Submission on the re-use of public sector information : with an emphasis on energy system datasets

Feedback on European Commission public consultation on directives 2003/98/EC and 2013/37/EU closing 12 December 2017

Coordinating author	Robbie Morrison <robbie.morrison@posteo.de>
Contributing authors	Tom Brown, Matteo De Felice
Submitters (39)	Raik Becker, Tom Brown, Matteo De Felice, Ulrich Frey, Clemens Gerbaulet, Graeme Hawker, Heidi Ursula Heinrichs, Philip Hiersemenzel, Simon Hilpert, Lion Hirth, Mathias Hofmann, Jonas Hörsch, Ludwig Hülk, Daniel Huppmann, Martin Jahn, Martin Klein, Hanns Koenig, Jochen Linssen, Casimir Lorenz, Barry McMullin, Craig Morris, Robbie Morrison, Berit Müller, Stefan Pfenninger, Sylvain Quoilin, Martin Robinius, Jonas Savelsberg, Ingmar Schlecht, Eva Schmid, Kais Siala, Konstantinos Syranidis, Tim Tröndle, Jens Weibezahn, Frauke Wiese, Clemens Wingenbach, Nikolas Wölfing, Jarad Wright, Florian Ziel, Thorsten Zoerner
Areas of expertise	Energy system models, energy sector databases, energy market analysis, all with a focus on public policy development
Organisation	Representing ourselves as individuals

Published	Berlin, Germany
Date	10 December 2017

Errata	Corrections and updates can be found on this web forum
PDF file name	`reuse-public-info-submit-openmod.09.pdf`

This statement accompanies our web-based public submission on the review of the directive on the re-use of public sector information (PSI directive) being undertaken by the European Commission Directorate-General for Communications Networks, Content, and Technology. Submissions closed on 12 December 2017. See the Commission website for details (European Commission 2017d).

1 Summary

We 39 submitters are from the open energy modelling community and allied research communities. Our 1
interest in this submission concerns energy sector datasets, be they public or private with public interest.

We comprise researchers and analysts from universities, research institutes, firms, consultancies, and the 2
interested public. Many of us build computer models of energy systems and run scenarios to assess how future energy systems might look and perform. Others undertake the econometric analysis of energy markets. We favour "open" methods wherever possible. These methods originate from open source software development and we are currently adapting them to suit our needs.

Energy system models require well-structured, high-quality, curated datasets. These datasets include 3
tables recording asset inventories and time series that describe weather, demand, and market status. Geolocation is of growing importance as energy models increase in spatial resolution. To this end:

- the legal context of public datasets must allow for both scientific research and open development
- data quality remains a key issue and some means by which missing or spurious data points can be flagged and fixed at their point of publication should be promoted and perhaps even mandated

More specifically, we are concerned with: 4

- copyright overreach and the unclear legal status of public sector datasets more generally
- clarification over the right to machine process lawfully obtained copyright-protected datasets
- republication of original, modified, and mixed-origin datasets
- data longevity and the right to archive
- privately held data of public interest, namely market information and engineering characterisations
- copyright law reform

Indeed, many of the datasets we require for energy modelling and analysis are either not available or 5
of poor quality (despite mandatory transparency requirements) or are ambiguously licensed or implicitly protected. In these latter cases, it may well not be lawful to use, repair, combine, and/or redistribute this data. The legal basis for circulating remixed datasets is especially problematic. We deal with such issues on a regular basis and will later recount our experiences to highlight the difficulties we face. The use of open data licenses and public domain dedications can alleviate many of these issues.

There has been a huge duplication of effort by energy system research teams preparing data for their 6
particular analysis and modelling. This time is closing and we will describe some notable open database projects within our domain which demonstrate the returns that derive from building common databanks.

While open licensing can address many of the issues we encounter, we believe that the law on copyright 7
urgently needs to be overhauled to deal effectively with public sector information (PSI), digital data, and the growing internet-mediated information commons.

We note that the European Commission has been moving towards transparent energy system models 8
(POTEnCIA), a comprehensive public energy database (IDEES), and open energy model development (Dispa-SET) and away from closed or protected models (PRIMES).

We use the term *open* and not *public* throughout much of this document to indicate that the information 9
under discussion must be able to be freely viewed, used, modified, and republished.[1] The right to inspect alone is not sufficient for our needs. Only genuine openness can guarantee research reproducibility and avoid the needless duplication of effort as each project necessarily assembles its own individual databank.

[1] These criteria are known as the "four freedoms". First articulated for software in 1986, they now apply to content and data.

2 Recommendations

We offer the following recommendations for consideration by the European Commission: 10

1. That a concerted effort be made to resolve the legal uncertainty surrounding data, datasets, and databases, by commissioning legal opinions, seeking declaratory judgements, and developing and proposing legislative changes. We note the copyright reform process (as described in European Commission 2016) is current and believe that positive PSI re-user rights under law would provide a better solution than that currently offered by open licensing (section 6.4).[2]

2. More generally, the Commission should create the requirement, either through policy or mandate, that all public sector information (in the absence of copyright law reform) be licensed under an established permissive license (CC BY 4.0) or public domain dedication (CC0 1.0) as appropriate (section 6.4).[3]

3. More specifically, that the Commission investigates standard systems for the assessment and correction of public sector datasets by users and issues good practice guidelines (section 7).

4. More specifically, that the Commission commissions a legal opinion on the machine-processing of copyright-protected digital datasets (sections 6.2 and 6.3).

5. More specifically, we believe that copyright overreach is a significant problem in our domain and we ask that the Commission develops and publishes clear guidelines on the circumstances under which routine datasets are copyrightable (section 6.6).

6. More specifically, article 13 of the proposed new copyright directive should be revised to account for the technical needs of open code and data hosting services located within Europe (section 6.10).

7. The European Commission Joint Research Centre (JRC) needs to finalise its policies on energy sector data licensing and metadata practices (section 6.9).

In addition, we offer the following measures in relation to the European electricity market transparency directive 543/2013 (sections 6.8 and 7.1). We ask that the Commission: 11

8. Creates a legal requirement that all data collected by a transparency platform be licensed under an established permissive license (CC BY 4.0) or public domain dedication (CC0 1.0) as appropriate (section 6.4) and that the terms of service and any website disclaimer align with the transparency directive (section 6.8).

9. Because much of the data on the electricity market transparency platform is provided to ENTSO-E by upstream third parties, create, if necessary, and enforce a binding obligation that such parties actually forward accurate, complete, and timely information (section 7.1).

10. Expands the scope of the current transparency regulation to include other relevant electricity sector datasets, such as power plant coordinates, historical market data, and engineering information on transmission substations, power lines, and similar assets.

11. Extends the minimum provision of datasets from the current five years (3§1) to five decades and make this obligation transferable and binding in the event of organisational restructuring.

12. Investigates the establishment of transparency platforms for energy commodities other than electricity, including but not limited to heat, gas, liquid fuels, and solid fuels.

[2] Tsiavos (2013:5–6) offers his view on the kinds of legislative change that could be useful.

[3] Our recommendation is more strongly worded than that found in re-use directive 2013/37/EU (European Commission 2013b:4) recital (26), namely that: "Member States should encourage the use of open licences".

3 Standing

We are part of the open energy modelling community and allied research communities and many of us participate in the Open Energy Modelling Initiative. The initiative itself is a network of individuals and is not incorporated under law. Its mailing list, established in October 2014, now numbers over 400. The initiative has held seven workshops in Europe and these, typically limited to 65 participants, attract researchers, private sector modellers, and on occasion the interested public. 12

The Open Energy Modelling Initiative has no process for canvassing and endorsing policy positions. Nor 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. 13

Background on individual submitters and their affiliations and open projects (with hyperlinks) follows:[4] 14

- Dr Raik Becker, researcher, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany: renewable energy integration and power system modelling.
- Dr Tom Brown, post-doctoral researcher, Frankfurt Institute for Advanced Studies (FIAS), University of Frankfurt: lead researcher for the PyPSA power system model.
- Dr Matteo De Felice, staff scientist, Climate Modelling Laboratory, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA): statistical modelling and machine learning techniques as applied to power systems and renewable generation.
- PD Dr Dr Ulrich Frey, researcher, German Aerospace Center (DLR): energy systems analysis.
- Dr Clemens Gerbaulet, post-doctoral researcher, Technische Universität Berlin: developer of the electricity sector model dynELMOD.
- Graeme Hawker, researcher, University of Strathclyde: energy systems analysis in association with the UK Energy Research Centre (UKERC).
- Dr-Ing Heidi Ursula Heinrichs, post-doctoral researcher, Department Process and Systems Analysis, Forschungszentrum Jülich (FZJ), Germany: energy system analysis and experienced with several energy models of differing levels of openness.
- Philip Hiersemenzel, strategic advisor, Younicos, Lumenion, and GRIPS Energy, Berlin: integrated energy and storage systems analysis and design.
- Simon Hilpert, analyst, University of Flensburg and energiekollektiv, Germany: energy system and data analysis and developer and user of the oemof model framework.
- Prof Dr Lion Hirth, Neon Neue Energieökonomik, Berlin, Hertie School of Governance, Germany, and Mercator Research Institute on Global Commons and Climate Change (MCC), Germany: author of the EMMA model and project leader for the OPSD database project.
- Mathias Hofmann, researcher, Technische Universität Berlin: efficient energy conversion and operation, including the use of cogeneration and thermal storage.
- Jonas Hörsch, doctoral researcher, Frankfurt Institute for Advanced Studies (FIAS), University of Frankfurt: developer and maintainer of the PyPSA power system model, being applied to investigate deep decarbonisation pathways for Europe.

[4] One submitter is resident outside the European Union. This is acceptable under Commission consultation guidelines.

- Ludwig Hülk, Reiner Lemoine Institute, Berlin: programmer, maintainer, user, and contributor to the OpenEnergy Database (oedb) as part the OEP (OpenEnergy Platform).
- Dr Daniel Huppmann, post-doctoral research scholar, International Institute for Applied Systems Analysis (IIASA), Austria: developer of the MESSAGEix integrated assessment model.
- Martin Jahn, researcher, University of Flensburg: developer and maintainer for the OPSD (Open Power System Data) database project.
- Martin Klein, doctoral candidate, German Aerospace Center (DLR): research on energy systems analysis and future energy scenarios.
- Hanns Koenig, Aurora Energy Research, Berlin, Germany and Oxford, United Kingdom: project leader and energy market analyst.
- Jochen Linssen, Head of Research Group "sector coupling", Department of Process and Systems Analysis, Forschungszentrum Jülich (FZJ), Germany: assessment of energy infrastructures.
- Dr Casimir Lorenz, research associate, Technische Universität Berlin, German Institute for Economic Research Berlin (DIW), and the German Advisory Council on the Environment (SRU): a research focus on climate and energy.
- Prof Barry McMullin, Dublin City University, Ireland: the modelling of complex systems across multiple domains with an emphasis on deep decarbonisation.
- Craig Morris, senior fellow, Institute for Advanced Sustainability Studies (IASS), Potsdam, Germany and author of Morris and Jungjohann (2016) on the German *Energiewende*.
- Robbie Morrison, Berlin, Germany: former maintainer of the deeco high-resolution model and voluntary contributor to the GENESYS2 project.
- Berit Müller, head of research group "transformation of energy systems", Reiner Lemoine Institute, Berlin: oversees several open source modelling and open data projects including open_eGo and open_Fred and a member of the oemof developer group.
- Dr Stefan Pfenninger, researcher, Department of Environmental Systems Science, ETH Zurich, Switzerland: project leader for the Calliope energy model and the Renewables.ninja data platform.
- Prof Sylvain Quoilin, University of Liège and KU Leuven, both Belgium: smart energy systems research and main developer of the open source Dispa-SET electricity sector model.
- Dr-Ing Martin Robinius, Head of Process and Systems Analysis Department, Forschungszentrum Jülich (FZJ), Germany: research into power-to-gas, fuel cell cars, and renewable energy sources.
- Jonas Savelsberg, doctoral student, University of Basel, Switzerland: the modelling of Swiss and European electricity markets.
- Dr Ingmar Schlecht, postdoctoral researcher, University of Basel, Switzerland: developer of the Swissmod electricity market model.
- Dr Eva Schmid, policy advisor, Germanwatch, Berlin and former modeller assessing the infrastructure requirements for the European energy transition: power grids and low-carbon policy.
- Kais Siala, doctoral student, Technical University of Munich (TUM): energy modeller and co-maintainer of the open source urbs electricity sector model.
- Konstantinos Syranidis, doctoral candidate, Jülich Research Center, Germany: modelling power-to-X potentials for Europe.

- Tim Tröndle, doctoral student, ETH Zurich, Switzerland: modelling and analyses of the European power system and a research interest in computational reproducibility.
- Jens Weibezahn, doctoral student, Technische Universität Berlin: energy sector modelling with sector coupling.
- Dr Frauke Wiese, post-doctoral researcher, Systems Analysis, Danish Technical University (DTU): author of the renpass model, co-initiator of the OPSD database project, and currently developing the Balmorel model.
- Clemens Wingenbach, analyst, University of Flensburg and energiekollektiv, Germany: energy system and data analysis and developer and user of the oemof model framework.
- Dr Nikolas Wölfing, researcher, Centre for European Economic Research (ZEW), Mannheim, Germany: econometric analysis of European energy markets.
- Jarad Wright, principal engineer, Council for Scientific and Industrial Research (CSIR), South Africa: power system markets.
- Prof Dr Florian Ziel, University of Duisburg-Essen, Germany: modelling and forecasting of energy systems and markets.
- Thorsten Zoerner, co-founder, Stromdao, Berlin: StromDAO blockchain-supported energy market communication and clearing software.

4 Introduction

The remainder of this document provides background to the points made in the summary (section 1) and 15
list of recommendations (section 2). It is structured conventionally and bookended by an introduction and a discussion.

Each year at least one trillion euros are spent on the supply of energy within Europe, with more still on 16
allied services like energy efficiency.[5] These systems are in the midst of a technological watershed and will need to be carbon free and substantially renewable by 2050.[6] This energy transition affects all aspects of national and social life. And transparency is vital to both the science behind this European *Energiewende* and to build the necessary public trust.

The energy models many of us develop and apply are primarily used to inform public policy. They vary 17
in their level of technical, market, and economic detail, their scope in terms of sector coverage, and the time horizons they span. Some address short-term issues like market design and system operations, while others investigate low-carbon energy systems in the mid-future. Detailed modelling is required to understand how Europe can collectively meet its greenhouse gas reduction targets while ensuring that energy-services remain affordable, market integration continues, and security of supply is maintained.

Open energy system modelling is a recent phenomena, with the first open model (Balmorel) making public 18
its source code in 2001. A mid-2017 survey shows that there are now 28 such projects visible worldwide, with 21 located in Europe and 11 in Germany (Morrison under review:§5). Open development offers the

[5] Based on a gross inland energy consumption of 75 EJ/a and a conservative unit price of 0.05 €/kWh (eurostat 2016).

[6] Integrated assessment modelling indicates the European energy sector will need to be fully carbon neutral by 2050 to offset continuing emissions from agriculture and other less tractable sectors. In relation to renewables share, the target for Germany is a 50 % reduction in primary energy consumption (base year 2008) by 2050 (BMWi 2015:4).

prospect of research networks forming organically around common codebases and databanks without the need for science funders to orchestrate cross-institutional collaborations.

Open data only became an issue with the advent of open modelling. Under the closed modelling paradigm, 19
research teams need not reveal their input or output datasets, just their post-processed interpretations.

Open energy system modelling is somewhat unusual because it assumes an open development ethos, 20
spans applied science and public policy, and is highly reliant on *privately held* data. For instance, the spot price and dispatch schedules determined by power exchanges (PX) have, until recently, been restricted to market participants or sold on as a commodity. That situation has fortunately changed with the launch of the ENTSO-E electricity market transparency platform (TP) in January 2015 (ENTSO-E 2015:11). For energy system modellers and market analysts, the TP is a remarkable step forward, although important and possibly precedent setting legal (section 6.8) and technical issues (section 7.1) remain unresolved.

We want to stress the importance of transparency in the public discourse on energy policy and prac- 21
tice, particularly given the current climate of distrust in experts. For example, for the last half decade, German electricity transmission system operators (TSOs) have consulted widely on their national Grid Development Plan (German TSOs 2017) and in doing so have increased public acceptance for new grid infrastructure significantly (Steinbach 2013:226–228).

Open energy data, together with open software and open toolchains, also allows citizens to build energy 22
system models and define and run their own scenarios to counter incumbent positions. That may sound fanciful, but the SEN non-governmental organisation in Western Australia has done exactly this with their open source SIREN model (Rose 2016, Sustainable Energy Now 2017).

One motivation of our community is to facilitate better energy and climate protection policy formation 23
through public engagement by adopting transparent and open methods. But we also want to stress that openness *is more than* transparency. While publishing information under all rights reserved copyright can suffice for public consultation, it does not provide sufficient legal context to enable scientific reproducibility. Nor for the next generation of energy system models to be built and run using open source development methods (Morrison under review). Which explains why this submission is so strongly focused on the need for open licensing and public domain dedications for the public sector information and the privately held information of public interest upon which we so necessarily depend (section 6.4).

Further background on our research domain can be found in Pfenninger (2017) and Pfenninger *et al* (2017). 24

5 Energy sector database projects

Since 2009, several open energy sector database projects have been launched, some explicitly designed 25
to meet the needs of open energy system modelling projects. A relatively complete list of such projects is available online on Wikipedia (2017). These projects do not produce primary data but rather act as information intermediaries. They collectively form part of the data ecosystem within which we work and it would be ill advised to evaluate public sector information re-use without considering this wider context. The following two projects have close ties to several of the submitters (as noted in section 3):

- OpenEnergy Platform (OEP) project: the project includes a data hosting platform with extensive versioning features and is aimed specifically at the open energy modelling community (a GitHub for data if you like).[7]
- Open Power System Data (OPSD) project: an aggregating site that assembles, cleanses, and curates electricity system datasets for use by open and closed energy modelling projects and various forms

[7] GitHub is a US-based source code hosting site, widely used by the energy modelling and analysis communities.

of energy analysis.

26 These projects, both in receipt of German science funding, are intended for community use. They are not affiliated with an official or semi-official organisation nor are they governed by statutory provisions.

27 A number of crowdsourced projects have also arisen, primarily as a response to the lack of available private data of public interest. Various reasons are cited for withholding this information, including commercial sensitivity and security concerns regarding critical infrastructure. OpenStreetMap (OSM) represents the most visible example of crowdsourcing in our domain. OSM has moved well beyond a mapping application and now records semantic information about surface features, including high voltage transmission assets and estimates of their engineering parameters.[8] Dedicated crowdsourced projects of particular note within our domain include:

- Enipedia (Delft University of Technology): a semantic wiki covering the energy sector and related sectors.
- OpenGridMap project (TU Munich): crowdsourced data, collected directly or obtained from OpenStreetMap, is used together with methods from statistics and graph theory to reverse engineer representative models of electricity networks. These functionally equivalent networks can then be analysed directly or used as input data for more general energy system models.

28 A third group of projects comprise on-demand data platforms. These serve, for instance, synthetic time series in response to user selections. For example:

- Renewables.ninja project (ICL, London and ETH Zurich): serves hourly wind and photovoltaic generation potentials by location using historical satellite weather data and calibrated engineering characterisations.

29 All the above projects guarantee stable URLs and provide documented APIs for programmatic access. Or at least for as long as the individual projects remain viable and maintained.

30 It cannot be stressed too much that these community projects provide a facility that was previously duplicated by every single energy modelling and analysis research group at a necessarily lower standard (Pfenninger *et al* 2017:213). Moreover, these sites provide mechanisms for the datasets they serve to be continually updated, for errors to be identified and corrected, and for these revisions to be readily propagated back through their communities. The projects themselves vary somewhat in their willingness to accept copyright-protected datasets and/or datasets under non-open licenses. The database projects listed above (except OEP which classes as a data host) rely heavily on public information or private information of public interest released under statutory provisions.

31 As can be seen, a broad range of web-based database projects have developed. But few, if any, make use of data collected and assembled in a way that involves intellectual creation or an imprint of personality — two criteria used in European legal jurisdictions to define the threshold for copyrightability. We return to this important question in section 6.6.

32 Big data and big data analytics have not been applied to energy system modelling and would appear to have little to offer. Energy system models generally require much cleaner and better resolved information than that provided by big data collection techniques.

33 For completeness, other notable open energy database projects include: the Energy Research Data Portal for South Africa, energydata.info (World Bank Group), OpenEI (US DoE), reegle (REEEP and REN21), and SMARD (German BNetzA).

[8] For instance, high tension line voltages can be inferred statistically from the physical spacing of transmission towers.

6 Legal context

The legal context is central to this discussion. In the absence of clear open license notices or public domain 34
dedications, the legal status of digital data in relation to its local use and modification and further re-use by downstream parties is, at best, ill defined. And in the case of open licensing, copyright overreach may still apply.

There is very little legal analysis available concerning the re-use of energy sector data. The legal opinion 35
by Jaeger (2017), commissioned by the OPSD project, on European electricity sector data stands out in this regard. The Commission should note this document.

6.1 Technical definitions

We adopt the following technical definitions to aid the discussion. A *dataset* is a collection of *data points* 36
or elements and may as such be eligible for copyright protection.[9] A dataset may be realised as a table or an array and existing datasets may be aggregated in various ways to yield new compilations. A *database* is a collection of datasets and may be eligible for protection under sui generis database rights if the server is located within Europe and a substantial effort was required to assemble the information and implement the site (section 6.7).[10] A dataset is normally the minimum useful entity for our purposes, possessing an identified semantics and optionally metadata about its provenance, authorship, and open licensing, if any. A database may vary in terms of its technology, structure, coherence, and ontological consistency.[11] A number of technologies are used by database systems. File servers transfer individual files, with the same information sometimes offered in different formats. Relational database systems provide tables in response to structured queries from clients. On-demand platforms serve synthetic time series in response to user selections (for example, the Renewables.ninja project described earlier). A *databank* loosely describes an assemblage of datasets and/or databases. The term *data* is generic and can be used, depending on the context, to cover any and all of the above. These definitions are important because the technical details can substantially influence the application of intellectual property law.

The term *re-use* also needs defining. Jaeger (2017:4) proposes that *energy data re-use* means (grammar 37
corrected) "any activity to copy, modify, publish or distribute energy data or communicate the data to the public subsequent to the publication by the data provider". Hence, re-use *includes* the legal ability to republish modified versions. But note that if the original material carried an open license, then the outbound conditions imposed by that license must also be honoured on distribution (section 6.4).

We use the phrase *all rights reserved* copyright to indicate that no license has been granted, either 38

[9] Jaeger (2017:15) analyses what constitutes a "collection" under regulation 543/2103 and under Germany copyright law. He notes that: "Neither the German Copyright Act nor Directive 96/9/EC contain a definition of what is considered a ”collection” of data. Courts provide examples but have not defined the boundaries of the term." And he goes on to opine that "information about the total load of electricity consumption in various countries [and] the generation of electricity per production type" can certainly be interpreted as collections under copyright law.

[10] Jaeger (2017:16–18) analyses the criteria for such protection and concludes conservatively that "any person who wants to reuse a database with electricity data has to assume that the database is protected by … [a] sui generis database right according to article 7 of directive 96/9/EC". That said, we have also encountered informal legal advice to the contrary: that the "substantial investment" threshold necessary for protection might not be met by sites that serve energy datasets that arise during the course of routine operations.

[11] Wiktionary defines *coherent* as "orderly, logical, and consistent" and an *ontology* as "a structure of concepts or entities within a domain, organised by relationships; a system model".

proprietary or open.

6.2 Machine processing of protected datasets

39 It remains unclear to us as to whether a lawfully obtained copyright-protected dataset can be *machine processed* at all without breach of copyright.[12] By machine processing, we mean read a structured data file into memory using a computer program, cast this information to native data types, and then manipulate it programmatically to yield useful output. We have advice, unfortunately not attributable, that such usage would "probably yes" be a breach of copyright under German law. The point is not that the suggestion lacks status (it clearly does), but that this very significant question remains largely unframed and completely unanswered.[13]

40 The Commission should commission a legal opinion on the machine processing of protected datasets in relation to copyright law and publicise the conclusions.

6.3 Debilitating nature of all rights reserved copyright

41 What is certain however is that protected datasets cannot (outside of a few exceptions) be republished, whether unaltered or in reformatted, corrected, modified, or combined form.[14] One possible work-around for web-based datasets is to circulate the original URLs instead. But these URLs readily suffer from typos, linkrot, and resource drift (Wren 2004), particularly when served from static websites that are periodically reworked. Moreover, datasets that have been formatted to suit a particular modelling project cannot be made public, nor can corrections and enhancements be readily propagated back upstream.

42 Public sector datasets released under all rights reserved copyright — in the absence of copyright overreach or the presence of specific statutory exemptions or a public interest defence — are simply not usable for open energy modelling. As noted, these datasets cannot be machine processed with legal certainty. Nor can they be republished in their original or modified state to support either scientific accountability or open development (Morrison under review).

6.4 Open data and open licensing

43 The open licensing of datasets that qualify for copyright protection is the key for securing the very substantial and self-evident returns to cooperation in our domain (and doubtless elsewhere). Open licensed data can be subject to processes whereby errors and omissions can be readily detected, flagged, and where possible corrected or estimated. The OPSD project provides an excellent example of how a research/policy community can assemble and curate its core datasets.

44 Navigating and managing the license compatibilities and complying with the legal requirements can be challenging. Table 1 lists common open data and database licenses and public domain dedications.[16]

[12] Grimmelmann (2016) contends that the machine reading of copyrighted text by robots does not infringe US copyright law. But it is not certain how the facts he presents might relate to the machine processing of numerical datasets under European law.

[13] In contrast, the situation for source code is clear. Under US and German law, no rights are granted to use, modify, or distribute the source code if no license notice is present (Meeker 2017:148) (Jaeger and Metzger 2016:129). See Morrison (under review:§6.1) for a verified translation of the relevant passage from Jaeger and Metzger concerning German law.

[14] In limited circumstances, academic freedom can justify copyright infringement. But the infringement must have minimal impact on the copyright holder, serve a public interest, and be such that there was no other reasonable or proportionate way to achieve the same outcome (District Court of Amsterdam 2015, Caspers 2016). Academic freedom in Europe is conveyed under article 13 of the charter of fundamental rights of the European Union (European Commission 2000:11). Adopting academic freedom as a defence means increasing one's exposure to the risk of litigation.

[15] Copyright 2017 Robbie Morrison. This diagram is licensed under a Creative Commons CC BY 4.0 International License.

[16] Less common licenses, such the Linux Foundation Community Data License Agreement CDLA-Permissive 1.0 and CDLA-Sharing 1.0 licenses and the Open Data Commons (ODC) Attribution License 1.0, are not shown on table 1 or figure 1. Nor is the

Abbreviation	SPDX identifier	Name
ODbL	ODbL-1.0	Open Data Commons Open Database License 1.0
CC BY-SA 3.0	CC-BY-SA-4.0	Creative Commons ShareAlike Attribution 3.0 license
CC BY-SA 4.0	CC-BY-SA-4.0	Creative Commons ShareAlike Attribution 4.0 license
CC BY 3.0	CC-BY-3.0	Creative Commons Attribution 3.0 license
CC BY 4.0	CC-BY-4.0	Creative Commons Attribution 4.0 license
CC0 1.0	CC0-1.0	Creative Commons Zero 1.0 Universal public domain dedication
PDDL	PDDL-1.0	Open Data Commons Public Domain Dedication and License 1.0

Table 1: Common open data and database licenses and dedications.

The SPDX identifiers form part of an emerging standard (Linux Foundation 2017).[17] Figure 1 shows compatibilities between common licenses and dedications. The restrictions imposed by open licenses are outbound and trigger *only* on distribution. Users can otherwise do as they please with the open data locally. Although managing license compliance is a client not provider issue, the choice of license can and does have major downstream use case implications. These implications are discussed next.

Dedications and license types fall into three camps (as indicated along the top of figure 1). Public domain 45
dedications (left side) place no conditions or restrictions on the use and distribution of the material beyond waiving liability. Next are the *permissive* licenses (centre) that require only that all copyright holders be acknowledged when distributing modified copies. Then come the *copyleft* licenses (right side) that guarantee that the data will forever remain in the digital commons. But which, because of their placement near the top of the compatibility graph require that, when recombined with more weakly (leftward) licensed data, that the new work adopt the stronger license when republished. Permissive licenses do not exhibit this trait and tend to be favoured for this reason. The CC0 1.0 and PDDL 1.0 public domain dedications fall back to maximally permissive copyright licenses in countries with civil law traditions, including Germany, France, and much of continental Europe (Kreutzer 2011:4,11).

Discriminatory licenses which preclude commercial usage are not technically open. They are also difficult 46
to delineate in terms of application and, for this reason, are not generally recommended (Tsiavos 2013:7). The most often applied non-commercial data licenses are the Creative Commons NC configurations (such as the CC BY-NC 4.0).

Figure 1 shows that the CC BY 4.0 license and CC0 1.0 dedication represent good choices for information. 47
They are applicable to content, data, datasets, and databases and are well positioned to interact with other licenses. These two instruments also have the advantage of being widely deployed already and of being generally well understood and trusted.

In summary, the combining of existing datasets to create new datasets raises a multitude of technical, 48
legal, and practical issues, including, but not limited to, detecting and avoiding non-compliant outcomes and honouring the notice requirements and any other restrictions placed on the new work. An amusing but nonetheless salient illustration of this problem concerns the distribution of bullfrogs in north America (Desmet 2013).

6.5 Copyright law reform

Open licensing nonetheless remains second best to a substantive reform of European copyright law to 49
support positive PSI re-user rights. We are in full agreement with Tsiavos (2013:5–6) in this regard:

EU reuse and copyright notice license (European Commission 2011a) included.

[17] The Free Software Foundation Europe (FSFE) recently issued its REUSE good practice guidelines for adding machine locatable copyright information and SPDX identifiers to sets of textual files, including source code (FSFE 2017b).

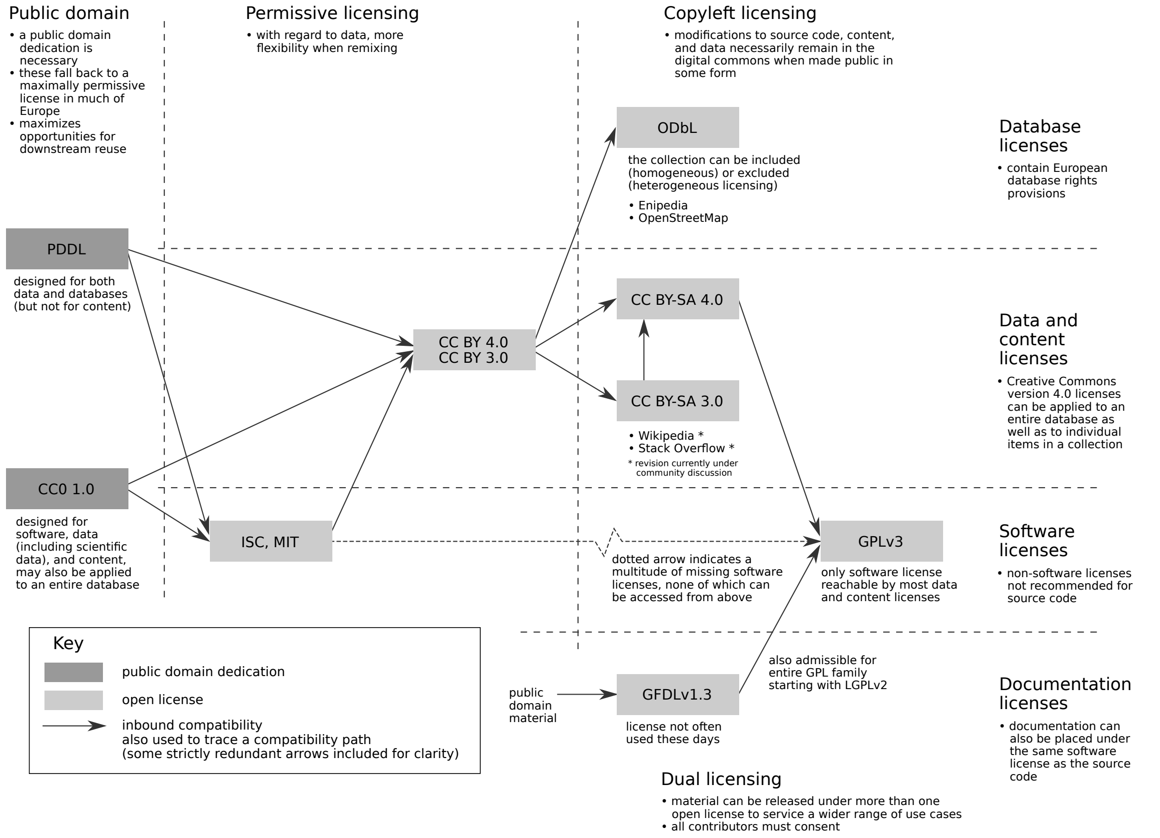


Figure 1: License compatibility graph for common open data and database licenses and dedications.[15]

> The adoption of a legislative solution, instead of licensing, could further reduce transaction costs, that is openness by default of PSI in the form of a law without further requirements of issuing or adopting of a specific license … (Recommendation 1(a))

The Creative Commons organisation holds a similar position on the need for legislative change (Creative Commons 2013): 50

> Our experience has reinforced our belief that to ensure the maximum benefits to both culture and the economy in this digital age, the scope and shape of copyright law need to be reviewed. However well-crafted a public licensing model may be, it can never fully achieve what a change in the law would do, which means that law reform remains a pressing topic. The public would benefit from more extensive rights to use the full body of human culture and knowledge for the public benefit. CC licenses are not a substitute for users' rights, and CC supports ongoing efforts to reform copyright law to strengthen users' rights and expand the public domain.

6.6 Copyright overreach

Copyright overreach is the practice of claiming copyright and perhaps adding open licenses to material that should *not rightfully* be under copyright. We are *not* referring to concerns in some circles that copyright has become too broadly defined (such as Brito *et al* 2012). Certain criteria apply when considering copyrightability in Europe outside of the United Kingdom and we believe that these criteria are not often 51 met in relation to public sector energy datasets.[18]

This section considers individual datasets and not databases. We presume here that the datasets being 52
used do not constitute a "substantial part" of the host database and that their download would therefore not be in breach of a sui generis database right.

The following analysis is based on the German statute covering copyright and other rights for authors 53
known as the *Urheberrechtsgesetz* (UrhG).[19] An official translation is available (Juris 2017).[20] UrhG §2(1) lists the types of works that are protected and *fails to mention* either data or datasets, digital or otherwise. UrhG §2(2) states that (emphasis added):

> Only the author's *own intellectual creations* constitute works within the meaning of this Act.

UrhG §4(1) covers collections, including collections of data (emphasis added):[21] 54

> Collections of works, data or other independent elements which *by reason of the selection or arrangement of the elements* constitute the author's *own intellectual creation* (collections) are protected as independent works without prejudice to an existing copyright or related right in one of the individual elements.

German law provides for exceptions in relation to certain forms of personal and non-commercial scientific 55
research (UrhG §52a, §87c). While these exceptions might be applicable and useful in some circumstances, they are not sufficiently general to be of much use for open energy modelling and analysis. For instance, a number of modelling projects are hosted on GitHub and may contain data that cannot be restricted to personal scientific research. Nor is there much interest in our communities for the use of discriminatory licenses (such as CC BY-NC 4.0) in order to disqualify commercial usage.

Much of the data under discussion is simply metered, logged, or recorded, so clearly no creativity is 56
involved in its collection and assembly. And even if subsequently processed using statistical or numerical techniques, these collections still lack any threshold of direct creativity.

Our lay interpretation is therefore that many, if not most, of the energy sector datasets released by official 57
and semi-official sources would *not meet* the threshold for copyrightability. We remain surprised that this issue has attracted such scant attention by policy makers and legal analysts alike.

We therefore recommend that a suitable public domain dedication (CC0 1.0) be applied to non-copyright- 58
able material to make its non-copyrightable status explicit and unambiguous. In practice, applying a permissive license (CC BY 4.0) to a dataset where not legally justified would not normally pose problems in our domain (yet may well do so in other domains). But failing to include a permissive open license or public domain dedication in such circumstances is completely unsatisfactory (section 6.3).

The Commission should therefore issue clear guidelines on the threshold that needs to be met before 59
routinely collectedly and processed energy sector data is eligible for copyright. The issue extends beyond public sector information and such guidelines may well have implications for publicly available commercial datasets in other domains.

[18] The United Kingdom has a lower threshold for copyright. The extent to which the UK will be subject to European intellectual property law beyond its planned Brexit in 2019 is not known.

[19] The ENTSO-E transparency platform, in contrast, operates under Belgium law (Jaeger 2017:26). But, one might presume, the choice of law that pertains to any copyright in a dataset served from the platform is determined by the location of the primary data provider.

[20] This version does not include revisions made on 30 June 2017. These changes primarily address text and data mining and are not especially relevant here (Jaeger 2017:44–45). But new provisions that relax the use of public databases for scientific research could be of interest (Jaeger 2017:43–44).

[21] This paragraph in the UrhG is the only time the statute traverses the concept of data or datasets. Moreover, the term *data* remains undefined within the act and the term *dataset* is never mentioned.

6.7 European database rights directive 96/9/EC

60 Sui generis database rights derive from the 1996 European regulation covering database protection for database servers located *within* Europe.[22] Article 7§1 requires that (European Union 1996:7–8) (emphasis added):

> Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a *substantial investment* in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a *substantial part*, evaluated qualitatively and/or quantitatively, of the contents of that database.

61 A database right is unlike a copyright in that it can be breached on download (extraction) as well as on distribution (a form of re-utilization). A database right is an economic right that can be transferred or waived by the rights holder. Like copyright overreach (section 6.6), database protection overreach is a known problem.[23]

62 Under German law, to infringe, all or a "substantial part" of a database must be downloaded, reconstructed, and used in a way that conflicts with the interests of the original database maker (UrhG §87b(1) Juris 2017:36). The meaning of "substantial" has yet to be determined through court rulings (Jaeger 2017:23).[24]

63 Database rights can therefore prove problematic for database projects which aggregate and host public sector data (such as the OPSD project). Database rights can also create difficulties for modelling projects because these same database hosts cannot be used to propagate corrections and enhancements to datasets obtained originally from public sector providers.

64 We therefore recommend that PSI providers be encouraged to expressly waive their database rights. We see no reason why database protection should apply to public sector information provision.

6.8 Electricity market transparency regulation 543/2013

65 The ENTSO-E transparency platform (TP) (ENTSO-E 2017) operates under European electricity market transparency regulation 543/2013 (European Commission 2013a).[25] The TP datasets are vital to our work and the open energy modelling community has been engaging with ENTSO-E to resolve the issues as we see them.[26]

66 Unlike datasets under all rights reserved copyright (section 6.2), transparency platform datasets can be machine processed with legal certainty. Jaeger (2017:5) opines that "the data provided … can be … used as input for computer models and analyzed in many ways for scientific and commercial purposes". This conclusion rests primarily on article 3 of 543/2013 which describes the establishment of a central information transparency platform (European Commission 2013a:19). An acknowledgement of ENTSO-E

[22] Database makers in the US can instead protect their collections by forming individual contracts with users (Kristof 2016:21).

[23] The European Court of Justice (ECJ) has significantly limited the scope of sui generis database rights through several of its judgements, including C-203/02. This implications of this case law are peripheral to the current discussion, but can be significant in other domains, see for instance Husovec (2014).

[24] One might surmise that 20 % would not be in breach.

[25] European regulation 543/2013 coexists with regulation 714/2009 covering cross-border network information (European Commission 2009) and regulation 1227/2011 covering wholesale market integrity and transparency (European Commission 2011b). These earlier regulations remain in effect and are reviewed by Jaeger (2017:8–10).

[26] Anecdotally, the in-house modellers at ENTSO-E face many of the same problems in relation to data quality as we do.

is required in scientific publications and reports that present results derived from TP datasets (Jaeger 2017:24).

67 But 543/2013 is a requirement to *publish* and not a requirement to *license* (Boecker 2016:14). As such, 543/2013 causes numerous problems that could easily be remedied by amending this regulation to expressly allow for all forms of re-use, including modification and republication. Regulation 543/2013 states that the "central information transparency platform shall be available to the public free of charge through the internet" but remains silent on the legal context (European Commission 2013a:4). Indeed Jaeger (2017:20) notes that "any other re-use beyond accessing, reading and downloading the data is not mentioned in regulation 543/2013". As of November 2017, the transparency platform terms and conditions do not provide a proper license and TP users must check with all relevant primary data owners in order to republish the data in either its original format or some modified state.[27] We view this as an unnecessary obstruction and one that goes against the spirit of 543/2013. Jaeger (2017:20) further notes that datasets on the TP site may be offered under legal notices that do not respect the requirements prescribed by 543/2013. This situation needs rectifying and members of our community have brought it to the attention of ENTSO-E.

68 It remains unclear to us as to whether the transparency platform is protected as a database under sui generis rights. Jaeger (2017:21–24) traverses this question but is unable to draw a conclusion.

69 It also remains unclear to us as to whether the direct (as opposed to interpreted) *results* from an energy system model run that utilised transparency platform datasets can later be published without restriction. Jaeger (2017:38–39) similarly examines this issue and cannot affirm any such right.

70 Jaeger (2017:7–12) lists the extensive reporting obligations under 543/2013 and other statutory requirements and these are considerable. The Commission should take heed of this succinct summary.

71 Datasets made available under European electricity market transparency regulation 543/2013 need only be made available for five years and can then go dark.[28] This requirement should instead be extended to five decades and steps taken to ensure continuity in the event of organisational restructuring.

72 It is imperative that the twin questions of open licensing and sui generis database rights for the transparency platform be resolved, so that researchers, the interested public, and others can work with and circulate data obtained from the platform at their discretion without having to operate in a legal grey zone. This is yet another salient example of where transparency considerations alone fall well short of the needs of energy system modellers and analysts.

73 The Commission needs to resolve both the legal context and the implementation of regulation 543/2013.

6.9 European Commission JRC data policy

74 The European Commission Joint Research Centre (JRC) plans to make part of its Integrated Database of the European Energy Sector (IDEES) public in late-2017 (Wiesenthal 2017). The database will initially span the years 2000–2018 for all member states. Dataset licensing is to be governed by the JRC policy on data, namely that the "acquisition of data by the JRC from third parties shall, where possible and feasible, be governed by the Open Data principles, and all efforts shall be made to avoid imposition of restrictions to their access and use by the JRC and subsequent users" (Doldirina *et al* 2015:6). The Open Data principles

[27] Jaeger (2017:23–26) provides an analysis of the June 2015 transparency platform terms and conditions and also its website disclaimer. He notes that this version of the agreement is unclear and confusing in multiple places and that the associated website disclaimer also contradicts its provisions.

[28] Regulation 543/2013 3§1 states (emphasis added): "The data shall be up to date, easily accessible, downloadable and available for *at least five years*. Data updates shall be time-stamped, archived and made available to the public." (European Commission 2013a:4). This means that transparency platform datasets can be legally removed after only five years.

however remain silent on the right of public users to distribute original and modified works (*ibid*:6). With regard to Commission-sourced data, some kind of attribution license, perhaps the EU reuse and copyright notice (European Commission 2011a), has been suggested (Zucker 2017). The Commission needs to finalise which open licenses it intends to use for these datasets. Metadata is to follow the JRC Data Policy Implementation Guidelines but, as of October 2017, these guidelines are not public.

6.10 Article 13 of the proposed European Copyright Directive

Article 13 of the proposed European Copyright Directive (European Commission 2016:29–30) will pose 75
difficulties for European data hosting sites as well as for European code hosting sites. The FSFE and others, for instance, have directed their criticism towards the problems that code hosting sites located in Europe will face (FSFE 2017a, OFE and FSFE 2017). But article 13 will have major implications for scientific data hosting sites (including the OEP open energy system database project). Data hosting sites are central to the scientific endeavour and it is likely that their number and importance will increase (Helmholtz Association 2016).

We suggest that the Commission revisit this provision in light of the difficulties it will doubtless cause code 76
and data hosting platforms with hardware located within the European Union.

7 Data quality and coverage

Energy system models require reliable high-quality data. This means datasets that are well-structured, 77
consistent and complete, and accompanied by suitable metadata and a documented provenance. The use of established and recognised *open technical standards* facilitates both portability and archivability. In addition, these datasets are best circulated within community-defined *data packages* which also embed human and machine-readable metadata (OKI 2017). The computer science details are not important here. Dataset and metadata standards are mostly matters for each domain to determine and not something that the Commission needs drive. Metadata must also be open licensed to be useful, which raises its own legal issues (Kreutzer 2011:6–10) and is something which the Commission should investigate and resolve.

Some process by which PSI data can be confirmed as reliable or alternatively flagged and repaired at 78
the point of publication is essential. If this facility is not offered by the primary data provider, then our community has (and will doubtless continue to do so) set up intermediary databases to assemble, cleanse, aggregate, and otherwise curate this information. The OPSD project provides an excellent example of just this facility.

Another example in our domain is the US Department of Energy OpenEI file server site. This site provides 79
user forums for each dataset together with a user-determined five star rating system (Brodt-Giles 2012:3).

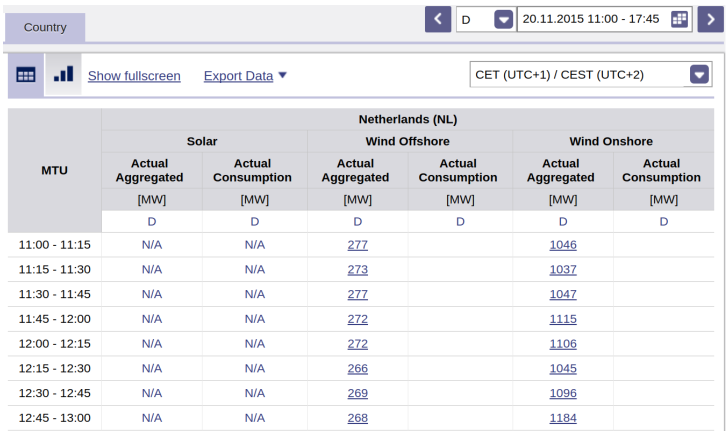
While much of this submission covers issues that arise from the electricity sector, energy system models 80
and market analysis also cover the gas and heat sectors and, on occasion, the transport sector and the built environment. Better information disclosure is needed in these other sectors. Particularly given the current research drive on sector coupling, energy storage, and end-use efficiency, split between technology development and market design and integration. Indeed, we believe that the electricity market transparency platform should be replicated for both gas and heat. Developing the necessary strategies and regulations is a task for the Commission.

Crowdsourcing falls outside the remit of this public consultation exercise. But we note, in passing, that 81 crowdsourced data can be surprisingly comprehensive and reliable.

7.1 ENTSO-E transparency platform

The ENTSO-E transparency platform (TP) is mandated under European regulation 543/2013 (European 82
Commission 2013a) as described in section 6.8. Most ENTSO-E data is sourced from its 43 member transmission system operators (TSO) who, in turn, may collect it from various power exchanges (PX) and market participants within their jurisdictions. Hence, ENTSO-E can face an uphill battle to obtain good quality primary data and has little or no ability to sanction substandard and/or late reporting by its members and their upstream partners.[29] In terms of open re-use, open licensing requires the consent of the primary data owner. In other words, the licensing web starts with the many tens of copyright holders who contribute data to the platform. We also refer to our earlier comments on copyrightability and overreach (section 6.6).

Figure 2 shows missing solar generation data on the platform. Users will need to repair these data points, 83
probably by substituting the missing values with renewables potentials estimates (such as those served by the Renewables.ninja project). Moreover, these corrections should be documented and circulated so that other projects can benefit from this effort.



MTU	Netherlands (NL) Solar Actual Aggregated [MW] D	Solar Actual Consumption [MW] D	Wind Offshore Actual Aggregated [MW] D	Wind Offshore Actual Consumption [MW] D	Wind Onshore Actual Aggregated [MW] D	Wind Onshore Actual Consumption [MW] D
11:00 - 11:15	N/A	N/A	277		1046	
11:15 - 11:30	N/A	N/A	273		1037	
11:30 - 11:45	N/A	N/A	277		1047	
11:45 - 12:00	N/A	N/A	272		1115	
12:00 - 12:15	N/A	N/A	272		1106	
12:15 - 12:30	N/A	N/A	266		1045	
12:30 - 12:45	N/A	N/A	269		1096	
12:45 - 13:00	N/A	N/A	268		1184	

Figure 2: Missing solar generation data (designated N/A) on the ENTSO-E transparency platform.

Data quality issues for the TP include but are not limited to: missing geo-coordinates for assets like 84
substations,[30] cryptic asset names, missing generator attributes, missing renewables data (figure 2), missing spatially resolved load time series, and missing engineering information for electricity grids. The potential five year lifespan is also of concern (section 6.8). This submission is not however the place to detail the current deficiencies of electricity market transparency platform. And as noted, members of our community have been raising these matters with ENTSO-E.

7.2 The Copernicus Climate Change Initiative

The Copernicus Climate Change Service (C3S) project uses state-of-the-art climate observations to provide 85
"authoritative and quality-assured information about the past, current and future states of the climate in

[29] This is the position that ENTSO-E adopts in our discussions on data quality. But this assessment may not necessarily be correct, nor does not preclude legal action by external parties under regulation 543/2013. Yet it seems extraordinary that data users would need to initiate public interest litigation to achieve what is intended by law, namely reliable, complete, and timely transparency information.

[30] Security concerns may be able to be addressed by releasing fuzzy values with say $\pm 10\,\text{km}$ white noise added.

Europe and worldwide" (Copernicus 2017). This climate information is being used to develop Sectoral Information Systems (SIS) to improve decision-making within selected sectors. There are currently seven services in the proof-of-concept phase, with two concerning the energy sector. ECEM investigates the impact of climate change on energy systems and CLIM4ENERGY provides tailored climate indicators. Energy sector analysis can clearly benefit from the use of high-quality meteorological information. And this becomes more important as climate variability increases and renewable generation capacity penetrates. The datasets produced by the weather and climate communities use long established data standards (such as the NetCDF CF metadata conventions) and these datasets need to be matched by similar quality information from the energy sector, covering, for example, supply and demand. But current energy sector datasets, downloaded from the ENTSO-E platform for instance, lack the quality needed and this places a significant impediment on the type and robustness of the analysis that can be undertaken in this nexus between climate change and energy system analysis.

8 Looking ahead

It is useful to spend a few moments looking ahead. We noted earlier that big data has not been seen as a 86
priority in our domain and that may well continue. Nor has information privacy been an issue either. But as models extend into the end-use domain and increase in resolution, it is likely that consumer privacy issues will need to be traversed.[31] That said, such data is unlikely to derive from public sector sources. Nor need it be exact, so techniques that anonymize the data and produce functionally equivalent datasets should generally suffice.

Beyond 2030 lies the prospect of carbon dioxide removal (CDR) technologies starting to be deployed at 87
scale (UNEP 2017:59). CDR technologies are widely mistrusted and their adoption will require a complete reversal in public opinion. The integrated assessment models (IAM) used to derive these technology projections are not without criticism. Shortcomings include pessimistic technology characterisations, including photovoltaics (Breyer *et al* 2017) and an absence of social preferencing. The point here is not whether the IAMs are sufficiently correct and/or subject to over-interpretation, but that the public debate would be very much better if the IAM models and data were open and at least some of the story lines and scenarios were formed through public consultation. As things stand, policymakers are reluctant to even mention CDR technologies in public (Geden and Löschel 2017:881).

Should CDR technologies become deployed at scale in Europe, then a new supercritical carbon dioxide 88
reticulation system may well develop. If so, all associated privately held data of public interest in relation to this network should be mandated as public and open from the outset.

9 Discussion

We would like to reiterate four key themes. First, that the potential for public sector digital data re-use 89
can best be realised by applying a suitable permissive license (CC BY 4.0) or public domain dedication (CC0 1.0). The latter designation is necessary for non-copyrightable material to make clear its legal status. A public domain dedication can also be used to explicitly waive copyright when the public sector copyright holder deems this either useful or appropriate.

Second, we want to reinforce again that transparency *is not* openness. Transparency can normally be sat- 90
isfied by publishing information under all rights reserved copyright. But only genuine openness, granted by open licenses and public domain dedications, can support scientific reproducibility and open development. This context explains why we are so concerned that the ENTSO-E transparency platform be required to make its datasets legally open and not merely public.

[31] Recent European right to data portability (RtDP) law will allow consumers to access their smart meter data (Graef *et al* 2017).

Third, there is a need to provide mechanisms by which users of public sector information can comment 91
on the quality of the information, flag omissions and mistakes, and propose corrections. This is not a form of crowdsourcing because the data provider will retain both the right and the responsibility to evaluate each issue and make the necessary edits and adjustments.

Fourth, the general legal context for public sector data needs clarifying. The digital commons community 92
has, over the past three decades, developed a range of sophisticated open licenses and public domain dedications. These instruments have become increasingly understood and trusted, in part due to a growing body of favourable case law. But a number of issues remain, the bulk of which can only be tackled through innovative public policy and legislative reform. These are governance matters that law makers need to resolve. Copyright law was not written with digital data in mind. Nor with the idea of an internet-mediated information commons of growing social and economic importance. So ultimately the problems we face in the open energy modelling community and allied communities can only be resolved through appropriate legislative change. We therefore offer our experiences and analysis for the Commission to consider in this endeavour.

Finally, we are, collectively and through the Open Energy Modelling Initiative, available to work together 93
with the Commission to progress these issues. We would also be happy to give evidence in person.

Acknowledgements

We wish to thank three lawyers, specialising in information technology and open source licensing and 94
necessarily unnamed, who willingly answered our questions, clarified and raised issues, and provided pointers to the legal and academic literature. Berit Erlach kindly offered feedback on earlier versions.

The document itself was prepared using free software (Linux, bash, git, LaTeX) and fonts (DejaVu Sans) 95
and the Overleaf online editing platform.

Abbreviations

API	application programming interface
ENTSO-E	European Network of Transmission System Operators for Electricity
FSFE	Free Software Foundation Europe
PSI	public sector information
TP	European electricity market transparency platform (ENTSO-E 2017)
TSO	transmission system operator
UrhG	*Urheberrechtsgesetz* (the German statute covering copyright)

Consultation documents

The key consultation documents are listed for completeness:

Original PSI re-use directive	European Commission (2003)
Revised PSI re-use directive	European Commission (2013b)
Review document	European Commission (2017a)
Annex to above	European Commission (2017b)
PSI directive webpage	European Commission (2017c)
Public consultation notice	European Commission (2017d)

References

BMWi. 2015. *The Energy of the Future: Fourth "Energy Transition" Monitoring Report — Summary*. Berlin, Germany: Federal Ministry for Economic Affairs and Energy (BMWi). November.

Boecker, Lina. 2016. *Energy databases: protection and licensing — Presentation*. Berlin, Germany: JBB Rechtsanwaelte. 21 November.

Breyer, Christian, Dmitrii Bogdanov, Ashish Gulagi, Arman Aghahosseini, Larissa Barbosa, Otto Koskinen, Maulidi Barasa, Upeksha Caldera, Svetlana Afanasyeva, Michael Child, Francisco Farfan Orozco, and Pasi Vainikka. 2017. On the role of solar photovoltaics in global energy transition scenarios. *Progress in Photovoltaics Research and Applications*. **25**(8): 727–745. 1 August. ISSN 1099-159X. doi 10.1002/pip.2885.

Brito, Jerry, Tom W Bell, Eli Dourado, Timothy B Lee, Christina Mulligan, David G Post, Patrick Ruffini, and Reihan Salam. 2012. *Copyright unbalanced: from incentive to excess*. Arlington, Virginia, USA: Mercatus Center at George Mason University. 29 November. ISBN 978-098360775-5.

Brodt-Giles, Debbie. 2012. *WREF 2012: OpenEI: an open energy data and information exchange for international audiences*. Golden, Colorado, USA: National Renewable Energy Laboratory (NREL).

Caspers, Marco. 2016. *The role of Anne Frank's diary and academic freedom for text and data mining*. Alphen aan den Rijn, The Netherlands: Kluwer Copyright Blog. 20 January.

Copernicus. 2017. *About C3S*. Webpage.

Creative Commons. 2013. *Copyright reform*. Mountain View, California, USA: Creative Commons. 16 October. Webpage. CC BY 4.0 licence.

Desmet, Peter. 2013. *Showing you this map of aggregated bullfrog occurrences would be illegal*. 17 October. Webpage.

District Court of Amsterdam. 2015. *ECLI:NL:RBAMS:2015:9312*. Amsterdam, The Netherlands: District Court of Amsterdam. 23 December. In Dutch.

Doldirina, Catherine, Anders Friis-Christensen, Nicole Ostlaender, Andrea Perego, Alessandro Annoni, Ioannis Kanellopoulos, Massimo Craglia, Lorenzino Vaccari, Giacinto Tartaglia, Fabrizio Bonato, Paul Triaille Jean, and Stefano Gentile. 2015. *JRC data policy — Report EUR 27163 EN*. Luxembourg: Publications Office of the European Union. ISBN 978-92-79-47104-9. doi 10.2788/607378.

European Commission. 2000. Charter of fundamental rights of the European Union — 2000/C 364/01. *Official Journal of the European Communities*. **C 364**: 1–22. 18 December.

European Commission. 2003. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. *Official Journal of the European Union*. **L 345**: 90–106. December.

European Commission. 2009. 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 (Text with EEA relevance). *Official Journal of the European Union*. **L 211**: 15–35. 14 August.

European Commission. 2011a. Commission decision of 12 December 2011 on the reuse of Commission documents — 2011/833/EU. *Official Journal of the European Union*. **L 330**: 39–42. 14 December.

European Commission. 2011b. 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. 8 December.

European Commission. 2013a. 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. 2013b. 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 (Text with EEA relevance). *Official Journal of the European Union*. **L 175**: 1–8. 27 June.

European Commission. 2016. *Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market — COM/2016/0593 final — 2016/0280 (COD)*. Brussels, Belgium: European Commission. 14 September.

European Commission. 2017a. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the mid-term review on the implementation of the Digital Single Market Strategy — a Connected Digital Single Market for All — COM/2017/0228 final*. Brussels, Belgium: European Commission.

European Commission. 2017b. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the mid-term review on the implementation of the Digital Single Market Strategy — a Connected Digital Single Market for All — COM/2017/0228 final — Annex*. Brussels, Belgium: European Commission.

European Commission. 2017c. *European legislation on re-use of public sector information*. Brussels, Belgium: European Commission. Webpage.

European Commission. 2017d. *Public consultation on the review of the directive on the re-use of Public Sector Information (PSI Directive)*. Brussels, Belgium: European Commission. Webpage.

European Union. 1996. *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*. WIPO Database of Intellectual Property.

eurostat. 2016. *Energy, transport and environment indicators — 2016 edition*. Luxembourg: Publications Office of the European Union. ISBN 978-92-79-60138-5. doi 10.2785/260003.

FSFE. 2017a. *European copyright reform hampers free software development*. Berlin, Germany: Free Software Foundation Europe (FSFE). 8 September. Webpage.

FSFE. 2017b. *REUSE initiative*. Berlin, Germany: Free Software Foundation Europe (FSFE). Website.

Geden, Oliver and Andreas Löschel. 2017. Define limits for temperature overshoot targets. *Nature Geoscience*. **10**(12): 881–882. December. ISSN 1752-0908. doi 10.1038/s41561-017-0026-z.

German TSOs. 2017. *Grid development plan — Power*. Germany: 50Hertz Transmission, Amprion, TenneT TSO, and TransnetBW. Website.

Graef, Inge, Martin Husovec, and Nadezhda Purtova. 2017. *Data portability and data control: lessons for an emerging concept in EU law*. 15 November. SSRN preprint.

Grimmelmann, James. 2016. Copyright for literate robots. *Iowa Law Review*. **101**(2): 657–681.

Helmholtz Association. 2016. *Making information resources more usable: a position paper on the management of research data in the Helmholtz Association*. Berlin, Germany: Helmholtz Association. 22 September.

Husovec, Martin. 2014. *The end of (meta) search engines in Europe? — Research paper no 14–15*. Munich, Germany: Max Planck Institute for Innovation and Competition. doi 10.2139/ssrn.2411917.

Jaeger, Till. 2017. *Legal aspects of European electricity data — Legal opinion*. Berlin, Germany: JBB Rechtsanwälte. 24 July.

Jaeger, Till and Axel Metzger. 2016. *Open Source Software: Rechtliche Rahmenbedingungen der Freien Software* [*Open source software: legal framework for free software*]. Fourth edition. Munich, Germany: CH Beck. 21 March. ISBN 978-3-406-67773-1. In German.

Juris. 2017. *Act on Copyright and Related Rights (Urheberrechtsgesetz, UrhG) — Amendments to 20 December 2016 — Official translation*. Saarbrücken, Germany: Juris. This version lacks the revisions enacted on 30 June 2017.

Kreutzer, Till. 2011. *Validity of the Creative Commons Zero 1.0 universal public domain dedication and its usability for bibliographic metadata from the perspective of German copyright law*. Berlin, Germany: Büro für Informationsrechtliche Expertise.

Kristof, Cindy. 2016. Data and copyright. *Bulletin of the Association for Information Science and Technology*. **42**(6): 20–22. August/September.

Linux Foundation. 2017. *Software Package Data Exchange (SPDX) license list*. San Francisco, California, USA: Linux Foundation. Webpage.

ENTSO-E. 2015. *ENTSO-E at a glance*. Brussels, Belgium: European Network of Transmission System Operators for Electricity. June.

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

Meeker, Heather. 2017. *Open (source) for business: a practical guide to open source software licensing*. Second edition. North Charleston, South Carolina, USA: CreateSpace Independent Publishing Platform. 4 April. ISBN 978-154473764-5.

Morris, Craig and Arne Jungjohann. 2016. *Energy democracy: Germany's Energiewende to renewables*. London, United Kingdom: Palgrave Macmillan. ISBN 978-3-319-31891-2. doi 10.1007/978-3-319-31891-2.

Morrison, Robbie. under review. Energy system modeling: public transparency, scientific reproducibility, and open development. *Energy Strategy Reviews*. ISSN 2211-467X. Submitted 1 August 2017.

OFE and FSFE. 2017. *European copyright reform: impact on free and open source software and developer communities — White paper*. OpenForum Europe (OFE) and the Free Software Foundation Europe (FSFE). September.

OKI. 2017. *The open data handbook*. Cambridge, United Kingdom: Open Knowledge International (OKI). Evolving web document.

Pfenninger, Stefan. 2017. Energy scientists must show their workings. *Nature*. **542**: 393. 23 February. ISSN 0028-0836. doi 10.1038/542393a.

Pfenninger, Stefan, Joseph DeCarolis, Lion Hirth, Sylvain Quoilin, and Iain Staffell. 2017. The importance of open data and software: is energy research lagging behind? *Energy Policy*. **101**: 211–215. February. ISSN 0301-4215. doi 10.1016/j.enpol.2016.11.046.

Rose, Ben. 2016. *Clean electricity Western Australia 2030: modelling renewable energy scenarios for the South West Integrated System*. West Perth, WA, Australia: Sustainable Energy Now (SEN). April.

Steinbach, Armin. 2013. Barriers and solutions for expansion of electricity grids: the German experience. *Energy Policy*. **63**: 224–229. 1 December. ISSN 0301-4215. doi 10.1016/j.enpol.2013.08.073.

Sustainable Energy Now. 2017. *Modelling overview: the SIREN toolkit and more*. Perth, Australia: Sustainable Energy Now (SEN). Webpage.

Tsiavos, Prodromos. 2013. *Licence interoperability report — Version 1.3*. European Union. Publication date from PDF metadata.

UNEP. 2017. *The emissions gap report 2017 — a UN environment synthesis report*. Nairobi, Kenya: United Nations Environment Programme (UNEP). November. ISBN 978-92-807-3673-1.

Wiesenthal, Tobias. 2017. *POTEnCIA and JRC-IDEES: a new modelling toolset for the European energy sector — Presentation*. Brussels, Belgium: EMP–E Meeting. 18 May.

Wikipedia. 2017. *Open energy system databases*. Webpage.

Wren, Jonathan D. 2004. 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics*. **20**(5): 668–672. 22 January. ISSN 1367-4803. doi 10.1093/bioinformatics/btg465. The 404 in the title is correct and is not a data scraping error.

Zucker, Andreas. 2017. *Data openness in JRC models — Presentation*. Brussels, Belgium: EMP–E Meeting. 17 May.

▢