SCIENCE AND SOCIETY

The Smart Electricity Grid and Scientific Research

Researchers have an interest in access to, and preservation of, data that will soon be collected by the smart electricity grid.

Jan Beyea

So-called "smart" meters and appliances have the potential to save energy, to shave peak electricity usage, and to reduce risks of blackouts (*1–6*). Typical smart meter designs include periodic transmission of current, phase, and frequency data from the user to the electricity distribution company. Utilities will use the data in billing calculations under time-of-day pricing, for load-management research, to provide customer feedback, and/or to adjust customer appliances.

However, there has been little discussion of the full scientific, economic, and historic potential of these data, whose usefulness may extend beyond the original purposes for which the data will be collected and stored. Input is needed from third-party researchers (neither customers nor utilities) at this early and critical stage of the smart grid's development, with legislation and regulations yet to come. Otherwise, it will be hard for regulators to reach a balanced decision on which high-resolution data should be stored, for how long, and to whom it should be made accessible inside and outside the collecting utility.

How Much Data?

The Obama Administration is projecting 40 million smart meters in the United States "over the next few years" (*7*). The European Union is planning to install 245 million within the next 10 years (*8*, *9*). Some systems in Europe are already transmitting at 1-s intervals (*10*). Computer algorithms are capable of converting 1-s interval data into a timeline of smart appliance use (*6*, *11–14*) by making use of libraries of characteristic appliance signals; performance degrades as the time resolution becomes coarser (*11*).

The cost of storing a customer's data would be trivial compared with a customer's annual electricity bill, even without data compression. A stream of four numbers (current, phase, frequency, and time) transmitted every second amounts to 0.5 gigabytes of data per year per customer, about as much data as are stored on one standard compact disc. The data from 100 million customers, roughly the amount of households in the United States, would amount to 50 petabytes per year before compression. About 0.05 petabytes would be needed to store data at 15-min intervals, reflective of many first-generation smart meters. The sample sizes potentially available for research are huge, but manageable, even without subsampling. Data sets in the multipetabyte range are already being analyzed in biology and physics (*15*). For example, about 2.5 petabytes of data are stored in mammograms each year in the United States (*15*). About 15 petabytes are projected to be stored each year at the Large Hadron Collider (*16*).

CREDIT: DEBBIE NODA/NEWSCOM

Privacy

Privacy protection is a major challenge for smart meter systems and a challenge for any use of social-spatial data by the scientific community at large, because it is possible to use high-resolution data to infer personal habits. The implications related to unwarranted government searches under the Fourth Amendment of the U.S. Constitution are sobering (*11*, *17*), and privacy guidelines are being developed and debated (*18–20*).

But if privacy-protecting access techniques are approved by regulators for energy management purposes, some variant should merit approval for use in third-party research. For instance, data may be confined to utility servers, with analysis algorithms transmitted from researchers and executed on the server. Also, many researchers have great experience dealing with the privacy of study subjects through service on, and interactions with, Institutional Review Boards (*21*, *22*). Consent could be obtained, just as it is in any human-subject study involving individualized data. However, some privacy risks will always remain (*20*), and some data will have to be off limits.

Promising Uses of Data

Even data that are stripped of all recorded personal identifiers to minimize the risk of privacy invasion (de-identified data) could help answer questions of interest to energy researchers and social scientists. Are there ways to spot impending electricity outages? How does energy usage correlate with current events, appliance standards, and price? Which utility programs work best to improve energy efficiency? How are appliance efficiencies changing over time? How varied is the usage of appliances from person to person, from region to region, and from decade to decade?

Even more useful would be de-identified data that retain some geographic information, such as county, town, census tract, and/or census block. Privacy risks would be minimal, and demographic data collected in the census, as well as local meteorological data, could be added to the electricity data set, thereby increasing its usefulness for group-level analysis. Economists and energy modelers could use the combined data to examine how electricity usage—even appliance by appliance, in some cases—is related to price, average household income, and average family size, as well as other aggregated variables available in census block data. The simultaneous responses of millions of customers to episodic weather changes, national tragedies, presidential proclamations, changes in laws,

Consulting in the Public Interest, Lambertville, NJ 08530, USA. E-mail: jbeyea@cipi.com

and cultural trends could be studied by social scientists. Comparisons across countries could be fascinating.

Electricity use in the home is of interest to epidemiologists. Appliance usage may be a main effect variable in epidemiologic studies of health impacts of electromagnetic fields (*23*). It may be a confounding variable in studies of air pollution or other causes of disease (*24*). Smart meters could provide direct data on use of some appliances (e.g., microwaves and electric blankets), as well as a check on usage data for other appliances obtained from the questionnaires normally used in such studies.

Other types of health studies (e.g., intervention studies) can benefit from access to detailed, individualized electricity data. Algorithms might be designed, for example, to infer how many times per day a refrigerator door was opened (relevant to dietary and obesity studies), the times of day a residence was occupied (relevant to air pollution health studies), and even the pattern of sleep (relevant to a variety of health studies).

Some uses of data may be unknown today, but important decades hence. Saved electricity data may help future generations solve unanticipated problems, just as 1960s satellite images collected before awareness of the climate problem assist us today (*25*).

Data Ownership and Access

Individual user data are considered to belong either to the customer or to the collecting utility, with access to the data in aggregate limited to the utility, subcontractors, and third-party energy-management companies. But other third-party researchers, who do not yet appear to be viewed as major potential users of data from smart meters, need to be added to the picture; otherwise, privacy regulations might inadvertently freeze them out (*1*–*6*, *8*, *11*, *17*, *26*). Even if access were expanded to include such outside researchers—under as yet undetermined restrictions—it is common practice for utilities to destroy customer data after a delay period chosen to allow for long billing disputes. This delay is often 7 years in the United States (*17*). This means that data would be lost to historians, unless specific regulations were implemented to allow data, or a subset of them, to be transferred to an archival repository.

How might a research project get approval to access data? A review board or staff at a utility commission might handle the duties of performing due-diligence assessments of research requests, working in consort with the Institutional Review Boards at the requesters' institutions. Funding of the project by government agencies that used a peer-review grant process might be a requirement that would ensure, for example, that sample size, measurement error, and other issues have been accounted for in any proposed use of specific appliance data.

Electricity data are only one example where the government has bargaining power (e.g., via public utility commissions) to facilitate access to data by qualified researchers, but taking action is timely, because this data revolution is in its very early stages. In addition, tax benefits may soon be given for installation of smart meter systems; sharing of data (with privacy protection mandated) could be made a prerequisite.

Furthermore, it is the customers as a group who will pay for the meters. As part of the "regulatory bargain" (*27*) that gives government control of prices, profitability, and service standards of monopoly utility companies, regulators will allow the electricity distribution company (a monopoly) to pass on direct costs, raise necessary capital, and make a limited, but guaranteed, rate of return on prudent investments. Under such arrangements, regulators could include (modest) costs needed to facilitate third-party research.

Planning for the Future

Now is the time for researchers to join discussions about data from smart meters. The National Institute of Standards and Technology is developing model standards in the United States for information management of smart grid devices (www.nist.gov/smartgrid/). The European Commission has a Smart Meter Coordination Group (www.cen.eu). Contact with state and local electricity regulators could be fruitful. Many utility companies will likely be helpful, if regulators consent to third-party research. Studies could be carried out in collaboration with industry research institutions such as the Electric Power Research Institute. Some utilities, however, may recoil at outside research, for example, on health effects from electromagnetic fields, because of concerns about possible future liability and/or bad publicity. Large samples can turn up associations that are very weak, but still statistically significant, which might be problematic for companies, depending on the presentation of results.

At this moment of transition for the electrical grid, researchers and their institutions can work with local utilities and local regulators, and scientific societies can work at the national level, to present the case for the benefits of access and research, to describe scientists' experience with privacy protection, and to help make sure that third-party researchers are not excluded by rules set without considering them. Reviews of third-party research concepts by a National Research Council panel that included privacy scholars would be useful, as would reviews by equivalent bodies outside the United States. Their reviews and recommendations could consider other massive data sets—e.g., cell-phone geospatial data, whose access might also be made feasible as part of regulatory and legislative bargains.

References

1. R. J. Meyers, E. D. Williams, H. S. Matthews, *Energy Build.* **42**, 563 (2010).
2. A. Faruqui, R. Hledik, S. Sergici, *Electr. J.* **22**, 55 (2009).
3. G. Strbac, *Energy Policy* **36**, 4419 (2008).
4. S. Roberts, *Energy Policy* **36**, 4552 (2008).
5. M. J. King, K. King, M. B. Rosenzweig, *Electr. J.* **20**, 38 (2007).
6. M. Newborough, P. Augood, *IEE Proc.* **146**, 283 (1999).
7. White House, *President Obama Announces $3.4 Billion Investment to Spur Transition to Smart Energy Grid* (White House, Washington, DC, 2009); www.whitehouse.gov/the-press-office/president-obama-announces-34-billion-investment-spur-transition-smart-energy-grid.
8. M. Venables, *Eng. Technol.* **4**, 10 (2009).
9. M. A. Lisovich, S. B. Wicker, *IEEE Proc. Power Syst.* **1**, 1 (2008).
10. M. K. El Mahrsi, S. Vignes, G. Hebrail, M.-L. Picard, *Proceedings of the Third International Conference on Research Challenges in Information Science* (RCIS 2009), Fez, Morocco, 22 to 24 April 2009 (IEEE, New York, 2009), pp. 395–402; http://ieeexplore.ieee.org/iel5/5073917/5089258/05089303.pdf%3Farnumber%3D5089303&authDecision=-203.
11. M. A. Lisovich, S. B. Wicker, *IEEE Proc. Power Syst.* **1**, 1 (2008).
12. A. Prudenzi, *IEEE Power Eng. Soc. Winter Meeting* **2**, 941 (2002).
13. S. Drenker, A. Kader, *Comput. Appl. Power IEEE* **12**, 47 (1999).
14. F. Sultanem, *IEEE Trans. Power Deliv.* **6**, 1380 (1991).
15. A. J. G. Hey, A. E. Trefethen, in *Grid Computing: Making the Global Infrastructure a Reality*, F. Berman, G. C. Fox, A. J. G. Hey, Eds. (Wiley, Hoboken, NJ, 2003), pp. 809–824.
16. The European Organization for Nuclear Research (CERN), *Worldwide LHC Computing Grid*; http://public.web.cern.ch/public/en/LHC/Computing-en.html.
17. J. I. Lerner, D. K. Mulligan, *Stanford Technol. Law. Rev.* **2008**, 3 (2008).
18. U.S. National Institute of Standards and Technology (NIST), *Smart Grid Cyber Security Strategy and Requirements* (Draft NISTIR 7628, NIST, Gaithersburg, MD, 2009); http://csrc.nist.gov/publications/drafts/nistir-7628/draft-nistir-7628.pdf.
19. European Committee for Standardization (CEN), Workshop on Data Protection and Privacy; www.cen.eu/cenorm/sectors/sectors/isss/activity/wsdpp.asp.
20. D. Lazer *et al.*, *Science* **323**, 721 (2009).
21. T. W. Rice, *Respir. Care* **53**, 1325 (2008).
22. K. B. Enfield, J. D. Truwit, *Respir. Care* **53**, 1330 (2008).
23. G. Mezei, M. Gadallah, L. Kheifets, *Epidemiology* **19**, 424 (2008).
24. M. D. Gammon *et al.*, *Breast Cancer Res. Treat.* **74**, 235 (2002).
25. H. Pringle, *Science* **327**, 1322 (2010).
26. Office of Science and Technology Policy Forum, Consumer Interface with the Smart Grid Blog, http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/OSTPBlogWeek2.
27. D. N. Jones, *Energy Law J.* **22**, 41 (2001).

10.1126/science.1189229