Open Source Software and Crowdsourcing for Energy Analysis

Morgan Bazilian\textsuperscript{a,b}, Andrew Rice\textsuperscript{c}, Juliana Rotich\textsuperscript{d}, Mark Howells\textsuperscript{e}, Joseph DeCarolis\textsuperscript{f}, Stuart Macmillan\textsuperscript{g,h}, Cameron Brooks\textsuperscript{i}, Florian Bauer\textsuperscript{j}, and Michael Liebreich\textsuperscript{k}

\textsuperscript{a}United Nations Industrial Development Organisation, Vienna, Austria
\textsuperscript{b}International Institute for Applied Systems Analysis, Laxenburg, Austria
\textsuperscript{c}Computer Laboratory, University of Cambridge, Cambridge, UK
\textsuperscript{d}Ushahidi, Nairobi, Kenya
\textsuperscript{e}Royal Swedish Institute of Technology, Stockholm, Sweden
\textsuperscript{f}North Carolina State University, North Carolina, USA
\textsuperscript{g}Stanford University, California, USA
\textsuperscript{h}National Renewable Energy Laboratory, Colorado, USA
\textsuperscript{i}Tendril Networks, Colorado, USA
\textsuperscript{j}Renewable Energy and Energy Efficiency Partnership, Vienna, Austria
\textsuperscript{k}Bloomberg New Energy Finance, London, UK

Abstract: Informed energy decision making requires effective software, high-quality input data, and a suitably trained user community. Developing these resources can be expensive and time consuming. Even when data and tools are intended for public re-use they often come with technical, legal, economic and social barriers that make them difficult to adopt, adapt and combine for use in new contexts. We focus on the promise of open, publicly accessible software and data as well as crowdsourcing techniques to develop robust energy analysis tools that can deliver crucial, policy-relevant insight, particularly in developing countries, where planning resources are highly constrained – and the need to adapt these resources and methods to the local context is high. We survey existing research, which argues that these techniques can produce high-quality results, and also explore the potential role that linked, open data can play in both supporting the modelling process and in enhancing public engagement with energy issues.

Keywords: Energy modelling; Open source data; Crowdsourced data

1. Introduction

Modern energy systems are characterised by increasingly complex interactions between energy supply, distributors, and consumer demand. High quality analytical data and tools\textsuperscript{1} are vital for predicting and understanding these interactions in order to enable informed energy system design, implementation and operation decisions. Effective energy analytics to inform public policy requires that three key conditions be met: (1) validated models must be available and appropriate for the target environment; (2) suitable data must be available for input into the model and for verifying model-based results; and (3) models must be operated by people trained in the use of the tools and in interpreting the outcomes for local conditions. Developing these resources can be expensive and time consuming. Even when data and tools are intended for public re-use they often come with technical, legal, economic and social barriers that make them difficult to adopt, adapt and combine for use in new contexts (Wilbanks and Macmillan, 2012).

\textsuperscript{1} We refer to tools in the broad sense including models, simulations and other analytical programs.
In this paper, we focus on how publically accessible software and data (i.e., open energy data, open source energy tools and open educational resources) can help modellers meet the above conditions. In April 2012, The President of the World Bank ‘tweeted’, “Open information, open data, and open access to knowledge may turn out to be the most important legacy of the past 5 years (19.4.2012 @worldbank)”. Still, the transformative impacts of applying open source software (OSS) and data as well as associated training tools are in the early stages of adoption in the area of energy system analysis. We argue that open modelling efforts can improve the utility and accessibility of energy models and also lower the cost of data collection and management. In addition, the directed application of crowdsourcing to push the development of open modelling tools and datasets could yield significant benefits to the international energy modelling community (Howe, 2006).

Energy is central to improved social and economic well-being, and is indispensable to most industrial and commercial wealth generation (Brew-Hammond and Kemausuor, 2009; Reddy, 2000). Adequate national capacity to track progress towards universal modern energy access represents a crucial element of energy poverty alleviation and sustainable development strategies. For developing countries - which frequently lack established infrastructure, data, and software tools - there are significant potential benefits from rigorous analyses enabled by open source tools and data (Bazilian, 2011). IEA (2009) cites, “sound statistical data…and a clear description of the [energy services] situation” as the first precondition for successful rural energy access policies. A reliable and comprehensive energy information base is required to set targets and monitor outcomes, to design strategies and policies, to make evidence-based decisions as well as to enable citizens to make informed choices. In addition, poor quality statistics limit multi-country analysis and undermine efforts to implement global or regional programmes.

Ultimately, these open energy resources, combined with open innovation processes, can be harnessed to better inform energy decision-making and rapidly develop low-cost, high-quality and localised energy resources.

2. Open data

Open data refers to that subset of data that is freely available to everyone to use without restrictions. Complex energy decisions often draw upon data from a wide range of economic, technical, policy and social sources. Portions of this data may be particular to a local context while other data sources may be less context-sensitive. Some of this data is available through open government initiatives, but increasingly it is coming from other institutions including non-profits, industry and research institutions. The ability to easily and effectively reuse a data set can eliminate considerable redundant work often associated with assembling datasets.

Governmental acceptance and adoption of open data has been growing rapidly with examples ranging from the USA5 and the UK6, to Kenya7 and Ghana (Alonso et al., 2011). As another example, the World Bank’s Open Data Initiative8 includes multiple platforms through which one can access and process data, including mobile ‘apps’, Application Programming Interfaces (APIs), catalogue listings of resources, data visualization tools, a knowledge repository, and development of metadata standards. The impetus for these open data government initiatives is transparency, accountability and

---

2 There are many open source and free tools in use in the energy domain – but their use is often limited to relatively small groups.
4 Such data might include: basic information on energy demand, outages, faults, fuel stocks, pricing, building occupancy, water levels, or weather.
5 http://www.data.gov
6 http://www.data.gov.uk
7 https://opendata.go.ke
8 http://data.worldbank.org
the belief that opening this data to the public will lower the barriers to innovations that will benefit society.

A large amount of data on US national energy consumption has been made available online. This information is presented in a variety of forms, often requiring pre-processing before use. More structured information can be made available through web services that provide machine-accessible mechanisms for retrieving data. The concept of linked open data takes this further and aims to provide information sources that can be easily combined together using standard tools (Bauer and Kaltenbock, 2012).

Machine-accessible access to linked open energy data has the potential to greatly enhance the productivity of modellers. One example of an open data repository of relevance to energy analysis is Megajoule. Such crowdsourced, open, and linked data could prove particularly beneficial by combining low-cost collection with low-cost use. Furthermore, the social good of creating or contributing to open energy data sets may be sufficient incentive for participation.

3. Crowdsourcing

Crowdsourcing has proven to be an effective and efficient way to generate and maintain valued datasets, tools and educational resources. It was first introduced as a term in a Wired Magazine article (Howe, 2006) and has since been adopted by a huge range of projects, which maintain common principles (Estelles-Arolas and Gonzalez-Ladron-de Guevara, 2012). Popular examples include Wikipedia, which provides an online encyclopaedia built from community volunteers, and Amazon’s Mechanical Turk, which provides outsourcing and a payment mechanism for small tasks. Since quality energy data is often not available for particular local needs, crowdsourcing can be an effective method for distributing the task among the broader community.

The application of crowdsourcing to data collection has several benefits including the potential for reduced cost, reduced time and higher quality given that contributors are often reviewers and users of the data. Acquiring empirical observations or measurements is costly because it requires the presence of observers at the physical (or temporal) location of interest. The intention with crowdsourcing is to minimize cost by making use of observations from community members who have access to model-relevant data that they are already collecting. This kind of information is available through sites such as Pachube, which provide access to concurrent data-streams from devices such as home energy meters or weather stations. An alternative form of contribution is to collect direct observations from individuals. The Ushahidi crowdsourcing tools have been used to collect information world-wide about energy shortages and power-cuts, whereas the OpenRoomMap project (Rice and Woodman, 2010) collects inventories and plans from buildings’ occupants. These uses of human observers to replace expensive sensing deployments can also be considered an instance of human computation (Quinn and Bederson, 2011). In the context of energy system tools, crowdsourcing can be used to direct the collection of geographically-specific input data by researchers, academics, and students with access to the relevant information.

Researchers have shown that crowdsourcing can produce results comparable to those produced by experts (Paolacci et al., 2010), but that in some cases this relies on correctly formulating and

---

10 http://www.eia.doe.gov
11 http://developer.nrel.gov
12 http://megajoule.org
13 This is an example of a transactional community (Sun et al., 2012), which encourages participation through payment of participants.
14 http://pachube.com
15 http://energyshortage.org
16 http://powercuts.in
structuring the requested task (Kittur et al., 2008). Crowdsourced input might also play a role in the larger view of policy-making or determining research directions. There are a number of tools for crowdsourcing collective decision making processes (Watkins and Rodriguez, 2008), which have been shown to outperform expert panels, particularly with regard to forecasting (Dalal et al., 2011). It is also possible to crowdsource a proprietary dataset (Brabham, 2008). In many cases this can be a desirable outcome; however, in the context of energy systems analysis, we believe that the largest benefits to the broader community accrue when crowdsourcing is used to direct the collection of data made available in the public domain. Public access to crowdsourced data also enables peer review, which over time can identify errors or inaccuracies in the aggregated data.

4. Open energy tools

The complexity of energy generation and distribution systems coupled with technological advances means that energy analysis tools must be continuously adapted and extended to new contexts. These uses apply to countries at all levels of socio-economic development. In developing countries, however, access to energy services remains a critical issue for large segments of the population. This requires the significant enhancement of, and adaptation to traditional energy planning tools (Bazilian et al., 2010).

The selection and sourcing of input data are key considerations for the modelling process. However, the collection and maintenance of these data sets is often a costly and time-consuming process. Data quality requirements for energy modelling can be measured along two dimensions. First, temporal fidelity is required. When contemplating electricity consumption for example, it is important to ask if the model requires measurements at high frequency resolution or whether long-term average values are acceptable. Second, one must identify the measurement accuracy required. Data with a high level of accuracy is unnecessary if the model itself represents a weak approximation. Increasing the data resolution will normally increase the cost of data collection. As such, researchers should aim for the minimum acceptable fidelity, which is especially important in developing economies where modelling resources are limited.

Energy modelling software can be proprietary or Open Source. Open Source Software is now a common paradigm for software distribution and is rapidly being adopted in many sectors (Gallego et al., 2008). The conventional form of OSS project has been developed by volunteers who contribute for reasons, which might include learning, career concerns or satisfying functional needs (Subramanyam and Xia, 2008), covering a wide range of platforms and programming languages (Sen et al., 2012). In addition, companies have also begun to adopt OSS. Researchers have identified economic incentives for the wholesale adoption of OSS (Bitzer, 2004); for example, in the context of competition to software monopolies (Raymond, 1999). There are also successful examples of hybrid strategies, which attempt to leverage the best of both open and proprietary solutions (West, 2003).

Regardless of the specific type of open source model, it has been shown that OSS can generally meet high standards with little or no difference in quality relative to proprietary software (Ajila and Wu, 2007), and the peer-reviewed nature of an open source project tends to produce software with high maintainability (Johnson, 2006). The potential benefits of low-adoption costs and the freedom to adapt the model make OSS particularly promising for use in developing countries. However, there remains a need for appropriate technical infrastructure and human resources (Yildirim and Ansal, 2011). We return to this issue in Section 5.

The Open Source Energy Modeling System (OSeMOSYS) provides a robust platform and community for open, accessible energy modelling17. OSeMOSYS was conceived to exploit the benefits of open

---

17 Related efforts exist. TEMOA (DeCarolis et al., 2010) provides an open database, and is developing a similar open effort based on a more computationally efficient, though less intuitive language. Other commonly used tools, such as MARKAL and TIMES have code bases open to members or affiliates of the International Energy
source development by being easy to update, modify, and extend to suit the needs of particular researchers. Customised OSeMOSYS models are constructed by combining a number of component blocks - each lending a particular functionality to the model - as required.

The OSeMOSYS structure has intuitive appeal and has proven successful, and researchers have already begun to publish results based on adaptations and extensions of the model, such as through: behavioural factors (such as social norms) governing the purchase of low carbon technologies (Warren, 2011a); consumer preferences (Warren, 2011b); and optimisation for energy security rather than energy cost (Howells, 2011). Although requiring alteration and extension of the core package, in each case the modular design of OSeMOSYS means that the scope of changes necessitated by these projects was tightly constrained. User interfaces have also been provided as OSeMOSYS has been integrated into larger models or served as an analysis tool in government projects. Specific examples include LEAP written by the Community for Energy, Environment and Development (COMMEND) for use by the South African Department of Energy for national planning. As OSeMOSYS demonstrates, OSS has the potential to produce high quality energy modelling tools while lowering the barriers to adoption relative to conventional proprietary solutions.

5. Education and capacity building

Wide-scale adoption of energy modelling will only be possible if suitable training and education is put in place. Low-cost communication over the Internet, crowdsourcing, and open source software have created new paradigms both for learning (Albors et al., 2008) and also for general philanthropic activity (Bernholz et al., 2010). These new paradigms have a role to play in three areas: creating demand and interest in energy modelling; providing the education required; and supporting the modelling process itself.

One way to create demand for educational resources related to energy modelling might be the growing prevalence of “Mashups”. Mashups are fostering more general engagement with data by allowing users to easily create new ways of visualising and relating data relevant to themselves (DiFranzo et al., 2011). OpenEnergyInfo provides open energy information, data, tools and models to enable its community. Innovation competitions, including “Hackathons”, held over very short periods, such as one weekend, provide another mechanism for stimulating activity. A related example is the Apps For the Energy programme in the USA.

This form of collaborative learning and exploration is also emerging in the form of “hackerspaces” (or hacklabs) in which a community space is created for members to learn and innovate (Hunsinger, 2011). OSS is also well suited to the educational process of teaching, learning and peer-review (Carmichael and Honour, 2002) and in some cases can deliver content more effectively than conventional methods (Martin et al., 2011). Distance learning courses can be an efficient means of meeting local demand for specialist educational content. Open source tools are showing promise in the support of distance education, although technical difficulties with installation, customisation and maintenance can be obstacles to adoption (Reilly and Williams, 2006). This finding has been supported through experience with learning management systems in developing countries. Further causes of failure were identified as high information and communication technology (ICT) illiteracy rates and poor marketing strategies to reach learners (Sekakubo et al., 2011).

Agency’s Energy Technology Systems Analysis Program (http://www.iea-etsap.org/web/Documentation.asp). When used by governments there are often obligations to make parts of that code base open to the public. Further MARKAL and TIMES depend on proprietary languages and solvers. Other tools such as RETScreen (www.retscreen.net) provide the user with transparent access to a large number of restrictive models via proprietary software.

19 MIT’s OpenCourseWare, http://ocw.mit.edu, cites a required investment of USD10-15,000 per course
20 http://en.openei.org
21 http://apssforenergy.challenge.gov
Crowdsourcing has been applied to support the scientific method in non-profit environments (Bucheler and Sieg, 2011) and for community engineering for specific innovation projects (Ebner et al., 2009). Support for using sophisticated modelling tools could be provided through technologies such as Legion (Lasecki et al., 2011), which crowdsourced desktop-level support or activities for computer users.

6. Conclusion

Energy analytics is essential to informing the design, implementation and operation of energy systems. This is particularly true in countries, which are undergoing rapid transformations of their energy systems. Developing these resources can be expensive and time consuming. In this context, open energy data and energy applications may induce better-informed energy decisions. These tools, models and applications are often easily adapted to local needs and can be a catalyst for innovation.

Open source energy models provide an easily extendable and open tool for adapting to new scenarios, and are already beginning to provide high quality, peer-reviewed software elements – such as OSeMOSYS. OSS and open data provide accessible, low-barrier platforms for widening the reach and adoption of energy modelling. Crowdsourcing is a now popular technique for outsourcing data collection to open communities of people, and with proper management, the output of such communities may be equivalent or better to strictly expert groups. It is important to ensure that the data collected is both open and linked in order to maximise its uptake and the innovations that follow.

However, it is also vital to maintain a focus on the capacity for using these platforms. Open educational resources create on demand expertise to address local energy challenges as well as help develop sustainable pools of local talent. Current tools for distance learning, community engineering, and open-innovation are highly promising, but more work is required to make them successful. In addition, further research is required to identify the conditions under which OSS projects or crowdsourced data initiatives might be most successful, and when they might fail.

Finally, the storage and maintenance of publically accessible datasets, software tools, and training materials require a long-term funding mechanism. This presents an important opportunity for the international community – especially in relation to developing countries.

Acknowledgements

The authors wish to thank: Holger Rogner (IAEA), Denise Recheis (REEEP), Leonida Mutuku (iHub Research), Ascha Pedersen, Ijeoma Onyeji, and Patrick Nussbaumer (UNIDO), and Dan Fredrickson (Tendril) for their useful input.

Disclaimer: This document represents work in progress and is intended to generate comment and discussion. It is not a fully polished publication. The views expressed herein are those of the author(s) and do not necessarily reflect the views of any organisation. This document has been produced without formal United Nations editing. The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries, or its economic system or degree of development. Designations such as ‘developed’, ‘industrialized’ and ‘developing’ are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process.
References


Brabham, D. C., Feb. 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies 14 (1), 75–90. URL http://con.sagepub.com/cgi/content/abstract/14/1/75


DiFranzo, D., Graves, A., Erickson, J. S., Ding, L., Michaelis, J., Lebo, T., Patton, E., Williams, G. T., Li, X., Zheng, J. G., Flores, J., McGuinness,


