



The importance of open data and software: Is energy research lagging behind?



Stefan Pfenninger^{a,*}, Joseph DeCarolis^b, Lion Hirth^c, Sylvain Quoilin^d, Iain Staffell^e

^a Climate Policy Group, Department of Environmental Systems Science, ETH Zürich, Switzerland

^b Department of Civil, Construction, and Environmental Engineering, North Carolina State University, USA

^c Neon Neue Energieökonomik GmbH (Neon), Germany

^d Institute for Energy and Transport, European Commission Joint Research Centre (JRC), Petten, Netherlands

^e Centre for Environmental Policy, Imperial College London, UK

ARTICLE INFO

Keywords:

Modelling
Data
Open source
Open data
Transparency
Reproducibility

ABSTRACT

Energy policy often builds on insights gained from quantitative energy models and their underlying data. As climate change mitigation and economic concerns drive a sustained transformation of the energy sector, transparent and well-founded analyses are more important than ever. We assert that models and their associated data must be openly available to facilitate higher quality science, greater productivity through less duplicated effort, and a more effective science-policy boundary. There are also valid reasons why data and code are not open: ethical and security concerns, unwanted exposure, additional workload, and institutional or personal inertia. Overall, energy policy research ostensibly lags behind other fields in promoting more open and reproducible science. We take stock of the status quo and propose actionable steps forward for the energy research community to ensure that it can better engage with decision-makers and continues to deliver robust policy advice in a transparent and reproducible way.

1. Introduction

For nearly a century, the global energy system has remained remarkably stable, powered largely by fossil fuel combustion. However, successfully addressing anthropogenic climate change with low-carbon technologies requires that we fundamentally alter energy supply and demand in the 21st century, yet the pathway and outcomes of this transformation are highly uncertain. For example, rapid improvements in solar photovoltaics and batteries coupled with information technology may point towards a more distributed energy system with its design actively shaped by consumers. Alternatively, large-scale technologies like nuclear, biomass, carbon capture and storage or wind may extend the dominance of a centralised power system.

Given the uncertainty and complexity of the energy system, quantitative models are one of the few available tools that allow analysts to explore alternative scenarios and help guide public policy. Quantitative analysis from energy models underpins much of academic research and energy policy-making (Strachan et al., 2009). Yet most models and data relied upon by utilities, consultancies and public research institutes remain inscrutable “black boxes” – whether economic models with a small number of parameters, or large linear

optimisation models with hundreds of thousands of input variables. In contrast to closed models, “open” models imply that anyone can freely access, use, modify, and share both model code and data for any purpose (Open Knowledge Foundation, 2015). Here, we (1) argue why energy data and models urgently need to become open; (2) discuss the key reasons why many are currently not; (3) examine whether energy research is lagging behind other fields in becoming more open; and finally (4) outline specific issues for individuals to consider and propose next steps for the energy research community.

2. Why models and data should be open

Given the critical guidance that energy models and data provide to decision makers, they should be made open and freely available to researchers as well as the general public. There are four specific reasons for this:

1. Improved quality of science. Fundamental scientific principles such as transparency, peer review, reproducibility and traceability are almost impossible to implement without access to models and data (DeCarolis et al., 2012; Nature, 2014). Better adherence to these principles leads to higher quality science. Researchers are

* Corresponding author.

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

fallible human beings and errors are inevitable under pressure to deliver. Such mistakes can have profound implications. For example, the Reinhart-Rogoff spreadsheet error arguably skewed the international debate on austerity (Herndon et al., 2014). Such incidents serve as warnings against poor programming practices, such as a lack of auditing as well as closed models and data: it was only through sharing the spreadsheet that the errors were discovered.

2. **More effective collaboration across the science-policy boundary.** Better and more transparent science ought to enable better policy outcomes, but the issue is more complex than that. Academic peer review routinely does not (and cannot) check model arithmetic and data validity, just that the analytical approach is appropriate. A separate process of quality assurance (QA) is required to verify and validate model mechanics and output. While mostly absent from academic practice, this is often implemented as a formal procedure in government (DECC, 2015). The reason for this is that unlike academics, governments, private companies and NGOs often model for numbers rather than insight. The specific numbers can be of great societal importance, such as the level at which to set subsidies or the cost of specific policies. Thus, in many cases, the most important aspect is the quality or transparency of input data, rather than the novelty of the modelling methodology. In large datasets used in government decision-making, traceability and referencing can become major problems, as civil servants developing models and data are often not trained scientists. Openly available, collaboratively developed datasets and reference models would allow the burden of this work to be shared more widely, and across both academia and government. There is a growing sense that the link between energy modelling and policy needs fundamental rethinking (Strachan et al., 2016), and opening up models and data will play a crucial role in enabling the transparency and better quality assurance necessary for this to happen.
3. **Increased productivity through collaborative burden sharing.** Collecting data, formulating models and writing code are resource-intensive. Research funding is limited and researcher time is a scarce resource. Society as a whole saves time and money if researchers avoid unnecessary duplication and learn from one another. Individual researchers gain more time to spend on pressing research questions rather than redundant work on model or dataset development. Furthermore, research only matters if it is seen and used, and open-access publishing has been shown to increase readership and citations (McCabe and Snyder, 2014). Since openly shared code or data is more likely to be known to others, it is more likely to be used and further improved. Not only does this benefit the original researcher through peer recognition and academic credit, but moves the research community as a whole forward.
4. **Profound relevance to societal debates.** Reengineering the energy landscape will affect everyone, producing winners and losers. A balanced societal and political debate requires transparent arguments based on scientific justifications, but escalating concern about reproducibility in some fields is shaking public confidence in scientific research (Goodman et al., 2016). Finally, besides the practical considerations outlined above, there remains the ethical argument that research funded by public money should be available to the public in its entirety.

3. Why models and data are mostly not open

Despite these arguments, we see four main reasons why closed models and data may remain attractive and rational in some cases:

1. There is a range of valid ethical and security concerns, particularly in the case of data. Researchers may have access to sensitive commercial data or to data containing personal information (particularly relevant when moving towards more decentralised smart grids with their focus on individual households). The aspiration to open up as much data as possible may give way to a more regulated approach to open data if individual researchers increasingly cross ethical boundaries, as in the recent release of personal data about users of a major online dating website (Resnick, 2016). Setbacks in the wider open data movement could also have repercussions on the use of information perceived as sensitive in the energy modelling context, e.g. data on energy consumer behaviour or on grid infrastructure.
2. Openly sharing details of models, analysis and data can create unwanted exposure. Flawed code or data can discredit research results and cause embarrassment to their authors, but only if they are visible. Indeed, a reluctance to share data was shown to be associated with weaker evidence (Wichert et al., 2011). Furthermore, there may be a fear that inexperienced researchers use an open model or open data to produce flawed analysis that reflects poorly on its original authors. There is also a policy dimension: government departments may choose to keep information closed precisely because of the potentially serious impact it may have on a country's economy and society, rather than opening the models and data to enable a more transparent political and societal discussion. For example, while the UK Department of Energy and Climate Change (DECC) are working with University College London to develop an open source UK TIMES energy system optimisation model (UCL, 2014), political sensitivities mean that code and data will not be released until its use in a major policy analysis (the UK's 5th carbon budget) is complete (Sargent, 2016).
3. It is time-consuming to write legible and reusable code, track data provenance and processing steps, document models and data and respond to feature requests or bug reports. Because model and dataset development are large investments, it is often rational for researchers and institutions to maintain "trade secrets" to compete in consulting work and third-party research funding. On the one hand, this can be seen as a classical collective action problem where individual actors are trapped in a suboptimal non-cooperative equilibrium. But, as discussed further below, the incentive structure that gives rise to this bargaining problem is also linked to institutional issues within academia, particularly the unrelenting pressure to publish ever-greater quantities of high-quality publications which underlies most academic career incentives and impact metrics (Sarewitz, 2016). A significant share of energy modelling is done in the private sector, in utilities, consulting firms, and financial institutions, where the need to protect the intellectual property within models and data is certainly more pressing than in academia. Nevertheless, where private sector modelling is used to inform public policy and/or where it is funded by public money, we believe the long-term goal should be for models and data to be open, even if this would challenge consultancies' established business practices. While examples of successful open-source businesses exist (e.g. RedHat or Canonical in the Linux world), it is clear that working business models can be difficult to find, especially in the energy field with the added difficulty of balancing commercial and academic principles. The "share alike" clause in licenses like the GPL (see below) may offer opportunities for companies here. Furthermore, private companies and consulting firms are also selling their expertise: energy models must be adapted for specific analyses, and the real value arguably comes from the application of judgement and expertise to adapt and apply the models in a way that produces useful insight.
4. Finally, there is simple institutional and personal inertia, often alongside complex and uncoordinated institutional setups. For example, energy models and datasets are developed and applied by

different agencies within the US government with no consistent and coordinated policy towards data and model availability. On the one hand, the federal government has placed emphasis on making data publicly available through a web portal.¹ With regard to energy data, this has included the openEI portal² and a listing of all open datasets within the Department of Energy.³ On the other hand, individual agencies have long-running practices that do not necessarily align with these high-level intentions. The National Energy Modeling System (NEMS) is publicly available, but its owners, the Energy Information Administration, ostensibly discourage its use by asserting that “Most people who have requested NEMS in the past have found out that it was too difficult or rigid to use” (EIA, 2016). The US Environmental Protection Agency employs closed source computable general equilibrium models for use in economic analysis of climate policy which are developed not by academic institutions but by private-sector consultants, which in turn must protect their own business interests (US EPA, 2016).

All of these factors are understandable from the perspective of individual actors, but collectively they engender a sense of mistrust in complex, impenetrable models and enigmatic datasets. For example, the European Commission faced criticism for using the proprietary PRIMES model to deliver key results for its Energy Roadmap 2050 (Helm et al., 2011). More significantly, the UK’s decarbonisation was arguably delayed for years by models that underestimated the scale of the challenge due to opaque and heroically optimistic cost assumptions for onshore wind (House of Lords, 2005).

4. Is openness in energy research lagging behind other fields?

In the face of such criticism, efforts are underway to make data and models more transparent. In the policy and government decision-making sphere, the UK Department of Energy and Climate Change (DECC) pioneered a simple open modelling approach through their 2050 Carbon Calculators (DECC, 2013), which is now being replicated elsewhere. Ongoing but incomplete plans to open up UK TIMES were discussed above, as well as open data efforts in the United States.

Legislation in Europe is evolving; for example, the EU Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) requires participants to publish electricity market data to thwart insider trading and market manipulation. Also in Europe, reference models and scenarios are facing increased scrutiny and criticism, but it is as yet unclear whether this will lead to fundamental changes in how policy analysis is performed. The recent controversy on the European Commission ignoring its own quantitative analysis when selecting emissions reductions targets has pushed the commission to publish a communication justifying the selection of discount rates in its impact assessments (European Commission, 2014). The standard grant agreement of the EU-funded H2020 projects requires open access to all peer-reviewed scientific publications relating to the results of the project. A similar requirement for research data is optional, however (European Commission, 2016a). Some software produced by Commission services is now open sourced under the European Union Public License (EUPL), but there are mixed signals. The closed-source PRIMES model is being replaced by POTEnCIA which was intended to be more transparent (European Commission, 2016b). However, while the database providing some of its input data will be made publicly available (Joint Research Centre, 2016), the new model’s code and actual model database will still remain closed.

The German government, on the other hand, has taken a leading

role. It sees open energy modelling as an important ingredient for high-quality scientific energy policy advice and has publicly committed itself to open software and data. Two flagship projects are funded by the federal government: SciGRID⁴ is developing an open-source transmission grid model based on OpenStreetMap data; while the Open Power System Data project⁵ is building an online data platform for free and open data for power system modelling, including data on power plant capacities and locations and renewable production time series from transmission system operators.

Amongst academic energy system models, BALMOREL (Ravn et al., 2001) is an early example with limited geographical scope, while more recent open models like OSeMOSYS (Howells et al., 2011) and Temoa (Hunter et al., 2013) were followed by a flourishing of activity. In 2014, the Open Energy Modelling Initiative was established, which now lists more than twenty open models of power networks, electricity markets and energy systems.⁶ However, the vast majority of energy research published in the peer-reviewed literature still makes neither code nor data openly available. The breadth of model development efforts underway underscores the fact that because of newly emerging modelling challenges, there is a window of opportunity for new, collaborative and open modelling and data gathering efforts to carve out new niches. For example, the accurate characterisation of renewable resources requires new kinds of datasets with high spatial and temporal resolution. Behavioural responses by consumers to prices and incentives will likely play an increasingly important role during the transition to a low-carbon energy system. Different sectors – such as electricity, heat and transportation – may become more tightly connected in the future with the deployment of heat pumps and electric vehicles, and therefore require new and more complex types of coupled models.

In contrast to monodisciplinary research (e.g., particle physics simulation) where the models and data are more standardised, energy research has the added challenge that model formulation and data require significant researcher judgment. Unlike the laws of physics, there are no universal governing principles guiding real-world energy systems development. Furthermore, the data used to parameterise energy models are heterogeneous, widely varying in terms of quality, and in some cases lacking altogether. Other applied fields share this issue. It is therefore not surprising that monodisciplinary fields are further ahead with collaboratively developed software packages and standard community databases. For example, in genetics, there are open software papers going back more than a decade, some garnering tens of thousands of citations, underlining their key importance for the research community (e.g., Altschul et al., 1990; Hunter et al., 2013). As open databases become more widely used, for example the GenBank sequence database of all publicly available nucleotide sequences and their protein translations (Benson et al., 2013), depositing results in such databases can be made a requirement by journals. While top journals in many monodisciplinary fields now have policies requiring the release of data, software, and other information required to replicate published results, Energy Economics is the only major energy journal to have put such policies in place.

Despite the efforts in energy research and policy outlined above, our field appears to be lagging behind even other applied fields. For example, in climate science there are coordinated efforts like the Community Earth Science Model (CESM, see Hurrell et al., 2013), which has a large body of peer-reviewed work extending, validating, and using it. The public health and biomedical research communities have over the last two decades progressed both on software (McDonald et al., 2003; Schindelin et al., 2012) and on data (Boulton et al., 2011), e.g. including the challenge of anonymising personally identifiable information (Lawlor and Stone, 2001). Funding bodies are an im-

¹ <http://en.openei.org/>

² <http://en.openei.org/>

³ <http://energy.gov/data/downloads/open-data-catalogue>

⁴ <http://www.scigrd.de/>

⁵ <http://open-power-system-data.org/>

⁶ http://wiki.openmod-initiative.org/wiki/Open_Models

portant enabling factor here: the Wellcome Trust is at the forefront of requiring and supporting open research (Kiley, 2016). Such efforts are important prerequisites to thinking about how open code and data work within the practical constraints of a specific field. For example, while making code and data available openly is an important first step towards higher quality, more transparent, and reproducible research, it is followed by difficult questions about how to truly achieve these goals – in particular for research results relying on very large datasets or computationally intensive analyses that require days or months to complete.

There is no reason to believe that energy research cannot rise to the challenge. For example, the U.S. Pacific Northwest National Laboratory recently released the Global Change Assessment Model (GCAM) to the public after 20 years of development (Joint Global Change Research Institute, 2015), demonstrating that even software coming out of large and complex research projects with a long legacy can be made openly available if there is the will to do so.

5. Policy recommendations

First, individual researchers and research groups need to understand the practicalities of open code and data, in particular:

1. **Consider the intended target audience.** For example, if the audience is other modellers, releasing code with a clean and documented interface might be the main priority. If it is policy-makers, attention to clearly documenting input data and assumptions and allowing reproducibility may well take precedence over code itself.
2. **Decide what and how much to publish openly.** Researchers should not feel forced to take a binary decision of open versus not open. Publishing parts of a codebase or a dataset is better than publishing nothing at all, and can be seen as part of an exploratory process in deciding what and how much to publish openly depending on the goals and contexts of specific analyses.
3. **Decide on a license and a distribution channel.** The first step is often to check who owns intellectual property rights, which is likely the researcher's employer. For licensing, the main decision is between a permissive license (such as the MIT license for code or CC-BY license for data) or a copyleft or "share-alike" license (such as GNU GPL or CC-BY-SA). This can be a difficult decision to make but much guidance exists (e.g. Morin et al., 2012). In addition, energy research often relies partly on commercial tools such as GAMS or Vensim. While this limits the choice of license (e.g. the GPL license is technically incompatible with a model relying on a closed-source environment to run), models built on such tools can still be openly licensed and made available. This is already common practice in the science and engineering fields, with numerous open projects written for commercial software such as Matlab. Finally, the choice of distribution (e.g. via an institutional website or an established platform like GitHub) also influences how a project is perceived and what kinds of community interactions will likely result.

It is important to realise that all of these choices come with both potential benefits and risks. In addition, existing energy modelling may rely on (commercial) software without even any code to share, and data in proprietary commercial data formats. With the high importance of policy decisions taken based on such work and the increasing availability of open-source alternatives, we believe it is necessary for the field to increasingly shift to those options. The Open Energy Modelling Initiative is developing an evolving working paper to help individual researchers tackle these issues (Pfenninger et al., 2016). Equally important, the energy research community as a whole needs to think about the hard problems around transparent research, open code and data. In particular, we believe the community needs to consider three issues:

1. **Reduce parallel efforts and duplication of work.** Despite an influx of new open data and software projects, the energy modelling landscape is still fragmented. There is a risk that different modelling and data collection efforts remain insular, and thus that the true benefits of openness are not realised. For example, while open code and data enable third party verification, they do not advance the state of modelling if no one tries to replicate results or directly build upon released code.
2. **Progressively consolidate from the top down and the bottom up.** Joint top-down coordination efforts could include the development of common datasets, community standards to ensure interoperability, and coordinated efforts to enable third-party verification of model-based results. There are example of community-building science projects spanning several fields such as rOpenSci (Boettiger et al., 2015) that show such efforts can succeed. Yet getting from a blank slate to working community efforts may face too much initial inertia to get started. There are many one-off analyses created for specific papers, or code that is written with the understanding that it will never be made public or see widespread use, and thus be poorly documented and structured. To bridge the gap between such barely re-usable code and well-maintained and documented community codebases, Varoquaux (2015) argues for progressive consolidation of released code as a collective undertaking within a given research community, thus letting community projects grow organically from the bottom up.
3. **Work towards changing incentives.** In parallel, the energy research community must engage with other stakeholders to ensure institutional and academic recognition for open energy models, and to start tackling the harder problems of transparent, reproducible analyses. There are currently various efforts across different scientific fields to give credit for data and software, but early and mid-career researchers in particular are still subject to the current system of academic credit. Open and transparent research is not currently incentivised: in fact, the opposite is often perceived as advantageous for scientific career advancement. Changing these incentives will require efforts not only from researchers themselves but also from their employers, from grant agencies, and other stakeholders like publishers (Nosek et al., 2015) – not all of which may have the same interests. Because of the blurred boundary between academic research and consulting work in the energy field, this coordination effort is likely to be more challenging than in other fields. This may require a cultural shift away from coveting the basic tool (be it an energy systems model or a scanning electron microscope) in favour of the expertise and insight required to use the tool effectively.

Science is an inherently collaborative endeavour. If every particle physics research group had to build their own accelerator, the scientific breakthroughs enabled by projects like CERN would not be possible. Open code and data put deeper and more meaningful collaboration within reach. Energy modelling is not basic science, it is uncertain and the decision stakes are high, putting it squarely within the realm of post-normal science (Funtowicz and Ravetz, 1993). For example, energy system models, alongside wider integrated assessment models, form a significant part of the analysis informing multinational agreements like the UNFCCC and binding national climate policies. They are thus a driving force behind policy targets that affect the future livelihoods of billions worldwide, produce economic winners and losers, and generate heated public discourse. Given the importance of rapid global coordinated action on climate mitigation and the clear benefits of shared research efforts and transparently reproducible policy analysis, openness in energy research should not be for the sake of having some code or data available on a website, but as an initial step towards fundamentally better ways to both conduct our research and engage decision-makers with models and the assumptions embedded within them.

Acknowledgements

Stefan Pfenninger thanks the European Research Council via grant StG 2012-313553 for funding.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. *Nucleic Acids Res.* 41, D36–D42. <http://dx.doi.org/10.1093/nar/gks1195>.
- Boettiger, C., Chamberlain, S., Hart, E., Ram, K., 2015. Building software, building community: lessons from the rOpenSci project. *J. Open Res. Softw.*, 3. <http://dx.doi.org/10.5334/jors.bu>.
- Boulton, G., Rawlins, M., Vallance, P., Walport, M., 2011. Science as a public enterprise: the case for open data. *Lancet* 377, 1633–1635. [http://dx.doi.org/10.1016/S0140-6736\(11\)60647-8](http://dx.doi.org/10.1016/S0140-6736(11)60647-8).
- DeCarolis, J.F., Hunter, K., Sreepathi, S., 2012. The case for repeatable analysis with energy economy optimization models. *Energy Econ.* 34, 1845–1853. <http://dx.doi.org/10.1016/j.eneco.2012.07.004>.
- DECC, 2013. 2050 Pathways - Detailed guidance. Available at: (<https://www.gov.uk/guidance/2050-pathways-analysis>) (accessed 14.03.16.).
- DECC, 2015. Quality Assurance tools and guidance in DECC. Available at: (<https://www.gov.uk/government/collections/quality-assurance-tools-and-guidance-in-decc>) (accessed 02.06.16.).
- EIA, 2016. Availability of the National Energy Modeling System (NEMS) Archive. Available at: (https://www.eia.gov/forecasts/aeo/info_nems_archive.cfm) (accessed 01.06.16.).
- European Commission, 2014. 2030 climate and energy framework: DG CLIMA reaction to unfounded claims about the Commission's economic analysis. Available at: (http://ec.europa.eu/clima/policies/strategies/2030/docs/2030_climate_and_energy_framework_reaction_to_claims_about_the_commission_economic_analysis.pdf) (accessed 07.06.16.).
- European Commission, 2016a. H2020 General Model Grant Agreement, Version 3.0. Available at: (https://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf) (accessed 07.10.16.).
- European Commission, 2016b. POTEnCIA model description - version 0.9 - EU Science Hub - European Commission. <http://dx.doi.org/10.2791/416465>
- Funtowicz, S.O., Ravetz, J.R., 1993. Science for the post-normal age. *Futures* 25, 739–755.
- Goodman, S.N., Fanelli, D., Ioannidis, J.P.A., 2016. What does research reproducibility mean? *Sci. Transl. Med.* 8, 341p12. <http://dx.doi.org/10.1126/scitranslmed.aaf5027>.
- Helm, D., Mandil, C., Vasconcelos, J., MacKay, D., Birol, F., Mogren, A., Hauge, F., Bach, B., van der Linde, C., Toczyłowski, E., Pérez-Arriaga, L., Kröger, W., Luciani, G., Matthes, F., 2011. Final Report of the Advisory Group on the Energy Roadmap 2050.
- Herdon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Camb. J. Econ.* 38, 257–279. <http://dx.doi.org/10.1093/cje/bet075>.
- House of Lords, 2005. The Economics of Climate Change. Vol. II. 2005, HL 12-II of 2005-06, QQ 407-408.
- Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M., Roehrl, A., 2011. OSeMOSYS: the open source energy modeling system: an introduction to its ethos, structure and development. *Energy Policy* 39, 5850–5870. <http://dx.doi.org/10.1016/j.enpol.2011.06.033>.
- Hunter, K., Sreepathi, S., DeCarolis, J.F., 2013. Modeling for insight using Tools for energy model optimization and analysis (Temoa). *Energy Econ.* 40, 339–349. <http://dx.doi.org/10.1016/j.eneco.2013.07.014>.
- Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J.-F., Large, W.G., Lawrence, D., Lindsay, K., Lipscomb, W.H., Long, M.C., Mahowald, N., Marsh, D.R., Neale, R.B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W.D., Hack, J.J., Kiehl, J., Marshall, S., 2013. The community earth system model: a framework for collaborative research. *Bull. Am. Meteorol. Soc.* 94, 1339–1360. <http://dx.doi.org/10.1175/BAMS-D-12-00121.1>.
- Joint Global Change Research Institute, 2015. Global Change Assessment Model | Joint Global Change Research Institute. Available at: (<http://www.globalchange.umd.edu/models/gcam/>) (accessed 15.03.16.).
- Joint Research Centre, 2016. Introducing the JRC-IDEES database. Available at: (https://ec.europa.eu/jrc/sites/default/files/JRC-IDEES_20160301.pdf) (accessed 15.06.16.).
- Kiley, R., 2016. Why We're Launching A New Publishing Platform | Wellcome. Available at (<https://wellcome.ac.uk/news/why-were-launching-new-publishing-platform>). (accessed 14.10.16.).
- Lawlor, D.A., Stone, T., 2001. Public health and data protection: an inevitable collision or potential for a meeting of minds? *Int. J. Epidemiol.* 30, 1221–1225. <http://dx.doi.org/10.1093/ije/30.6.1221>.
- McCabe, M.J., Snyder, C.M., 2014. Identifying the effect of open access on citations using a panel of science journals. *Econ. Inq.* 52, 1284–1300. <http://dx.doi.org/10.1111/ein.12064>.
- McDonald, C.J., Schadow, G., Barnes, M., Dexter, P., Overhage, J.M., Mamlin, B., McCoy, J.M., 2003. Open Source Software in Medical Informatics—why, How and What, in: Proceedings of the Working Conference on Health Information Systems, International Journal of Medical Informatics, 69, pp. 175–184. doi:[http://dx.doi.org/10.1016/S1386-5056\(02\)00104-1](http://dx.doi.org/10.1016/S1386-5056(02)00104-1).
- Morin, A., Urban, J., Sliz, P., 2012. A quick guide to software licensing for the scientist-programmer. *PLoS Comput. Biol.* 8, e1002598. <http://dx.doi.org/10.1371/journal.pcbi.1002598>.
- Nature, 2014. Journals unite for reproducibility. 515, 7. <http://dx.doi.org/10.1038/515007a>.
- Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. *Science* 348, 1422–1425. <http://dx.doi.org/10.1126/science.aab2374>.
- Open Knowledge Foundation, 2015. Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. Available at: (<http://opendefinition.org/od/2.1/en/>) (accessed 15.03.16.).
- Pfenninger, S., Schmid, E., Wiese, F., Hirth, L., Davis, C., DeCarolis, J.F., Fais, B., Krien, U., Matke, C., Momber, I., Müller, B., Pleßmann, G., Quolin, S., Reeg, M., Richstein, J.C., Schlecht, I., Shivakumar, A., Staffell, I., Tröndle, T., Wingenbach, C., 2016. Benefits, challenges and solutions for open energy modelling. Open Energy Modelling Initiative Working Paper. Available at: (<https://openmod-initiative.github.io/openmod-working-paper>) (accessed 01.07.16.).
- Ravn, H.F., Munksgaard, J., Ramskov, J., Grohnhøj, P.E., Larsen, H.V., 2001. Balmorel: a model for analyses of the electricity and CHP markets in the baltic sea region. Appendices. Elkraft System. Available at: (<http://www.eabalmorel.dk/files/download/Balmorel%20A%20Model%20for%20Analyses%20of%20the%20Electricity%20and%20CHP%20Markets%20in%20the%20Baltic%20Sea%20Region.pdf>). (accessed 10.03.16.).
- Resnick, B., 2016. Researchers just released profile data on 70,000 OkCupid users without permission. *Vox*. Available at: (<http://www.vox.com/2016/5/12/11666116/70000-okcupid-users-data-release>). (accessed 02.06.16.).
- Sarewitz, D., 2016. The pressure to publish pushes down quality. *Nature* 533, 147. <http://dx.doi.org/10.1038/533147a>.
- Sargent, P., 2016. How government uses energy models. Presented at the WholeSEM Workshop on Integrated Energy System Models Incorporating Spatial And Temporal Detail, Imperial College London, 24 May 2016.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <http://dx.doi.org/10.1038/nmeth.2019>.
- Strachan, N., Pye, S., Kannan, R., 2009. The iterative contribution and relevance of modelling to UK energy policy. *Energy Policy* 37, 850–860. <http://dx.doi.org/10.1016/j.enpol.2008.09.096>.
- Strachan, N., Fais, B., Daly, H., 2016. Reinventing the energy modelling—policy interface. *Nat. Energy* 1, 16012. <http://dx.doi.org/10.1038/energy.2016.12>.
- UCL, 2014. wholeSEM's UCL team to develop open source version of UK TIMES with the Department of Energy and Climate Change. Available at: (<http://www.wholesem.ac.uk/wholesem-news-publication/uktm-collaboration>) (accessed 26.02.16.).
- US EPA, 2016. Economic Climate Modeling. Available at: (<https://www3.epa.gov/climatechange/EPAactivities/economics/modeling.html>) (accessed 01.06.16.).
- Varoquaux, G., 2015. Software for Reproducible Science: Let's Not Have A Misunderstanding. *Computer/data / Brain Science*. Available at: (<http://gael-varoquaux.info/programming/software-for-reproducible-science-lets-not-have-a-misunderstanding.html>). (accessed 01.06.16.).
- Wicherts, J.M., Bakker, M., Molenaar, D., 2011. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE* 6, e26828. <http://dx.doi.org/10.1371/journal.pone.0026828>.