

# Show me the data

All papers published in *Nature Energy* must now contain clear statements on how to access the data contained within them or what restrictions prevent the data from being shared.

Sharing research data helps to increase the transparency and reproducibility of scientific publications, and enables others to build on those papers and take their findings further. It is a practice that is increasingly called for in academic communities. In recent years, making data publicly available has also become a condition of many funders. Of course, publishers have a role to play in supporting these important endeavours and so, following a recent trial here at Nature Research, we are now requiring all authors whose work is accepted for publication to provide a data availability statement in their paper.

These statements are intended to provide readers with information on where they can find the data relevant to understanding, reproducing and further developing a paper's findings. The full policy can be read at <http://go.nature.com/2eG4GcL>. Typical statements may provide links to publicly available datasets — either those owned by the authors<sup>1</sup> or by a third party — or explain any restrictions that inhibit the sharing of data (see, for example, ref. 2). At the very least, a data availability statement should state that all relevant data are available from the authors upon reasonable request.

This new policy comes together with changes in the way we cite and link to research data. We now encourage the citation of datasets that have been assigned digital object identifiers (DOIs) in reference lists.

These changes are intended to further the Nature journals' commitments to reproducibility and to increase the reuse of primary research data. We also hope they will work in tandem with — and hopefully augment — the requirements of the many funders calling for similar consideration of data. Research Councils UK (RCUK) consider data availability to be one of their key aims. Among their endeavours in this area, RCUK requires that research data they have funded be made openly available with as few restrictions as possible. They also expect that publications arising from research they fund will include information on how to access the data behind the study. Elsewhere, the US Department of Energy (DOE) and the European Commission have structures in place that encourage the sharing of data generated from their funding, requiring researchers to submit

data management plans with their funding proposals, which outline the researchers' responsibilities and what steps they will take to ensure access to data.

These policies of openness and availability bring many benefits. The DOE has launched a number of campaigns to increase public access to energy data in sectors including building performance and hydropower potential. By removing barriers to finding datasets, the US government hopes it can spur further understanding and innovation that will aid in the quest for cleaner and more sustainable energy systems.

More broadly, in the private sector, electricity network system operators are providing access to information on capacity and demand<sup>3</sup>. This type of large-scale data allows researchers to investigate the behaviour of real power systems and uncover important features that may not be picked up in laboratory simulations<sup>4</sup>. It also provides a more detailed characterization of energy demand profiles that — when coupled with additional socioeconomic and demographic data — can cast light on the various factors affecting residential energy consumption. Such insights could lead to better design of interventions or efficiency measures. Modern gas and electricity meters ('smart meters') now offer a range of intelligent functions, communicating energy data between suppliers and customers in real time. It is anticipated that these powerful data on energy use will increase our understanding of the determinants of energy consumption and our interaction with energy technology.

At the lab bench level, sharing data on device performance can also help with verification and replication. Meanwhile, providing access to data on material characteristics can cut down on re-measuring fundamental properties or on running computations repeatedly. Ventures like AFLOWlib (<http://www.aflowlib.org/>) or the Materials Project (<https://materialsproject.org/>) are designed to address this issue, offering researchers access to large databases of computed material structures and properties.

Of course, simply providing data without any context would make a lot of this process meaningless. As such, the establishment of best practices around data sharing will be increasingly important. Publishers have an

important role to play, and should expand their efforts to support researchers. Our publisher, Springer Nature, is working to introduce standardized research data policies across all its journals to enable different communities to encourage data sharing in a way that reflects their own needs and circumstances<sup>5</sup>. Ensuring that methods sections in publications are clear and contain all the necessary information is an obvious simple step, although it is usually limited to whatever data are shown in a paper. For data in repositories, inclusion of descriptive or supporting information offering essential information on the experimental or computational processes, the survey structure and methodology, the curation approach, and so on, is vital. Data journals like *Scientific Data* are a welcome and useful venue, in that regard, and are already providing valuable insights into energy-relevant domestic consumption datasets<sup>6</sup>; more tools that connect such documentation with repositories will help with the growth of data sharing practices.

The Nature journals' new data availability statement policy will naturally bring to light a variety of different approaches and flavours of statement, particularly for a journal like *Nature Energy*. Different disciplines — and communities within those disciplines — have different expectations about data sharing and different levels of awareness about their funder's or institution's requirements. We hope that where awareness is low, our new policy will improve knowledge and understanding. Where awareness is high, we hope the policy will galvanize efforts and improve standards. We will be working closely with our authors and communities to make sure we understand their needs and can provide transparent statements that increase the utility of papers. Ultimately, we believe that providing clear and constructive information about data is integral for research to be able to address the energy challenges we face today and in the future. □

## References

1. Crespo-Quesada, M. *et al.* *Nat. Commun.* **7**, 12555 (2016)
2. Popp, D. *Nat. Energy* **1**, 16020 (2016).
3. *Renewables Watch* (California ISO, 2016); <http://go.nature.com/2d9fzov>
4. Ji, C. *et al.* *Nat. Energy* **1**, 16052 (2016).
5. *Promoting Research Data Sharing at Springer Nature* (Springer Nature, 2016); <http://go.nature.com/2by6l6x>
6. Makonin, S., Ellert, B. & Popwich, F. *Sci. Data* **3**, 160037 (2016).