Open Source and Open Data in energy system modelling:

Why NGOs should engage with the topic

Studies on the transformation of the energy system are often opaque: nobody can access the underlying data and assumptions, and important variables that strongly affect the results (e.g. uptake of electric vehicles, discount rates, cost decreases for solar) are not disclosed. Therefore, it is not surprising that the call for disclosure of data is becoming louder and louder. In an effort to open up and share their insights with broader society, the EU, ENTSO-E, ACER and others have already started to disclose more and more data.

The practice of disclosing data and code has become established in areas where complex interrelationships can only be investigated with computer models (e.g. climate systems). The advantage here being that it is cost-efficient to share laboriously collected data with others and enable people from different institutions and fields to collectively use and further develop the results of one's work. This means that results can be automatically checked or peer-reviewed and that, by combining individual modules, models can be further developed in a considerably shorter time.

The case of energy system modelling

The transformation of energy generation is a complex task, which is by no means simply a question of shifting from fossil to renewable energy sources. The models currently in use need to be further developed when it comes to mapping decentralised energy sources and flexible consumers. On top of that, the interactions which must be included increase exponentially when mapping energy usage across *all* sectors. In order to plan and establish a cost-efficient energy system, changes in energy usage, for example in the mobility sector, must also be considered. Modelling must be able to take these interactions into account.

In addition to the energy system becoming more complex, rate of related questions arising from politics and society is also on a steep increase. Simultaneously, given advancing climate change, the time to wait for results is running out. Modelling must therefore become more comprehensive and able to react more quickly. In this context, the usage of the open source approach seems to be the only way to generate answers to urgent questions faster – despite increasing complexity.

However, the majority of NGOs have difficulties in providing data and code of energy models with open source licences. One reason for this is that the responsible institutions regard their data collections and model code as a unique selling point, which they wish to protect out of economic interest. This sometimes creates a situation in which not even the institutions that paid for the study can verify the data. This makes it hard to understand what drives the results and to compare different studies. Furthermore, if an NGO commissions a study and the data and code remain with the consultant, they are locked in to using the same consultant for all related future studies.

Another reason for not open sharing data, is that work in the political sphere demands that results be ready at a particular point in time, meaning sometimes quality takes second place. In these cases, it is perceived as advantageous that data is not openly

available, and a control check cannot take place. A third reason for cautiousness is a certain concern that data could be abused by adversaries, in order to discredit studies or support contradictory interpretations. While these concerns are not entirely unfounded, resolute implementation of data disclosure would also force parties with conflicting interests to disclose their data, if they want to be seen as credible. Given that the facts speak for climate protection and sustainability, there is no reason for NGOs to fear transparency, but every reason to embrace it.

Finally, a lack of knowledge about data licensing and how to deal with the resulting technical and legal issues can also prevent the provision of data and source code. All this is actually easier than one might think. See a quick guide below.

Call for the establishment of an open source culture

If you have a good hand of cards, you shouldn't fear playing them face up – in fact, on the contrary. Forcing those who seek to counteract the fight against climate change to reveal their cards (data) could prove to be an effective way to reduce delays in system transformation.

Open source not only offers transparency, but also helps us to react quickly, as multiple actors can work on one issue at the same time, thus perfecting the same data set. In order to be able to jointly use model building blocks, it is necessary to design the interfaces so that they can easily be integrated into other models. The open source community has already developed a number of such building blocks, and the standards used for this purpose are constantly being developed further.

We require an efficient use of all available informational resources in order to be able to deal with the many questions which arise around the topic of climate change. Therefore, NGOs in particular should place themselves at the forefront of the open source movement.

The PAC scenario as an example of good practice

Under the umbrella of the PAC project (Paris Agreement Compatible Scenarios for Energy Infrastructure), the European Environmental Bureau and Climate Action Network Europe endeavoured jointly in a bottom-up, collective research exercise to develop the key assumptions for a European-wide energy scenario that is aligned with the objective to limit global warming to 1.5°C degree. The PAC scenario suggests a trajectory with at least 65% greenhouse gas emission reductions by the year 2030 and net-zero emissions in a 100% renewable energy system by 2040.

Both EEB and CAN-E support the open source modelling community and decided to publish the PAC scenario under an open data licence. After publication, the data was immediately taken up by different research institutions.

The first study results comparing the PAC scenario data are already available. A joint study by the TU Berlin/ DIW[1] was able to show the macroeconomic benefits of the decarbonisation path described in the PAC scenario, make statements about total emissions up to 2050 and prove that a stable energy supply with the generation mix assumed in the PAC scenario and with the inclusion of flexible consumers is possible even in extreme situations. The researchers also highlighted their interest in comparing PAC scenario findings with ongoing scenario building for 100% renewable energy systems as well as potentially feeding PAC scenario data into their GENeSYS-MOD modelling suite.

In another project, the Karlsruhe Institute of Technology and Aarhus University have worked with the PAC scenarios by feeding the PAC scenario data into their open source model PyPSA-Eur-Sec[2]. This has provided new insights into the reliability of supply at country level, intra-European energy flows on an hourly basis and the optimal balance between renewables energy supply, storage, electrolysers and other loads.

Advantages of open source data and code at a glance:

- **Transparency:** A fundamental quality of open source software and open source data is transparency. Transparent structures enable results to be traced, thus increasing trust in the necessity of measures and therefore **acceptability.** This is a central issue, for example when dealing with infrastructural measures such as grid development for the energy transition.
- **Promotion of decision-making processes:** If the reasoning behind measures is recognised and accepted by various actors, then results can be discussed more easily, and compromises found more quickly.
- **Increase in quality:** Open source enables developers from different institutions and elsewhere to work simultaneously on the expansion of a model system and the joint use of the software modules developed by the various teams. This way, errors can be identified more quickly, and simulations improved. This increases the speed of further developments and helps make simulations more robust.
- **Cost savings:** At present, similar modelling questions are repeatedly investigated using public funds. With open source software, parts of programmes can be used repeatedly, further developed and adapted. This leads to reduced development costs.
- **Avoidance of consultant lock-in:** If studies are commissioned based on open data and open code, the data can be reused by the commissioning entity or a different consultant in the future.

---

[1] ("Make the European Green Deal Real, Hainsch, et al. 2020)

[2] See this presentation by KIT and Aarhus University at the 4th PAC Scenario Workshop

Licensing is simple! What you should do to be open

- **Establish who owns the copyright before you distribute under an open licence.** Only the copyright-holder can publish a document under an open licence. If you own the data just add a sentence saying, "This document is distributed under licence X".
- **Release any data or code you want to share under an open licence**. A licence is a document which tells other users what they are allowed to do with the data or code, e.g. copying it, sharing it, etc. There are several standard licences you can use. See an overview of common licences below.
- **When commissioning studies from consultants, insist on open data and code**. This way you can reuse it, as can other NGOs, and you can benefit from the open work of others.
- **Help campaign for an open energy modelling ecosystem**. The more people work openly, the more benefits everyone reaps – network effects!
- **Making code public** is a bit more complicated but most programmers know what to be aware of.

Worth knowing in open source / open source licences:

- **A licence tells users what they can do with your data or code.** For example, whether you can reuse it, modify it, or distribute it. You can also make requirements of users, for example that they credit you as the source ("attribution") or that they release modifications under a similar licence ("share-alike"). Without a licence, default intellectual property restrictions apply.
- **There are standard licences which most people use.** Don't reinvent the wheel! Standard licences have been written by experts and legally tested in court.
- **For data, the Creative Commons Licence is often used.** For example, the European Commission recommends the "Creative Commons Attribution 4.0 International Licence". This licence has also been used by the German Federal Network Agency (BNetzA for the SMARD platform), ENTSO-E for TYNDP 2020 Scenario Data, and many others.

Which open licences do people typically use?

A selection of open data projects with their licences

dataset/report	licence	link to dedication
European Commission data	CC BY 4.0	announcement

dataset/report	licence	link to dedication
Bundesnetzagentur SMARD platform for electricity market data	CC BY 4.0	SMARD Datennutzung
eurostat	custom	custom re-use notice
French energy network data (RTE, GRTgaz, etc.)	Licence Ouverte	Licence Ouverte
JRC-EU-TIMES	CC BY 4.0	zenodo record
JRC report on net-zero scenarios	CC BY 4.0	JRC report
OpenStreetMap	ODbL	Copyright page
Wikipedia	CC BY SA 3.0	Wikipedia License

CC BY 4.0 = Creative Commons Attribution 4.0 International License

CC BY SA 3.0 = Creative Commons Attribution-ShareAlike 3.0 Unported License

ODbL = Open Data Commons Open Database License

Here is a selection of open model code projects with their licences:

model	licence	link to dedication
calliope	Apache 2.0	LICENSE
oemof	MIT	LICENSE
SWITCH	Apache 2.0	LICENSE
TIMES	GPL 3.0	LICENSE
PyPSA	GPL 3.0	LICENSE

Further links:

- Open Data for Electricity Modeling Report on behalf of the Federal Ministry for Economic Affairs and Energy – Neon Neue Energieökonomik
https://www.bmwi.de/Redaktion/EN/Publikationen/Studien/open-Data-for-electricity-modeling.html
- Open Mod – Open Energy Modelling Initiative. Forum FAQ for people new to open modelling
https://forum.openmod-initiative.org/t/openmod-faq-for-people-new-to-open-modelling/1016