Champion the energy data revolution

Access to rich, high-quality datasets is widely considered to be vital for energy research and public policy. While smart metering has the potential to revolutionise access to energy consumption data, coordinated efforts are needed from government, funding bodies and researchers to overcome the barriers to data access.

Ellen Webborn*, Tadj Oreszczyn*

*UCL Energy Institute, Central House, University College London, WC1H 0NN, UK

Correspondence to e.webborn@ucl.ac.uk

Online searches, social media, location tracking, virtual personal assistants – data are everywhere and in abundance on previously unimaginable scales. The energy sector has been slow to catch up, but smart meter rollouts, smart heating controls, micro-generation and electric vehicles, plus the developments of research fields such as artificial intelligence [1], energy epidemiology [2], and peer-to-peer energy trading [3], signal the big changes underway.

Historically, energy data was dominated by the supply-side, with large fossil-fuel or nuclear generators providing a nation's highly predictable energy needs. Small-scale energy consumers were typically only metered a few times per year. The development of cheap, powerful modelling tools combined with high monitoring costs – and hence a scarcity of consumption data – produced many large, complex models designed to guestimate unverifiable outputs that were then used for policy-making with little appreciation (or quantification) of uncertainty [4].

Smart meters are now being rolled out in many countries, enabling the measurement and reporting of property-level energy consumption on the order of minutes. By 2016, Italy had achieved 100% penetration (thanks to mandates), the United States had 73% penetration, and Canada had achieved 49% [5]. By the end of 2017, there were over 800 million installed smart meters worldwide, of which around 500 million were in China [6]. The potential implications for research are huge, but there are important questions about data ownership, privacy, access and linking that must be asked in order to protect individuals and permit research in the public interest.

The need for data access for research

Tackling climate change necessitates a better understanding of why and how energy is used, as well as the implications of possible policy decisions, new technologies, and new tariffs. This can be achieved with models of energy systems and consumer behaviour – but the value of model outputs is always limited by the quality of its inputs.

The first component of quality input data is high-resolution energy consumption data – both spatially and temporally. High spatial resolution (ideally at postcode level) is needed to understand and plan for localised requirements. High temporal resolution (at least half-hourly, as provided by smart meters in most countries) reveals within-day fluctuations and features, such as peak consumption, correlation between consumption and domestic solar generation, and heating patterns. Half-hourly monitoring will create new opportunities for consumers to buy and sell energy, for example through peer-to-peer trading, switching supplier for different half hours, or 'device as a service' markets. These opportunities are exciting, but to maximise the benefits for consumers and system operators and mitigate against negative consequences, researchers and policy-makers need access to half-hourly consumption data to inform their models.

Related to consumption are generation data, which often exist at the national scale but less commonly at the domestic level. Smart meters typically transmit net energy exported and imported each half hour, but without behind-the-meter monitoring of micro-generation (such as solar panels) it is impossible to know how much energy is actually being generated, and also being self-consumed, thereby displacing demand from the grid. Yet, such information is important for determining the value of investing in micro-generation without a feed-in tariff, the value of schemes such as peer-to-peer trading [7], or for attributing reductions in gross demand to energy efficiency or local generation.

Although smart meter data enable new research avenues, energy data alone are insufficient to make real progress. The second (and no less important) component of quality data is contextual data: factors that contribute to energy consumption patterns or consumer responses to, for example, new tariffs or technologies. This might be information about a dwelling's inhabitants, such as age, affluence, and attitudes towards saving energy. In the non-domestic sector contextual data include component activities such as catering or open-plan working in an office. Possibly the most important type of contextual data for explaining consumption, which is still heating and cooling dominated, is building information [8], such as size, wall and roof composition, and window types. Weather and solar irradiance also play a role, particularly for heating patterns and solar generation.

Linking contextual data to energy consumption data at the building level makes it possible to explain the underlying drivers of consumption and hence how it may change in the future [9]. For example, wall composition and window types partially explain heating patterns and can be used to evaluate the efficacy of energy efficiency measures such as cavity wall insulation. The combination of smart meter and medical data has been used to unobtrusively monitor people with dementia, supporting independent living by detecting changes in routines that may signal help is needed [10].

Energy feedback and tailored advice for consumers can help reduce demand and energy bills, but the effectiveness of such a service depends on the type of information provided and the quality of tailoring to each household/business. Different buildings with different occupants in different locations naturally have different needs. An understanding of the context in which behaviours develop and energy is converted into outcomes such as heat, light and entertainment is important to improve the lives of the fuel poor, address the grid-operation challenges presented by mass-adoption of electric vehicles, and identify home and appliance improvements for saving energy.

The third requirement for quality data is quantity. Any sample needs to be sufficiently large (in terms of number of buildings or households) to be representative of the wider population of interest and robust to statistical significance tests. The data also need sufficient longevity (ideally several years) to reveal seasonal variations (highly influential in energy data), capture long-term trends, and include rarer events that drive changes in energy use, such as extreme weather or socio-economic events. Large datasets are also important for minimising the problems caused by missing and erroneous data points, which inevitably occur in any detailed real-world dataset.

Barriers to access

Despite the many advantages, opening up datasets, even for academic research in the public interest, has not historically been championed in the energy sector. In many countries this is due to the privatisation of energy data, which means access requires legislation or goodwill on the part of utility companies. Another factor is the lack of value that has typically been attributed to energy data, partly due to the lack of urgency and prioritisation with which climate change has historically been treated.

Data governance and privacy issues have risen up the public agenda in the wake of high-profile data scandals and legislation such as the General Data Protection Regulations (GDPR). Consumers are becoming more aware of the value of their data and the potential for companies to profit from them without their knowledge or fully-informed consent. It remains to be seen whether this will mean fewer people consent to share their energy and contextual data with researchers, but fears about spying have resulted in some public pushback to smart meter installations, such as the "Stop Linky" campaign in France which is preparing a class action against smart meter installers.

Half-hourly energy consumption data is a relatively new phenomenon so there is not yet a clear consensus on how sensitive the data are, or are perceived by the public to be. This is an evolving conversation and has naturally (and reasonably) led many to err on the side of caution in their willingness to share data. A strong desire not to be accused of spying on citizens has led to some governments refusing to access smart meter data and delaying the rollout of smart meters. In the Netherlands the rollout was delayed because smart meters could infer information about households' religion (classified as private information) from energy consumption patterns during Ramadan [11]. In Great Britain access to smart meter data is controlled by each consumer (who can also refuse the installation), so access to identifiable energy consumption data requires opt-in consent from each individual, as well as technical and legal infrastructure to access and store the data – a sizable task for any research organisation.

Although governments often have (or could have) access to many large datasets, such as smart meter data, national surveys, or administrative data, they rarely have the resource or political will to combine and explore them to the depth possible by the research community. Contextual data are often collected for specific administrative purposes such as taxation, and/or with specific legislative issues. They may be held by different government departments who do not value their data being used for applications beyond those initially intended, and who are hesitant to allow property-level linking.

There are also barriers to linking or sharing data collected by researchers. It is rarely possible to make data attained through collaboration with private companies available to other researchers, even after anonymisation. Publicly-available datasets often lack mechanisms for address-level linking with new datasets, because address data is understandably kept private. Some journals facilitate, mandate or encourage data sharing, but even when this is possible there is often little incentive for researchers to make use of it. Too often data provision is an after-thought, without dedicated resource or planning from a project's conception.

Potential solutions

Given sufficient motivation and safe-guarding protocols, existing datasets could be opened up by government for research deemed in the public interest. University research ethics committees and research council funding boards play a key role in public interest decisions in academia while organisations like the Smart Meter Energy Data Public Interest Advisory Group [12] are emerging to inform discourse and establish principles more broadly. Once a public interest case has been established, access to raw data could be granted in a secure environment to approved researchers, as part of an independent department to authorise and promote use of data across government departments.

Action can be taken to recognise and develop the shared interests of government and industry in improving data availability. In the UK an Energy Data Taskforce has been created to explore how better data availability and transparency between industry and the public sector can improve competition, innovation and markets in the energy sector [13]. This model may be a valuable example for other institutions to follow. Academics also have an important role to clearly communicate the value of the data sharing via initiatives such as the IEA EBC Annex 70 Building Energy Epidemiology [14].

The energy sector could learn a lot from progress made with health data in countries where extremely sensitive personal data are being made available to accredited researchers because of their perceived importance for health research. The UK's National Health Service (NHS) offers a Data Access Request Service – operated by Health Data Research UK, an independent, non-profit national institute for health data science – for clinicians, researchers and commissioners to request data to help improve NHS services. The UK Medical Research Council also provides a Regulatory Support Centre that offers guidance and training materials for regulators and researchers.

Privacy concerns need to be carefully considered before widening access to energy data but evidence suggests that progress is possible. A 2018 UK survey found that 63% of participants considered medical records to be one of the three most sensitive types of data, compared with only 4% for half-hourly energy consumption data [15]. Fears that households might be identifiable once datasets are linked are not unfounded, and consumers have specific concerns if data are combined for profiling activities for commercial services, resulting in their becoming rationally disengaged [16]. Data privacy and protection must not be taken lightly [17], but many data privacy concerns can be resolved with appropriate safeguarding and statistical disclosure procedures such as the "5 Safes" protocol used by the UK Data Archive. It is also important to keep consumers fully informed and in control of access to their data. For example, in Estonia any consumer can view their dwelling's smart meter data and control which organisations are granted access via an online portal [18].

Researchers should be incentivised – and supported – to manage and share data by default, with requirements for data management plans from a project's outset. Such practices will enable researchers to better appreciate the wider benefits of good data curation and sharing, which include quality-checking from replication and the enabling of new research. Publishers and funding bodies can drive this culture forward by requiring data curation, data referencing and sharing costs to be built into research proposals. In 2017 the Netherlands launched the National Plan Open Science with the ambitions of 100% open access publishing, research data optimally suited for reuse, and corresponding systems to recognise and reward researchers [19]. This is related to a broader EU open science action plan, which includes the European Open Science Cloud for data sharing.

The Smart Energy Research Lab (SERL) [20] in the UK is an example of a research project designed with the wider research community in mind. With opt-in consent, SERL will collect half-hourly smart meter data from a target of 10,000 homes by the end of 2020 and link them with participant survey responses and other datasets such as local weather and energy performance certificates. UK university researchers will be able to apply for access in a secure environment provided that the research is in the public interest and that strict ethical and security standards are met. This example could be replicated in other countries, but challenges still exist. For example, large-scale, reliable datasets of internal building temperatures remain scarce, despite their importance for understanding and modelling energy demand. Smart thermostats may make key contributions here in future.

The changes needed for access to energy and contextual data require strong leadership to not only drive forward high-level strategy but to overcome practical and very detailed problems, normally at an operational level within organisations with the keepers of the data. Moreover, despite efforts from initiatives like the Open Data Institute and the Administrative Data Research Network, it can take years to get permission to link existing data for research. Time is running out to meet carbon targets and transform the energy system – we cannot afford long delays in accessing and linking data.

Acknowledgements

This piece is supported by EPSRC grant EP/P032761/1. The authors are grateful to Simon Elam, Eoghan McKenna and Ian Hamilton at the UCL Energy Institute for useful discussions and feedback.

References

[1] S. Kalogirou, Artificial Intelligence in energy and renewable energy systems, Nova Publishers, 2007.

[2] I. Hamilton, A. Summerfield, T. Oreszczyn and P. Ruyssevelt, "Using epidemiological methods in energy and buildings research to achieve carbon emission targets," *Energy and Buildings,* vol. 154, pp. 188-197, 2017.

[3] C. Zhang, J. Wu, C. Long and M. Cheng, "Review of Existing Peer-to-Peer Energy Trading Projects," *Energy Procedia,* vol. 105, pp. 2563-2568, 2017.

[4] M.-H. Laurent, B. Allibe, T. Oreszczyn, I. Hamilton, C. Tigchelaar and R. Galvin, "Back to reality: How domestic energy efficiency policies in four European countries can be improved by using empirical data instead of normative calculation," in *European Council for an Energy Efficient Economy (ECEEE) Summer Study Proceedings*, 2013.

[5] EXL Utilities Academy, “Smart metering: what the U.K. can learn from other countries,” EXL, 2016.

[6] International Energy Agency, “Perspectives for the Clean Energy Transition: The Critical Role of Buildings,” 2019.

[7] E. McKenna, E. Webborn, P. Leicester and S. Elam, “Analysis of international residential solar PV self-consumption,” in *European Council for an Energy Efficient Economy (ECEEE) Summer Study Proceedings*, 2019.

[8] G. M. Huebner, I. Hamilton, Z. Chalabi, D. Shipworth and T. Oreszczyn, “Explaining domestic energy consumption - The comparative contribution of building factors, socio-demographics, behaviours and attitudes,” *Applied Energy,* vol. 159, pp. 589-600, 2015.

[9] I. Hamilton, T. Oreszczyn, A. Summerfield, P. Steadman, S. Elam and A. Smith, “Co-benefits of Energy and Buildings Data: The Case For supporting Data Access to Achieve a Sustainable Built Environment,” *Procedia Engineering,* vol. 118, pp. 958-968, 2015.

[10] C. Chalmers, W. M. M. Hurst and P. Fergus, “Smart meter profiling for health applications,” in *International Joint Conference on Neural Networks (IJCNN)*, 2015.

[11] C. Cuipers and B.-J. Koops, “Smart Metering and Privacy in Europe: Lessons from the Dutch Case,” in *European Data Protection: Coming of Age*, S. Gutwirth, R. Leenes, P. De Hert and Y. Poullet, Eds., Springer Science & Business Media, 2012.

[12] Centre for Sustainable Energy, “Smart Meter Energy Data Public Interest Advisory Group,” 2019. [Online]. Available: https://www.smartenergydatapiag.org.uk/. [Accessed 1 May 2019].

[13] Energy Data Taskforce, “Terms of reference for the energy data taskforce,” GOV.UK, 2018.

[14] RCUK Centre for Energy Epidemiology, “IEA EBC Annex 70 – Building Energy Epidemiology,” 2017. [Online]. Available: https://energyepidemiology.org/. [Accessed 11 March 2019].

[15] Ofgem, “Consumer views on sharing half-hourly settlement data,” Ofgem, London, 2018.

[16] Which?, “Control, Alt or Delete?,” BritainThinks, 2018.

[17] C. Véliz and P. Grunewald, “Protecting data privacy is key to a smart energy future,” *Nature Energy ,* vol. 3, pp. 702-704, 2018.

[18] Elering, “e-elering,” 2019. [Online]. Available: https://elering.ee/en. [Accessed 10 April 2019].

[19] T. T. Chan and I. Meijer, “The Netherlands' plan on open science: Open science monitor case study,” Publications Office of the European Union, 2019.

[20] UCL Energy Institute, “Smart Energy Research Lab (SERL),” 2019. [Online]. Available: https://www.ucl.ac.uk/bartlett/energy/research/research-projects/smart-energy-research-lab-serl. [Accessed 11 March 2019].