EERA data

Winter 2023

Harvesting the value of low carbon energy research data

Policies towards better strategies for open and FAIR energy research data

POLICY BRIEF

KEY IDEA

The **prospective potential** from the digitalisation of the energy sector for a just, green and economically viable transition has previously been stressed [1]. Sustainable energy innovation could be propelled in many ways, being able to lead into a new technology revolution for the betterment of society.

However, **these scenarios will not be realised unless the challenges of data accessibility, data interoperability and data reusability are solved. Substantial intellectual and financial investments are required** to make data fit for the digital age with autonomously acting agents in the energy sector.

TOWARDS FUNCTIONING ENERGY DATA MARKETS

Despite optimistic prospects, functioning energy data markets are lacking and business models for FAIR and open energy data are scarce. The problem starts with a lack of methods to estimate the value that lies in energy data, in particular if shared, as the economic value of the data is not immediately obvious.

As a matter of principle, data valuation can be derived from the cost of either reproducing a dataset or replacing its utility through alternatives. The value of data can also be based on the added value that is expected to arise from data use or from saving costs by using them. For example, when knowing the preferences of an end-user of an energy service, pricing solutions can be tailored to the customer and thereby become more profitable for the customer and the service provider.

As regards shared data, a consortium not only can profit from previously endorsed work but also from shared overhead costs connected to data governance and data applications. Moreover, its members can take advantage of being part of a community of experts with common interests. Knowing the value of data is a precondition for being able to identify business models and establish markets for energy data.

---

[1] Realising a human-centred digitalization of the energy sector; https://www.eeradata.eu/component/attachments/?task=download&id=953; November 2022.

THE LICENCE: THE BASIS FOR OPEN DATA

By default most data are currently not open, even if researchers have had the intention to do so. The reason is a lack of knowledge and practice of proper licensing. According to the Open Data Institute (ODI) , *"there are lots of aspects to openness, but at its most fundamental, the key is how the data is licensed. Data that doesn't explicitly have an open licence is not open data"* [2].

A created piece, either an article, a photograph or a database, belongs to the person (or the organisation) who puts the effort into developing it and owns the rights over it. To protect the rights, licences are created, which usually address *"the manner of delivery, data security policies, as well as maintenance and control over data"* [3]. Moreover, the licence not only contains the licensor's ownership of the data but also the licensee's use of the data. This means that licences how the owner explicitly gives someone else permission to use the work.

THE IMPORTANCE OF LICENCING FOR OPEN DATA RESEARCH

Promoting open science and open data requires correct licencing, but the problem is that the data are either not licenced or the default licence is often the wrong choice. An open licence entails that *"anyone can freely access, use, modify, and share [data and content] for any purpose (subject, at most, to requirements that preserve provenance and openness)"* [4]. As such, quoting ODI [5], an open licence allows others to do things like:

- republish the content or data on their own website
- derive new content or data from the one created
- make money by selling products that use that content or data
- republish the content or data while charging a fee for access

However, there are **two restrictions that an open licence can still place:**

1. Reusers of the data or content must give attribution to the source.
2. Reusers must publish any derived content or data under the same licence (share-alike).

Given the conditional nature of these restrictions, an open licence might in-

[2] Publisher's Guide to Open Data Licensing; https://theodi.org/article/publishers-guide-to-open-data-licensing/; 15 December 2013.
[3] What is data licensing? https://www.infoclutch.com/infographic/what-is-data-licensing; 15 October 2019.
[4] The Open Definition http://opendefinition.org/. For the full open definition look at: Open Definition 2.1 http://opendefinition.org/od/2.1/en/.
[5] Publisher's Guide to Open Data Licensing; https://theodi.org/article/publishers-guide-to-open-data-licensing/; 15 December 2013.

-corporate both conditions, one of the two or neither of them. This entails that the content or data could be made available under one of the following levels of licence:

- **Public domain:** the licence has no type of restrictions. It will be similar to waiving the ownership rights over the content or data.
- **Attribution licence:** the licence entails that reusers must give attribution to the source.
- **Attribution and share-alike licence:** the licence requests reusers to attribute the derived content or data to the original source and to share it under the same licence.

Regarding the type of open licences, there are two sets: 1) Open licences for creative content and 2) Open licences for databases. ODI suggests users select one of these sets rather than creating a custom licence. Why? Besides the fact that it requires less work from the owner, it ensures that the legal aspects of the licence are correct and "makes it easier for reusers to know what they can do with your data" [6].

Zooming into open licences for databases, ODI recommends Creative Commons 4.0 licence for data. It is also possible to use a similar set of licences created specifically for databases on the Open Data Commons. There are three levels to choose from:

Level of licence	Open Data Commons Licence
Public domain	PDDL
Attribution	ODC-by
Attribution & share-alike	ODbL

BEYOND HUMAN READABILITY

As EERAdata has been advocating throughout the project, quality open data require that humans can understand them and machines can manipulate them. Therefore, the licensee should indicate the licence using both a human-readable description and computer-readable metadata. The human-readable descriptions and marks are suggested on the Creative Commons and Open Data Commons websites:

[6] Publisher's Guide to Open Data Licensing; https://theodi.org/article/publishers-guide-to-open-data-licensing/; 15 December 2013.

- Creative Commons licence chooser
- Open Data Commons licences

For the computer-readable metadata, the Publisher's Guide to the Open Data Rights Statement Vocabulary describes the process to do this.

However, the open licence alone is not enough to guarantee the usability of the data. According to the data.europa.eu e-learning programme, community-driven standards define the legal, practical, technical and social requirements for an open dataset to be useful:

Legal requirements

- Guarantee the protection of sensitive information such as personal data.
- Preserve the rights of the data owners.
- Promote the correct use of data.

Practical requirements

Publishers must:

- Provide links to the data from their website.
- Update the data regularly if they change.
- Commit to continue to make the data available.

Technical requirements

- The form in which the data are published.
- The structure of the data.
- The channels through which the data are available.

Social requirements

This requisite entails the importance of having an engaged community of users for the sustainability of the data. In this sense, datasets should have:

- Active support channels.
- Discussion groups and forums.
- Published how-to guides on working with the data.

BETTER STRATEGIES FOR OPEN AND FAIR ENERGY RESEARCH DATA

Current approaches to overcome data that are siloed, not interoperable and exclusive fall short in three ways:

1 First, the energy research domain is characterized by largely isolated projects that work on overcoming the FAIR and open data challenges. Together with top-down supporting initiatives set up by, for example EOSC, they find solutions that are unfortu-

-nately not interoperable or too generic to be domain applicable. This perpetuates the need for enlarging the community of practice for FAIR and open energy data and at the same time highlights the potential for moving towards community-wide data standards. Still, for most researchers, the implementation of FAIR and open data principles continues to be an expensive side activity with few accompanying career merits.

2 Second, best practices from open data collaboration in the energy industries such as Open Subsurface Data Universe™ show that successful platforms are able to attract relevant stakeholders and businesses within the scope of 10 years and when provided with proper resource commitment. For low carbon energy research, such critical mass composed of individual researchers, universities and research institutions is still lacking. Links to practitioners and the private sector are also weak. This results in lost potential for scaling up valuable lessons learnt from manifold use cases and isolated projects.

3 Third, funding policies currently limit the possibilities to investigate uncharted territory due to the need of minimizing and mitigating any research project risks. Consequently, the time that it takes to elicit research insights and bring innovative ideas for FAIR and open data services to the market is prolonged unnecessarily.

POLICY RECOMMENDATIONS TO REAP THE VALUE OF ENERGY RESEARCH DATA

Realising the value that lies in low carbon research data rests upon the ability to **quantify the business potential of FAIR and open data, attract a critical mass of low carbon research stakeholders and raise critical resources to finance collaborative platforms and the necessary data infrastructure.** The following policy measures could support this community building process.

- Fund the identification, experimentation and implementation of open energy data business models and open innovation in general, hedging the connected risks of such endeavours.

- Fund shared innovation consortia and digital community networks that connect European researchers to private sector actors, as well as energy cooperatives and businesses that commit to open-source data and tools.

- Steer and enforce the setting of standards for open and FAIR data, but also their implementation.
- Support the establishment of markets for open technologies, including the enforcement of open metadata by and large.
- Focus licencing on access rights rather than property rights.

ABOUT EERADATA

The EERAdata project develops, explores, and tests a FAIR (findable, accessible, interoperable, re-usable) and open data ecosystem in the low-carbon energy field. The new data infrastructure will be established through broad community involvement and applied in four use cases.

EERAdata also implements an open platform for seamless access to energy data and establishes a pool of experts and data stewards to facilitate a mental shift towards FAIR and open data practices.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 883823.

EERA data