Submission on a proposed Data Act for the European Union from the perspective of open energy system analysis

European Commission public consultation closing 25 June 2021

Coordinating author	Robbie Morrison <robbie.morrison@posteo.de>
Published	Berlin, Germany
Date	25 June 2021
Submitters (16)	Cruz E Borges, Tom Brown, Matthew Deakin, Lion Hirth, Jack Kelly, Barry McMullin, Robbie Morrison, Fabian Neumann, Sylvain Quoilin, Malte Schäfer, Hans Schäfers, Wolf-Peter Schill, Ingmar Schlecht, Adam Stein, Jens Weibezahn, Frauke Wiese
Areas of expertise	Energy system models, energy sector data, energy market analysis, and rapid decarbonization, all with a focus on public policy development
Organisation	Representing ourselves as individuals (while noting that the European Commission might classify us an “informal organisation” for their internal purposes)

Public release	The statement may be made public without redacting personal details
Errata	Corrections and updates can be found on this web forum
DOI	Document to be uploaded to Zenodo upon submission and hence gain a DOI
PDF file name	`morrison-european-data-act-inception-submit.07.pdf`

This statement accompanies our web-based public submission on a proposed Data Act being undertaken by the Data Policy and Innovation Unit G.1 of the European Commission Directorate-General CNECT. Submissions close on 25 June 2021. The underpinning consultation document is European Commission (2021). See the Commission website for more details.

Contents

1 Requests

This section details points that we believe the European Commission should consider when developing their next tranche of data policy and supporting legislation. Some of the key issues given here are later expanded upon. 1

Our remarks are made in the context of our domain, which is **public-interest energy system analysis** with an emphasis on open source development, genuinely open data, and scientific cooperation. We presume that the data under discussion should be machine-readable. 2

References to "our community" relate to an informal network of **energy system modellers** committed to **open analysis** in all its dimensions (more in section 2.2). Whereas the term "we" refers solely to the listed submitters (more in section 2.1). 3

1.1 Core requests

On the basis of the inception impact assessment report issued by the Commission (European Commission 2021), the submitters advocate the following: 4

1. that all information made public under **statutory reporting** — including, but not limited to, system status, market monitoring, and mandatory public disclosure — be **machine-readable and explicitly open licensed** (for instance, disclosure reporting by electricity market exchanges appears to be implemented so as to intentionally inhibit data reuse)
2. that the **period of retention be extended** and preferably made permanent subject to operational considerations — because under current legislation much of the material under statutory reporting can be **legally withdrawn from public access after five years** (and we have one related example of this occurring)
3. that **REMIT energy market transparency information**[1] be either centrally catalogued or centrally provided and that reporting standards, programmatic access, and the reporting itself be more consistent and served free of charge and without screening recipients (see also point 1)
4. that **closed public-interest databases** (such as the ENTSO-E sponsored Pan European Market Modelling Database) be made public and explicitly open licensed
5. that **business-to-government** (B2G) information, suitably aggregated, anonymized where necessary, and open licensed would substantially improve the accuracy and legitimacy of energy system analysis and resultant policy advice
6. further to the above, that domain-relevant but generalised **cost accounting information** for various technologies and services, together with **metrics covering their technical performance, operational profiles, deployment and uptake**, and similar be identified, tracked, and reported to enable more transparent and defensible public interest analysis (with status and trend reporting from the US Energy Information Administration representing good practice in this regard)
7. that the Commission review legislative options for the voluntary **dedication of numerical information to the public domain** or some equivalent status (we observe that CC0–1.0 is not a legislative option)[2]

[1] REMIT is the Regulation on Wholesale Energy Market Integrity and Transparency. Section 3.1 provides background on the statutory reporting mandated under this regulation.

[2] We use software package data exchange (SPDX) license identifiers where these exist.

We note as well:

8. that **96/9/EC database protection** has outlived any usefulness and should be repealed — indeed, database protection in our realm provides only legal uncertainty, inhibits risk-averse institutions and researchers but not others, is routinely ignored by market participants and consultants conducting in-house modelling, and is also disregarded by data portals operating from beyond Europe

Legal questions related to the public licensing of data also need to be resolved, hence we ask:

9. that the Commission provide **legal opinions on the interoperability** of the growing number of **national data licenses** within the European Union (including the German dl–de/by–2.0 and French Licence Ouverte licenses and also the United Kingdom OGL–UK–3.0 license, noting that the UK National Grid continues to supply data to the ENTSO-E Transparency Platform post Brexit) with regard to the Creative Commons CC–BY–4.0 license — or alternatively legislate to prevent a proliferation of **data silos** that would naturally arise from the use of legally incompatible national data licenses

1.2 Additional requests

These points are over and above the nine requests just listed. Some of the suggestions here may well be rendered unnecessary if the Commission decides to address the informing concerns at a deeper level. Particular points either address specific issues and/or represent fallback measures should suitable open licensing not be made mandatory on statutory reporting. We therefore conditionally advocate that the Commission: 5

10. issue a determination on whether human authorship is a necessary condition for copyright
11. issue a determination on whether the machine processing of a legitimately held copy of a collection of atomic data under copyright constitutes civil or criminal infringement
12. amend the 96/9/EC Database Directive to better reflect the intention of lawmakers regarding thresholds and scope and provide guidance on what constitutes “substantial investment” and “substantial extraction” and how that information should be transmitted to users
13. revise the statutory definition for “reuse”[3] provided in the 2019/1024 Open Data Directive under 2§11 — currently and perversely remapped to “use” — so as to provide a definition of “reuse” consistent with the remainder of that statute, including recital 16
14. either clarify or remove the term “primary owner of the data” from Regulation 534/2013 at 2§23, which covers the statutory reporting of electricity market information
15. explicitly waive 96/9/EC database protection by default on public sector information

1.3 Wider picture

We note the concept of a novel **industrial property right** (IPR) that the Commission was earlier pursing seems to have fallen by the wayside (Stepanov 2020:80). 6

We note that the notion of **high-value datasets** (HVDS), as given in the 1024/2019 Open Data Directive, does not include the energy sector. This issue was raised in 2020 (Morrison 2020) and should be revisited by the Commission as a priority. 7

And we note the Commission will release its “fit for 55” **energy and climate package** on 14 July 2021. That package will doubtless draw very heavily on the kind of numerical data we are discussing here. And 8

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

thereby underscores our concerns that much of the necessary data is either privately held and thereby unobtainable or inadequately licensed for transparent and reproducible analysis.

1.4 Commentary

Our community has been **formally engaging with the European Commission** on these matters since 2017 (Morrison *et al* 2017, Morrison 2020). And also with various providers of information under statutory reporting within the European energy sector since somewhat earlier. Indeed, almost none of the issues raised here are new. What is novel is the increasingly deep community-wide cooperation on data management within our domain, a dynamic that has only really become evident in the past two years (see section 4.1). 9

Our community is increasingly desperate for access to reasonable quality **legally-reusable public-interest data** covering the European energy sector — indeed it is hard to overstate how debilitating the current legal regime is for those pursuing independent public-interest energy systems analysis. For us, "legally-reusable" means "explicitly licensed under Creative Commons CC–BY–4.0, CC0–1.0, or something inbound compatible".[4] That is because any other choice of license, even if technically data-capable and approved as open,[5] will necessarily create **data silos** due to legally incompatible licensing provisions (see section 3.6). We also note somewhat dated discretionary policy from the Commission on choice of instrument (for which our recommendations broadly overlap) and core requirements for datasets (ditto also) (Commission 2014:2,4–6). 10

We note with dismay, the practice (as best we can identify) that some statutory entities publish **data under statutory reporting** in the least reusable manner possible in order to **on-sell full numerical access** to willing third parties. The Commission should investigate this practice and amend the underlying legislation as required. 11

As indicated, our own **data management paradigm** is shifting from one of classical databases and schemas, whether local or web-accessible, to domain-wide data management and use (see section 4.1). This paradigm shift is contingent on two factors. First, sufficient community agreement on and adoption of high-level data "specifications" covering specifically: structured semantics (and preferably a formal ontology), metadata practices, collection protocols, and information exchange formats. And second, the application of open licenses as articulated in the previous paragraph. 12

Our community generally only deals with **non-personal information** that has been or can be **legitimately published** — hence issues of personal and commercial privacy are not material. 13

We will later provide examples of the hindrances we collectively encounter that derive primarily from the **current statutory environment** in Europe. These issues can, therefore, be best remedied by legislative reform. In which case, we welcome the opportunity to offer input to a **proposed Data Act** based on our experiences. We will also cite some current instances of good practice to indicate how things could and should be (see section 4.5). 14

Many in our community also work for **risk-averse intuitions** and a presumption of free-to-republish in the absence of suitable open licensing is not an option. Furthermore, many academic journals require an explicit declaration that such rights are held. These rights are then often routinely transferred to the publisher although the authors may also select from a range of Creative Commons licenses.[6] 15

[4] We use software package data exchange (SPDX) license identifiers where these exist.

[5] The United Kingdom-based Open Knowledge Foundation (OKF) act as the licence steward for data-capable open licenses and adjudicates on conformance with its Open Definition 2.1.

[6] This practice of copyright assignment no doubt falls outside of the scope of a Data Act, but is nonetheless something that the Commission should examine and remedy in the context of open science policy.

Analysis from **trusted and reliable sources** is critically important for the rapid and complete decarbonization of our society and that analysis is, in turn, critically dependent on good quality usable and reusable data. Moreover science is facing a **reproducibility crisis** and the measures we advocate here should also assist. 16

If the proposed Data Act actions **only one of our concerns**, let it be this: 17

- mandate that all information under statutory reporting be licensed Creative Commons CC–BY–4.0

2 Standing

We submitters are part of an open energy modelling community and allied research communities and many of us participate in the Open Energy Modelling Initiative (see section 2.2). And 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 previous submissions in 2017 and 2020. 18

2.1 Submitters

Background on individual submitters, their affiliations, research interests, and projects, as follows:[7] 19

- Dr Cruz E Borges, Bilbao, Spain: coordinator of the EU WHY project (GA 891943) and researcher with 10 years experience in the energy sector.
- Prof Dr Tom Brown, Technische Universität Berlin, Germany: Professor of Digital Transformation in Energy Systems, energy system researcher, and co-maintainer of the PyPSA energy/electricity system models.
- Dr Matthew Deakin, Newcastle University, United Kingdom: whole system modelling of energy system networks across a range of network scales and energy vectors.
- Prof Dr Lion Hirth, Hertie School, Germany: Assistant Professor of Governance of Digitalization and Energy Policy and has worked extensively on open data for energy system analyses.
- Dr Jack Kelly, Open Climate Fix, United Kingdom: a machine learning researcher who is also terrified by climate change.
- 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.
- Fabian Neumann, Karlsruhe Institute of Technology (KIT), Germany: researcher and co-maintainer of the PyPSA energy/electricity system models.
- Prof Sylvain Quoilin, University of Liège and KU Leuven, Belgium: head of the "Integrated and Sustainable Energy Systems" research unit.
- Malte Schäfer, doctoral student, TU Braunschweig, Germany: researcher assessing the environmental impact of grid electricity based on public energy system data.

[7] Three submitters are resident outside the European Union. This is acceptable under Commission consultation guidelines.

- Prof Dr Hans Schäfers, University of Applied Sciences Hamburg (HAW Hamburg), Germany: Professor for intelligent energy systems and energy efficiency, deputy head of the HAW's competence centre for renewable energies and energy efficiency (CC4E).
- Dr Wolf-Peter Schill, DIW Berlin, Germany: energy modeller and Deputy Head of the Department Energy, Transportation, Environment at DIW Berlin.
- Dr Ingmar Schlecht, Director of Neon Neue Energieökonomik GmbH, Berlin, Germany, a boutique energy economics consultancy firm and also postdoctoral researcher at ZHAW Zurich University of Applied Sciences.
- Dr Adam Stein, The Breakthrough Institute, United States: researcher in energy systems, risk, and resilience.
- Dr Jens Weibezahn, Workgroup for Infrastructure Policy (WIP), Technische Universität Berlin (TU Berlin), Germany and Copenhagen School of Energy Infrastructure (CSEI), Copenhagen Business School (CBS), Denmark: electricity markets, electricity transmission systems, and decentralized energy transformation.
- Assoc-Prof Frauke Wiese, Associate Professor, Europa-Universität Flensburg, Germany: research domain "Transformation of Energy Systems".

2.2 Open Energy Modelling Initiative

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, open science, and transparent policy analysis. The openmod was established six 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, eastern Europe, and Russia, and more latterly the United Kingdom The organisation itself is not incorporated under law. 20

The openmod mailing list established in October 2014 now numbers over 800 and our discussion forum over 700. The majority of participants reside in Europe. 21

The openmod has held 13 workshops in Europe and elsewhere and these attract researchers, private sector modellers, and on occasion the interested public. The last physical meeting, pre-Covid, took place in Berlin in January 2020, spanned three days, and brought together 190 participants. Planning has begun for the next physical workshop, contingent on pandemic control measures becoming sufficiently relaxed. 22

The openmod does not endorse individual projects, nor does it form and advocate policy positions. This being an unwritten but nonetheless negotiated community norm (Morrison 2019). This submission is therefore solely in the name of the listed submitters. 23

3 Legal issues

This section reviews specific legal issues that we believe should be expressly traversed and resolved by the Commission where possible. Background and context is also provided where this might assist. 24

3.1 Information under statutory reporting

Much of the information that energy system modellers rely upon derives from statutory reporting — also referred to as mandatory public disclosure when dealing with financial information. Reporting divides into system-related and market-related, depending on the implementing statute and its objectives. This 25

section briefly describes the entities involved and then traverses issues of licensing, longevity, and access as they relate to data made public by mandate.

3.1.1 Entities and regulations

The ENTSO-E and ENTSOG umbrella organisations span the electricity and gas transmission system operators (TSO) respectively. Both organisations are required by law to collect and make available system and price information. This written submission deals solely with ENTSO-E although broadly the same matters would apply to ENTSOG. More specifically ENTSO-E is required to operate a "central information transparency platform", known as the Transparency Platform, under Regulation 543/2013 (European Commission 2013). The Platform itself was more recently described and reviewed by Hirth *et al* (2018). This statutory reporting is designed to enhance system security, uncover latent business opportunities, encourage new entrants, and allow customers to better evaluate the supply risks they might face. 26

Similarly, the respective market operators are required to publicly disclose information that could potentially or unduly distort the operations of the various physical, ancillary services, and derivatives markets that fall within the single energy market framework of the Europe Union. The informing statute is Regulation 1227/2011 (European Commission 2011), commonly referred to as REMIT. The Agency for the Cooperation of Energy Regulators (ACER) issues guidelines on applying REMIT provisions (ACER 2016). 27

More specifically, market actors must disclose inside information in an effective way. ACER considers that requirement satisfied once published on an inside information platform (IIP). ACER also sets conditions for these platforms and maintains a list, currently numbering about 15.[8] The European Energy Exchange (EEX) Transparency Platform is one such IIP and is especially restrictive from a public-interest perspective, because the EEX not only operates the platform but also monetises bulk access to its contents. 28

Regulations 1227/2011 and 543/2013 are the key statutes governing transparency and statutory reporting in our domain, but other regulations also mandate the provision of public information and related processes. Our representations are therefore not limited to the two regulations just cited. 29

We requested earlier (point 3 in section 1.1) that IIP reporting be either centrally catalogued or centrally provided and access be offered programmatically, free of charge, and without screening. 30

Section 4.4 further describes community efforts to engage with ENTSO-E and EEX to remedy the lack of open licensing on statutory reporting and related issues. 31

3.1.2 Public licensing

The underlying problem is that the legislation that mandates statutory reporting is silent on public licensing. Thus although the information is publicly available at some minimal level, it remains legally encumbered — or at least, potentially so. The main issue is the uncertain presence of 96/9/EC database protection (Hirth 2020:10). Giannopoulou (2018:5) reviews the status of 96/9/EC database rights in public sector information (PSI) and concludes there are good reasons to believe that no such protection applies. A careful reading of recital 41 of the Database Directive could easily arrive at the same view in regard to information under statutory reporting. Our understanding of the present situation is as follows: 32

> EXAMPLE 1 **ENTSO-E Transparency Platform licensing** ENTSO-E is not able to generally add CC–BY–4.0 licenses to Transparency Platform datasets because there is no consensus among member organisations to do so. This lack of licensing is inhibiting efforts to create a reusable data pool for open energy system analysis. Participants from the Open Energy Modelling Initiative have been liaising with ENTSO-E for some years on this question but have yet to make tangible progress.

[8] https://www.acer-remit.eu/portal/list-inside-platforms

EXAMPLE 2 **EEX public disclosure licensing** Participants from the Open Energy Modelling Initiative have attempted to liaise with EEX on open data licensing to no avail. The submitters believe that the EEX may have a financial interest to deny meaningful public access to statutory reporting, so that fully-featured access can be sold as a commercial service. A requirement to instead apply CC–BY–4.0 licensing would fix this problem in one short paragraph of legislative text.

We now turn attention to platforms operating outside of the European Economic Area (EEA). Another instance shows how a US-based portal is able to harvest from the Transparency Platform and effectively sidestep key legal protections deriving from Europe Union law. The World Resources Institute (WRI) draw from the Transparency Platform to stock their Power Explorer portal and then offer that data under CC–BY–4.0 licensing. We understand ENTSO-E did contact WRI and were told that the Database Directive does not apply to servers operating from within the United States. 33

Further background on open licensing is given in section 3.6, including a graph of license compatibilities. 34

We believe that the Commission should propose that the European Union make CC–BY–4.0 licensing mandatory on all material published under statutory reporting. 35

3.1.3 Data longevity

Turning to data longevity, European Commission Regulation 543/2013 states at 3§1 (European Commission 2013:4) (emphasis added): 36

> "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."

So after five years, datasets hosted by ENTSO-E can go dark and there is little or nothing that energy system analysts could do about that under present conditions. Hence our request for an extended life. Moreover information storage costs are both falling and negligible relative to the social value of transparent data for research and commercial purposes. 37

We have no examples of datasets being removed from the ENTSO-E Transparency Platform due to the above provision. However ENTSO-E has in the past withdrawn historical datasets without prior notice. ENTSO-E Power Statistics, a uniquely high quality data pool of verified historical data, is no longer provided. Hence adding statutory guarantees on greater longevity would seem prudent. 38

3.1.4 Data accessibility

Data access is crucial and datasets must be available in machine-readable formats. Here is a problem our community currently faces in relation to public-interest market reporting: 39

EXAMPLE 3 **EEX public disclosure is viewable but not otherwise recoverable** The European Energy Exchange (EEX) market operator serves public disclosure and other statutory reporting in such a way that it can be viewed on screen but is not machine-readable in any sense. Even selecting individual numbers using mouse clicks is prevented on their website so as to inhibit data usage. In other words, that data is not recoverable or numerically usable in any meaningful context. We have no idea if this practice is legally compliant, but it must certainly be counter to the spirit and intent of the legislation that mandates disclosure in support of the public interest.

Repeated attempts within our community to liaise with the EEX market operator have so far yielded nothing substantive. It would appear that EEX policy is to release as little information as possible to the public under the most restrictive conditions possible. 40

One reason for this conduct, in our view, is that EEX also offers paid bulk access to the data it holds. The fact that information published under statutory reporting is even monetised in this way is regrettable and clearly detrimental for public-interest research that seeks to understand and explore energy systems and energy transition pathways. 41

3.2 Business-to-government (B2G) information publication

Our community has begun to consider the option of obtaining key metrics via mandate, suitably aggregated, anonymized where necessary, and made public under suitable open licensing. Indeed, we recently wrote the following in a community interview in April this year (Cooper 2021): 42

> "Cost information is necessarily estimated in most cases because this information is normally commercially sensitive. Notwithstanding, the European Commission, as well as other governing agencies around the world, could collect cost and performance information under a public interest rationale and make key metrics available in generic form. Future costs and performance projections, which can also be subject to technological learning, are necessarily speculative."

This topic is covered extensively in a High-Level Expert Group on B2G report (HLEG on B2G 2020) which also makes a number of references to "aggregated, anonymized" information. Notwithstanding, the authors barely mention energy and decarbonization data in this exercise. Indeed, it seems that our domain is not much considered when investigating this kind of reporting (nor for HVDS for that matter). 43

In the United States, the Energy Information Administration (EIA) collect and publish this kind of aggregated data — with one excellent example being a US EIA (2020) report on market trends for grid-scale battery storage. It is often very difficult for researchers to obtain such information with this level of fidelity. Moreover, sales and installation figures are usually regarded as commercially sensitive and would certainly not normally be available for republication as open data in support of a public policy study. 44

A Data Act is probably not the place to detail the types of collection and reporting we seek. But the Act should nonetheless make provision for other European agencies or the Commission to define and specify the particulars. 45

Before leaving the topic of upwards B2G information flow, another source of aggregated data is not internal to firms but arises from within emerging smart grid systems. One example could be the Green Energy Hub being developed by Danish national TSO Energinet. This hub will support smart meters and, looking forward, digital twin concepts as well. Being able to recover suitably anonymous policy-relevant information derived from the detailed operational data transiting hubs of this type could prove invaluable. 46

3.3 Closed public-interest databases

We use the term "closed public-interest databases" to describe databases developed in a regulatory context that are not made available to independent researchers under explicit open licensing — as outlined in request 3. Such databases are usually only on offer to incumbent organisations. 47

> EXAMPLE **4** **PEMMDB database remains closed** The Pan-European Market Model Database (PEMMDB) is not accessible as best we can determine.[9] Moreover, results using this particular database are often quoted in presentations aimed at influencing public policy. And yet independent researchers cannot gain access. The United Kingdom CLEARHEADS project would like to connect but has been unable to do so.

[9] The ENTSO-E legal department was contacted on this matter on 11 June 2021 and that appears to be the correct status at that juncture.

We view the practice of closed public-interest databases as deficient, an inefficient use of public resources, and counter to the norms of both science and public policy development. Nor is the PEMMDB database a sole example. 48

3.4 Public records

We simply raise this issue but make no recommendations. As energy system models become more resolved in spatial terms, some are starting to represent individual buildings and streets. As a result, these models draw on municipal public records or associated commercial databases covering such facilities. No strictly personal information is involved, but this analysis does mark a transition regarding the use of detailed public records. It also raises questions about the legal status of such records and what measures are in place or should be in place to protect personal privacy and private life. 49

3.5 Artificial intelligence

Some in our community work with artificial intelligence. Notwithstanding, the focus of this submission is on primary data. That is the kind of data that would also normally be used to create training sets. 50

3.6 License choice and license compatibility

The debate on the best choice of data license or set of licenses rarely looks at the question of legal interoperability — meaning can material under one licence be mixed with material under another license and republished. Rather, the merits of individual classes of license are debated and then the merits of individual licenses. This same discussion takes place within our community too. But this approach tackles the problem from the wrong end. Instead, licence interoperability must be a paramount consideration. For background, please see this recent thread: 51

- Which open data license [10]

The open data-capable license compatibility graph (figure 1) summarises the key issues.[11] 52

One conclusion is that the following subset of open data-capable licenses should be used exclusively: 53

- Creative Commons CC–BY–4.0 license
- Creative Commons CC0–1.0 license
- any license inbound compatible with the above

If the Commission wish to encode this recommendation, the following abbreviated phrase should suffice: “use of CC–BY–4.0, CC0–1.0, or something inbound compatible”. We generally believe CC–BY–4.0 should be favoured over CC0–1.0 because the former instrument encourages scientific attribution and can contribute to provenance tracking. Attribution and provenance are particularly important for information of public interest. That said, we regard the choice of license as indicated as a decision for the primary data provider to make. And there are clearly circumstances where attribution would not be appropriate. Metadata should be licensed under CC0–1.0 to facilitate flexibility of reuse (Kreutzer 2011). 54

3.7 Of limited relevance

For completeness, the following topics are of limited relevance to our community. The bulk of proposed measures in the inception report (European Commission 2021) are aimed at improving business-to-business (B2B) flows while simultaneously protecting private communications. Or promoting fairness 55

[10] https://forum.openmod.org/t/2611

[11] Various national data licenses, including those mentioned in section 1.1, are to be added in due course.

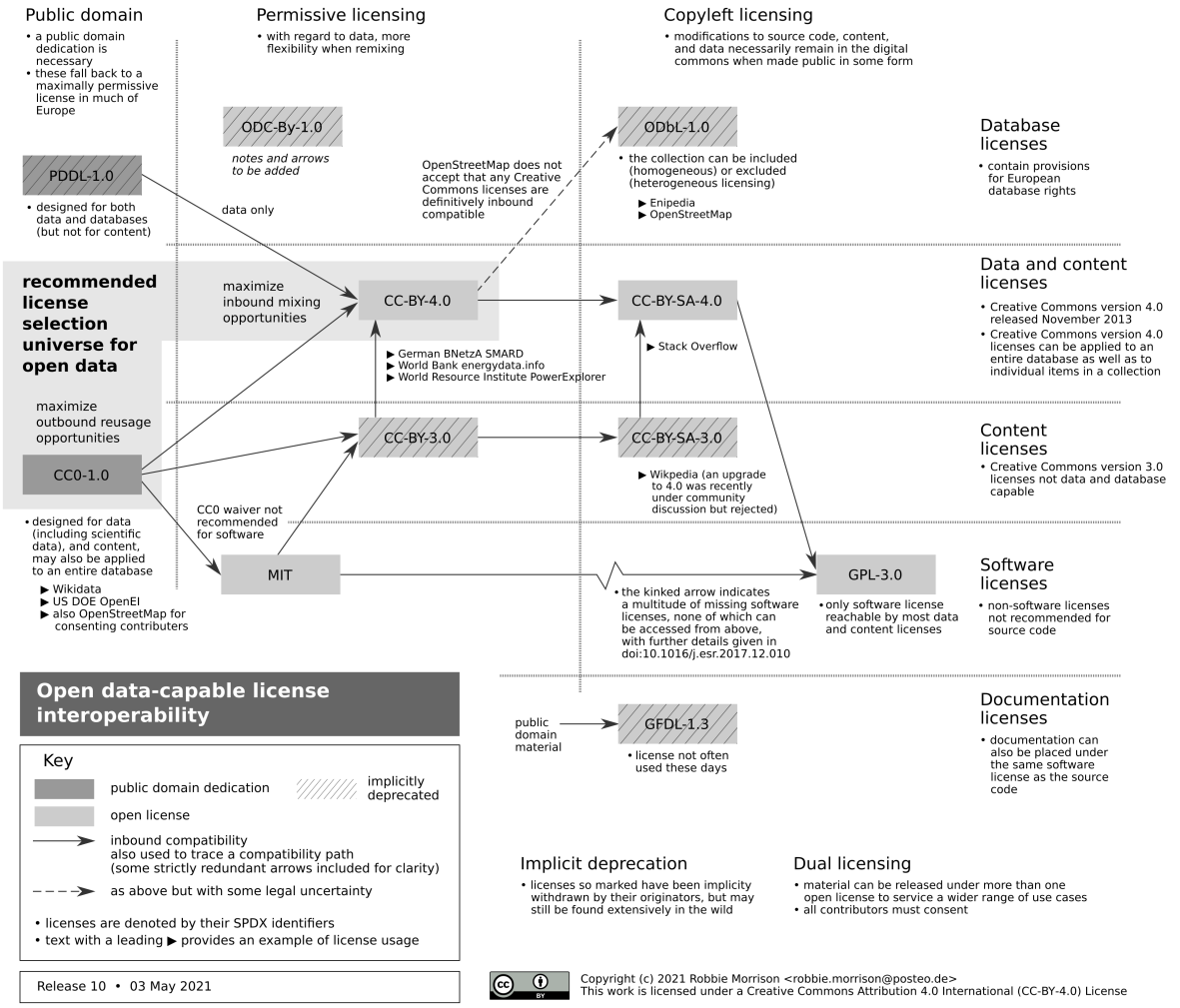

Figure 1: Open data-capable license compatibility graph.

in regards access to B2B information and participation in nascent digital markets.

4 Data management within our domain

This section describes some relevant themes that arise from within our domain of energy system modelling and analysis. This material is primarily intended to provide context. 56

4.1 A paradigm shift

Data practices within our domain appear to be midway through a shift in paradigm from localised management to domain-wide cooperation. The idea of local, as used here, extends to databases and their schemas accessible from the web but not otherwise coordinated.[12] This change in practice has been evident in our community for about two years now and involves a number of component projects implicitly coalescing to facilitate this larger vista. Relevant projects are summarised in this thread: 57

[12] A database schema describes the structure of a given database in a formal way. Broadly a schema may cover the types of data that can be stored, the meanings of that data in a consistent manner, the internal organisation of the database itself, and any other relevant rules, constraints, and conventions that may apply.

- Domain-wide data projects [13]

This switch in data practices necessarily parallels and lags an earlier transition in modelling practices from closed source to open source (Morrison 2018). And just as robust and established open software licensing was material to that coding revolution, recognised and accepted open data licensing is a necessary condition for the new data management paradigm being described. 58

Looking ahead, a logical endpoint would be for energy system data to form part of a common pool under community curation. Data semantics would move from a proliferation of localised schemas, as indicated, to an agreed formal ontology. Metadata would likewise be standardised. New model-agnostic and semi-smart "data systems" would offer coherent and complete datasets covering particular geographic regions and would be widely supported and shared. Data portals would be augmented by a graph of semantic triples relationships, some referencing individual datasets, to realise the concepts behind linked open data (LOD). Intelligent "data buses" would interface between analysts and these web-based resources to address operational deficiencies in the internet, such as inadequate guarantees on persistence. Data provenance and metrics on data quality would be tracked. Future scenarios, used in sustainability analysis, would be managed collectively and the outputs from different models compared and dissected. 59

As indicated, quite a few of the building-block projects needed are currently being realised. To this end, two key high-level projects are covered shortly in sections 4.2 and 4.3. 60

Outside our community, the interface between policy development and numerical analysis is also trending toward open data. Süsser *et al* (2021:1) review that territory and opine: "Our study implies that greater transparency, including open-source code and open data, and transdisciplinary elements in modelling could increase model legitimacy and impact in policymaking". 61

4.2 Data semantics

One key part of the paradigm shift under discussion is a shared semantics. Our community is currently developing an Open Energy Ontology (OEO) (Booshehri *et al* 2021). That particular effort could have built on earlier scientific work but was prevented from doing so through a lack of suitable public licensing, thus: 62

> EXAMPLE **5** **The EnArgus ontology** The EnArgus ontology was developed with German public funding to support public decision-makers but remains proprietary nonetheless (Booshehri *et al* 2021:10). As a result, none of the substantial work invested could be recycled by the OEO project, causing both delays and expense that our society can ill afford at present.

It is worth noting too that similar work on ontologies is being undertaken by the smart systems community under the rubric of "digital twins". Our community informally coordinates with these efforts via the Linux Foundation LF Energy data architecture working group. 63

4.3 Metadata

Metadata policy and practices are being pursued by the EERAdata metadata project (Wierling *et al* under review). This initiative also covers non-open data and extends well beyond our community. The details of that project are not relevant here but what is notable is that the process strives to be inclusive and consensus-driven. 64

Entities providing information under statutory reporting should also adopt these metadata policies and 65

[13] https://forum.openmod.org/t/2553

practices where applicable and practicable.

4.4 Interacting with statutory entities

Participants of the Open Energy Modelling Initiative have been interacting with ENTSO-E for quite some time. For instance, this thread was initiated in October 2017 and explores various opportunities for cooperation: 66

- Cooperating with ENTSO-E on data [14]

We understand that ENTSO-E is committed to open data but that they are a membership-based organisation and that all members would need to consent to a general policy of CC–BY–4.0 licensing before embarking on such. An alternative is, of course, that the European Union mandate such licensing. And after at least four years of engagement with ENTSO-E with no apparent movement, that is exactly what we are asking the Commission to propose. 67

In defence of ENTSO-E, the organisation has generally been very supportive of opening up information. ENTSO-E offer a large number of truly open datasets on the Transparency Platform. The non-open datasets tend to be sourced from outside the TSO domain and instead from the power exchanges (PX). Examples here include price data coming from nominated electricity market operators (NEMO), including the EEX itself, and renewables data deriving from market parties. 68

All attempts within our community to engage with the EEX energy market operator, who also operate an ACER-approved inside information platform (IIP), have to date proved fruitless. Again, mandating suitable open licensing would seem the most sensible path forward. 69

4.5 Examples of good practice

An example of good practice is the recently released database of European energy storage technologies and facilities, provided by the European Directorate-General for Energy (DG ENER). The database is published on the data.europa.eu portal under CC–BY–4.0 licensing and accompanied by documentation. 70

- announcement posting on the Open Energy Modelling Initiative forum [15]
- Database of the European energy storage technologies and facilities [16]

The Open Power System Database (OPSD) is another example of good practice (Wiese *et al* 2019). The OPSD draws from ENTSO-E Transparency Platform, curates that material in a transparent, reproducible, and open fashion, and then republishes it for use by researchers, analysts, and others. The legal status of the datasets served remains nonetheless “grey”. Hence: 71

> EXAMPLE **6** **OPSD portal licensing** The OPSD project (see section 4.5) also serves data drawn from the Transparency Platform but is unable to apply CC–BY–4.0 licensing for the reasons indicated in section 3.1.2.

This legal “greyness” can impede republishing and therefore prevent downstream usage, particularly for risk-averse researchers or researchers working within risk-averse institutions. 72

The German BNetzA SMARD site also sources from the Transparency Platform, but then offers the data served under CC–BY–4.0 licensing. BNetzA does so on a basis of a legal opinion.[17] 73

[14] https://forum.openmod.org/t/367
[15] https://forum.openmod.org/t/2689
[16] https://data.europa.eu/data/datasets/database-of-the-european-energy-storage-technologies-and-facilities?locale=en
[17] One of the submitters requested a copy of that opinion but was politely turned down. That opinion might nevertheless be of interest to the Commission? The overarching message here is, once again, that the current statutory context is inconsistent and unsatisfactory.

EXAMPLE 7 **German BNetzA SMARD portal licensing** The SMARD portal is able to draw material from the ENTSO-E Transparency Platform and make it available under CC–BY–4.0 licensing.

Finally, the Commission could consider transferring and cleaning core material from the ENTSO-E Transparency Platform and EEX reporting to the data.europa.eu portal for reasons of visibility, access, persistence, and licensing. 74

4.6 Community curation and related projects

The paradigm shift mentioned earlier (see section 4.1) is as much social as it is technical. And one of the consequences is a growing sense that the community "takes ownership" the data is collects, curates, and uses. 75

In that regard, the OPSD is an example of a conventional portal (Wiese *et al* 2019). But as indicated, the OPSD has invested considerable effort in cleaning datasets sourced from the ENTSO-E Transparency Platform.[18] A 2018 review found significant shortcomings in data quality on that platform (Hirth *et al* 2018:1056–1061). 76

EXAMPLE 8 **Power plant lists** Obtaining a list of power plants in Europe has proved difficult. Indeed researchers have had to locate lists from several sources and then algorithmically identify the least inconsistent combination (Gotzens *et al* 2019). There surely has to be a better way.

This forum thread provides further background in relation to this topic. It also indicates how scattered and disjoint even basic system information can be to piece together: 77

- Lists of power plants [19]

Several projects, many of which derive from the United States, have been assembling coherent and complete datasets specifically for energy system modelling purposes. These project are more than databases or portals — they contain sophisticated code to check the integrity and completeness of the inventory, perform basic network calculations, provide extensible interfaces for various model frameworks, provide standardised reporting, and more.[20] US-based projects include PowerGenome, PowerSystems.jl and its associated data libraries, and the Open Energy Outlook database. The Spine Toolbox from Finland represents another example. 78

To reinforce the point being made, these new domain-specific "data systems" provide joined-up data that can service any number of projects and modelling teams working with the same local, national, or regional scope. Moreover, these data systems encourage cooperation on the collection and maintenance of the assembled information and facilitate cross-model comparisons. Both aspects are vital for robust analysis. We cannot help but speculate that many of these developments originated in the United States because the bulk of their energy sector data is legally unencumbered.[21] Or expressed in the converse, very little of 79

[18] Indeed, we understand that some member organisations return spreadsheets to the Transparency Platform via email with the column identifiers unilaterally altered.

[19] https://forum.openmod.org/t/2526

[20] The term "model framework" is used within our community to separate the underlying application software from specific instances that have been populated with data. These instances are then generally known as "models" and a specific model may then be used to run any number of exploratory "scenarios". These scenarios are then interpreted relative to some stated "reference case" scenario, often but not necessarily some expectation of business-as-usual. The overall technique is known as comparative analysis.

[21] Work produced by federal employees is automatically public domain, datasets lacking creative input are not subject to copyright, and there has never been legislated database protection.

this effort is legally feasible within Europe without the primary data first being released under CC–BY–4.0 or CC0–1.0 licensing.

Another community resource, more technical than informational, is the Open Energy Platform (OEP).[22] The OEP offers backend services to energy modelling teams and can be used to define and host framework-independent future scenario sets for use across any number of open projects (Reder *et al* 2019). 80

We note in passing, that the Commission has recently undertaken exploratory work for what may be a somewhat similar platform (Gaschnig *et al* 2020). 81

4.7 Looking forwards

It can be expected that some of the themes articulated in the submission will continue. Speculatively: 82

- that data is increasingly viewed and treated as a domain-wide common pool resource
- the shift from classical schemas to semantic triples graphs with embedded datasets will continue
- parallel work on a community-agreed ontology and on metadata practices will proceed
- the ethos of open science and open policy analysis will gain ground, be it for ethics or pragmatics
- conversely, non-open science and non-open policy analysis will become increasingly untenable

5 Discussion

We wish to underscore again how debilitating it is to not have good access to important public-interest datasets licensed under Creative Commons CC–BY–4.0. 83

Moreover, the adoption of that particular data license looks likely to unlock a new data management paradigm within the domain of open energy system modelling based on linked open data (LOD) concepts. This new paradigm has the potential to radically improve public policy development in complex areas like energy systems analysis. In addition, this emerging data revolution parallels, lags, and is synergistic with a similar revolution underway in model development, namely the uptake of open source models. Taken together, these two changes in paradigm should bring substantial benefits to society through higher quality, more extensive, more transparent, and more rapid public policy analysis and advice. 84

We again highlight the need to fully decarbonize our energy systems in three short decades. Indeed, energy systems, broadly defined, may well need to be substantially net-negative in order to counteract emissions from hard-to-decarbonize sectors such as agriculture. 85

We note too that, in general, the transmission system operators have been much easier to engage with than the market operators. And reiterate the point about the EEX serving statutory information in formats that are viewable but not recoverable or numerically usable without significant payment. 86

Furthermore, to reinforce that much of the public information available is in relatively poor state and often lacks provenance. 87

The Commission should not simply introduce a new Data Act, but also take the opportunity to repair other statutes including the 1024/2019 Open Data Directive as indicated earlier (point 13 in section 1.2). 88

The submitters would be happy to provide the Commission with further information, sources, and examples on request. And many in our community would doubtless be willing to engage with the Commission and explain trends, expand on their experiences, and/or discuss their research interests in the context of open data, open science, and public good. 89

[22] The Open Energy Platform should not be confused with a energy sector data brokerage service with a similar name being developed by United Kingdom-based Icebreaker One.

Acknowledgements

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 open fonts.

Abbreviations

ACER	Agency for the Cooperation of Energy Regulators
API	application programming interface
B2G	business-to-government information flows
EEA	European Economic Area
EEX	European Energy Exchange
ENTSO-E	European Network of Transmission System Operators for Electricity
ENTSOG	European Network of Transmission System Operators for Gas
IDR	industrial data right (speculative)
IIP	inside information platform
HVDS	high-value dataset (European Commission 2019)
ICT	information and communications technology
LOD	linked open data
NEMO	nominated electricity market operator
ODD	2019/1024 Open Data Directive (European Commission 2019)
OEP	Open Energy Platform
openmod	Open Energy Modelling Initiative
PEMMDB	Pan-European Market Model Database
PSI	public sector information
PX	power exchange
REMIT	Regulation on Wholesale Energy Market Integrity and Transparency
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)
US EIA	United States Energy Information Administration
WRI	World Resources Institute

References

ACER. 2016. *ACER guidance on the application of REMIT — Updated 4th edition: guidance on the application of Regulation (EU) No 1227/2011 of the European Parliament and of the Council of 25 October 2011 on wholesale energy market integrity and transparency*. Fourth edition. Ljubljana, Slovenia: Agency for the Cooperation of Energy Regulators (ACER). June.

Booshehri, Meisam, Lukas Emele, Simon Flügel, Hannah Förster, Johannes Frey, Ulrich Frey, Martin Glauer, Janna Hastings, Christian Hofmann, Carsten Hoyer-Klick, Ludwig Hülk, Anna Kleinau, Kevin Knosala, Leander Kotzur, Patrick Kuckertz, Till Mossakowski, Christoph Muschner, Fabian Neuhaus, Michaja Pehl, Martin Robinius, Vera Sehn, and Mirjam Stappel. 2021. Introducing the Open Energy Ontology: enhancing data interpretation and interfacing in energy systems analysis. *Energy and AI*. **5**: 100074. September. ISSN 2666-5468. doi 10.1016/j.egyai.2021.100074.

Commission, European. 2014. Commission notice: guidelines on recommended standard licences, datasets and charging for the reuse of documents. *Official Journal of the European Union*. **C 240**: 1–10. July.

Cooper, Danielle. 2021. Emergent data community spotlight: an interview about energy modeling with the Open Energy Modelling Initiative. April. Creative Commons CC–BY–4.0 license.

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. 8 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. 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. 2021. *Inception impact assessment: Data Act (including the review of the Directive 96/9/EC on the legal protection of databases) — Ares(2021)3527151*. Brussels, Belgium: European Commission. May. Lead DG: CNECT/G1. Landing page for download given. Download name: 090166e5ddb6bc31.pdf.

Gaschnig, Hannes, Diana Süsser, Andrzej Ceglarz, Vassilis Stavrakas, George Giannakidis, Alexandros Flamos, Antina Sander, and Johan Lilliestam. 2020. *User needs for an energy system modeling platform for the European energy transition — Deliverable 1.2. Sustainable Energy Transitions Laboratory (SENTINEL) project*. Potsdam, Germany: Institute for Advanced Sustainability Studies (IASS). November. doi 10.48481/iass.2020.059. Creative Commons CC–BY–4.0 license.

Giannopoulou, Alexandra. 2018. Understanding open data regulation: an analysis of the licensing landscape. Chapter 6, pp. 101–125. In: *Open data exposed*. The Hague, the Netherlands: TMC Asser Press. ISBN 978-94-6265-261-3. doi 10.1007/978-94-6265-261-3_6.

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.

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. Creative Commons CC–BY–4.0 license.

Hirth, Lion, Jonathan Mühlenpfordt, and Marisa Bulkeley. 2018. The ENTSO-E Transparency Platform: a review of Europe's most ambitious electricity data platform. *Applied Energy*. **225**: 1054–1067. September. ISSN 0306-2619. doi 10.1016/j.apenergy.2018.04.048. Creative Commons CC–BY–4.0 license.

HLEG on B2G. 2020. *Towards a European strategy on business-to-government data sharing for the public interest — Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing — KK-01-19-750-EN-N*. Luxembourg: Publications Office of the European Union. ISBN 978-92-76-11422-2. doi 10.2759/731415. Licenced under European Commission reuse policy.

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.

ENTSO-E. 2017. *ENTSO-E Transparency Platform*. Belgium, Brussels: European Network of Transmission System Operators for Electricity. 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. Creative Commons CC–BY–4.0 license.

Morrison, Robbie. 2020. *Submission on a European strategy for data with an emphasis on energy sector datasets — Release 08*. Berlin, Germany. 30 May. Creative Commons CC–BY–4.0 license.

Morrison, Robbie. 2021. *Submission on a proposed Data Act for the European Union from the perspective of energy system analysis — Release 07*. Berlin, Germany. 25 June. A reference to the current release this document. Creative Commons CC–BY–4.0 license.

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. Creative Commons CC–BY–4.0 license.

Open Energy Modelling Initiative. ongoing. *openmod forum*. Open Energy Modelling Initiative. Discussion server.

Reder, Klara, Carsten Pape, Mirjam Stappel, Hannah Förster, Lukas Emele, Christian Winger, Ludwig Hülk, Christian Hofmann, Editha Kötter, Martin Glauer, and Till Mossakowski. 2019. *Scenario data on the Open Energy Platform (SzenarienDB on the OEP): a web-platform to improve transparency and reproducibility of energy system analyses — Poster*. Kassel, Germany: Fraunhofer Institute for Energy Economics and Energy System Technology (IEE). Creative Commons CC–BY–4.0 license.

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. Creative Commons CC–BY–4.0 license.

Süsser, Diana, Andrzej Ceglarz, Hannes Gaschnig, Vassilis Stavrakas, Alexandros Flamos, George Giannakidis, and Johan Lilliestam. 2021. Model-based policymaking or policy-based modelling? How energy models and energy policy interact. *Energy Research and Social Science*. **75**: 101984. May. ISSN 2214-6296. doi 10.1016/j.erss.2021.101984. Creative Commons CC–BY–4.0 license.

US EIA. 2020. *Battery storage in the United States: an update on market trends*. Washington DC, USA: US Energy Information Administration (US EIA). July.

Wierling, August, Valeria Jana Schwanitz, Sebnem Altinci, Maria Balazińska, Michael J Barber, Mehmet Efe Biresselioglu, Christopher Burger-Scheidlin, Massimo Celino, Muhittin Hakan Demir, Richard Dennis, Nicolas Dintzner, Adel el Gammal, Carlos M Fernández-Peruchena, Winston Gilcrease, Pawel Gladysz, Carsten Hoyer-Klick, Kevin Josho, Mariusz Kruczek, David Lacroix, Malgorzata Markowska, Rafael Mayo-García, Robbie Morrison, Manfred Paier, Giuseppe Peronato, and Mahendranath Ramakrishnan. under review. Advancing FAIR metadata standards for low carbon energy research. *Energy Strategy Reviews*. ISSN 2211-467X. Code ESR-D-21-00036. To be open licensed.

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 : local. doi 10.1016/j.apenergy.2018.11.097. Post-print.

▢