefits, which can be, but is not restricted to, monetary compensation. It is the concrete use case rather than the identity of the user that matters.[15] Nevertheless, it is obvious that most uses of publicly available electricity data by firms such as network operators, utilities or consulting firms will be considered commercial. Other uses are clearly non-commercial, such as private personal use.

[15]While most legal scholars agree on this point, a diverging court ruling exists, see LG Köln, Urt. v. 5.3.2014 – 28 O 232/13.

Shades of grey. Defining "non-commercial" precisely has turned out to be notoriously difficult. In Germany, a court assessed the use of a picture licensed for non-commercial use by Deutschlandradio, a not-for-profit public radio station (LG Köln, Urt. v. 5.3.2014 – 28 O 232/13). The court ruled the use to be a license breach and defined "non-commercial" as "purely private use" (see Jaeger (2014) for further discussion of the case). As a consequence, it seems likely that contract research – even if conducted by universities – would classify as "commercial use". See Klimpel (2012) for an extensive discussion of the matter.

Most energy data are used commercially. It is important to note that much, probably most, use of electricity data is commercial in nature. As a consequence, the statutory exemptions have little relevance in this area.

3.2 An application to electricity data

In the following, we discuss the lawfulness of using electricity data for modeling. In other words, we discuss (a) to what extent IPR restricts the possibility to use electricity data for modeling and (b) to what extent statutory exemptions or contractual permissions would allow a modeler to do so.

First, we discuss which type of IPR applies (only the *sui generis* database right does). Second, we discuss if using data for modeling violates the exclusive right of the database maker (it probably does, at least if the use is commercial in nature). We then review licenses granted by data providers (they often disallow commercial use). Finally, we discuss the consequences of publication obligations and the surprisingly difficult task of establishing ownership, i.e. determining who holds the rights.

3.2.1 Which IPR applies to electricity data (if any)?

IPR on electricity data. What kind of intellectual property right applies to electricity data? In other words, do electricity databases fulfill the requirements for protection? Answering this requires the clarification of the following preliminary questions:

- Does the data concerned constitute a database?
- Does (classical) copyright apply?
- Does the *sui generis* database right apply?

Database. It seems evident that much of the relevant data available for energy modeling qualifies as a database in the legal sense: it is arranged systematically and can be accessed individually (a detailed discussion can be found in Jaeger 2017).

Classical copyright. For most energy data, "classical" copyright protection seems unlikely, as the selection and arrangement of data requires no intellectual effort. Most electricity data is offered complete and arranged alphabetically or chronologically, thereby lacking intellectual effort and unlikely to attract classical copyright is unlikely to apply.[16] For the ENTSO-E Transparency Platform, an additional argument applies: as the transparency regulations specify which data has to be selected for publication and how to arrange it, there is little scope left for ENTSO-E's to exercise creativity.

Sui generis database right. As most energy databases appear to require substantial investment to obtain, verify or present, they are likely to be protected by the *sui generis* database right (Jaeger 2018). Moreover, based on public information it is often virtually impossible to assess if such a substantial investment has been made; hence a cautious user may necessarily presume that this is the case. In the following section we assume that only *sui generis* database rights pertain.

3.2.2 Lawful use for modeling

Assessing legality of use. Can electricity data be legally used if protected under the *sui generis* database right? The answer depends on the type of use. Hereafter, we discuss using electricity data as input for modeling. We consider two use cases as relevant: (i) downloading data and using it as an input to a model; (ii) processing, reformatting, amending or repairing data and sharing it with someone else, e.g. through a public website. Any assessment of the legality of a use case has to take three steps:

- Is this act of using the data covered by the exclusive rights of the database maker? The answer to this question depends on the type of use.
- Are any statutory limitations to the *sui generis* right applicable, i.e. a limitation derived from law? The answer to this question depends on the purpose of the use.
- Is this act of using the data allowed under the license provided by the rightsholder, i.e. a permission derived from a contract? The answer to this question depends on the license granted; it cannot be answered by studying the law alone.

Exclusive right? The maker of the database has the exclusive right to reproduce and distribute the database and to make it available to the public. Downloading data implies creating a copy. This act of reproduction affects the rights of the holder of the *sui generis* right if the entire or a substantial part of the database is copied. Also making it available to the public is an exclusive right of the maker of the database. In contrast, what is not protected is any use that (a) either involves less than a substantial part of the database or (b) does not involve creating a copy, e.g. just inspecting data in a web browser. For electricity system modeling, it is regularly required to download substantial parts of or entire databases, which is an exclusive right.

Statutory exceptions? In Germany, relevant statutory limitations and exceptions exist for non-commercial personal scientific research and teaching. As electricity modeling is often done for commercial purposes (e.g., trading, investment decision, consulting, contract research), those limitations are not very relevant.

---

[16] In a different context, the U.S. supreme court has called an alphabetical ordering "devoid of even the slightest trace of creativitiy." (Feist Publications Inc. v. Rural Telephone Service, Co., 499 U.S. 340 (1991))

Allowed under license granted? As a consequence, for commercial energy data users, the only way to legally download data and making them available to the public is to be granted a license by the rightsholder. In this context, "license" means the grant of a right to use the database with the scope of the right specified in the license. Most data sources grant quite restrictive licenses at best, as we discuss in the following.

3.2.3 Licenses used today

This section reviews the licenses offered (if any) by the data providers listed in section 2.3. The license texts themselves are provided in the Appendix. The order of sources resembles Table 1.

Eurostat. Eurostat provides a liberal license, which explicitly allows commercial use. Using Eurostat data for modeling is unproblematic.

Statistisches Bundesamt. Germany's statistical office provides data under the "Data license Germany – attribution – version 2.0", which explicitly allows commercial use. Its use is not problematic for modelers. This is an open data license, see the overview of open standard licenses 3.3.

ENTSO-E Transparency Platform. The "General Terms and Conditions for the Use of the ENTSO-E Transparency Platform" governs the use of data published at ENTSO-E's website. This is not a classical license agreement but it stipulates which uses shall be allowed. There is no explicit clause for a grant of rights, but one could argue that certain use is implicitly assumed ("*when using of the Transparency Platform Data for any purpose whatsoever*"). However, it remains unclear to which extent a reuse is permitted. As a consequence, most users cannot be sure that they are currently using the TP database legally. ENTSO-E does not provide users with the right of making download data available to the public, so this must be considered an exclusive right of the database maker (for a more detailed discussion of the Transparency Platform see Jaeger 2018).

ENTSO-E Power Statistics. ENTSO-E permits non-commercial, personal use of its "Power Statistics". Downloading data for commercial use is not allowed. Making downloaded data available to the public for either commercial or non-commercial use is explicitly excluded.

EEX Transparency. EEX does not provide any data license for non-paying users. A commercial license is available for paying customers.

Bundesnetzagentur. In general, Germany's regulator provides data under the "Data license Germany – attribution – version 2.0", just as Statistisches Bundesamt. Individual publications and databases might be published under different terms.

SMARD. The data published on SMARD, a platform operated by Bundesnetzagentur, is licensed under CC-BY-4.0 International, a widely used open data license (see 3.3). Commercial use and making data available to the public is permitted.

Umweltbundesamt. Germany's environmental protection agency Umweltbundesamt authorizes only non-commercial, personal use.

Platts. The company Platts permits only non-commercial, personal use of its data. Presumably commercial licenses are available for paying customers.

TSOs. TSO 50Hertz disallows "automated download" of data, but is silent on other forms of download. Besides this, the four German TSOs do not provide any data license. Any use beyond statutory exemptions therefore has to be considered a breach of the *sui generis* database right. This is in stark contrast to Finland's Fingrid, France's RTE or Denmark's Energinet.dk, who grant permission for commercial use and to make data available to the public.

Regelleistung.net and Netztransparenz.de. Germany's TSO operate two data publication platforms, www.regelleistung.net (for balancing energy) and www.netztransparenz.de (for various kinds of data they are obliged to publish under German law). Neither of the two websites provides a data license, hence statutory norms apply. Any use beyond statutory exemptions has to be considered a breach.

EPEX SPOT. The largest European spot market for electricity does not permit commercial use of the data provided. Presumably commercial licenses are available for paying customers.

BDEW. The German Association of Energy and Water Industries does not provide any data license. If databases are protected, any use beyond statutory exemptions therefore has to be considered a breach of the *sui generis* database right.

AG Energiebilanzen. AG Energiebilanzen, a working group for German energy statistics, provides an interesting license. It permits non-commercial use including making data available to the public, but prohibits modification of the database. It also prohibits any commercial use beyond what is granted by law.

BMWi. The German Federal Ministry for Economic Affairs and energy does not provide a data license, hence statutory norms apply.

EEA. The European Environment Agency provides a custom license that authorizes commercial and non-commercial use, including making data available to the public.

NASA. For the MERRA2-based weather data, NASA does provide a license, according to our reading. One might deduce implicit permission from the general principles stated on the website.

ECMWF. Climate reanalysis data from the ERA5 model operated by the European Center for Mid-Range Weather Forecast is available under an open license, despite the fact that commercial use is not explicitly mentioned.

Summary. Table 3 provides an overview of the license agreements discussed above. For each source, it is indicated if the license permits (i) creating a copy for commercial use and (ii) modify the database and making it available to the public. Non-commercial personal use is allowed under all above licenses (if any are granted). For both (i) and (ii) it was assumed that at least a substantial part of the database is concerned. The table provides also the type of license; standard licenses are given by name.

Table 3. License conditions of selected electricity data sources

Source	Data license	Allowed to create a copy? (commercial use, substantial part of database)	Allowed to modify and make available to public? (commercial or not, substantial part)
Eurostat	Custom	Yes	Yes
Statistisches Bundesamt	dl-de/by-2-0	Yes	Yes
ENTSO-E Transparency Platform	Custom	Possibly (implicit)	No
ENTSO-E Power Statistics	Custom	No	No
Bundesnetzagentur	dl-de/by-2-0	Yes	Yes
SMARD platform (BNetzA)	CC-BY-4.0	Yes	Yes
Umweltbundesamt	Custom	No	No
Platts world electric power plants database	Custom	No	No
German TSOs	-	No	No
RTE	Custom	Yes	Yes
Energinet.dk	Custom	Yes	Yes
Fingrid	CC-BY-4.0	Yes	Yes
Regelleistung.net and Netztransparenz.de	-	No	No
EPEX SPOT	Custom	No	No
BDEW	-	No	No
AG Energiebilanzen	Custom	No	No (only if non-commercial and unmodified)
BMWi Energiedaten	-	No	No
European Environmental Agency	Custom	Yes	Yes
NASA's MERRA-2	Custom	Possibly (implicit)	Possibly (implicit)
ECMWF's ERA5	Custom	Yes	yes

Why do not more institutions provide an open license? Some commercial entities refrain from open licenses in order not to compromise the ability to sell data. According to our experience however, for most data providers this is not an important motivation. The main reasons include lack of clarity as to who holds the rights, lack of awareness that simply putting data online does not constitute a release into the public domain, and unwillingness to skate on thin legal ice where rightsholders had themselves originally assembled their information from multiple sources.

3.2.4 Publication obligations

A number of European laws exist that mandate system operators and market participants to disclose a wide range of data, in particular the Transparency Regulation and REMIT.[17] This section first introduces these obligations and then discusses their relationship with the *sui generis* database right.

Transparency Regulation. The Transparency Regulation (Commission Regulation (EU) No 543/2013), amending Regulation (EC) No 714/2009, requires TSOs and market actors to publish a wide range of detailed specified market data through a common platform, the ENTSO-E Transparency Platform (see Hirth et al. (2018) for a review of the platform). Its purpose is to serve market participants, such as generators, retailers and traders, in particular new entrants and smaller providers. The Transparency Regulation stipulates that the data has to "be made available to market participants" (Article 1) and "to the public" (Article 3). It shall be "easily accessible, downloadable", and "free of charge" (Article 3).

REMIT. The purpose of the Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) (Regulation (EU) No. 1227/2011) is to increase the transparency and stability of European energy markets while combating insider trading and market manipulation. Article 4 stipulates that "market participants shall *publicly disclose* in an effective and timely manner inside information which they possess" (emphasis added).

IPRs are not mentioned. These disclosure obligations seem to conflict with the exclusive right to make the database available to the public. None of the three regulations mentions the *sui generis* database right or other IPRs. They regulate *what* information is to be published, *how* it ought to be published, *when* it needs to be published and *who* is responsible, but do not touch upon the relationship with IPRs.

IPR and publication mandates. How are the publication obligations then related to intellectual property rights? Two possible interpretations are that they form a statutory exemption to IPR or that they imply an obligation to license.

- Statutory limitation? Do the publication obligations constitute a statutory exemption to exclusive rights granted by the *sui generis* database right? At least in Germany, this is not the case. Under German legal doctrine, the limitations in German Copyright Act constitute a definitive list, such that the publication obligations cannot be understood as statutory limitations to the exclusive rights granted by the *sui generis* database right.
- Obligation to license? Jaeger (2018) argues that one might understand the publication obligations to be a requirement to provide a license that allows the intended use of the data. The Transparency regulation mentions two types of reuse: access and download. It also explicitly mentions commercial entities ("market actors") as users.

[17] Also national regulations include such publication mandates, e.g. Germany's EEG, EnWG, AnlRegV, and StromNZV. Also those norms do not mention intellectual property rights, and the following arguments do in principle apply also to German publication mandates.

As a consequence, the rightsholders are obliged to provide a license that allows creating a copy of the database for commercial purposes, which is currently not the case (recall 3.2.3). The Transparency Regulation does not mention making data available to the public as a type of reuse.

3.2.5 Establishing ownership

Establishing ownership is imperative. Only the maker of a database – the rightsholder – (or someone with a right or license to act on their behalf) can grant a license. Establishing ownership – determining who holds the rights – is therefore an essential step before a license can be granted.

Multiple rightsholders. It is common for energy data to pass through multiple institutions before being published. A good example is the ENTSO-E Transparency Platform: some parameter may be measured by a distribution system operator, reported to a transmission system operator, aggregated by a power exchange, and transmitted to ENTSO-E before being published on the Transparency Platform. Later, certain data might be published on yet another platform such as SMARD. In this case, it may be that multiple parties – maybe all of the ones mentioned above – hold *sui generis* database rights. It is our understanding that anyone within that chain who spent a substantial investment in either the obtaining, verification or presentation of the contents holds this right. However, there is considerable legal uncertainty when it comes to aggregating, combining and linking of database. It is often not clear where *sui generis* database rights and new rights begin.

The user perspective. In many cases, it is not transparent to data users how data was sourced and which entities might hold rights to it. Since it is usually not publicly documented where the data originated from and who handled it, it is impossible for users to identify the rightsholder(s). For example, ENTSO-E does not disclose who provides a certain data item. In practice, this is a major obstacle to the legal use of energy data, because it makes it impossible to request a license.[18]

Proposal. Every provider of data should track down potential ownership and secure the right to license it to data users in an unrestricted way, i.e. including open licenses. Data providers should make the data sources transparent to users.

3.3 Open data licenses

Above the legal framework of IPR on databases was introduced and applied to data required as input for electricity system modeling. It turned out that (a) most data is likely to be protected under the *sui generis* database right, that (b) statutory exemptions do not apply, and

[18] In 2015, we requested a list of primary data sources from ENTSO-E. It took months to receive this list. Even with this list, it remained unclear if all right holders were included.

that (c) the licenses granted by most data sources are insufficient for most modelers. In our view, the best way forward is to provide data under open licenses, which are reviewed in this section.

3.3.1 The principles of open data licenses

Permit different types of use. For effective and high-quality modeling, it is not sufficient to have the right to simply inspect data. Users need to be legally able to retrieve data (create a local copy); machine-process data in order to clean, repair, combine or otherwise amend it; and make original and amended data available to other parties. Both commercial and non-commercial use must be enabled.

Open licenses. Licenses are the usual way of communicating permissions to potential users of data. An open license grants users of any type the right to freely use, modify and share data for any purpose.

Waivers. A waiver is a legal instrument for giving up one's rights to a resource. Waivers release intellectual property to the public domain. Under German doctrine, some rights cannot be waived and the concept of public domain does not exist. A functional equivalence to waivers are public domain licenses is discussed below.

3.3.2 Different types of open data licenses

Custom versus standard licenses. Most electricity data platforms provide a custom, or bespoke, license. Writing a data license is not trivial, in particular when considering the peculiarities of different national IPR systems. Custom licenses therefore often come at the price of legal uncertainty, as authors fail to consider certain aspects. In addition, bespoke licenses require users to work with many different licenses, which considerably increases transaction costs, an issue sometimes dubbed "license proliferation". For these reasons, the EU Commission Guidelines on recommended standard licenses (2014/C 240/01) favors open standard licenses.

Standard licenses. Legal uncertainty can be greatly reduced by using one of several "standard" open data licenses. These are licenses that were developed by an international community of intellectual property experts, are widely used, have been scrutinized for years in litigation, and consider the peculiarities of all major legal systems world-wide (such as the European *sui generis* database right or the American concept of public domain). These licenses share the following characteristics:

- They are royalty-free.
- They are irrevocable such that they can be terminated only by expiry of the licensor's intellectual property rights.

Flavors of open licenses. Three categories of open licenses exist that differ in the conditions they attach: public domain, attribution, and share-alike licenses. While the first attaches no condition to use, the last two grant permission to use and distribute on condition that certain terms are met.

- **Public domain dedications.** These legal instruments are waivers for copyright and *sui generis* database right. In jurisdiction where waiving copyrights is not possible – as in many European countries – a maximally permissive license agreement replaces the waiver. The license is unconditional and permits use by anyone for any purpose.
- **Attribution licenses.** Attribution licenses require one to give due credit for the maker of a database when it is distributed or used to derive a new work.
- **Share-alike licenses.** Copyleft or share-alike licenses require that databases, if amended or modified, are licensed under the same license as the original database if they are republished. They typically also require attribution. Some share-alike licenses require also non-database derivative works, such as a figure generated from data, to be licensed under the same license.

Attribution versus citation. Note that attribution of authors, required by certain licenses as a condition to lawfully use content that is protected as intellectual property, is different from scientific citation. The latter, being a social normal and a core principle of good scientific practice, is independent from intellectual property right. In scientific publications, authors are expected to cite theirs sources, including data, regardless of license requirements.

3.3.3 Creative Commons and Open Data Commons

CC and ODC families. For each of the three categories, the "Creative Commons" and the "Open Data Commons" families of licenses offer an option. Creative Commons licenses are developed for all kind of content but include clarifications for databases from version 4.0 onwards. Creative Commons 4.0 licenses are used by Wikipedia and Wikimedia. Open Data Commons licenses have been specifically developed for databases, for example they are used by OpenStreetMap. Many countries have developed "national" open data licenses, often intended for the use by public institutions, such as the "Data license Germany". Table 4 summarizes the three license families and gives examples of use in electricity data.

Table 4. Overview of open licenses and examples for electricity data.

Type	Creative Commons	Open Data Commons	Data license Germany
Public domain dedication ("zero licenses")	Creative Commons Zero (CC0-1.0) • OpenEI (U.S. DOE)	Open Data Commons Public Domain Dedication and License (PDDL)	Data license Germany - Zero - Version 2.0 (dl-de/zero-2-0)
Attribution	Creative Commons Attribution (CC-BY-4.0) • SMARD (BNetzA) • Energydata.info (World Bank)	Open Data Commons Attribution (ODC-BY)	Data license Germany – attribution – version 2.0 (dl-de/by-2-0) • BNetzA • StatBA
Share-alike (copyleft)	Creative Commons Attribution Share-Alike (CC-BY-SA-4.0)	Open Data Commons Database License (OdbL) • Enipedia • SciGRID	-

Non-commercial licenses. Licenses that exclude commercial use are not considered open because they discriminate against certain application domains (OKI 2015b).

Issues when combining databases. When combining or merging databases, certain types of open licenses create issues. Attribution licenses can lead to “attribution stacking” if the list of contributors becomes impracticably long. More problematic is license incompatibility which arises when combining two datasets with different share-alike licenses: the derived database would not be able to satisfy both sets of license terms simultaneously.

Proposal. We recommend that electricity data that are *not* protected as intellectual copyright should be marked as such. With the Public Domain Mark this can be easily done. Databases that *are* protected should be licensed under an open license that permits reuse and derivative works. From a user perspective, the fewer restrictions that apply, the better. A public domain dedication such as Creative Commons “CC0 Public Domain Dedication” is the preferred option, which removes nearly all restrictions on further reuse of so-licensed data. The second best would be an attribution license like “Create Commons Attribution 4.0”. Since research is international, an internationally known license such as Creative Commons creates more legal certainty than does the “Datenlizenz Deutschland 2.0”.

3.4 Misuse of data

Concerns of “misuse”. In our conversations with data providers, they occasionally have reported concerns of “loss of control” and “misuse of data”. According to our understanding, this refers to the concern that someone who republishes or otherwise reuses data might intentionally or unintentionally publish “wrong” data or interpret data incorrectly (think of misunderstood data definitions). The fear is that this misrepresentation might come back to the original data source, e.g. in form of press requests or negative publicity.

Misuse is not an open data issue. While this might or might not be a realistic scenario, it does not have much to do with open data, but rather with the publication of data itself. Misuse of data is not a question of intellectual property right.

- Intentional misinformation can easily be addressed by basic civil law: you can prevent anyone from claiming being you. If someone breaks such basic civil law, he is unlikely to be stopped by intellectual property rights.
- Data under open licenses are not more easily misused that any other data that is published under restrictive licenses (or no license at all). Attribution licenses, by enforcing attribution of sources, might even help avoiding confusion among readers about original versus secondary sources.

4 Technical usability of electricity data

This section discusses aspects of technical usability (or "user friendliness") of electricity data, such as access options, documentation and version control.

Sources. This section is informed by our own experience as modelers, our experience from constructing the Open Power System Data platform, interactions with the data community of the Open Energy Modeling Initiative, and a previous assessment of ENTSO-E Transparency Platform we conducted (Hirth et al. 2018). It is also informed by publications, in particular the OECD (2015), European Statistical System Committee (2017) and European Commission (2014) guidelines and recommendations on good statistical practice, recommendations by data repositories[19] as well as community standards and recommendations such as the "Frictionless Data" standard (OKI 2018a).

Eight topics. We discuss eight issues of technical usability in the following:

1. Access (download) options
2. Permanence and version control
3. Machine readability of data
4. Metadata
5. Machine readability of metadata
6. Data documentation
7. Data quality reporting and user involvement
8. Combining data sources and centralized data platforms

4.1 Access options: frequent and in bulk

Context. Some types of electricity system data are large in size and multidimensional in the sense that they may span many different countries / years / technologies / units. In addition, some power system data is updated frequently, e.g. market data.

Different needs. One can differentiate three stylized types of data users, for simplicity called "citizens", "researchers" and "market actors". They face different problems.

- Citizens are not looking for large quantities of raw data, but would like to be informed about individual data points or particular events. Often, visualization is key to satisfying their needs. The needs of this type of user, not being involved in modeling, will not be discussed further.
- Researchers access data relatively infrequently, maybe once or a few times per year. They often download larger chunks of data at once. Reproducibility of scientific studies require data to be permanently available.

[19] Relevant data repositories include the Open Science Framework and Zenodo. See https://www.nature.com/sdata/policies/repositories (Generalist repositories).

- Market actors access data very often, maybe several times per day, downloading only the most recent data.

Problem. Many data providers do not provide convenient access for researchers and market actors. Some sources restrict the data that can be downloaded manually, e.g. offering not more than one month of data at once (e.g., 50Hertz, TenneT, TransnetBW) or not more than one country and one year at once (ENTSO-E Transparency Platform). Some sources make data available only through web interfaces that require human interaction (e.g. CEPS, ENTSO-E Power Statistics) which makes it unnecessarily tedious to access larger volumes of data or to download updated data on an ongoing basis. These restrictions make the access to larger quantities of data a time-consuming exercise.

Best practice examples. Positive examples with respect to allowing users to download large quantities of data with a few clicks include ENTSO-E Power Statistics and Amprion. Other sources facilitate script-based downloading by either providing a web API (RTE, Fingrid, ENTSO-E Transparency), an FTP server (ENTSO-E Transparency) or at least stable URLs following a predictable pattern (i.e. German TSOs construct URLs from the start- and/or end-dates of the period covered in a dataset).

Proposal. The needs and requirements of "researchers" and "market actors" are to some extent incompatible. When adding new data at high frequency to a large dataset, it is impossible to create a new version of the dataset with every addition. We therefore recommend to provide two alternative access options for users, designed along the requirements of researchers and market actors, a proposal also backed by the Open Data Handbook (OKI 2015a). They should build on one uniform backend solution, such as an internal database

- Web APIs. Users who need frequent access to the latest data should be granted access through a web API. This allows programmers to select specific portions of the data, rather than providing all of the data in bulk as a large file. Web APIs are typically connected to a database which is being updated in real-time. This means that making information available via an API can ensure that it is up to date. The web API itself needs to be fully documented. Not all data sources might require a web API; it is most relevant for frequently updated data. For other data sources, the relatively high costs of web APIs (they must be continuously maintained and fed from a database) might not be justified.
- Packaged data. To accommodate users who need bulk access, sets of the data could regularly be "frozen" (all data is saved at one point in time) in specified intervals and offered under version control. The user interface should allow manual download of all available data in bulk, i.e. with one or a few clicks. In the case of time series data, the entire timespan should be available in one file (not only individual weeks/months/years), all variables should be available in one file (not wind and solar separately) and all countries should be available in one file as well. Most arbitrary limits on the amount of data that can be downloaded at once are technically unnecessary, especially when the volume of data is of the order of magnitude of kilobytes to megabytes. The URLs for download files should be stable (i.e. not change after publication) and follow a clearly identifiable pattern, so that it is possible to

programmatically download data files. Data should be version-controlled (see 4.2) and packaged with machine-readable metadata (4.3).

4.2 Permanence and version control

Context. For reproducibility of scientific results, it is important that data sources can be uniquely identified and that they remain available over time.

Problem. Many datasets do not remain stable over time. Some energy statistics undergo revisions, as more and better estimates and measurements for the same data are collected over time, such as the generation by fuel type data on the ENTSO-E Transparency platform. Today, most institutions overwrite previous estimates. This creates confusion and makes it difficult to reproduce model results using legacy datasets. It also makes it impossible to trace revisions over time.

Best practice examples. A data provider that addresses this problem is Open Power System Data, which assigns a version identifier to each dataset published and keeps an archive of all previous versions.

Proposal. We propose that data providers use version control and offer stable identifiers. This is more easily done with packaged data, which gives another reason to provide data in bulk. Older versions of the revised data should remain available with a version identifier – such as the publication date – attached. Each version should include a change log describing which changes have been made to the data. Specific versions of a dataset should receive a unique identifier, such as a permanent URL (i.e. a URL which is always accessible, even when the website design or layout changes) or – better – a Digital Object Identifier (DOI). Unlike URLs, DOIs can be easily remapped. In addition, a hash (checksum) of each data file should be provided to confirm integrity. New versions should be provided in monthly to yearly intervals. It is not necessary to provide a new version of a dataset every time a single value is updated. For the ENTSO-E Transparency Platform for example, a quarterly snapshot of the files provided on the FTP server (currently ~50 GB) could be achieved provided in zipped format in order to reduce storage requirements.

4.3 Machine readability of data

Context. Models require large amounts of data from various sources that are regularly updated. Frictionless processing requires data to be machine readable, meaning that data should be available in formats that can be easily read by computers and integrated into data analysis workflows. Machine readability is one of the core recommendations of OECD (2015) and European Commission (2011).

Problem. Although many datasets are now machine readable, their format is not standardized. Until a few years ago, it was very common to provide data in PDF files, from which it is extremely difficult to extract tabular data; this can still be found in some cases. Most large data providers today supply data in XLSX, CSV or XML files that can be automatically processed, so the minimum requirement for machine readability is fulfilled. However, the organization of data within files is not standardized, which means that considerable effort is required to parse the data contained in files, especially if many different data sources are processed. For example, solar and wind generation data is organized differently by every TSO that provides such data (e.g., organization in rows vs. columns, labelling of columns, number and order of columns). Moreover, the same source often changes the organization of data in files over time. This not only adds a burden to data users, it also increases the risk of analysis being based on flawed data.

Best practice examples. Central data platforms that aggregate information from multiple sources such as ENTSO-E's Power Statistics and Transparency Platform help to address the problem of machine readability of data, since they provide a vast amount of data in identically structured data files.

Proposal. Data standards exist and should be adhered to. Data should be provided in CSV files, and possibly also in XLSX files. Data providers should provide data in a consistent format for all data they provide; in particular, if file types and formatting changes, all previous data should be made available in the new setup so that all data are available in a consistent format. Values in the data files should be formatted according to established formatting standards, e.g. timestamps conforming to ISO standard 8601.

4.4 Metadata

Context. In order to correctly interpret a dataset, documentation describing its structure and contents is required. Certain information requires extensive documentation, such as estimation and measurement techniques (we discuss data documentation in section 4.6 below). Other information such as sources, units of measurement or contact persons can, and indeed should, be reported in a structured fashion, i.e. as a list of nested key-value pairs. This kind of structured data about data is called "metadata". The keys denominate a standardized set of properties (i.e. "publishing date"), while the values contain the actual information (i.e. "2018-11-06"). This structure, applicable not only to datasets, is formalized by the Resource Description Framework (RDF; W3C 2014b) and has emerged as a common framework for expressing metadata on the internet. Ideally, metadata provide a quick but comprehensive overview of a dataset's content and structure without needing to inspect the data itself.

Problem. There are many providers of energy data who do not supply sufficient or even any metadata. Often, information on measurement units are only apparent from field names or are lacking altogether, license information is almost never provided, and also detailed information on the source and provenance is often lacking.

Best practice example. The data platform the French TSO RTE publishes its data on, Open Data Réseaux Énergies, provides metadata for their datasets, including a short description of the content of a dataset as well as information on publication date, license, and datatype definitions for each column (in French).

Proposal. Metadata should be at least provided on the "landing page". This is the page users see when accessing the URL or DOI. The landing page should include a brief description of the dataset, metadata, and download links to access the data in different formats. Metadata should be rich and have a plurality of accurate and relevant attributes. At the very least, the metadata should allow to cite or give attribution to the dataset with regard to title, author, source and license (TASL). Table 5 provides a list of suggested metadata items that should be provided. The table lines out metadata *content*. See section 4.5 for suggestions regarding metadata *formatting*.

Table 5. Recommended metadata properties for a dataset. Depending on the dataset, spatial scope should be specified either at the level of the dataset or individually for each variable/column.

Key	Value
Metadata properties for each dataset	
Name	A name or a short title of the dataset.
ID	A globally unique identifier, e.g. Universal Unique Identifiers (UUID) or Digital Object Identifiers (DOI).
Homepage	A URL for the home on the web that is related to the dataset ("landing page"). The URL should be permanent.
Licenses	The license(s) under which the dataset is provided as name, URL and SPDX identifier, such as: Creative Commons Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/, CC BY 4.0. If the license maintains intellectual property rights, it should be specified how the rightsholder should be attributed.
Description	A description of the dataset. A short paragraph providing some context information.
Documentation	A link to the further documentation of the dataset
Spatial scope	The geographical area the dataset represents, e.g. "50Hertz control area"
Temporal scope and resolution	Either a reference date indicating which point in time a dataset represents e.g. "2017-12-31" or, in the case of time series data, the beginning and end of the total period covered, e.g. "2008-01-01T00:00:00 to 2018-05-04T:23:59:59" For time series data, also indicate the duration between time steps as well as whether data is associated with the beginning or end of a reported period.
Publication date	The publishing date of the dataset, such as 2018-04-05.
Sources	The raw sources for the dataset. Where the publisher of the dataset is not the original creator of the data, proper reference should be given to the original source. For example, in the case of the ENTSO-E Transparency Platform, the contact details of the Primary Data Owner(s) as well as the Data Provider(s) should be provided.
Contact	A point of contact, such as a forum, or ideally a person available to contact for questions and feedback regarding the data.
Technical specifications	Information required to parse the data file(s) by the processing software, i.e. for CSV files, the character encoding (UTF8), column separator (,) and decimal character (.) and for time series, the ISO 8601 datetime format (YYYY-MM-DDThh:mm:ssZ).
Metadata properties for each variable (e.g. one column/attribute in a list of power plants)	
Name	This property should correspond to the name of the column in the data file. As such it should be unique.

Description	A description for the column.
Unit	The unit of measurement, e.g. MW.
Type	The data type (e.g. ”string”, ”number”, ”datetime”) and format of the column, e.g. for a datetime ”YYYY-MM-DDThh:mm:ssZ”
Spatial scope	The spatial entity represented by the column, e.g. “50Hertz control area”.

4.5 Machine readability of metadata

Context. Metadata only play out their strengths if they are provided in a machine readable and standardized format, allowing scripts from users and aggregators to build on it i.e. when combining data from multiple sources. This includes the detection of datasets by search engines. Metadata are machine readable if they are defined using a standardized vocabulary and an established serialization format.

Metadata vocabulary. A metadata vocabulary defines the domain of expected properties and acceptable keys and values. Common vocabularies include Dublin Core (Dublin Core metadata initiative 1999), the Data Catalog Vocabulary (DCAT; W3C 2014a) and schema.org (2018). With DCAT-AP and DCAT-AP.de, the European Commission (2018) and German governmental data portal GOVDATA (Sklarß 2018) are extending the DCAT vocabulary in order to standardize the description of public sector datasets in Europe and Germany. Google Dataset Search builds on the schema.org vocabulary (Google 2018). The frictionless data vocabulary developed as part of the Tabular Data Package Standard (OKI 2018c) is a particularly lightweight vocabulary tailored for CSV-datasets.

Serialization format. A serialization format defines a syntax for the metadata. Widespread minimal formats that are readable for machines as well as humans include JSON and JSON-LD.

Problem. Metadata are often not available or don’t follow the established standards.

Best practice examples. The problem of supplying metadata along with the measurements is solved by Open Power System Data by following the Tabular Data Package standard.

Proposal. Metadata should be provided in a standardized, machine-readable, structured form in a dedicated file that accompanies the data, such as the “datapackage.json” file prescribed by the Data Package (OKI 2018c) standard.

4.6 Data documentation

Context. In order to use data for energy system analysis, it needs to be described. Certain information such as sources, units of measurement or contact persons should be provided as structured metadata. Other information requires more extensive description, such as estimation and measurement techniques, the individual steps of data processing, interpolation and extrapolation, and/or errors and confidence intervals. This description, which may include equations, tables, figures and references to the literature, we call “data documentation.”[20]

An example. Take the example of load: depending on the source, load may or may not include electricity consumption incurred by (pumped hydro) storage, small scale self-produced electricity, industrial self-produced electricity, power consumption by railroads; and it might include or exclude certain geographic areas like Austria’s *Kleinwalsertal* which is connected to the German grid only. Even the term “load” itself might require definition, as some sources treat it as synonymous with “consumption”, while others do not.

Problem. Today, many sources do not provide a detailed data documentation. In fact, many sources do not provide any documentation at all. To the extent that a description is available, it is often difficult to find.

Best practice examples. The “Detailed Data Descriptions” of ENTSO-E’s Transparency Platform (ENTSO-E 2014) are an example of an attempted to define data clearly and consistently. Eurostat’s Energy Balances (Eurostat 2018) provide a rich documentation including data definitions, estimation methodology and relationships between variables. The European Environmental Agency publishes a “data viewer manual” (EEA 2018): accompanying the national emissions reported to the UNFCCC and to the EU Greenhouse Gas Monitoring Mechanism, which addresses questions about coverage and calculation methods. Outside the electricity sector, statistical offices often provide extensive data documentation. Take, for example, the system of national accounts: Germany’s Statistisches Bundesamt provides a dedicated website (Destatis 2018) plus an entire series of technical papers on estimation methods and calculations procedures.

Proposal. All data sources should provide a detailed, up-to-date documentation of data definitions and estimation methods. Depending on the type of data, this should include but not be limited to:

- The definition of the data item and its scope
- A description of the measurement or estimation technique
- A description of data processing steps, including possible extrapolation (in case of sampling) or interpolation (in case of time series data)
- Any changes to data availability or to collection and estimation methods over time
- Relationship to other data, e.g. known inconsistencies with similar data (say, in the case of a system operator, a definition that is divergent from other system operators)

[20] The terms “metadata”, “data documentation” and “data description” are not well defined in the literature and are sometimes used interchangeably. We think it make sense to discuss metadata (which can be structured and machine-readable) seperately from the data documentation.

4.7 Data quality reporting and user involvement

Context. Data is never perfect. Mistakes, gaps and inaccuracy will always exist. Therefore, it is important for users such as electricity system modelers to remain informed about the reliability and known issues with the data.

Problem. As of today, there is often little information available to users about data quality. This makes it difficult to determine the trustworthiness of analysis when there is uncertainty about the quality of the input data. It also requires that users check individually for data quality, causing redundant work.

Best practice examples. Energinet.dk operates a data platform that allows public comments by users.[21] Open Power System Data uses GitHub Issues to track questions and errors regarding datasets.

Proposal 1. Data providers should regularly validate and assess data quality and the assessment reports should be made public. For data that are published due to statutory obligations, such assessments should be made obligatory. For datasets containing large data volumes, such as time series, a structured format of marking possibly corrupt or otherwise problematic data should be adopted. Depending on the type of data, this can take different forms; some validation and plausibility checks could be automatic. Often gaps in data are much easier to identify than incorrect entries, so data gap reporting might be a first step. In addition, alternative data sources may exist and it can be helpful to compare these in order to identify inconsistencies. Take the example of load data: one could regularly report on consistency between hour-by-hour data and more aggregated data from Eurostat, ENTSO-E or national statistical offices and discuss deviations. (It is clear that they will never exactly match – the point is rather to equip users with an indication about the order of magnitude of differences, together with potential reasons and changes over time.)

Proposal 2. In addition, we propose to harvest the knowledge of data users by establishing a "public data error log" for each major data source. In its simplest form, it could be implemented as a forum on the landing page of the data item. Registered users should be able to post a comment if they encounter issues. The data provider as well as other users can respond; all comments are public. Once the issue is solved the service desk flags the item as "resolved". The posting and comments remain online. A user contributed public data log has multiple benefits:

1. Users are warned about issues and can use data with additional care. This reduces the likelihood of analyses being based on flawed data.
2. Data providers are warned immediately about issues and have the chance to respond quickly. They also can declare immediately that there is not an issue if that is the case.

[21] E.g., see the discussion on spot prices on https://www.energidataservice.dk/en/dataset/elspot-prices.

3. Other users can post solutions or explanations. This can help data providers to identify the cause of the issue, and thereby reduce the workload for data providers.

4. A log creates transparency about structural problems and hence provides an incentive for data providers to improve the quality of their data and processes.

4.8 The Tabular Data Package Standard

We are not the first to identify bulk access, machine readability and (structured) metadata as important characteristics of frictionless data. Standard formats for data and metadata exist that address multiple problems simultaneously.

Data Package standard. One example is the "Tabular Data Package" standard, developed and maintained by Open Knowledge International (Fowler et al. 2018, OKI 2018c). Put simply, each Tabular Data Package comprises one or more CSV files that contain the data itself, plus a file that contains structured metadata in the JSON format, which is both machine-readable and human-readable. Metadata are specified in a lightweight vocabulary particularly suited to CSV data. Values in the data files are formatted according to established formatting standards, i.e. timestamps are formatted in the ISO standard ISO 8601. The tabular data package is suitable for both tabular and time series data. Table 6 summarizes the problems identified: four out of eight are addressed by the Tabular Data Package standard.

Table 6. Technical usability problems and proposed solutions

Problem / Need	Suggested solution	Addressed by Tabular Data Package Standard
Bulk access (researchers)	Packaged data, e.g. one big file	✓
Fast access (market actors)	Web API	✗
Reproducibility, permanent availability	Version control + permanent link, e.g. DOI	✗
Machine readability of data	Common standard of data file formats / structure / naming	✓
Minimum list of metadata item	Common standard, e.g. Table 5	✓
Structured (machine readable) metadata	Common standard for metadata, e.g. JSON/ISO	✓
Data documentation	Detailed documentation	✗
Data quality reporting and user involvement	Regular quality assessments, crowd-sourced issue list	✗

Proposal. We recommend that all published data should follow this or a similar standard. This would solve the problem of bulk access (as data is packaged and can be downloaded with one click), of machine readability of data (as all data is structured in a compatible way), of the provision of rich metadata (as the standard requires certain metadata to be published), and of metadata machine-readability (as the standard requires to provide metadata in structured form).

Not enough. Not all the issues described are addressed by the Data Package standard. In particular, web API access and the provision of DOIs are not part of the standard. Neither are data documentation, quality reporting and user involvement included.

4.9 Centralized data platforms

Context. An obvious requirement for data to be usable is that users know it exists and where to find it. However, given the multitude of data providers, this is no trivial task.

REMIT. This particularly concerns information on outages of production, generation and transmission assets as required by REMIT (Regulation (EU) No. 1227/2011) Such information is disclosed via so called “Urgent Market Messages” (UMMs), which are specified by the European Agency for the Cooperation of Energy Regulators (ACER 2018). REMIT does not require a common data platform, although the ACER Guidance encourages the use of centralized “Inside Information Platforms”. In practice, common Inside Information Platforms exist for the CWE region and the Nordic countries (EEX Transparency and Nordpool REMIT UMM) as well as for individual countries (Great Britain, Hungary, Italy, Poland, Portugal). For other countries however, no such platforms exist, meaning that data has to be collected from individual operator’s websites. The ENTSO-E Transparency platform also provides UMMs for assets >100 MW, but is not recognized as an Inside Information Platform since it does not fulfill the timeliness requirements set out by REMIT. In some cases, the same UMMs might appear on some or all of the websites mentioned above. The nature of the data on outages (which might be planned or unplanned, cancelled or rescheduled) requires that UMMs are updated frequently. Since market participants need to ensure to always have the most recent information, they often monitor all the different sources in parallel.

Problem 1. Some data types are dispersed across numerous providers that are hard to locate without in-depth and country level knowledge.

Problem 2. Each data provider usually establishes their own formatting conventions, increasing the workload for users looking to combine different sources.

Problem 3 Combining data from different sources is particularly burdensome for power plant data, since they have to be matched along a common identifier.

Best practice example. The ENTSO-E Transparency Platform bundles many types of data for the whole EU and some further European countries in one place using the same harmonized data format. The German TSOs operate the joint platforms regelleistung.net and netztransparenz.de in order to fulfil their publication mandates. Eurostat bundles data from national statistical offices in one place using a common format.

Data aggregation platforms. In the absence of centralized platforms, search engines such as Google Dataset Search or curated lists of data sources such as Open Power System data (2018) can help users find the data they need.

Proposal. Any regulation establishing publication obligation should mandate the use of centralized platforms. In the case of REMIT outage data, one central platform should gather all UMMs from all data providers, not just the ones already publishing on EEX or Nordpool UMM. Any power plant level data sets should include the Energy Identification Codes (EIC) – an identifier – in order to allow matching with other sources

5 Recommendations

From the above analysis, we summarize recommendations for data providing institutions and for policy makers.

5.1 Recommendations for data providers

This section summarizes practical recommendations from chapters 0 and 4 for how data-providing institutions can facilitate the use of electricity data for modeling. It constitutes a "best practice guide" for institutions that provide electricity system data, such as governments, regulators, statistical offices, other authorities, TSOs, DSOs, power exchanges, market participants, trade associations, and researchers.

These are the recommendations in short:

1. Establish ownership. Data providers, including platforms that publish data supplied by different institutions, should indicate who holds intellectual property rights. It must be transparent to users as to who holds the intellectual property rights.
2. Provide an established open license. To facilitate analysis and reuse, data should be published under an established open license. We recommend a public domain dedication; where this is not possible, an attribution license is the second best choice, such as Creative Commons Attribution 4.0.
3. Provide packaged data that includes metadata. Data should be made available through "packages" of data plus metadata. Those packages should be large to permit download of data in bulk, be version controlled, and be permanently available, preferably through a Digital Object Identifier.
4. Follow the Tabular Data Package standard. Packages should follow the Tabular Data Package standard, distributing data in CSV files that come with structured metadata in JSON format.
5. Provide access through web APIs where necessary. Market data and other data that is frequently updated should be available through a well-documented web API. In addition, it should also be published periodically in packages.
6. Provide detailed and up-to-date documentation. The documentation should include detailed information about measurement and estimation techniques.

7. Assess and report data quality. Data quality assessment and forums for public user feedback can help improve data quality.

5.2 Recommendations for policy makers

Options for policy-makers at the state, national and European level to improve data availability and quality and public usability.

1. Require open data when funding research. All data generated in publicly funded research projects should be available to the public and the modeling community for reuse. This means it should be available in machine-readable formats, accompanied with metadata, well documented, and published under an open license.
2. Public sector information should be open data. Data provided by public authorities such as ministries and agencies should be open. In particular, we recommend waving copyright and/or the *sui generis* database right on databases by default along the lines of the proposed amendment of the PSI directive. For clarification, a public domain dedication (CC0-1.0 or PDDL) could be provided.
3. Publication obligations should be improved. The obligations in European and national laws to publish data, including REMIT and the Transparency Regulation for power system data, could be improved in several ways, in particular:
    - Require open licensing if intellectual property rights apply; for clarification, require a public domain mark if intellectual property rights do not apply.
    - Make the Best Practice Guide of section 5.1 a requirement
    - Require publication through centralized platforms (as the Transparency Regulation already does, but REMIT does not)
    - Set the right incentives for data providers (and publication platforms) to provide well-curated data (i.e., accurate, timely, user-friendly, well documented, validated)
    - Legal analysis and statutory reforms: Sui generis database rights should be automatically waived for public sector information and legally mandated data publication.
4. Intellectual property right reform. It seems that a broad IPR reform is warranted. In particular, it seems sensible to abolish the *sui generis* database right altogether or at least to remove it for public sector information and mandated data publications at both at the EU and national levels.[22] .
5. Exceptions and Limitations to copyright. Make sure that exceptions for official works in national copyright law are as broadly applicable as possible and also encompass the *sui generis* database right.

[22] See ongoing work including public consultation on PSI directive: https://ec.europa.eu/info/consultations/public-consultation-review-directive-re-use-public-sector-information-psi-directive_en

6 References

ACER (2018): Guidance on the application of Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency. https://documents.acer-remit.eu/category/guidance-on-remit/

Ball, Alex (2014): *How to License Research Data*, a Digital Curation Cenre and JISC Legal 'working level' guide, http://www.dcc.ac.uk/resources/how-guides/license-research-data

Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009 of the European Parliament and of the Council. http://eur-lex.europa.eu/eli/reg/2013/543/oj

Davison, Mark J (2008). *The legal protection of databases*. Cambridge, United Kingdom: Cambridge University Press. ISBN 978-0-521-04945-0

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases ("Database Directive"). http://data.europa.eu/eli/dir/1996/9/oj

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. http://data.europa.eu/eli/dir/2003/98/oj

Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. http://data.europa.eu/eli/dir/2013/37/oj

Dublin Core metadata initiative (1999). *Dublin Core metadata element set, version 1.1*. July 1999. Dublin Core recommendation. http://dublincore.org/documents/dcmi-terms/

ENTSO-E (2014): *Detailed Data Descriptions, Version 1, Release 4*. https://docstore.entsoe.eu/Documents/MC%20documents/Transparency%20Platform/MOP/DetailedDescriptionDocument.pdf

European Commission (2005): *First Evaluation of Directive 96/9/EC on the legal protection of databases.* http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf

European Commission (2014): *Commission Notice. Guidelines on recommended standard licences, datasets and charging for the reuse of documents* 2014/C 240/01. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2014.240.01.0001.01.ENG

European Commission (2018): *About DCAT application profile for data portals in Europe*. https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/about

European Environmental Agency (EEA) (2018): *EEA greenhouse gas data viewer manual*. https://www.eea.europa.eu/data-and-maps/data/national-emissions-reported-to-the-unfccc-and-to-the-eu-greenhouse-gas-monitoring-mechanism-14#tab-additional-information

European Statistical System Committee (2017): *European Statistics Code of Practice (ECoP).* https://doi.org/10.2785/798269

Eurostat (2018): *Energy Balances*. http://ec.europa.eu/eurostat/web/energy/data/energy-balances

Fowler, Dan, Jo Barratt, & Paul Walsh (2017). "Frictionless data: making research data quality visible". *International Journal of Digital Curation*. 12 (2): 274–285. doi:10.2218/ijdc.v12i2.577

Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz - EGovG) vom 25.07.2013 i. d. F. vom 05.07.2017. http://www.gesetze-im-internet.de/egovg

Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetzt - UrhG) vom 09.09.1965 i. d. F. vom 01.09.2017. https://www.gesetze-im-internet.de/urhg

Gesetz über die Weiterverwendung von Informationen öffentlicher Stellen (Informationsweiterverwendungsgesetz – IWG) vom 13.12.2006 i.d.F. vom 08.07.2015. https://www.gesetze-im-internet.de/iwg

Google 2018: *Dataset*. https://developers.google.com/search/docs/data-types/dataset

Hirth, Lion, Jonathan Mühlenpfordt & Marisa Bulkeley (2018): "The ENTSO-E Transparency Platform – A review of Europe's most ambitious electricity data platform". *Applied Energy* 225, 1054-1067. https://doi.org/10.1016/j.apenergy.2018.04.048

Jaeger, Till (2014): *LG Köln beschränkt CC-BY-NC auf rein private Nutzungen. ifrOSS News, 19.03.2014.* http://www.ifross.org/artikel/lg-k-ln-beschr-nkt-cc-nc-rein-private-nutzungen

Jaeger, Till (2017): *Legal Opinion – Legal Aspects of European Energy Data*, http://open-power-system-data.org/legal-opinion

Jaeger, Till (2018): *Legal Opinion – Legal Aspects of European Energy Data*, appendix to the *Study on the quality of electricity market data of transmission system operators, electricity supply disruptions, and their impact on the European electricity markets.* https://ec.europa.eu/energy/sites/ener/files/documents/legal_aspects.pdf

Klimpel, Paul (2012): *Freies Wissen dank Creative-Commons-Lizenzen. Folgen, Risiken und Nebenwirkungen der Bedingung »nicht-kommerziell – NC«*. https://irights.info/wp-content/uploads/userfiles/CC-NC_Leitfaden_web.pdf

Morrison, Robbie, Tom Brown & Matteo De Felice (2017): *Submission on the reuse of public sector information: with an emphasis on energy system datasets*, https://okfn.de/files/blog/2018/01/reuse-PSI-openmod.pdf

Morrison, Robbie (2018): "Energy system modeling: Public transparency, scientific reproducibility, and open development", *Energy Strategy Reviews* 20, 49-63. https://doi.org/10.1016/j.esr.2017.12.010

OECD (2015): *Recommendation of the OECD Council on good statistical practice.* http://www.oecd.org/statistics/good-practice-toolkit

Open Knowledge international (OKI) (2015a): *The Open Data Handbook: Guides, case studies and resources for government & civil society on the 'what, why & how' of open data.* http://opendatahandbook.org

Open Knowledge international (OKI) (2015b): *The Open Definition.* https://opendefinition.org

Open Knowledge international (OKI) (2018a): *Frictionless Data. Specifications and Software*. https://frictionlessdata.io

Open Knowledge international (OKI) (2018b): *Data Package*. https://frictionlessdata.io/specs/data-package

Open Knowledge international (OKI) (2018c): *Tabular Data Package*. https://frictionlessdata.io/specs/tabular-data-package

Open Power System Data (2018): *Data sources*. http://open-power-system-data.org/data-sources

Pfenninger, Stefan, Lion Hirth, Ingmar Schlecht, Eva Schmid, Frauke Wiese, Tom Brown, Chris Davis, Matthew Gidden, Heidi Heinrichs & Clara Heuberger (2018): "Opening the black box of energy modeling: strategies and lessons learned" *Energy Strategy Reviews* 19, 63-71, https://doi.org/10.1016/j.esr.2017.12.002

Pfenninger, Stefan, Joseph DeCarolis, Lion Hirth, Sylvain Quoilin & Iain Staffell (2017): "The importance of open data and software: is energy research lagging behind?", *Energy Policy* 101, 211-215, https://doi.org/10.1016/j.enpol.2016.11.046

Regulation (EC) No 714/2009 of the European Parliament and of the Council of 13 July 2009 on conditions for access to the network for cross-border exchanges in electricity and repealing Regulation (EC) No 1228/2003. http://data.europa.eu/eli/reg/2009/714/oj

Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency (REMIT). http://data.europa.eu/eli/reg/2011/1227/oj

Schema.org (2018):*Dataset*. Schema.org version 3.4, 15 June 2018. https://schema.org/Dataset

Sklarß, Sebastian, Dietmar Gattwinkel, Antje Göldner, Christian Horn, Christian Wittig & Ogeigha Koroyin (2018): *DCAT-AP.de Spezifikation.* Deutsche Adaption des „Data Catalogue Application Profile" (DCAT-AP) für Datenportale in Europa. Version: 1.0.1. 14 June 2018. https://www.dcat-ap.de/def

Statistisches Bundesamt (Destatis) (2018): *Erläuterungen zu den Volkswirtschaftlichen Gesamtrechnungen.* https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Methoden/methodisches.html

W3C (2014a): *Data Catalog Vocabulary (DCAT)*. W3C Recommendation 16 January 2014. https://www.w3.org/TR/vocab-dcat

W3C (2014b): *RDF 1.1 Primer*. W3C Working Group Note, 25 February 2014. http://www.w3.org/TR/rdf11-primer

Wilkinson, Mark D. et al. (2016): "The FAIR Guiding Principles for scientific data management and stewardship", *Nature Scientific Data,* 3, https://doi.org/10.1038/sdata.2016.18

Glossary

6.1 Legal glossary

Copyright. Intellectual property right for creative works. Databases can attract copyright if the selection or arrangement of their contents is sufficiently creative (Directive 96/9/EC, Article 3). Chapter II of the Database Directive specifies this right.

Data. An individual piece of information, observation, number, or fact. Data alone cannot attract an intellectual property right, but collections of data in different circumstances may. See →Database.

Database. In legal terms, a "collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means" (Directive 96/9/EC, Article 1). The structured data itself is the legal entity for protection, not the software or hardware used to store that information. The legal definition of a database is far broader than its technical definition (Davison 2008).

Dataset. A term not used or defined in relevant directives or statutes. See →Database.

Reuse. In this study we follow Jaeger (2018) to define reuse as "any activity to copy, modify, publish or distribute energy data or communicate the data to the public subsequently to the publication by the data provider". In this context, the Database Directive defines "re-utilization" as "making available to the public". The Directive does not contain the term "reuse".

Database right. See →*Sui generis* database right.

***Sui generis* right**. See →*Sui generis* database right.

***Sui generis* database right**. The intellectual property right granted to the maker of a database who expends a "substantial investment in either the obtaining, verification or presentation of the contents". This right protects against "the extraction and/or re-utilization of the whole or of a substantial part, evaluated quantitatively and/or qualitatively" (Directive 96/9/EC, Article 7). Chapter III of the Database Directive specifies the right, although there is very little case law on exactly how these definitions should be interpreted. The *sui generis* database right exists under European Union law and has no equivalent in other legal jurisdictions, including the United States.

6.2 Technical glossary

API. An application programming interface (API) provides a bridge to software components which offer services using a set of well-documented functions which the client software can call. In the context of this paper, we refer to web APIs, which are described further below.

Data Package. A bundle of data files and associated metadata as defined by the specification of the same name by Open Knowledge international (OKI 2018b).

Metadata. Additional machine-readable information which accompanies data files and which documents their structure, provenance, semantics, legal context, and similar. See → Data Package.

SPDX identifier. A standardized identifier for common open licenses developed by the Linux Foundation to aid the automated parsing of legal information in metadata.

Tabular Data Package. A specific dialect (or more precisely: profile) of the Data Package standard that defines Data Packages containing tabular data. The standard prescribes data to be published in CSV file formats conforming to certain conventions (OKI 2018c).

Landing page. In general, a landing page refers to the initial page users see on a web site. The term is used in this report to describe the main web page where a dataset is described and where the download link is visible.

Web API. An API (see → API) which can be accessed over the web. Usually, web APIs define a set of URL parameters that can interact with a server or database.

Appendix: license texts

For all data sources mentioned in section 2.3, this appendix provides the relevant license text. If multiple languages are available, the English text is reproduced. For each data source, the source URL is provided, followed by the license agreement. Only the parts of the text relevant for intellectual property right on databases is reproduced (excluding, e.g. statements on trade marks of photographies as well as liability disclaimers); please visit the respective sites for full text. Emphasize is added by us.

Eurostat

http://ec.europa.eu/eurostat/about/policies/copyright

Copyright notice and free reuse of data

Eurostat has a policy of encouraging free reuse of its data, both for non-commercial and commercial purposes. All statistical data, metadata, content of web pages or other dissemination tools, official publications and other documents published on its website, with the exceptions listed below, can be reused without any payment or written license provided that:

1. the source is indicated as Eurostat;
2. when reuse involves modifications to the data or text, this must be stated clearly to the end user of the information.

Statistisches Bundesamt

https://www.destatis.de/EN/Service/Imprint/CopyrightGENESISOnlineDatabase.html

Copyright for the Genesis-Online database (the Download page of Statistisches Bundesamt)

The following applies for the content of the Genesis-online database:

Data license Germany – attribution – version 2.0

The granting of rights under this data license corresponds to the scope of usage rights given under Copyright-General.

https://www.destatis.de/EN/Service/Imprint/CopyrightGeneral.html

Copyright - general

Reproduction and distribution, also of parts, are permitted provided that the source is mentioned.

This applies for any redistribution of texts, data and charts by third parties where the Federal Statistical Office is the copyright owner and holds exclusive publishing rights. Reuse is permitted both for commercial and non-commercial purposes. There are also no restrictions on distribution by electronic means, via the Internet or in printed form. No express authorisation is required from the Federal Statistical Office. It is, however, required to provide a source note. The Statistisches Bundesamt (Destatis) shall be named as the publisher in the list of references. Supplementary provisions apply for the use/redistribution of content from Genesis-Online.

Amendments, deletions/abridgements or omissions, redesigns or other modifications shall be identified as such, or a note shall be included in the references stating that the data have been changed, have only been used as the basis for calculation or have been presented in a different form.

ENTSO-E Transparency Platform

https://www.entsoe.eu/Documents/MC%20documents/Transparency%20Platform/150615_ENTSOE_Transparency_Terms_Conditions_Full.pdf

5. Use of the Transparency Platform Data

In accordance with the applicable legislation, the Data User shall, when using of the Transparency Platform Data for any purpose whatsoever:

-use the Transparency Platform Data in good faith and always comply with good business

practices regarding the reuse of publicly available data;

-mention the ENTSO-E Transparency Platform as the source of publication of the data, in accordance with good industry practices and comply with all reasonable requests from ENTSO-E regarding the visibility of the ENTSO-E Transparency Platform origin of the reused Transparency Platform Data;

-be only allowed to make reference to the ENTSO-E Transparency Platform as the source of publication of the reused data. It is therefore expressly prohibited to use the ENTSO-E Transparency Platform name or the ENTSO-E name in any manner that is likely to cause confusion regarding the possible existence of any kind of sponsorship or of endorsement of any use of the Transparency Platform Databy the Data User;

-not cause prejudice to the copyright or related right on a Transparency Platform Data, which may be owned by the concerned Primary Owner of Data. In case of a risk to cause prejudice to said right, the Data User shall seek the prior agreement of the holder of the copyright or related right.

ENTSO-E Power Statistics

https://www.entsoe.eu/disclaimer/Pages/default.aspx

Disclaimer

2. Ownership of Content

The Site and all of its contents including, but not limited to, all text, graphs and images ("Content") are owned and copyrighted by ENTSO-E or others with all rights reserved unless otherwise noted. Your use of any Content, except as provided in these Terms of Use, without the written permission of ENTSO-E is strictly prohibited.

3. Your Use of the Site

The ENTSO-E grants you permission to use the Site as follows:

1. with the exception of images of people or places that are located outside of the "News" section of the Site, *you may download Content, but only for non-commercial, personal use* and provided that you also retain all copyright and other proprietary notices contained on the Content; […]

4. *you may not distribute, modify, copy (except as set forth above), transmit, display, reuse, reproduce, publish, license, create derivative works from, transfer, sell or otherwise use Content* without ENTSO-E's written permission;

Bundesnetzagentur: generell

https://www.bundesnetzagentur.de/DE/Service/Impressum/impressum_node.html

Rechtliche Hinweise

4. An allen Seiten dieses Internetauftritts inklusive Layout, Quelltext, Software und deren Inhalten besitzt die Bundesnetzagentur das Urheberrecht und sonstige Schutzrechte. Urheberrechtshinweise und Markenbezeichnungen dürfen weder verändert noch beseitigt werden. *Die auf der Internetseite verwendeten Daten in den Formaten .xls, .xlsx und .csv stehen, falls nicht anders gekennzeichnet, unter der Datenlizenz Deutschland – Namensnennung – Version 2.0.* Alle darüber hinausgehenden Handlungen bedürfen einer vorherigen schriftlichen Zustimmung der Bundesnetzagentur.

EEX Transparency

https://www.eex.com/en/market-data/market-data-download/subscription/general-terms-of-contract

13. INFO-USER – RIGHTS OF USE

EEX AG grants the Subscriber of the Info-User product the right to use, save and process the available information in any form and in any way – however, he shall exclusively be entitled to use such for his own purposes.

Any dissemination, publication and other commercial use of the data is not permitted. The right to use the information cannot be transferred.

[These rights of use pertain to paying customers only. There is no information regarding the rights of non-paying users]

Bundesnetzagentur: SMARD

https://www.smard.de/blueprint/servlet/page/home/datennutzung/666

Datennutzung

Die auf dieser Platform veröffentlichten Daten gelten gemäß § 111d EnWG als für die Öffentlichkeit zur freien Verfügung und Verwendung bereitgestellt. Hierfür können die Daten kostenfrei heruntergeladen und gespeichert werden. Eine Haftung der Bundesnetzagentur für die Richtigkeit und Vollständigkeit der Daten wird ausgeschlossen. *Die Daten aus dem Bereich Marktdaten visualisieren sind lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz*.

Umweltbundesamt

https://www.umweltbundesamt.de/dokument/datenbank-kraftwerke-in-deutschland

Das Umweltbundesamt weist darauf hin, dass die Datenbank „Kraftwerke in Deutschland" dem Urheberrecht unterliegt und *nur zur nichtkommerziellen Nutzung verwendet werden darf*. Eine anderweitige Verwendung, Weitergabe oder Verwertung der Daten ist nicht gestattet und bedarf der gesonderten, schriftlichen Zustimmung des Umweltbundesamtes.

Platts

https://www.platts.com/terms

1. USE OF THE WEBSITE

Subject to these Terms of Use, Platts grants you a personal, revocable, non-exclusive, non-transferable, limited license to access and use the Website and its Content (as defined below) for the fees, if applicable, and under the terms set forth below. The Website and the content, including, but not limited to, text, data, reports, opinions, images, photos, graphics, graphs, charts, animations and video (the "Content"), displayed on the Website, *may be used only for your personal and non-commercial use and conditioned upon your compliance with this "Terms of Use" agreement*. Except as otherwise permitted under these Terms of Use, you agree not to copy, download, reproduce, modify, create derivative works from, or store any Content, in whole or in part, from the Website or to display, perform, publish, distribute, transmit, broadcast or circulate any Content to anyone, or for any public or commercial purpose, without the express prior written consent of Platts. You may not commingle any portion of the Website with any other information and you may not edit, modify, or alter any portion of the Website.

TSOs (selection)

50Hertz

Legal Disclaimer - Use of Internet pages

http://www.50hertz.com/en/Imprint

The content and design of these Internet pages are subject to copyright. Unless stipulated otherwise, *it is forbidden to reproduce these pages or their content or automated download of any kind of data* without the prior written consent of 50Hertz Transmission GmbH.

Amprion

https://www.amprion.net/Imprint.html

[No information on intellectual property right on data]

TenneT

http://www.tennettso.de/site/en/Common/disclaimer

[No information on intellectual property right on data]

TransnetBW

https://www.transnetbw.com/en/conditions-of-use

[No information on intellectual property right on data]

RTE

https://www.etalab.gouv.fr/wp-content/uploads/2014/05/Open_Licence.pdf

Licence Ouverte / Open License

You may re-use the « Information » made available by the « Producer » under the freedoms and the conditions specified by this licence.

Re-use of information under this licence

The « Producer » grants the « Re-user » a worldwide, perpetual, free of charge, non-exclusive, personal right to use the « Information » subject to this licence under the freedoms and the conditions set out below.

You are free to Re-use the « information » :

- To reproduce, copy, publish and transmit the « Information » ;
- To disseminate and redistribute the « Information » ;
- To adapt, modify, transform and extract from the « Information », for instance to build upon it in order to create « Derivative information » ;
- To exploit the « Information » commercially, for example, by combining it with other « Information », or by including it in your own product or application.

You must, where You do any of the above :

- Attribute the « Information » by acknowledging its source (at least the name of the « Producer ») and the date on which it was last updated.
  The « Re-user » may fulfil this condition by providing one or more hypertext links (URL) referring to the « Information » and effectively acknowledging its source.
  This attribution shall not suggest any official status or endorsement, by the « Producer » or any other public entity, of the « Re-user » or the re-use of the « Information ».

[…]

Compatibility of the licence

To facilitate the re-use of the « Information », this licence has been designed to be compatible with any licence which requires at least the attribution of the « Information ». For instance, it is compatible with the « Open Government Licence » (OGL) of the United Kingdom, the « Creative Commons Attribution 2.0 » (CC-BY 2.0) licence of Creative Commons and the « Open Data Commons Attribution » (ODC-BY) licence of the Open Knowledge Foundation.

[…]

About the open licence

Etalab is the task force under the French Prime Minister's authority leading Open Government Data policy for France. Etalab introduces the Open Licence to facilitate and encourage easy reuse of public sector information – as defined by French Law – free of charge and as broadly as possible,

Within the scope of their public service missions, public sector bodies produce and receive public sector information, which may be re-used by any natural or legal person.

Under the terms of French Law, are not considered public sector information : information the communication of which is not a right under information access legislation ; information contained in documents produced or received by public sector bodies exercising a public service of industrial or commercial character ; and information contained in documents over which third parties hold intellectual property rights,

Information which contains personal data is not considered to be public sector information re-usable under the terms of French Law – except where persons on which data is collected have agreed to its reuse, where this data has been rendered anonymous by the public sector bodies, or where a legal or statutory provision permits its re-use (in these three cases, re-use is subject to compliance with French privacy protection legislation).

This licence is version 1.0 of the Open Licence. Etalab may, from time to time, offer new versions of the Open Licence. However, re-users may continue to re-use information available under this licence if they wish to do so.

Energinet.dk

https://www.energidataservice.dk/Conditions_for_use_of_Danish_public_sector_data-License_for_use_of_data_in_ED.pdf

Conditions for use of danish public-sector data

1. General information

Energinet, (the 'Licensor'), which owns the intellectual property rights to data published on the Energy Data Service portal (www.energidataservice.dk) marked as being published under this license (the 'Data'), hereby grants you a license to use the Data subject to the conditions below.

You accept these conditions as soon as you use the Data in any way. These conditions are subject to regulation under Danish law.

2. Right of use

The Licensor grants you a worldwide, free, non-exclusive and otherwise unrestricted license to use the Data, which you are free to e.g.

- *copy, distribute and publish;*
- *adapt and combine with other material;*
- *exploit commercially and non-commercially.*

3. Conditions

The Data may not be used in a way which suggests that the Licensor endorses supports, recommends or markets the user of the Data, or the services or products of the user of the Data. You must ensure the use of the Data is in accordance with Danish law.

4. Acknowledgement of source

When using the Data the following may be stated: Contains data used pursuant to 'Conditions for use of Danish public-sector data'from the Energi Data Service portal (www.energidataservice.dk).

5.Rights and responsibilities of the Licensor

The Data is provided 'as is' and the Licensor shall not be liable for content, origin, errors or omissions in the Data and shall not be liable for any loss or damage caused by its use.

The Licensor provides no guarantee for the continued availability of the Data, and may at all times change the right of use of the Data and the conditions for its use.

Regelleistung.net

https://www.regelleistung.net/ext/static/imprint?lang=en

Use of this website

Contents and design of this website are protected by copyright. Reproduction of the website or parts thereof including but not limited to the contents of individual pages requires prior written permission from the German TSOs unless reproduction is authorised by law.

Netztransparenz.de

https://www.netztransparenz.de/Impressum

NUTZUNG DER INTERNET-SEITEN

Inhalt und Gestaltung der Internet-Seiten sind urheberrechtlich geschützt. Eine Vervielfältigung der Seiten oder ihrer Inhalte bedarf der vorherigen schriftlichen Zustimmung per E-Mail der deutschen Übertragungsnetzbetreiber, soweit die Vervielfältigung nicht ohnehin gesetzlich gestattet ist.

EPEX SPOT

https://www.epexspot.com/en/extras/general_conditions_website-elearning

Use of data

EPEX SPOT SE has the exclusive control on any data (in particular market data such as prices) published on its internet website. *The user of this website may not download, reproduce, license, cease or use it for commercial purposes* (inter alia for indexation purposes) without the explicit prior approval of EPEX SPOT SE. The information found on this website is *for the personal use of individual visitors only*.

BDEW

https://www.bdew.de/impressum/

Rechtshinweis samt Haftungsausschluss

Alle Informationen dienen ausschließlich zur Information der Besucher des Onlineangebotes.

AG Energiebilanzen

https://ag-energiebilanzen.de/24-0-Impressum.html

Rechtshinweise

Alle Nutzer dieser Seiten erklären ihr Einverständnis, dass die Nutzung auf eigenes Risiko erfolgt. Das Layout der Homepage, die verwendeten Grafiken sowie alle auf den Seiten enthaltenen Beiträge und Datensammlungen sind urheberrechtlich geschützt. Die Arbeitsgemeinschaft Energiebilanzen *gewährt Ihnen jedoch das Recht, den auf dieser Website bereitgestellten Daten ganz oder ausschnittsweise zu speichern und zu vervielfältigen. Die Daten dürfen nicht zum kommerziellen Gebrauch vervielfältigt oder veröffentlicht, Änderungen nicht vorgenommen werden*.

BMWi

https://www.bmwi.de/Redaktion/DE/Artikel/Service/impressum.html

[No information on copyright provided]

European Environmental Agency

https://www.eea.europa.eu/legal/copyright

Copyright notice

Unless otherwise indicated, the European Environment Agency (EEA) is the owner of copyrights and database rights in this website and its contents.

Information, documents and material available on this website and for which the EEA holds the rights of use are public and may be reused without prior permission, free of charge, for commercial or non-commercial purposes, provided that the EEA is always acknowledged as the original source of the material and that the original meaning or message of the content is not distorted. Such acknowledgment must be included in each copy of the material. The reuse of the content on the EEA website covers the reproduction, adaptation and/or distribution, irrespective of the means and/or the format used. The reuse of certain data may be subject to different conditions, and if so the item concerned is accompanied by a copyright mark or other mention of the specific conditions relating to it. The above mentioned permissions do not apply to content supplied by third parties. Therefore, for documents where the copyright lies with a third party, permission for reproduction must be obtained from the copyright holder.

The EEA reuse policy follows:

- Directive 2003/98/EC of the European Parliament and the Council on the reuse of public sector information throughout the European Union
- Commission Decision of 12 December 2011 on reuse of Commission documents
- Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the reuse of public sector information

NASA MERRA-2

https://disc.sci.gsfc.nasa.gov/citing

Data Policy

Distribution of data from the Goddard Earth Sciences Data and Information Services Center (GES DISC) is funded by NASA's Science Mission Directorate (SMD). Consistent with NASA Earth Science Data and Information Policy [see below], data from the GES DISC archive are available free to the user community.

https://science.nasa.gov/earth-science/earth-science-data/data-information-policy/

Data & Information Policy

NASA's Earth Science program was established to use the advanced technology of NASA to understand and protect our home planet by using our view from space to study the Earth system and improve prediction of Earth system change. To meet this challenge, NASA promotes the full and open sharing of all data with the research and applications communities, private industry, academia, and the general public. The greater the availability of the data, the more quickly and effectively the user communities can utilize the information to address basic Earth science questions and provide the basis for developing innovative practical applications to benefit the general public.

A common set of carefully crafted data exchange and access principles was created by the Japanese, European and U.S. International Earth Observing System (IEOS) partners during the 1990s and the early years of the 21st century. From these principles, NASA has adopted the following data policy (in this context the term 'data' includes observation data, metadata, products, information, algorithms, including scientific source code, documentation, models, images, and research results):

- NASA will plan and follow data acquisition policies that ensure the collection of long-term data sets needed to satisfy the research requirements of NASA's Earth science program.
- NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available.

[...]

ECMWF ERA5

http://apps.ecmwf.int/datasets/licences/copernicus/

Licence Agreement

I. Licence to Use Copernicus Products

1. Definitions

1.1. ‘Licensor’ means the European Union, represented by the European Centre for Medium-Range Weather Forecasts (ECMWF).

1.2. ‘Licensee’ means all natural or legal persons who agree to the terms of this Licence.

1.3. ‘Licence’ means this license agreement between the Licensor and the Licensee as amended from time to time.

1.4. ‘Copernicus Services’ means:

1.4.1. the Copernicus atmosphere monitoring service (CAMS), which is to provide information on air quality on a local, national, and European scale, and the chemical composition of the atmosphere on a global scale.

1.4.2. the Copernicus climate change service (C3S), which is to provide information to increase the knowledge base to support policies on adaptation to and mitigation of Climate Change

1.5. ‘Copernicus Products’ means all products listed in the C3S or CAMS Service Product Specification or any other items available through an ECMWF Copernicus portal, except those items which are labelled/flagged as being subject to their own separate terms of use.

1.6. ‘Intellectual Property Rights’ refers to intellectual property rights of all kinds,

1.6.1. Including: all patents; rights to inventions; copyright and related rights; moral rights; trademarks and service marks; trade names and domain names; rights in get-up; rights to goodwill or to sue for passing off or unfair competition; rights in designs; rights in computer software; database rights; rights in confidential information (including know-how and trade secrets); any other rights in the nature of intellectual property rights;

1.6.2. In each case whether registered or unregistered and including all applications (or rights to apply) for, and renewals or extensions of, such rights and all similar or equivalent rights or forms of protection which subsist or will subsist now or in the future in any part of the world together with all rights of action in relation to the infringement of any of the above.

1.7. ‘Copernicus Contractor’ refers to providers of Copernicus related goods and services to ECMWF, including Information and Data, to the Licensor and/or to the users.

1.8. ‘Copernicus Regulations’ refers to Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme.

1.9. ‘ECMWF Agreement’ refers to the agreement between the European Commission and ECMWF dated 11 November 2014 on the implementation of CAMS and C3S.

2. Introduction

Copernicus is funded under the Copernicus Regulation and operated by ECMWF under the ECMWF Agreement. Access to all Copernicus (previously known as GMES or Global Monitoring for Environment and Security) Information and Data is regulated under Regulation (EU) No 1159/2013 of the European Parliament and of the Council of 12 July 2013 on the European

Earth monitoring programme, under the ECMWF Agreement and under the European Commission's Terms and Conditions. Access to all Copernicus information is regulated under Regulation (EU) No 1159/2013 and under the ECMWF Agreement.

3. Terms of the Licence

This Licence sets out the terms for use of Copernicus Products. By agreeing to these terms the Licensee agrees to abide by all of the terms and conditions in this Licence for the use of Copernicus Products.

4. Licence Permission

4.1. This Licence is free of charge, worldwide, non-exclusive, royalty free and perpetual.

4.2. Access to Copernicus Products is given for the purpose of the following use in so far as it is lawful: reproduction; distribution; communication to the public; adaptation, modification and combination with other data and information; or any combination of the foregoing.

5. Attribution

5.1. All users of Copernicus Products must provide clear and visible attribution to the Copernicus programme. The Licensee will communicate to the public the source of the Copernicus Products by crediting the Copernicus Climate Change and Atmosphere Monitoring Services:

5.1.1. Where the Licensee communicates or distributes Copernicus Products to the public, the Licensee shall inform the recipients of the source by using the following or any similar notice:

- 'Generated using Copernicus Climate Change Service information [Year]' and/or
- 'Generated using Copernicus Atmosphere Monitoring Service information [Year]'.

5.1.2. Where the Licensee makes or contributes to a publication or distribution containing adapted or modified Copernicus Products, the Licensee shall provide the following or any similar notice:

- 'Contains modified Copernicus Climate Change Service information [Year]'; and/or
- 'Contains modified Copernicus Atmosphere Monitoring Service information [Year]'

5.1.3. Any such publication or distribution covered by clauses 5.1.1 and 5.1.2 shall state that neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus Information or Data it contains.

6. Intellectual Property Rights

6.1. All Intellectual Property Rights in the Copernicus Products belong, and will continue to belong, to the European Union.

6.2. All Intellectual Property Rights of new items created as a result of modifying or adapting the Copernicus Products through the applications and workflows accessible on the ECMWF Copernicus portals will belong to the European Union.

6.3. All other new Intellectual Property Rights created as a result of modifying or adapting the Copernicus Information will be owned by the creator.

7. Provision of Third Party Information and Data

This Licence only covers Copernicus Products. Access to third party products, information, and data related to Copernicus information to which the Licensee is directed or which can be directly accessed through any Copernicus portal will be subject to different licence terms.

8. Disclaimers

8.1. Neither the Licensor nor ECMWF warrant that Copernicus Products will be free from errors or omissions or that such errors or omissions can or will be rectified, or that the Licensee will have uninterrupted, continuous, or timely access to Copernicus Products. 8.2. The Licensor, as well as ECMWF, exclude all warranties, conditions, terms, undertakings, obligations whether express or implied by statute including but not limited to the implied warranties of satisfactory quality and fitness for a particular purpose or otherwise to the fullest extent permitted by law.

9. Liabilities

Neither the Licensor nor ECMWF will accept liability for any damage, loss whether direct, indirect or consequential resulting from the Licensee's use of the Copernicus Products.

10. Termination of and Changes to this Licence

The Licensor may terminate this licence if the Licensee breaches its obligations under these terms. The Licensor may revise this Licence at any time and will notify the Licensee of any revisions. 11. Arbitration Clause and Governing Law In the event of a dispute arising in connection with this License, the parties shall attempt to settle their differences in an amicable manner. If any dispute cannot be so settled, it shall be settled under the Rules of Conciliation and Arbitration of the International Chamber of Commerce by one arbitrator appointed in accordance with the said rules sitting in London, United Kingdom. The proceedings shall be in the English language. The right of appeal by either party to regular Courts on a question of law arising in the course of any arbitral proceedings or out of an award made in any arbitral proceedings is hereby agreed to be excluded.

It is the intention of the parties that this License shall comprehensively govern the legal relations between the parties to the Licence, without interference or contradiction by any unspecified law. However, where a matter is not specifically covered by these terms or a provision of the Licence terms is ambiguous or unclear, resolution shall be found by reference to the laws of England and Wales, including any relevant law of the European Union.

Nothing stated in this License shall be construed as a waiver of any privileges or immunities of the Licensor or of ECMWF.