Energy Strategy Reviews 27 (2020) 100433

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: http://www.elsevier.com/locate/esr

Open data for electricity modeling: Legal aspects

Lion Hirth [a,b,c]

[a] Neon Neue Energieökonomik GmbH (Neon), Germany
[b] Hertie School, Germany
[c] Mercator Research Institute on Global Commons and Climate Change (MCC), Germany

ARTICLE INFO

Keywords:
Open data
Electricity system analysis
sui generis database right
Copyright

ABSTRACT

Power system modeling is data intensive. In Europe, electricity system data is often available from sources such as statistical offices or system operators. However, it is often unclear if these data can be legally used for modeling, and in particular if such use infringes intellectual property rights. This article reviews the legal status of power system data, both as a guide for data users and for data publishers. It is based on interpretation of the law, a review of the secondary literature, an analysis of the licenses used by major data distributors, expert interviews, and a series of workshops. A core finding is that in many cases the legality of current practices is doubtful: in fact, it seems likely that modelers infringe intellectual property rights quite regularly. This is true for industry analysis but also academic researchers. A straightforward solution is open data – the idea that data can be freely used, modified, and shared by anyone for any purpose. To be open, it is not sufficient for data to be accessible free of cost, it must also come with an open data license, the most common types of which are also reviewed in this paper.

1. Introduction

Computer-based numerical modeling is ubiquitous in the energy sector. This is not only true for research, but also energy policy making, which heavily relies on model-based assessments, and the private sector, which uses numerical models for energy trading, dispatching power stations, investment decisions and due diligence.

In recent years, more and more researchers have turned to open modeling [1–3]. 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 through improved reproducibility and greater scrutiny, increase productivity by allowing reuse and collaborative development, increase credibility and legitimacy 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.

Electricity system modeling requires large quantities of input data, such as hour-by-hour load and generation data, information on existing power plants, fuel cost, and electricity prices. Such data are often provided by official or quasi-official institutions such as governments, public authorities, trade associations, system operators, and power exchanges. Most data concern fundamental, well-defined information about the power system, measured or estimated, and concern the present or the past. Efficient and robust modeling requires data not only to be of high quality, but also to be accessible without legal or technical barriers – the data need to be open. This article discusses the legal aspect of "openness" of data for electricity system modeling. In another report [4,5], we discuss technical aspects of data accessibility.

Open Knowledge Foundation [6] defines data to be open if it "can be freely used, modified, and shared by anyone for any purpose", in other words: "open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)". When it comes to electricity system 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 other sources. Often, it is combining databases into larger collections where the real value lies.
- Pass on such amended data to other users.

E-mail address: hirth@neon-energie.de.

https://doi.org/10.1016/j.esr.2019.100433
Received 12 March 2019; Received in revised form 11 October 2019; Accepted 23 November 2019
Available online 11 December 2019

- Publish original and amended data online for scientific reproducibility. This is particularly important as original sources are sometimes taken offline or are updated.

In other words, modelers require the right to copy, use, modify, distribute, or publish the data. Such activities are called "reuse" in legal jargon.[1] The right to modify and publish is essential in particular for the "downstream data ecosystem" of institutions and initiatives that collect and republish data from primary sources, catering to the specific needs of different user groups, one example being the Open Power System Data platform [7].

Unlike often assumed, "public" in the sense of "available free of charge" is not identical to "open" in the above sense. Unrestricted access, yes, is a precondition for open data, but openness includes the right to copy, use, modify and share. 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 reuse by others. In Europe, structured data can be protected as intellectual property. For electricity data, the so-called "*sui generis* database right" is the most relevant type intellectual property right. As with other property rights, it gives the data owner – within certain limitations – an exclusive right to use the property concerned. Just publishing data online without restricting access does *not* mean these rights have been waived.[2] Therefore, one can only use data legally if the rightsholder grants permission to do so. Such permission is called a "license". Open data hence means that either the rightsholder has granted a license that allows the data to be freely used, modified and shared by anyone, or that such freedoms are granted by law.

It is common practice for energy modelers to use freely-available data anyway without being aware of license conditions and the legal framework. In fact, not only users but also many data-providing intuitions I have interviewed for this research seem to be unaware of the fact that data use is legally restricted by intellectual property rights. Despite the topic being often silently ignored, given that nearly all empirical energy system research relies on publicly available data, the question if using such data is actually legal is pretty relevant.

One reason is the scarce literature on the topic. Ball [8] 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. [9] and the legal opinions by Jaeger [10,11]. Comprehensive legal background is provided by Davison [12]. The Open Data Handbook [13] provides a good overview of open data, covering many practical aspects. Yet none of these publications provides a comprehensive assessment of data used for electricity system modeling.

This paper 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. The license conditions of two dozen major data sources – ranging from statistical offices to trade associations and climate research institutes – were reviewed. While some legal uncertainty remains, it seems likely that most cases of commercial data use – which includes contract research – constitutes an infringement of intellectual property rights. This is an unfortunate situation, and probably not even intended by most of the data-providing institutions. It could be easily overcome if data were provided under open data licenses. The paper reviews the most important open data licenses and recommends a public domain dedication or an attribution license such as Creative Commons Attribution 4.0.

In a related report [4,5], we discuss technical aspects of using data, such as access options, file formats, documentation and version control. This, when well resolved, is sometimes described as "frictionless data".

The following section outlines methods and scope. Section 4 summarizes the legal framework, which is applied in the subsequent section to electricity system data. Section 6 reviews open data licenses. Section 7 discusses potential "misuse" of open data. Section 8 concludes.

2. Methods and scope

This article is based on legal analysis, expert interviews, and a series of workshops. The legal analysis included both an interpretation of the primary European and German law as well as a review of the secondary legal literature. During 2014–19, I interviewed about 20 experts in the fields of protection of databases and energy system modeling from academic institutions, think tanks, ENTSO-E, the European Commission, utilities and NGOs. During the same time, various sessions at workshops of the Open Energy Modeling Initiative on open data were conduction. In 2017 and 2018, separate expert workshops were held in Brussels and Berlin. The paper is also informed from my own experience during years of modeling. I have published my open source power market model EMMA[3] in 2013, along with all inputs as open data.[4]

The field of intellectual property rights on structured data is relatively young. Few court cases exist, and I am not aware of any litigation that concerned electricity system data.

The scope of this study is defined as follows:

- Type of data: this article covers "official" data sources, but does not cover crowd sourced data or research data, i.e. data generated by electricity system models.[5]
- Geography: intellectual property rights are very diverse across jurisdictions. It is therefore necessary to limit the scope of this paper to the European Economic Area.
- Legal scope: the data within the scope of this study is non-personal data, i.e. it does not relate to individuals. This is different in other parts of energy research, 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. Data as a trade secret is not discussed.

3. Data for electricity system modeling

This section outlines what kind of data one needs for power system modeling, and who provides such data.

3.1. Types of data

Electricity system models include a range of model types such as power market models, grid expansion planning models, and grid models.

[1] The EU Open Data Directive provides the following definition: "'re-use' means the use by persons or legal entities of documents held by public sector bodies … ".

[2] 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.

[3] See https://neon-energie.de/emma-documentation.pdf.

[4] For applications of EMMA in peer-reviewed outlets see Hirth [24]; Hirth [25]; Hirth & Steckel [26]; Hirth & Müller [27]; Hirth [28]; Hirth [29]; Hirth & Ueckerdt [30]; Hirth [31].

[5] 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. [2] on energy and Ball [8] 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.

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.

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 and other generation technologies, including installed capacity and heat rate
- Fuel costs and CO_2 certificate prices

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.

Some models and applications require significantly 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 hydropower 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.)

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 and the cartography of power infrastructure

For brevity, the above data collectively will be referred to as "electricity data".

3.2. Data sources

Electricity system modeling regularly requires obtaining input data from multiple sources, such as public authorities, grid operators, market actors, power exchanges, and trade associations. Table 1 lists a selection of important sources. This list is meant to give a flavor of data items and sources, not a comprehensive enumeration. For example, of all national statistical offices, only Germany's is included.

As evidenced by Table 1, it is not only governmental agencies that make energy data available, but also a number of commercial entities. These commercial firms usually publish data because regulation requires them to do so (section 5.4 discusses such regulations). Public sector information and information released under such statutory provisions comprise the bulk of input data required for energy system modeling.

4. The legal framework

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[6] (for creative works). Structured data can be a creative work and thereby attract copyright. In addition, a specific intellectual property right for certain structured data exists in Europe, the so called "*sui generis* database right". Both rights will be discussed in the following, as well as specific regulation governing data published by public authorities.

4.1. Intellectual property rights on databases

The EU Database Directive [14] defines the statutory framework for the legal protection of structured data in Europe.[7]

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.[8]

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 data*, 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.

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 in 1996, 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

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

[7] 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.

[8] Possibly it could be protected as a trade secret.

[9] The content of a database does not need to be data, but "a collection of independent works, data or other materials" (Database Directive). 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 might call a collection of data points a "dataset". However, that terms is neither used nor defined in the relevant law (it has been introduced, however, with the 2019 EU Open Data Directive).

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

Source	Type of data provided (selection)	Free of cost	Open license	Why?
Eurostat [40]	• 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
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 (most)	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 [38]	• 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
OpenStreetMap	• Geolocations of plants and network elements	Yes	Yes	Voluntary

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. They are discussed in turn.

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 [15]; Davidson [12] and Fisher et al. [57] recall the rationale for introducing the Database Directive.

4.2. Copyright on databases

A database attracts copyright if its selection and arrangement is sufficiently creative; it is then called an "original" database. 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 "non-original" database) does not attract copyright. As most electricity databases lack originality (see 5.1 below), copyright will not be considered further.

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 • Making available to the public	Whole or substantial part: • Extraction (copying) • Re-utilization (making available to the public)
Duration	70 years after death of author	15 years

4.3. *Sui generis database right*

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 automatically attracts a so-called "*sui generis* database right".[11] This right was newly introduced in 1996 with the Database Directive. The *sui generis* database 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 is protected under the *sui generis* database right, while not being subject to copyright.

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 Ref. [14]; Article 7), which refers primarily to online publication. In other words, re-utilization here means "redistribution and online transmission".[12]

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

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

Given that all relevant electricity databases are accessed by electronic means, the first exemption is of no practical relevance here. The second exemption is also of little relevance in the context of this paper. The third exemption allows to use and publish up to 15%[14] of a database, if research is "non-commercial". Section 4.6 will discuss this and concludes that most use for modeling is actually commercia. Summing up, the statutory exemptions do not apply to most cases that are of interest to use.

A European Court of Justice ruling in 2004 considerably restricted the scope of the *sui generis* database right.[15] 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 [15]; 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')".

4.4. *Specific regulations for the public sector*

A significant amount of electricity data is published by public sector bodies, such as federal or state-level departments, authorities and agencies (recall section 3.2). On a political level, open data strategies have been adopted such as the G8 Open Data Charter [16].

At the European level, the Open Data Directive (Directive (EU) 2019/1024),[16] regulates the reuse of data published by public sector bodies with the goal of facilitating the reuse of such data. Since 2019, Article 1 of the directive requires public sector bodies (but not public undertakings) to refrain from exercising their *sui generis* database right. Article 2 defines such bodies as "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, being regulated entities and subject to publication obligations, may qualify as public undertakings, but not as public sector bodies. The same is true for the European Network of Transmission System Operators for Electricity [36].

Article 8 of the Open Data Directive recommends the use of standard licenses. Separately from the directive, the EU Commission Guidelines on Recommended Standard Licenses (2014/C 240/01) recommends a Creative Commons Zero license for public sector information.

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

[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".

[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] For „own", non-commercial research up to 75% can be used, but cannot be published.

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

[16] Formerly called the Public Sector Information Directive [32]; amended by Ref. [33].

According to the German copyright law (§5 UrhG), official works do not enjoy copyright protection. This is likely to also apply to electricity databases: the would-be landmark case "Sächsischer Ausschreibungsdienst" at the Federal Court of Justice 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.

4.5. Licensing data

With all the complexity and ambiguity surrounding the legal protection of databases, users need clear guidance from information 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 are legal instruments for a rightsholder to permit someone to do things that would otherwise infringe on the rights held. In other words, licenses are use of data agreements; they can be one-to-one (private) or one-to-many (public) agreements. Under German legal 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 typically include 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. With power system data, they are often found as "Terms of Use" in the "Disclaimer" section of a website.

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 rights provided by copyright and the *sui generis* database right. Should those rights apply, using such data beyond statutory exemptions will constitute an infringement.

4.6. Commercial versus non-commercial use

As mentioned above in section 4.3, only non-commercial research and teaching are granted relevant statutory exemptions to use otherwise protected databases. In addition, most electricity data providers provide licenses that permit non-commercial usage only (section 5.3). This makes a precise definition of "non-commercial" imperative for energy system modeling using such data.

"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.[17] 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.

Defining "non-commercial" precisely has turned out to be notoriously difficult Klimpel [17]. 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 Ref. [18] for further discussion of the case). As a consequence, it seems likely that contract research – even that conducted by universities – would classify as "commercial use".

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.

5. An application to electricity data

This section discusses the lawfulness of using electricity data for modeling, assessing to what extent (i) IPR restricts the possibility to use electricity data for modeling, and (ii) statutory exemptions or contractual permissions would allow a modeler to do so. Finally, the consequences of publication obligations and the surprisingly difficult task of establishing ownership are touched upon.

5.1. Which IPR applies to electricity data (if any)?

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 questions:

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

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 Ref. [10].

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 copyright.[18] 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.

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 [11]. Based on public information it is often impossible to judge, hence a cautious user may necessarily presume that this is the case. In the following section 1 assume that the *sui generis* database right pertains while copyright does not.

5.2. Lawful use for modeling

Can electricity data be legally used if protected under the *sui generis* database right? The answer depends on the type of use. For this paper, we are interested in electricity data as input for modeling. Two cases are considered: (i) downloading data and using it as an input to a model; (ii) processing, reformatting, amending or repairing data and sharing it with an unrelated third party, e.g. through a public website. Any assessment of the legality of a use case takes 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.*

[17] 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.

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

- 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* provided; it cannot be answered by studying the law alone.

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 infringes the rights of the holder of the *sui generis* right if the entire or a substantial part of the database is copied. Also making more than an insubstantial amount 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. It seems likely that electricity modelers need to download substantial parts of databases, which is an exclusive right.

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.

As a consequence, for commercial energy data users, the only way to legally download data or making them available to the public is to be granted a license by the rightsholder. It will turn out that most data sources grant quite restrictive licenses at best.

5.3. Licenses used today

This section reviews the licenses offered (if any) by the data providers listed in section 3.2. The license texts themselves are provided in the Annex. 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.

Destatis. 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 6.

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 Ref. [11].

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 6). 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 50 Hz 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.

OpenStreetMap. OpenStreetMap data is licensed under the Open Data Commons Open Database License, an open data license.

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, assuming a substantial part of the database is concerned. Non-commercial personal use is allowed under all above licenses (if any are granted) – but this is covered by statutory exemptions anyway, at least in Germany.

A few commercial entities refrain from open licenses in order not to compromise their 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 database makers had themselves originally assembled their information from multiple sources.

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, substantial part)
Eurostat [40]	Custom	Yes	Yes
Statistisches Bundesamt [53]	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
OpenStreetMap	OdbL	Yes	Yes

5.4. 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.[19] This section first introduces these obligations and then discusses their relationship with the *sui generis* database right.

The Transparency Regulation [19], amending Regulation (EC) No 714/2009 [20]; 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 Refs. [4,5] for a review of the platform). Its purpose is to serve market participants, such as generators, retailers and traders, and in particular small firms. 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).

The purpose of the Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) [21] 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".

These disclosure obligations seem to conflict with the exclusive right to make the database available to the public. None of these 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 are silent on the need to license.

How are the publication obligations then related to intellectual property rights? According to my understanding, there is no legal consensus on this question. There are at least two possible interpretations:

1. Statutory limitation: One might argue that the publication obligations constitute a statutory exemption to exclusive rights granted by the *sui generis* database right. At least in Germany, this does not seem to be 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 additional statutory exemptions.
2. Obligation to license: Jaeger [11] 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 (but not make further available to the public). It also explicitly mentions commercial entities ("market actors") as users. 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 done (recall 5.3).

5.5. Establishing ownership

Only the maker of a database – the rightsholder – can grant a license. Establishing ownership – determining who holds the rights – is therefore an essential step before a license can be granted.

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 further published on yet another platform such as SMARD. In this case, it may be that multiple parties – perhaps all of the ones mentioned above – hold *sui generis* database rights. It is my 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.

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.[20]

6. The way forward: open data licenses

Above an assessment of the legal status of using data for electricity system modeling was conducted. 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, at least not broadly enough, and that (c) the licenses granted by most data sources are insufficient for most modelers. In my view, the best way forward is to provide data under open licenses.

[19] 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.

[20] 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.

Such licenses are reviewed in this section.

6.1. The principles of open data licenses

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; use data numerically in various types of computer models; and make original and amended data available to other parties. Both commercial and non-commercial use must be allowed.

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. Hence it is a one-to-many (or, actually, a one-to-any) license. Similarly, a "waiver" is a legal instrument for surrendering one's rights to a resource. Waivers release intellectual property to the public domain. Under German legal doctrine, some rights cannot be waived and the concept of public domain does not exist. A functional equivalent to waivers are public domain dedications is discussed below.

6.2. Different types of open data 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 national legal doctrines. 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".

Legal uncertainty can be greatly reduced by using one of several "standard" open data licenses. This is not only recommended by open data experts [22] but also the European Commission [23]. 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 have in common that they allow any type of use by anyone, including commercial entities, are royalty-free and are irrevocable such that they are terminated only by expiry of the licensor's intellectual property rights. Licenses that exclude commercial use are not considered open [6].

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.

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 their sources, including data, regardless of license requirements.

6.3. Creative Commons and Open Data Commons

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 kinds of content but include clarifications for databases from version 4.0 onwards. Creative Commons licenses are used by Wikipedia and Wikimedia. Open Data Commons licenses have been specifically developed for databases; 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 their use in electricity data.

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.

In line with Lämmerhirt Lämmerhirt (2017), I recommend that electricity data that are *not* protected as intellectual property 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. 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 the 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 is preferred to national licenses.

7. Misuse of data

In interviews and on workshops, data providers occasionally have reported concerns related to "loss of control" and "misuse of data" when moving to open licenses. According to my 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. The fear seems to be that this misrepresentation might negatively impact the original data source in some way, e.g. in form of press requests or negative publicity and loss of reputation.

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

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	–

misinformation can readily be addressed by basic civil and commercial law: you can prevent anyone from claiming to be you, for example. 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 to avoid confusion among readers about original versus secondary sources.

8. Conclusions and recommendations

The legal assessment of data for electricity system modeling presented in this paper suggests six conclusions. First, most relevant data is likely to fall outside copyright but does fall under the *sui generis* database right. Second, the vast majority of data sources reviewed do not provide a widely-used open data license. Most of licenses provided explicitly exclude commercial use, which includes use by for-profit companies but also in contract research. Third, as a consequence, many ways of using data for electricity system modeling are likely to infringe upon the exclusive rights of the rightsholder. Fourth, for users, it is often not obvious who that rightsholder might be. The reasons for this surprising finding is that data is often processed through a chain of institutions before being published. Fifth, not only are most researchers unaware of all this, also many data-publishing institutions seem to be. Many seem to believe that "data are provided free of charge, so anyone can use it" – but this is incorrect. Finally, a number of European and national laws mandate firms to publish data, but none of these obligations refers explicitly to intellectual property rights.

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). The situation is also a severe impediment for open modeling, since the input data used by researchers cannot be published under an open license for replication, as required by an increasing number of academic journals, including Nature Scientific Data.

Data-providing institutions can facilitate the use of electricity data. First, they should establish ownership. Data providers, including platforms that publish data supplied by different institutions, should indicate who holds intellectual property rights. Second, they should provide data under an established open license. The preferred option is a public domain dedication. An attribution license such as Creative Commons Attribution 4.0 is a good alternative. Bespoke or national licenses are not recommended.

Options for policy-makers at the state, national and European level to foster open data include the following. First, research funding should be tied to open data provision. 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. Second, public sector information should be open data. Data provided by public authorities such as ministries and agencies should be open, ideally by law. For clarification, a public domain dedication could be provided. Third, publication obligations should explicitly require open licensing if intellectual property rights apply and, for clarification, a public domain mark if they do not. Fourth, a broader intellectual property right reform might be warranted. 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. Finally, exceptions and limitations to copyright should be clarified and possibly extended. It should be made sure that exceptions for official works in national copyright law are as broadly applicable as possible and also encompass the *sui generis* database right.

Declaration of competing interest

I do not have any conflict of interest.

Acknowledgement

This research was conducted for the project "Open Source Modellierung und Open-Data für quantitative Analysen des Stromsystems im Rahmen der Energiewende" (project no. 060/17), funded by Germany's Federal Ministry for Economic Affairs and Energy and implemented by Neon in cooperation with DIW Berlin, TU Berlin, ETH Zürich, JBB, and Tom Brown. A previous version of this paper was published as "Open Data for Electricity Modeling" [4,5]. I would like to thank Ingmar Schlecht, Jonathan Mühlenpfordt, Till Jäger, John Weitzman, Robbie Morrison, Tom Brown, Juliane Reimann, Ludwig Hülk, Matteo De Felice, Frauke Braun, Ralf Uttich, Stefan Pfenninger, Wladimir Tiderko, Nils Börnsen, Felix Arnold, the participants of the project workshop for inspiring discussions and the anonymous reviewer for helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.esr.2019.100433.

References

[1] Stefan Pfenninger, DeCarolis Joseph, Lion Hirth, Sylvain Quoilin, Iain Staffell, The importance of open data and software: is energy research lagging behind? Energy Policy 101 (2017) 211–215, https://doi.org/10.1016/j.enpol.2016.11.046.
[2] Stefan Pfenninger, Lion Hirth, Ingmar Schlecht, Eva Schmid, Frauke Wiese, Tom Brown, Chris Davis, Matthew Gidden, Heidi Heinrichs, Clara Heuberger, Opening the black box of energy modeling: strategies and lessons learned, Energy Strategy Rev. 19 (2018) 63–71, https://doi.org/10.1016/j.esr.2017.12.002.
[3] Robbie Morrison, Energy system modeling: public transparency, scientific reproducibility, and open development, Energy Strategy Rev. 20 (2018) 49–63, https://doi.org/10.1016/j.esr.2017.12.010.
[4] Lion Hirth, Ingmar Schlecht, Jonathan Mühlenpfordt, Open Data for Electricity Modeling, Report on Behalf of the Federal Ministry for Economic Affairs and Energy, 2018. https://www.bmwi.de/Redaktion/EN/Publikationen/Studien/open-Data-for-electricity-modeling.html.
[5] Lion Hirth, Jonathan Mühlenpfordt, Marisa Bulkeley, "The ENTSO-E Transparency Platform – a review of Europe's most ambitious electricity data platform", Appl. Energy 225 (2018) 1054–1067, https://doi.org/10.1016/j.apenergy.2018.04.048.
[6] Open Knowledge Foundation (OKFN), The Open Definition, 2015. https://opendefinition.org.
[7] Frauke Wiese, Ingmar Schlecht, Juliane Reimann, Clemens Gerbaulet, Martin Jahn, Jonathan Mühlenpfordt, Friedrich Kunz, Wolf-Peter Schill, Casimir Lorenz, Lion Hirth, Open power system data - frictionless data for electricity system modelling, Appl. Energy 236 (2019) 401–409, https://doi.org/10.1016/j.apenergy.2018.11.097.
[8] Alex Ball, How to License Research Data, a Digital Curation Centre and JISC Legal 'working Level' Guide, 2014. http://www.dcc.ac.uk/resources/how-guides/license-research-data.
[9] Robbie Morrison, Tom Brown, Matteo De Felice, Submission on the Reuse of Public Sector Information: with an Emphasis on Energy System Datasets, 2017. https://okfn.de/files/blog/2018/01/reuse-PSI-openmod.pdf.
[10] Till Jaeger, Legal Opinion – Legal Aspects of European Energy Data, 2017. http://open-power-system-data.org/legal-opinion.
[11] Till Jaeger, 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, 2018. https://ec.europa.eu/energy/sites/ener/files/documents/legal_aspects.pdf.
[12] Mark J. Davison, The Legal Protection of Databases, Cambridge University Press, Cambridge, United Kingdom, 2008, ISBN 978-0-521-04945-0.
[13] Open Knowledge Foundation (OKFN, The Open Data Handbook: Guides, Case Studies and Resources for Government & Civil Society on the 'what, Why & How' of Open Data, 2015. http://opendatahandbook.org.
[14] 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.
[15] European Commission, First Evaluation of Directive 96/9/EC on the Legal Protection of Databases, 2005. http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf.
[16] G8 Open Data Charter. https://opendatacharter.net/g8-open-data-charter/.

[17] Paul Klimpel, Freies Wissen dank Creative-Commons-Lizenzen. Folgen, Risiken und Nebenwirkungen der Bedingung »nicht-kommerziell – NC, 2012. https://irights.info/wp-content/uploads/userfiles/CC-NC_Leitfaden_web.pdf.
[18] Till Jaeger, LG Köln beschränkt CC-BY-NC auf rein private Nutzungen, ifrOSS News, 2014, 19.03.2014, http://www.ifross.org/artikel/lg-k-ln-beschr-nkt-cc-nc-rein-private-nutzungen.
[19] 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.
[20] 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.
[21] 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.
[22] Danny Lämmerhirt, Avoiding Data Use Silos: How Governments Can Simplify the Open Licensing Landscape. Open Knowledge Foundation, Webpage, Cambridge, United Kingdom, 2017.
[23] European Commission, Commission Notice. Guidelines on Recommended Standard Licences, Datasets and Charging for the Reuse of Documents 2014/C 240/01, 2014. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2014.240.01.0001.01.ENG.
[24] Lion Hirth, What caused the drop of European electricity prices? A factor decomposition analysis, Energy J. 39 (1) (2018), https://doi.org/10.5547/01956574.39.1.lhir.
[25] Lion Hirth, The benefits of flexibility: the value of wind energy with hydropower, Appl. Energy 181 (2016) 2010–2223, https://doi.org/10.1016/j.apenergy.2016.07.039.
[26] Lion Hirth, Jan Steckel, The role of capital costs for decarbonizing the electricity sector, Environ. Res. Lett. 11 (2016), https://doi.org/10.1088/1748-9326/11/11/114010.
[27] Lion Hirth, Simon Müller, System-friendly wind power, Energy Econ. 56 (2016) 51–63, https://doi.org/10.1016/j.eneco.2016.02.016.
[28] Lion Hirth, The optimal share of variable renewables: how the variability of wind and solar power affects their welfare-optimal deployment, Energy J. 36 (1) (2015) 127–162, https://doi.org/10.5547/01956574.36.1.6.
[29] Lion Hirth, The market value of solar power: is photovoltaics cost-competitive? IET Renew. Power Gener. 9 (1) (2015) 37–45, https://doi.org/10.1049/iet-rpg.2014.0101.
[30] Lion Hirth, Falko Ueckerdt, Redistribution effects of energy and climate policy: the electricity market, Energy Policy 62 (2013) 934–947, https://doi.org/10.1016/j.enpol.2013.07.055.
[31] Lion Hirth, The Market Value of Variable Renewables: the effect of solar wind power variability on their relative price, Energy Econ. 38 (2013) 218–236, https://doi.org/10.1016/j.eneco.2013.02.004.
[32] 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.
[33] 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.
[36] ENTSO-E, Detailed Data Descriptions, vol. 4, 2014. Release, 1. https://docstore.entsoe.eu/Documents/MC%20documents/Transparency%20Platform/MOP/DetailedDescriptionDocument.pdf.
[38] European Environmental Agency (EEA), EEA Greenhouse Gas Data Viewer Manual, 2018. 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.
[40] Eurostat, Energy Balances, 2018. http://ec.europa.eu/eurostat/web/energy/data/energy-balances.
[53] Statistisches Bundesamt (Destatis), Erläuterungen zu den Volkswirtschaftlichen Gesamtrechnungen, 2018. https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Methoden/methodisches.html.
[57] R. Fisher, Julien Chicot. https://ec.europa.eu/digital-single-market/en/news/study-support-evaluation-database-directive.