Energy Strategy Reviews 19 (2018) 63–71

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.ees.elsevier.com/esr

Opening the black box of energy modelling: Strategies and lessons learned

Stefan Pfenninger [a,*], Lion Hirth [b,c,d], Ingmar Schlecht [e], Eva Schmid [f,g], Frauke Wiese [h], Tom Brown [i], Chris Davis [j], Matthew Gidden [k], Heidi Heinrichs [l], Clara Heuberger [m], Simon Hilpert [n], Uwe Krien [o], Carsten Matke [p], Arjuna Nebel [q], Robbie Morrison [r], Berit Müller [o], Guido Pleßmann [o], Matthias Reeg [s], Jörn C. Richstein [t], Abhishek Shivakumar [u], Iain Staffell [m], Tim Tröndle [v], Clemens Wingenbach [n]

[a] ETH Zürich, Switzerland
[b] Neon Neue Energieökonomik GmbH, Berlin, Germany
[c] Hertie School of Governance, Berlin, Germany
[d] Mercator Research Institute on Global Commons and Climate Change (MCC), Berlin, Germany
[e] University of Basel, Switzerland
[f] GERMANWATCH, Berlin, Germany
[g] PIK, Potsdam, Germany
[h] Danish Technical University, Denmark
[i] FIAS, University of Frankfurt, Germany
[j] University of Groningen, The Netherlands
[k] IIASA, Laxenburg, Austria
[l] Institute of Energy and Climate Research, FZ Jülich, Germany
[m] Imperial College London, UK
[n] University of Flensburg, Germany
[o] Reiner Lemoine Institute, Germany
[p] DLR Institute of Networked Energy Systems, Oldenburg, Germany
[q] Wuppertal Institute, Germany
[r] Energy Consultant, Berlin, Germany
[s] DLR, Germany
[t] DIW, Berlin, Germany
[u] KTH Royal Institute of Technology, Stockholm, Sweden
[v] Cambridge University, UK

ARTICLE INFO

Article history:
Received 20 July 2017
Received in revised form
6 November 2017
Accepted 10 December 2017
Available online 27 December 2017

Keywords:
Open source
Open data
Energy modelling

ABSTRACT

The global energy system is undergoing a major transition, and in energy planning and decision-making across governments, industry and academia, models play a crucial role. Because of their policy relevance and contested nature, the transparency and open availability of energy models and data are of particular importance. Here we provide a practical how-to guide based on the collective experience of members of the Open Energy Modelling Initiative (Openmod). We discuss key steps to consider when opening code and data, including determining intellectual property ownership, choosing a licence and appropriate modelling languages, distributing code and data, and providing support and building communities. After illustrating these decisions with examples and lessons learned from the community, we conclude that even though individual researchers' choices are important, institutional changes are still also necessary for more openness and transparency in energy research.

1. Introduction

The history of energy system planning is primarily closed and proprietary, having been pursued by research institutions,

* Corresponding author.
E-mail address: stefan.pfenninger@usys.ethz.ch (S. Pfenninger).

https://doi.org/10.1016/j.esr.2017.12.002

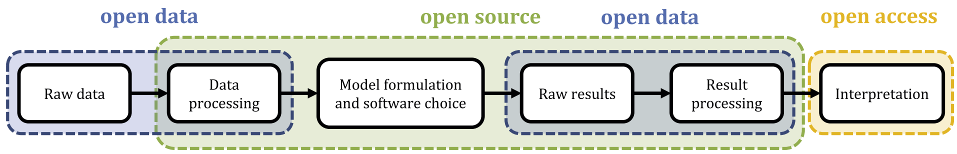

Fig. 1. Open data, open source, and open access in relation to the energy modelling process.

government agencies, and large, vertically-integrated utilities that were under no obligation to reveal their modelling assumptions or methodologies. This may have been acceptable in a conventional energy system with only a few players, but the requirements on energy system planning are changing significantly driven by the advent of liberalised, regulated markets and the need for deep reductions in greenhouse gas emissions [1]. In addition, the rapid deployment of wind and photovoltaics (PV) and growing importance of energy storage require models with sufficient spatial and temporal resolution to accurately assess these new technologies.

Open energy modelling is desirable for many reasons [2]. First, open code and data improve scientific quality if they lead to more transparency and reproducibility, and thus permit effective collaboration across the science-policy boundary. This is particularly important in energy policy, an urgent and highly contested topic. More transparent modelling is desirable from a regulatory and political perspective, as opening up decision processes and the reasoning behind them may lessen public opposition to new legislation and infrastructure. By reducing parallel efforts and allowing researchers to collaborate and share the burden of developing and maintaining large codebases and datasets, openness also enables increased productivity. We believe research funded with public money should be freely available to the public.

The open code, open data and open science movements are only slowly leading to models and data being opened up [3]. In order to classify as "open", the data or model code needs to be both accessible and legally usable. Hence, we pragmatically define open code, open data, and open models as artefacts that are available under commonly used licences, which allows re-use without undue restrictions.[1] The history of open energy modelling can be traced to the release of the model Balmorel in March 2001 [5] (albeit without a licence until 2017), followed by early attempts with a now abandoned GPL-licensed model called deeco in 2004 [6]. After a long pause, the release of the modelling frameworks OSeMOSYS in 2009 and TEMOA in 2010 [7,8] has spearheaded several dozen open projects., In 2014, a group of modellers founded the Open Energy Modelling Initiative[2] in order to better coordinate the further development of open models and data. Nevertheless, the mainstream approach to energy modelling is often still proprietary and opaque, even where it directly feeds into policy [9]. Underlying reasons are manifold; however, commercial sensitivity, lock-in to proprietary models, lack of awareness, institutional and personal inertia, and fear of losing competitive advantage are some factors at work in academic research [2]. In addition, the boundary between academic research and commercial consulting is blurred in the energy modelling field, so different actors have different aims and incentives (e.g., selling data or software to customers, or publishing policy-relevant results).

Based on our experience, this article addresses what we perceive to be the crucial factors limiting openness of energy data and models: the lack of practical knowledge as well as personal and institutional inertia. It is not intended as a review of energy modelling methods or tools, but a how-to guide for researchers considering to open up their model code and data. The article proceeds as follows. First, we briefly introduce the key steps in energy modelling and how they link to aspects of openness. We then walk through the practical steps energy modellers must think about and choices they must make when deciding to go open. Finally, we describe four examples that provide further context for these key choices, before concluding with the most important challenges that remain to be overcome in the institutions that shape the research landscape.

2. The energy modelling process

The kinds of models we discuss here can be summarized by two overarching terms: energy system models and power system models. The former group includes in particular long-term planning models that generate deployment scenarios over decades into the future, and has recently seen a renaissance with an increased focus on variable renewable generation. The latter group has a clear focus on electricity and often includes more detailed power grid models also used in an operational setting by utility companies. Due to the increasing importance of electricity as a clean energy carrier, power system models and energy system models are increasingly converging.

In this section we outline how data and code relate to the energy modelling process, introducing the concepts for the following sections. We limit our discussion to open data and open source code. The process of publishing results in the literature and the conditions under which the resulting articles are distributed go beyond the scope of this article (see Fig. 1).

2.1. Data

Data is both an input and output of the modelling process. Raw data in the energy field is spread widely and of varying quality, coming from academic sources, non-governmental bodies, markets, individuals and commercial entities. An obvious impediment to openness is the widespread use of non-disclosure agreements under which commercially sensitive data may be shared. A less obvious impediment is that in many cases, no explicit licence is attached to input data. Contrary to common practice, this does not imply the legal permission to use and share data, as discussed below. This is of crucial importance since the degree of openness and the licensing conditions of input data influence the degree to which a model based on them can be made open.

Raw data from various sources and in different formats must first be processed to become accessible to a model. Three kinds of data processing are usually necessary: (1) time series data: creating simulated or processing measured intermittent renewable generation and demand data, (2) geographic data (e.g. installed generation capacities): aggregation or disaggregation and other forms of geospatial analysis, (3) simple "tabular" data (e.g., costs of technologies or fuels): varied manual processing such as making assumptions where values are missing or converting currencies.

[1] This is akin to the Open Source Definition used by the Open Source Initiative [4], to which all commonly used licences for code conform. The full list of OSI-approved licences is at https://opensource.org/licenses.

[2] http://www.openmod-initiative.org/.