# Submission on a European strategy for data with an emphasis on energy sector datasets

## European Commission public consultation closing 31 May 2020


| | |
|---|---|
| Coordinating author | Robbie Morrison <robbie.morrison@posteo.de> |
| Published | Berlin, Germany |
| Date | 30 May 2020 |


| | |
|---|---|
| Submitters (6) | Charitha Buddhika Heendeniya, Sebastian Hellmann, Michal Hetmanski, Barry McMullin, Robbie Morrison, Berit Müller |
| Areas of expertise | Energy system models, energy sector data, and energy market analysis, all with a focus on public policy development |
| Organisation | Representing ourselves as individuals (the European Commission classified us as an "informal organisation" for the purposes of a previous submission) |
| Copyright (c) | 2020 Robbie Morrison |
| License | This work is licensed under a Creative Commons Attribution 4.0 International License |
| Public release | The statement may be made public without redacting personal details |
| Errata | Corrections and updates can be found on this web forum |
| PDF file name | `morrison-european-data-strategy-submit.08.pdf` |

This statement accompanies our web-based public submission on a European strategy for data being undertaken by the Data Policy and Innovation Unit G.1 of the European Commission Directorate-General CONNECT. Submissions close on 31 May 2020. See the Commission website for further details (Anon 2020).

# 1 Summary

This submission advocates the following: 1

1. The current legal regime must be both clarified and made more supportive of open data:
   (a) determination on whether human authorship is a necessary condition for copyright
   (b) determination on whether the machine processing of a legitimately held copy of a collection of atomic data under copyright constitutes infringement
   (c) improve legislative support for the dedication of information to the public domain
   (d) amendments to the 96/9/EC database directive to better reflect the intention of lawmakers regarding thresholds and scope and guidance on what constitutes “substantial investment” and “substantial extraction” and how this information should be transmitted to users
   (e) open licensing by default for market and system information published under statutory reporting (such information usually intended to mitigate market failure and/or facilitate system security)
   (f) additionally require the above reporting to be provided in machine-readable format
   (g) correct and develop the statutory definition for “reuse” provided in the 2019/1024 open data directive (§2.11) in order to align with legal and common understandings of the term (which is currently and perversely remapped to “use”)[1]
   (h) either clarify or remove the term “primary owner of the data” from regulation 534/2013 (§2.23) which covers the statutory reporting of electricity market information
   (i) waive 96/9/EC database protection by default on public sector information
2. Extend existing statutory requirements to publish privately held data of significant public interest and require such data to be in machine-readable format.
3. Designate energy sector information as an 2019/1024 open data directive high-value dataset category and additionally specify key datasets for that new category.
4. Central and community data standards have useful and complementary roles to play.

This submission is ambivalent as to whether a new industrial data right (IDR) for commercial data transactions would be useful,[2] but stands firm on the idea that privately-held data of significant public interest must become a common resource, unencumbered and available to solve collective problems collaboratively. 2

This public interest information, together with public sector and citizen-generated information, allows all stakeholders — be they energy planners, system operators, infrastructure developers, academics, policymakers, NGOs, journalists, or the interested public — to work with the same underlying information for improved understanding, cooperation, and transparency. 3

Conversely we argue strongly that if public interest information is not easily available and licensed for reuse, much needed reforms and projects within the energy sector are less likely to gain public acceptance. Indeed, in our view, genuinely open public interest data is a necessary condition for progress on important economic, social, and environmental transformations within Europe. 4

---

[1] Both spellings of “reuse” and “re-use” appear in Commission documents. This submission adopts the former variant.

[2] We presume that “rights to use data from smart machines” in the online survey seeks a point of entry to an IDR regime. Notwithstanding, any such IDR must be compatible with current open data licenses and particularly CC BY 4.0.

## 2 Standing

We are part of the open energy modelling community and allied research communities and many of us 5
participate in the Open Energy Modelling Initiative (see section 4.2). The initiative itself is a network of individuals and is not incorporated under law. Its mailing list, established in October 2014, now numbers over 700 and its discussion forum numbers about 600. The initiative has held eleven workshops in Europe and elsewhere and these, often limited to 65 participants, attract researchers, private sector modellers, and on occasion the interested public.

The Open Energy Modelling Initiative has no process for canvassing and endorsing policy positions. Nor 6
has it legal standing. So while some of the material in this submission was discussed on initiative forums, the views expressed here are solely those of the named submitters. The European Commission classified us as an "informal organisation" for the purposes of a previous submission, Morrison *et al* (2017).

Background on individual submitters and their affiliations and open projects (with hyperlinks) follows: 7

- Charitha Buddhika Heendeniya, University of Applied Sciences, Stuttgart, Germany and Universitat Politècnica de Catalunya, Barcelona, Spain: research scientist and doctoral candidate examining local energy systems, smart grids, and district multi-energy systems modelling and analysis.
- Dr Sebastian Hellmann, Universität Leipzig, Germany: Director of DBpedia Association and the KILT Competence Center at InfAI.
- Michal Hetmanski, Instrat, Poland: researcher at Warsaw-based think tank Instrat working on coal exit and just transition policy and initiator of newly launched energy.instrat.pl open energy data platform serving information on the Polish power market and coal sector and on climate impacts.
- Prof Barry McMullin, Dublin City University, Ireland: the modelling of complex systems across multiple domains with an emphasis on deep decarbonization.
- Robbie Morrison, Berlin, Germany: former maintainer of the deeco high-resolution model and contributor to the open energy modelling community.
- Berit Müller, speaker of the Working Group Transparency at the Research Network Energy of the German Federal Ministry of Economics and managing director of the Berlin-Brandenburg section of the German Association of Solar Energy (DGS).

## 3 Introduction

This submission welcomes the development of an **European-wide strategy for data**. The focus of this 8
submission is one particular genre of data management — namely community-curated open data. And one particular application domain — namely the energy sector in Europe.

The European Union needs genuinely **open data** because treating data as a common resource and not as 9
an information good offers clear social and economic advantages. Moreover, the European Union needs to facilitate **community curation** because that model is among the most effective for establishing and maintaining useful, high-quality, diverse, traceable information.

For several reasons, the **energy sector** provides a useful case study. Reliable and affordable energy 10
supply and complete and rapid decarbonization are clear policy imperatives. The energy sector is large, turning over about one trillion euros per annum.[3] The sector contributes close to 80% of Europe's greenhouse gas emissions (European Commission 2018:57). The sector is currently at the centre of a technical

[3] Based on informal information from European Commission JRC staff. Using 2011 prices and reporting in USD, the value of total production for the energy sector as a whole for the EU28 for the year 2015 is USD 1530 billion. The individual sectors

revolution as renewables displace fossil fuels, market liberalisation proceeds, and smart technologies and architectures realise and deploy. In addition, most of these changes require **public acceptance** and particularly so if a premium for public resistance is to be avoided.[4]

The submitters have backgrounds in energy analysis and contribute regularly to the open energy mod- 11
elling community. We believe that some of our experiences in dealing with open data and community curation are material and can indicate useful ways forward.

We stress that this submission does *not* cover information governed by **personal or commercial privacy**. 12
We also want to stress also that "public" and "open" are not synonyms and that the touchstone definition for open data is that from the Open Knowledge Foundation (Open Knowledge Foundation no date).

We discuss **privately held data of public interest** in some detail because energy analysis is particularly 13
dependent on this class of information.

Without **well-resolved legislative support** and widely-understood and adopted legal instruments for 14
open data, ambiguities regarding intellectual property are normally and necessarily resolved to the benefit of automatic rights holders. This situation causes endless problems for the use and reuse of public interest datasets, where collection copyright and 96/9/EC database protection intersect to generate legal uncertainty. Under these circumstances, genuine and explicit open licensing offers the only route forward and open science proponents expend considerable effort advocating such licensing.

# 4 Background and context

We note the consultation document from the European Commission (2020a) and the accompanying fact- 15
sheet European Commission (2020b).

This submission builds on a similar submission on PSI reform as Morrison *et al* (2017). 16

This document adopts the following conventions. The term "96/9/EC database" indicates the data store in 17
question qualifies under the very wide scope of the 96/9/EC database directive. The term "open license" is taken to include public domain dedications, unless circumstances indicate otherwise (see also section 7).

## 4.1 Open science

The application of **open science** principles to the domain of **energy system analysis** has now some 18
track record. Our experiences can offer insights related to the public interest use of energy sector data. In the context of energy system analysis, namely:

- the right to reuse data is key
- the community is highly dependent on privately held data of public interest, much of it released under statutory reporting but not suitably licensed
- much of our work is policy relevant and potentially contentious
- much of our work is undertaken within risk-averse institutional environments and published in journals which require that the rights to all intellectual property be explicitly confirmed

The Commission examined "**privately held data which are of public interest**" during the recent PSI 19
reforms but no legislative measures eventuated. European Statistical System (2017) provides background

tallied comprise coal, gas, oil, refineries, and electricity. Allied sectors like biofuels are not accounted but the result is reasonable to a first order. Converting USD at an historical rate of 0.72 yields EUR 1100 billion per annum. These results were later officially published as Rey *et al* (2019) and Rey Los Santos *et al* (2018).

[4] The REMod pathways study from Fraunhofer ISE, Germany investigated the very considerable additional cost of encountering either strong resistance to the private uptake of new technologies or strong resistance to new infrastructure projects (Sterchele *et al* 2020a,b). See also an informal blog in english (Morrison 2020a).

on such data. Statutory reporting is one legal mechanism for providing such data under suitable reuse conditions and the question is considered further in section 5.3.

20 Researchers often work in **risk-averse environments** that require known provenance and proper licensing of all published information.

21 The principles of **FAIR data**, first articulated by Wilkinson *et al* (2016), provide necessary conditions for open science. But FAIR deals with data licensing only in the most general terms, noting simply that conditional permission is required. Specific licenses are not discussed, nor are any recommendations offered in this regard.

22 Burgelman *et al* (2019) review developments within the **European Union** to promote open science. We wholeheartedly support those developments.

23 One prism for assessing open science is **transparency** and the open energy modelling community has applied transparency checklists to energy systems analysis in order to reveal shortcomings and identify opportunities for improvement (Hülk *et al* 2018). Open science is additionally a precondition for open public policy analysis in the energy domain (Morrison 2018). Allied disciplines are also developing transparency policies for numerical research, for instance the American Economic Association (2019). Transparency help can build much needed public trust (Wiese *et al* 2014, Müller *et al* 2018).

24 Stodden (2009) recognised the need to apply suitable open licenses within the numerical sciences a decade back. Creative Commons later released its version 4.0 licenses in 2013 with explicit coverage of the legal rights potentially held in data and databases. We support the use of the CC BY 4.0 license in this context.

25 The key message is that open science *requires* the application of **data-capable open licenses** to data and software-capable licenses to code to enable transparency and reproducibility. Under current legislative arrangements and international treaties (including the Berne Convention, TRIPS, and the WIPO Copyright Treaty), there is no alternative but to apply public licenses to each and every work.

26 Before detailing some of the community-wide efforts to improve our digital commons, we first need to describe the underlying community a little more.

## 4.2 Open Energy Modelling Initiative

27 As noted, the Open Energy Modelling Initiative, shortened to "openmod", is an informal network of energy system modellers and analysts committed to open source software, genuinely open data, and open science and open policy analysis practices. The openmod was established five years ago in Berlin primarily by German researchers together with researchers from Denmark, Austria, and Switzerland. The community is now active in the United States and Canada, with interest from India, Africa, South America, and eastern Europe. The openmod works solely in english.

28 The openmod mailing list and discussion forum number around 700 participants and 600 participants, respectively. The openmod holds regular workshops, with the most recent being a 3-day event in Berlin in January 2020 which attracted 190 researchers. The openmod has since been holding a series of online mini-workshops during the coronavirus crisis, with three to date and a fourth in planning.

29 The openmod does not endorse individual projects, nor does it form and advocate policy positions. This is an unwritten but nonetheless negotiated community norm (Morrison 2019). This submission is therefore solely in the name of the submitters.

30 A recently example collaborative science in our community involved modellers from diverse locations working on the highly novel HELM method for solving AC power flow problems. As recorded in this recent

thread headed: The new HELM Powerflow in GridCal.

Clearly, one thing that brings the openmod community together is data. Data has been described as our 31
"social glue" (Morrison 2019). There are now community-wide efforts in the following areas, which each point expanded upon shortly.

- data semantics
- community-curated data
- distributed data architectures

But first, some comments on artificial intelligence. 32

### 4.3 Artificial intelligence

We note the AI white paper and accompanying report from the Commission (European Commission 2020c,d). 33

Machine learning blurs the traditional computer science demarcation between code and data. That clas- 34
sification also exists in established law with dedicated provisions for computer programs, collections of data, and 96/9/EC databases. Clearly, one purpose of the current consultation is to re-examine these boundaries and provisions.

Artificial intelligence techniques have not traditionally played much part in energy system analysis — 35
which has instead relied on classical methods using relatively literal numerical models and very clean data.[5] But that is starting to change — particularly for more peripheral issues like synthetic weather, demand estimation, and cataloguing system assets using remotely-sensed information (Jin *et al* 2020). It remains an open question whether machine learning will make much inroad into core analysis without the development of hybrid AI techniques, which combine physical constraints and machine learning systems.

Our community has recently begun to make contact with AI researchers and in particular the Climate 36
Change AI community.

AI naturally raises the question of machine authorship in relation to copyright. Ramalho (2018) observes 37
that "international treaties do not provide a definition of authorship" but that "an argument could be made that the international norms are crafted to cater for a human author". This is clearly a matter that could be resolved through simple legislative change.

### 4.4 Data semantics

One potentially useful example we can provide as a community involves the development of common 38
data semantics within energy system analysis. There are two parallel processes underway at present, a shared data glossary and a domain-specific ontology.

The Horizon 2020 openENTRANCE project is developing a simplified **data glossary** as one of its work 39
streams. While the Open Energy Ontology (OEO) project is developing an extensive **domain-specific ontology**. Both projects discuss their evolving work programs in order to avoid duplicated or inconsistent results.

We believe that communities of practitioners are better placed to develop and find consensus on data 40

[5] Most energy system models are sensitive to the "lightswitch effect" where one small change to the system boundary conditions, arising say from one unit of increased demand, may provoke a complete revision of the system state (Outhred and Kaye 1996). This is why clean data is required and why such extensive efforts are made to assure its provenance and accuracy.

semantics than centralised standards setting processes and organisations.[6]

### 4.5 Community curation

Two established projects are worth raising in the context of community-curated data. 41

The first is the **Open Power System Data** (OPSD) (link) project, established in 2010, which serves 42
community-curated electricity sector datasets covering Europe (Wiese *et al* 2019). Extensive efforts have been made to cleanse served data in a transparent way, with full revision histories and all processing scripts available. Much of the data contained has now been transferred to US portals without having first sought consent (see section 5.4).

The second project is the **Open Energy Platform** (link) which provides a sophisticated backend resource 43
for community database development.

An example of community curation follows. An important dataset for energy system modellers is an 44
inventory of electricity generation assets together with key engineering and financial characteristics — also known as a **power plant database**. One might imagine establishing such a list is straightforward, but it is surprisingly difficult, even for Europe. Gotzens *et al* (2019) describe the powerplantmatching python code they developed to combine disparate power plant fleet databases (vertical matching) and then compare and flag discrepancies (horizontal matching) for further processing. The module will also optionally select the median value when faced with conflicting plant attributes. We again argue that this kind of work is best undertaken by practitioners.

The power plant example is only one of many similar exercises within our community. Wikipedia has a 45
review of open energy system databases that might be of interest to the Commission.

We see community curation as occurring in parallel with centralised data provision such as the proposed 46
Common European energy data space (European Commission 2020a:22,31).

Finally to note that citizen-generated data is part of the data landscape Lämmerhirt *et al* (2018). Energy 47
analysts use this kind of data as well, often sourced from OpenStreetMap or less commonly from dedicated citizen science projects.

### 4.6 Distributed data architectures

Two decades ago, energy system models typically ran from sets of hand-edited structured text files 48
(Groscurth 1995). These files were later replaced by local SQL databases. Due to the scale, diversity, and churn of information today, **distributed data architectures** (DDA) are now increasingly required, together with support for reproducible workflows and cross-model data interoperability. Reproducible workflows not only benefit scientific research, they are also a necessary condition for transparent public policy analysis.

The technical details of these developments fall outside the scope of this submission but legal interoper- 49
ability is paramount and it is open data licenses that provide that facility. The UK-based Open Knowledge Foundation (OKF) describe open licensing as one component of their "frictionless data" concept.

A number of projects within the open energy modelling community are currently developing components 50
of this new architecture. Figure 1 indicates schematically how these various components and projects fit together. Data semantics (see section 4.4) are also an intrinsic part of this architecture.

[6] These are not the only two possibilities for setting these kinds of standards. Biddle (2020) describes how software consortia have evolved to compliment official standards setting processes. Of relevance here is the European LF Energy project seeking to develop shared interoperability protocols and open source implementations of same for smart energy systems. We should stress that the LF Energy initiative has limited overlap with energy system analysis, the latter being the subject of this submission.

The DBpedia Databus project is developing a **sophisticated communications system** — technically 51
a data bus — for managing, storing, archiving, and curating data communally so that later users can access and benefit from prior efforts. The data bus also provides a secure gateway to changeable web databases that often do not retain their histories, storing copies when required. The data bus maintains data catalogues to assist findability and notify of new changesets. These features relieve data users of many tedious and error-prone tasks, provide support for reproducible workflows, facilitate cooperation, and enable more complex data integration to occur. (Databus ongoing) (Hellmann 2019)

The Spine project is developing a **system for translating data** between data sources and energy mod- 52
elling frameworks and between different frameworks.[7] The project utilises a high-level intermediate data model with source and framework-specific translators to effect the incoming and outgoing transformations. For Douglas Adams fans, this would be a babel fish for energy models. (Spine ongoing)

Protocols and libraries to retrieve and confirm **machine-parsable license information** are also being 53
developed (DALICC ongoing). The Commission should consider supporting this work.

The key message is that the intrinsic value of data is not simply something created by data producers and 54
served to data consumers together with mechanisms for flagging and correcting errors. Rather this value can be leveraged to positively impact on societal welfare when data of public interest is truly accessible, coexists within that society (see section 4.5), and is resident within a rich technical ecosystems (this section).[8]

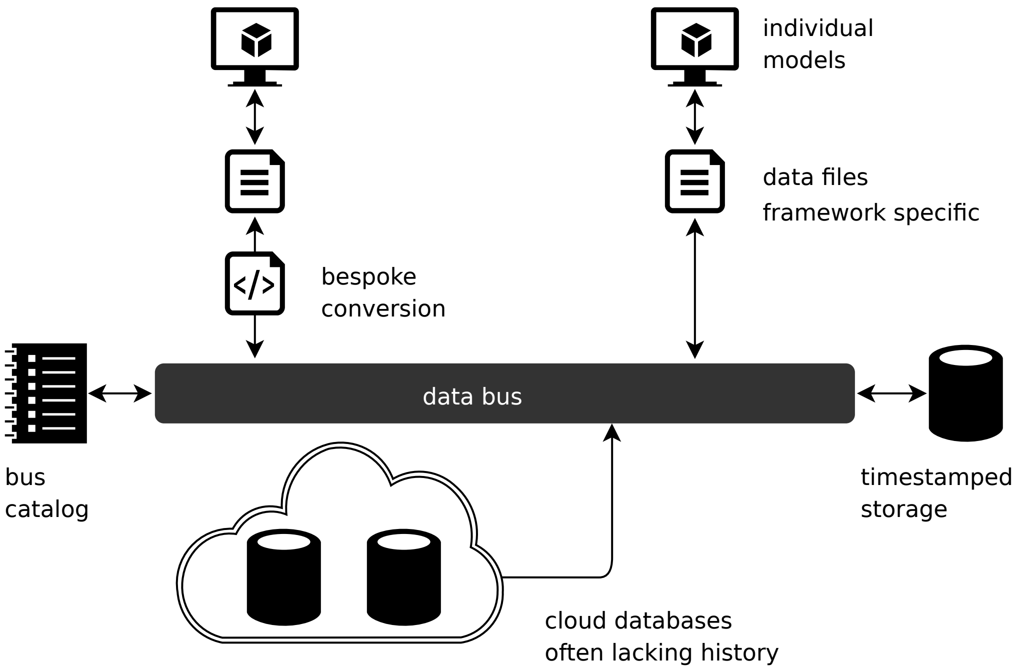


**Figure 1:** Schematic showing various components of a distributed data architecture (DDA) being developed within the open energy modelling community. Legal interoperability and shared data semantics are two essential requirements. *Credit*: Genaro Longoria and Robbie Morrison.

[7] Energy system modellers sometimes distinguish between an "energy modelling framework" and an "energy model" *per se*. The former refers to the software and development environment and the latter additionally to the data used to populate a specific instance. However, the term "model" is frequently employed to cover both concepts.

[8] Other projects not mentioned in this section but part of the open energy sector data landscape include openENTRANCE (see section 4.5), the Open Energy Platform (see section 4.5), and SENTINEL.

### 4.7 A digital commons

Many of the issues raised in this submission can be seen in context of an emerging **digital commons** 55
(Fromhold-Eisebith *et al* 2019). This WGBU report from Germany explores the relationship between digitalization and sustainability and argues that the former can assist the latter, as follows:

> According to the WBGU's understanding — and following the idea of commons in general — digital commons are digitalized goods in the fields of data, information and knowledge which, as non-rival resources, are made as broadly, that is, publicly, accessible as possible in the common interest. Examples include open education, free knowledge and open data, or digitalized natural and cultural heritage. Technically they should be provided via public-service ICT and must therefore be protected from exclusion, privatization and under-use. To achieve this, fundamental organizational, regulatory and financial decisions, for example obligations to provide information, are necessary to develop a public-welfare orientation via [a] digital commons. (p12)

We naturally support these views and believe they must underpin any future European data strategy. 56

### 4.8 Stakeholder engagement, social acceptance, and consumer cooperation

The submitters believe that open policy analysis is a precondition for **social acceptance**. And that open 57
data, in turn, is a precondition for open policy analysis. Similar sentiments were recently expressed within our community as follows (Gotzens *et al* 2019):

> As sustainability is not only limited to technically or economically feasible solutions, it requires social feasibility in aspects like justice or acceptance as well. Therefore, it is crucial for a successful energy transition to discuss different competing pathways with varying benefits for different groups in society openly with all stakeholders. This represents a highly complex task suited to be addressed by modelling exercises. (p1)

The mobility sector, in particular, needs common and accepted datasets so that stakeholders and end- 58
users can cooperate efficiently to facilitate a rapid and sensible transformation of that sector.

### 4.9 Industrial data right

The Commission white paper (European Commission 2020c:6) mentions a "new legal instrument" but 59
does not elaborate. The online questionnaire also hints at a new intellectual property right for data.

There is a growing literature on a novel industrial data right (IDR), including Kerber (2016), Wiebe (2016), 60
Duch-Brown *et al* (2017), Drexl (2017), Wiebe (2017), Negreiro (2018), Stepanov (2020). Of particular note is the European Commission Joint Research Centre (JRC) paper by Duch-Brown *et al* (2017) and a more recent paper by Stepanov (2020).

Any new IPR should be negotiated internationally. The 96/9/EC database right never took hold much be- 61
yond Europe despite several attempts to pass similar legislation in the United States (Davidson 2008:213). 96/9/EC database protection is generally viewed as a failure (refer to section 5.4 for one current example). It would be a mistake to repeat the experience of enacting European-bounded rights which can be readily ignored in other jurisdictions.

Any new IPR *must* be compatible with current open data licences, particularly the CC0 1.0 and CC BY 4.0. 62

## 5 Concrete issues

This section examines some specific issues that we would like the Commission to consider. These are real 63 issues in the sense that our community has had to confront them and variously find solutions, settle on less-than-ideal work-arounds or retreat altogether, at least for the time being.

### 5.1 PSI 'reuse' definition

The terms of public sector information reuse are contained in numerous bespoke licenses, terms of use 64
(ToU), and terms of service (ToS) but all ultimately rely on a single definition in the ODD (European Commission 2019:70) (Hirth 2020). Quoting from §2.11 and also running clauses (a) and (b) together for brevity (emphasis added):

> 're-use' means the **use by persons or legal entities** of documents held by … public sector bodies / public undertakings … for commercial or non-commercial purposes other than the initial purpose … within the public task / of providing services in the general interest …

Here we see the term "reuse" being mapped to mere "use" without further elaboration except that com- 65
mercial purposes are included. The legal concept of "use" of PSI documents has a well established meaning under copyright law, including via doctrines like fair use. Any judicial interpretation is therefore likely to rest heavily on these meanings and highly unlikely to extend to the concept of reuse as established under open licensing regimes and also as broadly understood by the lay public.

We therefore ask that the term "reuse" and its scope in the context of public sector information be clarified 66
as a matter of urgency.

### 5.2 96/9/EC database protection

The 96/9/EC database directive (European Parliament and European Council 1996) causes no end of prob- 67
lems for energy system analysts (Hirth 2020). The principle of substantiality without supporting case law is debilitating. Much of intellectual property law rests on the test of "substantial". But in the case of 96/9/EC database investment, the user has no information on which to base their usage decisions.[9] In our case, substantial investment is the problem. Yet even if there was clear jurisprudence on this matter, it cannot be known by users whether the investment in a particular database was substantial nor what would not constitute a substantial extraction. These are serious defects.

During the recent PSI reform process, the waiving of database protection for PSI was considered. Our 68
community submitted why this would be beneficial at that time (Morrison and Hirth 2018).

### 5.3 Statutory reporting

Statutory reporting in the electricity sector is governed by regulations 543/2013 (European Commission 69
2013) and 1227/2011 European Commission (2011). Such reporting is an attempt to reduce market failure and improve security outcomes. Both regulations mandate publication but are silent on licensing and not sufficiently specific on technical provision or interoperability.

Regulation 543/2013 also introduces the notion of data ownership *without* specifying which legal regime 70
applies namely §2.23 (emphasis added):

> 'primary owner of the data' means the entity which **creates the data** (p3)

Given the nature of the data covered by regulation 543/2013, this terminology is legal fiction. There is no 71
data ownership in measured or inferred power flows or in market clearance information for that matter.[10]

---

[9] It would be a distinct improvement if database protection was sought, approved, and communicated in the same way that trademarks are conditionally granted.

[10] The United Kingdom has a lower threshold for copyright than continental Europe but will not be subject to Union law following the completion of its withdrawal from the European Union, presumably this year.

In our experience, some power exchanges (PX) go to significant lengths to make their reported data 72
technically difficult to extract. This includes webserving numerical information in a form that cannot be manually copied or otherwise saved.[11] Such practices lie clearly outside the spirit, if not the letter of the regulation. Users have contacted the PXs in question, but have been unable to influence these practices. The issue needs to be addressed by the Commission as part of this consultation process.

The ENTSO-E Transparency Platform gets tangled up in 96/9/EC database protection. A lay reading of 73
recital (41) of 96/9/EC directive would suggest that it is unlikely *sui generis* rights would apply to a public database mandated under Union law (emphasis added):

> … whereas the maker of a database is the **person who takes the initiative and the risk of investing**; whereas this excludes subcontractors in particular from the definition of maker (p23)

In the case of the Transparency Platform, there is no initiative present and arguably no substantive risk 74
either. Moreover, the TSOs and other entities contributing data (often via XLS spreadsheets) could be seen in a role similar to subcontractors and therefore ineligible to share in any *sui generis* rights in any case. We therefore seek urgent clarification on these matters. That may require that the Commission seeks a declaratory judgement.

Our community has a long engagement with ENTSO-E to resolve the open licensing issues on their Trans- 75
parency Platform. We understand that there is general support for CC BY 4.0 within the organisation itself but that not all member system operators agree and that unanimous consent would be required.

It is worth noting that the French (RTE) and Finnish (Fingrid) transmission system operators webserve all 76
their public data under CC BY 4.0 or CC BY 4.0 compatible licenses.

### 5.4 Data portals located in the United States

There are large efforts in the United States to provide domain-specific information repositories. Indeed, 77
data is sometimes labelled the new oil. The World Resources Institute (WRI) is developing its Power Explorer portal with support from United States tech giants. Of particular note is that US portals do not believe that European 96/9/EC database protection extends to the United States — a legal position that has not been tested in law, nor is likely to be. This means data portals based in Europe need to comply with 96/9/EC database protection of indeterminate scope, while projects elsewhere can simply ignore these protections and extract and license content as they wish.

Indeed, some of the Transparency Platform inventory is now available on the WRI portal just mentioned 78
and licensed CC BY 4.0.[12] This is also the case for OPSD inventory too. If nothing else, these events provide an example of how localised intellectual property instruments are fraught, given the globalised online world.

## 6 High-value datasets

That the ODD does not schedule a category for energy seems a legislative oversight.[13] This should be 79
rectified by the Commission or the European Parliament.

The following suggestions were canvassed on the openmod mailing list during the week beginning 25 May 80

[11] For example: https://www.eex-transparency.com/en/power/de/production/usage/

[12] One submitter has discussed this matter with ENTSO-E's legal department.

[13] Scheduled HVDS categories currently comprise: geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility.

2020.

### 6.1 HVDS mobility

81 The future growth of electromobility poses challenges to local distribution networks that need to be considered by network planners, operators, infrastructure developers, and researchers, the key stakeholders involved in driving the growth of electromobility and the reliability of electricity supply. There are four main types of data that are critical to make robust analyses and decisions regarding the integration of large-scale electromobility into local electricity supply systems:

1. statistical data corresponding to the use of electromobility (for instance, number and type of electric vehicles registered in each district, usage statistics of public charging locations)
2. statistical data corresponding to the performance of electric vehicles, batteries, and charging technologies (for example, mileage, efficiencies, and charging/discharging patterns)
3. geospatial data covering the locations of the existing and planned vehicle charging infrastructures
4. geospatial data corresponding to the locations and layouts of existing electricity distribution networks and grid connected energy infrastructures (for instance, storage, renewable generators, and commercial and industrial loads)

82 Publicly available datasets on the above-mentioned data categories would enable network planners, operators, infrastructure developers, and researchers to understand:

- how the electric mobility future will look like
- how electric mobility will affect the local distribution network in the future
- where the potential “hot-spots” might emerge
- what are the technologically and economically feasible strategies that could support the fast growth of electric mobility while ensuring the necessary security and reliability thresholds of the distribution grid

### 6.2 HVDS electricity sector

83 Arguably, all the datasets on the OPSD platform (see section 4.5) meet the ODD requirements for support as high-value datasets (refer section 4.5). Much of the primary data derives from the ENTSO-E Transparency Platform but has been extensively curated. Notwithstanding, the question of licensing remains unresolved and HVDS status would certainly assist in this regard.

84 The OPSD platform does not cover medium and low voltage distribution networks and further consideration needs to be given to that kind of information too. This is particularly relevant for electromobility and smart consumption architectures.

## 7 Open data licensing

85 The details of open data licensing were set aside earlier. This section returns to the subject because the choice of license can be highly material.

86 Figure 2 examines the compatibilities between different data licenses. The diagram is speculative because some of the relationships are unknown and would probably require court rulings to provide definitive answers.

87 The submitters advocate just two licenses should be applied to datasets: CC BY 4.0 and CC0 1.0. This view is supported by Lämmerhirt (2017) who recommends against the ODbL and in favour of CC BY 4.0

for public sector information because the use of ODbL creates a data silo[14] — as can be seen in figure 2.

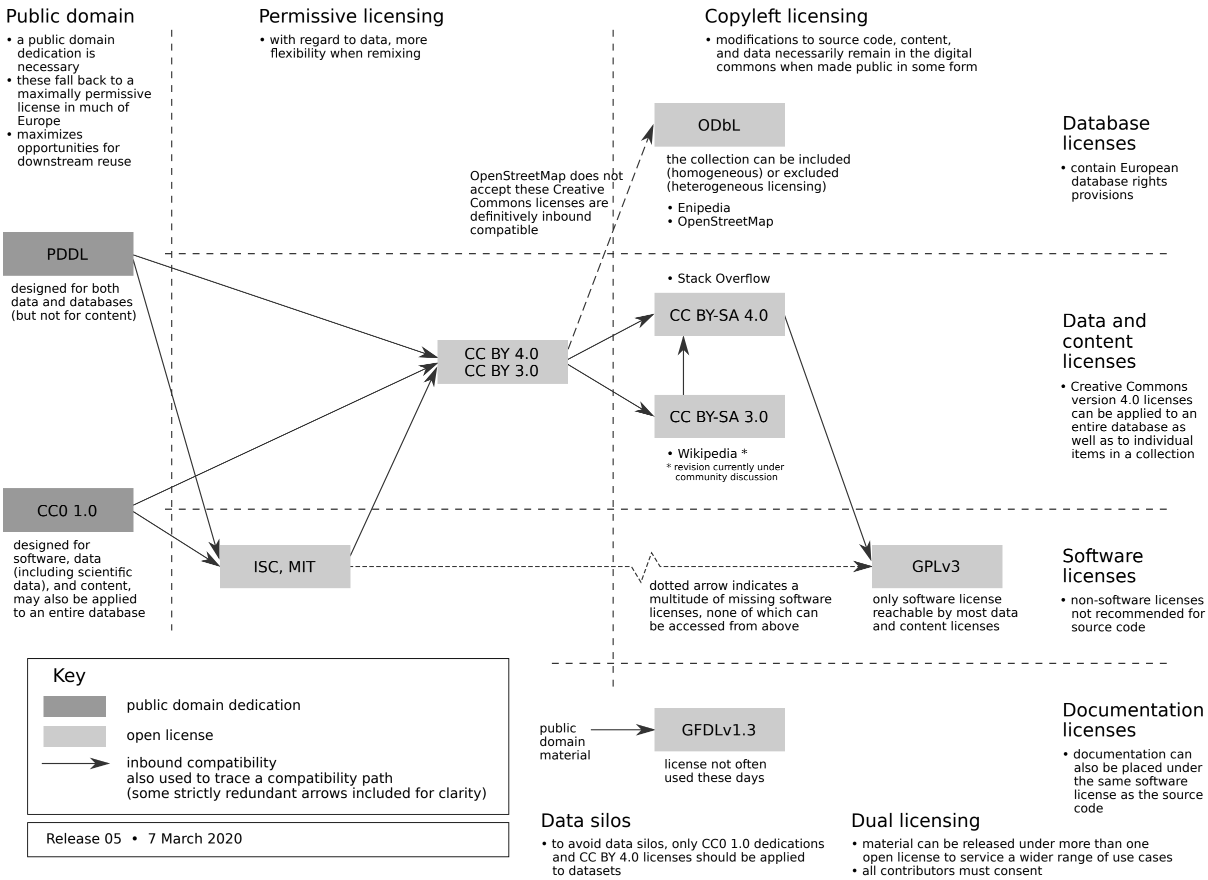


**Figure 2:** Data license compatibility chart. An inbound arrow represents an inbound compatibility. *Credit*: Robbie Morrison.

# 8 Discussion

Intellectual property law is clearly not fit-for-purpose when applied outside of traditional creative endeavours such as literature and music. While computer programs have been formally subject to copyright since 1980 in the United States and somewhat later elsewhere, it has taken enormous effort to create an open source ecosystem where developers can easily leverage from the state-of-the-art and, should they so choose, contribute back to this common heritage. Estimates vary but around 80% of code in use today is open source.[15] 88

There are now three decades of open source license development with countless expert working groups carefully crafting licenses while attempting to second guess how civil courts and copyright transgressors might react. Furthermore, during that time, the regulatory environment has only become more complex and difficult. 89

Similar institutional and structural processes are now being repeated for machine-readable data. On one hand, copyright and copyright-like protections have been extended in an attempt to establish clear 90

[14] More specifically, by not mentioning the ODC-By and ODbL licenses, we can infer that the Open Knowledge Foundation (OKF) no longer recommends their use for public sector information. It should be noted too that Lämmerhirt is employed by OKF who originally drafted the ODbL. The OKF was, for some time, known as Open Knowledge International (OKI).

[15] Assessing the economic contribution is difficult but analysis includes Eghbal (2016) and Robbins *et al* (2018).

and secure intellectual property rights for collections of data and structured databases. On the other hand, an emerging open data movement is attempting to secure a digital commons modelled on the aforementioned software commons.

Society is currently in danger of repeating the exact same tragedy for open data as occurred earlier 91
for open source software. Difficult-to-scope property rights now automatically attach to collections of data and to databases — with the European 96/9/EC database right being especially open-ended and problematic to assess.

Of particular concern is 96/9/EC database protection in relation to public sector information — this *sui* 92
*generis* right should have been waived by default with the 2019 open data directive. Furthermore, a number of legal clarifications concerning the extent of copyright resident in datasets would be straightforward to enact. Another legislative improvement would be explicit support for the elective committal of information to the public domain.[16]

In our experience, with regard to energy system analysis, open data licenses do not provide data users 93
with permissions, because collection copyright and 96/9/EC database protection only rarely attach. Rather such licenses provide data users with certainty in terms of use and reuse. In short, **an open license in this context simply confirms rather than grants openness**.

We believe the Commission should first repair existing legislation before embarking on novel intellectual 94
property rights. We outlined key issues in section 1 and elaborated on these throughout this submission.

Speaking generally, the open energy modelling community is particularly interested in working with the 95
Commission on the specification and selection of high-value datasets. We also believe that high-value datasets cannot be separated from their technical and social ecosystems and that their genuine open licensing is non-negotiable.

The submitters would be happy to provide the Commission with further information, sources, and exam- 96
ples on request.

## Acknowledgements

The submitters wish to thank Genaro Longoria, Ingmar Schlecht, Matthias Weitzel, and Andrew Wilson 97
for their input. Berit Erlach kindly proofread a final version. All errors remain the responsibility of the submitters.

The document itself was prepared using free software (Linux, bash, git, LaTeX, Inkscape) and fonts (De- 98
jaVu Sans).

[16] India has straightforward legislation in this regard, which was amended in 2012 to allow a simple public notice to effect transfer of an identified work to the public domain. See the Copyright Act (1957) §21 covering the right of the author to relinquish copyright and the Copyright Rules (2013) rule 5.2 detailing the information that a public notice must provide.

## Abbreviations

| | |
|---|---|
| API | application programming interface |
| DDA | distributed data architecture |
| ENTSO-E | European Network of Transmission System Operators for Electricity |
| FSFE | Free Software Foundation Europe |
| IDR | industrial data right (speculative) |
| HELM | holomorphic embedding load flow method |
| HVDS | high-value dataset (European Commission 2019) |
| ICT | information and communications technology |
| ODD | 2019/1024 open data directive (European Commission 2019) |
| PSI | public sector information |
| SPDX | software package data exchange |
| TP | European electricity market transparency platform (ENTSO-E 2017) |
| TSO | transmission system operator |
| UrhG | *Urheberrechtsgesetz* (the German statute covering copyright) |

## References

American Economic Association. 2019. *Data and code availability policy*. Nashville, Tennessee, USA. 10 July. Website.

Anon. 2020. *Online consultation: a European strategy for data*. Brussels, Belgium: European Commission. February. Website.

Biddle, C Bradford. 2020. Linux Foundation is eating the world. *Journal of Open Law, Technology and Society*. **11**(1). May. ISSN 2666-8106. doi 105033/jolts.v11i1.137. Open access.

Burgelman, Jean, Corina Pascu, Katarzyna Szkuta, Rene Schomberg, Athanasios Karalopoulos, Konstantinos Repanas, and Michel Schouppe. 2019. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data*. **2**: 43. 10 December. doi 10.3389/fdata.2019.00043. Open access.

DALICC. ongoing. *DALICC project*. Austria: Data Licenses Clearance Center. Website.

Databus. ongoing. *Databus: global and unified access to data*. Databus. Website.

Davidson, Mark J. 2008. *The legal protection of databases*. Cambridge, United Kingdom: Cambridge University Press. January. ISBN 978-0-521-04945-0. Paperback edition.

Drexl, Josef. 2017. Designing competitive markets for industrial data: between propertisation and access. *Journal of Intellectual Property, Information Technology and E-Commerce Law (JIPITEC)*. **8**(1): 257–292. ISSN 2190-3387.

Duch-Brown, Nestor, Bertin Martens, and Frank Mueller-Langer. 2017. *The economics of ownership, access and trade in digital data — JRC Digital Economy Working Paper 2017-01*. Seville, Spain: European Commission JRC (Joint Research Centre) Technical Reports. Document generated 2 March 2017.

Eghbal, Nadia. 2016. *Roads and bridges: the unseen labor behind our digital infrastructure*. New York, USA: Ford Foundation. July.

European Commission. 2011. Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency (text with EEA relevance). *Official Journal of the European Union*. **L 326**: 1–16. December.

European Commission. 2013. 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. *Official Journal of the European Union*. **L 163**: 1–12. 15 June.

European Commission. 2018. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, the Committee of the Regions and the European Investment Bank. A clean planet for all — a European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy — COM (2018) 773 final*. Brussels, Belgium: European Commission. November. Not ratified as of May 2020.

European Commission. 2019. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information — PE/28/2019/REV/1. *Official Journal of the European Union*. **L 172**: 56–83. June.

European Commission. 2020a. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: a European strategy for data — COM (2020) 66 final*. Brussels, Belgium: European Commission. 19 February.

European Commission. 2020b. *Factsheet: the European data strategy*. Brussels, Belgium: European Commission. 19 February.

European Commission. 2020c. *On artificial intelligence: a European approach to excellence and trust — White paper — COM (2020) 65 final*. Brussels, Belgium: European Commission. February.

European Commission. 2020d. *Report on the safety and liability implications of artificial intelligence, the internet of things and robotics — COM (2020) 64 final*. Brussels, Belgium: European Commission. February. Report accompanying white paper.

European Parliament and European Council. 1996. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. *Official Journal of the European Union*. **L 77**: 20–28. 27 March.

European Statistical System. 2017. *Position paper on access to privately held data which are of public interest: opening up new data sources for a new generation of official statistics: in light of the growing European Digital Single Market and the revision of the Public Sector Information Directive*. Luxembourg: European Statistical System (ESS). November.

Fromhold-Eisebith, Martina, Ulrike Grote, Ellen Matthies, Dirk Messner, Karen Pittel, Hans Joachim Schellnhuber, Ina Schieferdecker, Sabine Schlacke, and Uwe Schneidewind. 2019. *Digital momentum for the UN Sustainability Agenda in the 21st century — Policy paper 10*. Berlin, Germany: German Advisory Council on Global Change (WBGU). July. ISBN 978-3-946830-23-8. German version also.

Gotzens, Fabian, Heidi Heinrichs, Jonas Hörsch, and Fabian Hofmann. 2019. Performing energy modelling exercises in a transparent way: the issue of data quality in power plant databases. *Energy Strategy Reviews*. **23**: 1–12. January. ISSN 2211-467X. doi 10.1016/j.esr.2018.11.004.

Groscurth, Helmuth-M. 1995. Design and management of energy databases. *Energy Sources*. **17**(4): 445–457. ISSN 0090-8312. doi 10.1080/00908319508946093.

Hellmann, Sebastian. 2019. *DBpedia's Databus and strategic initiative to facilitate "1 billion derived knowledge graphs by and for consumers" until 2025*. Leipzig, Germany: DBpedia. September.

Hirth, Lion. 2020. Open data for electricity modeling: legal aspects. *Energy Strategy Reviews*. **27**: 100433. January. ISSN 2211-467X. doi 10.1016/j.esr.2019.100433.

Hülk, Ludwig, Berit Müller, Martin Glauer, Elisa Förster, and Birgit Schachler. 2018. Transparency, reproducibility, and quality of energy system analyses: a process to improve scientific work. *Energy Strategy Reviews*. **22**: 264–269. 1 November. ISSN 2211-467X. doi 10.1016/j.esr.2018.08.014. Open access.

Jin, Donghan, Raffaella Ocone, Kui Jiao, and Jin Xuan. 2020. Energy and AI. *Energy and AI*. **1**: 100002. August. ISSN 2666-5468. doi 10.1016/j.egyai.2020.100002. First issue statement. Open access.

Kerber, Wolfgang. 2016. *A new (intellectual) property right for non-personal data?: an economic analysis — Joint discussion paper 37-2016*. Marburg, Germany: Philipps-University Marburg. September. Forthcoming in Gewerblicher Rechtsschutz und Urheberrecht. Internationaler Teil (GRUR Int), 11/2016.

Lämmerhirt, Danny. 2017. *Avoiding data use silos: how governments can simplify the open licensing landscape*. Cambridge, United Kingdom: Open Knowledge International. December. Webpage. Organization now named Open Knowledge Foundation.

Lämmerhirt, Danny, Jonathan Gray, Tommaso Venturini, and Axel Meunier. 2018. *Advancing sustainability together? citizen-generated data and the Sustainable Development Goals*. Global Partnership for Sustainable Development Data, Open Knowledge International, Public Data Lab. December.

ENTSO-E. 2017. *ENTSO-E transparency platform*. Belgium, Brussels. Website.

Morrison, Robbie. 2018. Energy system modeling: public transparency, scientific reproducibility, and open development. *Energy Strategy Reviews*. **20**: 49–63. April. ISSN 2211-467X. doi 10.1016/j.esr.2017.12.010. Open access.

Morrison, Robbie. 2019. *An open energy system modeling community*. Hannover, Germany: Leibniz Research Alliance Open Science. November. doi 10.25815/ff3b-d154. Blog.

Morrison, Robbie. 2020a. *REMod pathways study from Fraunhofer ISE, Germany*. Germany: Open Energy Modelling Initiative. February. Posting.

Morrison, Robbie. 2020b. *Submission on a European strategy for data with an emphasis on energy sector datasets — Release 08*. Berlin, Germany. 30 May. A reference to the current release this document.

Morrison, Robbie, Tom Brown, and Matteo De Felice. 2017. *Submission on the re-use of public sector information: with an emphasis on energy system datasets — Release 09*. Berlin, Germany. 10 December. A reference to release 09 this document.

Morrison, Robbie and Lion Hirth. 2018. *Follow-up to the public consultation on the re-use of public sector information: issues surrounding sui generis database rights — Release 03*. Berlin, Germany. February.

Müller, Berit, Jens Weibezahn, and Frauke Wiese. 2018. Energy modelling: a quest for a more open and transparent approach. *European Energy Journal*. **8**(2): 18–24. December. ISSN 2211-9175.

Negreiro, Mar. 2018. *Free flow of non-personal data in the European Union — Briefing — PE 614.628*. Brussels, Belgium: European Parliamentary Research Service (EPRS). September.

Open Energy Modelling Initiative. ongoing. *openmod forum*. Open Energy Modelling Initiative. Website.

Open Knowledge Foundation. no date. *Open Definition 2.1 — Defining open in open data, open content and open knowledge*. Cambridge, United Kingdom: Open Knowledge Foundation (OKF). Website.

Outhred, Hugh R and R John Kaye. 1996. Incorporating network effects in a competitive electricity industry: an Australian perspective. Chapter 9, pp. 207–228. In: *Electricity transmission pricing and technology*. Einhorn, Michael A and Riaz Siddiqi (editors). Boston, Massachusetts, USA: Kluwer. ISBN 978-0-7923-9643-7.

Ramalho, Ana. 2018. *Ex machina, ex auctore? Machines that create and how EU copyright law views them*. Alphen aan den Rijn, the Netherlands. November. Website.

Rey, Luis, Krzysztof Wojtowicz, Bert Saveyn, Matthias Weitzel, Toon Vandyck, Marie Tamba, and Umed Temursho. 2019. *Baseline GECO 11/2018*. Brussels, Belgium: European Commission Joint Research Center (JRC). 14 January. PID: 3ffc59a1-edff-491f-8894-3147d2202e42. GECO is Global Energy and Climate Outlook.

Rey Los Santos, Luis, Krzysztof Wojtowicz, Marie Tamba, Toon Vandyck, Matthias Weitzel, Bert Saveyn, and Umed Temursho. 2018. *Global macroeconomic balances for mid-century climate analyses — Supplementary material to global energy and climate outlook 2018 — EUR 29490 EN*. Luxembourg: Publications Office of the European Union. ISBN 978-92-79-98168-5. doi 10.2760/858469. JRC113981.

Robbins, Carol A, Gizem Korkmaz, José Bayoán Santiago Calderón, Daniel Chen, Claire Kelling, Stephanie Shipp, and Sallie Keller. 2018. *Open source software as intangible capital: measuring the cost and impact of free digital tools — Preliminary Draft*. October.

Spine. ongoing. *Spine*. Espoo, Finland: VTT Technical Research Centre of Finland. Website.

Stepanov, Ivan. 2020. Introducing a property right over data in the EU: the data producer's right — an evaluation. *International Review of Law, Computers and Technology*. **34**(1): 65–86. January. ISSN 1360-0869. doi 10.1080/13600869.2019.1631621. Open access.

Sterchele, Philip, Julian Brandes, Judith Heilig, Danien Wrede, Christoph Kost, Thomas Schlegl, Andreas Bett, and Hans-Martin Henning. 2020a. *Wege zu einem Klimaneutralen Energiesysem: Die deutsche Energiewende im Kontext gesellschaftlicher Verhaltensweisen*. Freiburg, Germany: Fraunhofer ISE. February.

Sterchele, Philip, Julian Brandes, Judith Heilig, Danien Wrede, Christoph Kost, Thomas Schlegl, Andreas Bett, and Hans-Martin Henning. 2020b. *Wege zu einem Klimaneutralen Energiesysem: Die deutsche Energiewende im Kontext gesellschaftlicher Verhaltensweisen — Anhang zur Studie*. Freiburg, Germany: Fraunhofer ISE. February. Supplementary material.

Stodden, Victoria. 2009. The legal framework for reproducible scientific research: licensing and copyright. *Computing in Science Engineering*. **11**(1): 35–40. January. ISSN 1521-9615. doi 10.1109/MCSE.2009.19.

Wiebe, Andreas. 2016. *Who owns non-personal data?: legal aspects and challenges related to the creation of new "industrial data rights" — Presentation*. Göttingen, Germany: University of Goettingen. Slides presented at the GRUR conference on data ownership, Brussels.

Wiebe, Andreas. 2017. Protection of industrial data: a new property right for the digital economy? *Journal of Intellectual Property Law and Practice*. **12**(1): 62–71. January. ISSN 1747-1532. doi 10.1093/jiplp/jpw175.

Wiese, Frauke, Gesine Bökenkamp, Clemens Wingenbach, and Olav Hohmeyer. 2014. An open source energy system simulation model as an instrument for public participation in the development of strategies for a sustainable future. *Wiley Interdisciplinary Reviews: Energy and Environment*. **3**(5): 490–504. September. ISSN 2041-840X. doi 10.1002/wene.109.

Wiese, Frauke, Ingmar Schlecht, Wolf-Dieter Bunke, Clemens Gerbaulet, Lion Hirth, Martin Jahn, Friedrich Kunz, Casimir Lorenz, Jonathan Mühlenpfordt, Juliane Reimann, and Wolf-Peter Schill. 2019. Open Power System Data: frictionless data for electricity system modelling. *Applied Energy*. **236**: 401–409. February. ISSN 0306-2619. doi 10.1016/j.apenergy.2018.11.097.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie

Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair JG Gray, Paul Groth, Carole Goble, Jeffrey S Grethe, and Jaap Heringa. 2016. The FAIR Guiding Principles for scientific data management and stewardship — Comment. *Scientific Data*. **3**: 160018. March. doi 10.1038/sdata.2016.18. Open access.

▢