Energy Data Access: Who Wants the Data?

Mikhail Haramati, Opinion Dynamics

ABSTRACT

We are in the midst of a data revolution. Companies like Google and its competitors have changed our relationship to all kinds of information and our expectations about privacy and data access are rapidly changing. This revolution is being widely felt in the building energy arena where smart meters collect high-resolution usage data and data analytics companies scrutinize it for savings opportunities. From grid operations to evaluation results, as the world becomes increasingly full of data on all aspects of our lives we must decide how to organize it and who should have access. For energy data this means understanding who is interested, what they want to use it for, and at what level the data can be provided.

This paper presents the results of a survey of 108 energy professionals in California in 2013 who expressed an interest in using energy data. Respondents represented government at various levels, consultants, non-profits, academics, utilities, and those with products in the private sector. Promoting the indirect outcomes of open access to data was cited as the top concern regarding data accessibility and four out of five reported being interested in customer billing data followed closely by energy efficiency information. The paper describes the methods employed, respondent demographics, and survey findings. The research was conducted in conjunction with a qualitative evaluation effort on energy data access that included in depth stakeholder interviews and an examination of policy options for increasing access to energy data in California.

Introduction

Never before have we been able to collect and analyze as much information about the world. The amount of feedback that we can now get about our own lives is extraordinary. Yet because this quantity of information is new, we, as a society, are still trying to decide where to draw the line between protecting the privacy and security of the data and realizing its beneficial uses. The fact that public discourse on the topic is often philosophical indicates that we haven't yet come to agreement on how to apply our values to this new arena. Big data is increasingly becoming a part of everything we do, and the energy field will need to adhere to any rules that are established for data in the larger sphere.

The critical issue among those concerned with sharing energy data is protecting individual privacy. There is particular sensitivity around residential energy usage information because it is reflective of the activities of occupants within the home. Privacy of the home holds a special status in the US supported by the Fourth Amendment to the Constitution[1] which protects individuals from unwarranted government search and seizure. There are also concerns

[1] The Fourth Amendment to the United States Constitution requires probably cause for search and seizure of property: "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

that providing information on habits of occupants will make them more vulnerable to crime and therefore compromise their safety.

The trade-off between use and privacy is difficult and exists in every area where issues of big data are being explored and debated. From sharing of medical records, to use of data in law-enforcement, to the online publishing of public records, we're still deciding how much information should be kept private, who should have access to it, and when it should and should not be used. The rules currently vary by country, by state, and by local jurisdiction.

Understanding how to balance the risks of data sharing with the potential benefits of its use was a primary focus of this research. For energy data, some states provide access to data for all users online for everyone to see so that those with high energy use can be called out for it. Other jurisdictions only publish the energy use of public buildings.

This paper presents the results of research conducted in California in the spring of 2013 to develop a better understanding of who is interested in the energy data held by the California Investor Owned Utilities (IOUs) and what they would like to do with it. The primary research was designed to provide insight on the type of data, users, and potential applications.

The client for the research was the California Governor's office, and the work was conducted in part to inform the policy questions identified by the California Public Utilities Commission (CPUC) in Phase 3 of its Smart Grid Proceeding, R.08-12-009. In this proceeding, the CPUC sought to resolve the matter of how to provide access to data collected from utility-installed smart-meters, while still protecting the privacy of those customers and following existing state and federal rules regarding data privacy. Party comments in the rulemaking referenced various datasets and possible purposes for the use of those datasets. While anecdotal cases were presented in the proceeding, little was known about how representative the party concerns were relative to the population of potential data users. This research was designed to help inform the discussion.

Guiding research questions included:

- Who is trying to access this data?
- What data are potential users interested in?
- For what purposes do they want to use it?
    - If they have tried to gain access in the past, what has been their experience and what are the pain points?

Stakeholder Interviews

Methods

Stakeholder interviews were utilized throughout the research process to develop an understanding of the complexities of energy data access in California, and were conducted in two phases. Phase I Interviews were used for concept formation and to develop the questions for the Energy Data User Survey. The interviews also facilitated the identification of stakeholder concerns, and energy data and smart meter vocabulary to aid in the formation of the user survey. Phase II interviews were conducted after the development of the User Survey, and were used to explore themes and experiences identified in the survey, discuss possible solutions, and react to

survey results.[2] All interviews took place between January and May 2013 and were conducted in person or by phone.

Results

Common themes in the stakeholder interviews included difficulty understanding the differing utility data request process,[3] frustration with utility responsiveness in general, confusion about what datasets exist, and the need to protect customer privacy while allowing access to the data. Stakeholder concerns centered on privacy, access to existing data sets, and the need for a clear, standard process for requesting data. All stakeholders were in agreement that the data is useful and should enter the public realm in some form. The list below summarizes the potential benefits and concerns.

Potential benefits

- Maintaining transparency of utility operations
- Maintaining transparency of the state, the local government, or the regulator
- Providing open access to data
- Promoting the indirect outcomes of open access to data such as programmatic efficiencies and cost savings, improving program offerings, reducing procurement costs, and leading to more equitable program delivery
- Reducing fraud from energy contractors who provide services to homeowners and renters
- Keeping costs low for utility ratepayers

Concerns

- Maintaining utility customer privacy
- Concern about government possession of data

Experiences Accessing Data

Many respondents spoke about past experiences attempting to access utility data. The descriptions were largely negative and highlighted frustration and confusion with data access rules and requirements. Pain points were encountered at nearly every step in the data access process. Procedural difficulties involved filling out the utility data request form, following up with the utilities, and receiving the data. Respondents reported technical issues with cleaning and organizing the data once received and the difficulty in matching it to other data sets.

[2] Depth interview participants included the following organizations: UCLA, The Energy Institute at Haas, UC Berkeley, CPUC—Policy and Planning Division; Energy Division; Office of President Peevey; Office of Commissioner Ferron; Office of Commissioner Florio, California Governor's office, Division of Ratepayer Advocates (DRA), California Energy Commission (CEC), Climate Protection Initiative (CPI), Environmental Defense Fund (EDF), California Center for Sustainable Energy (CCSE), Lawrence Berkeley National Laboratory, OPOWER, ICLEI, A small energy storage company in the San Francisco Bay Area, Efficiency.org, DRA, CEC, The Utility Reform Network (TURN), and The Urban Institute at UC Berkeley School of Law.

[3] While the Green Button initiative makes it easy for customers to access their own data, there was not a standard process in place for requesting data from groups of customers.

Respondents were asked about possible technical and regulatory solutions to address the needs of data requesters. Most believed that a technical solution was possible, but cited frustration with the current process, and did not have specific ideas for a solution. Some were aware of the national, DOE-led Green Button[4] initiative that allows customer of participating utilities to download their own energy usage data; two respondents reported trouble receiving client data through the initiative.

Other observations

1. *Confusion about applicability of Privacy Rules.* There appeared to be general confusion about what is allowed under existing privacy rules. The CPUC's Office of Ratepayer Advocates (ORA)[5] reported that on utility has been providing smart meter data under a non-disclosure agreement (NDA) but that this likely violates the consent requirement of the Privacy Rules.
2. *Lack of technical background among stakeholders*. Most stakeholders and parties involved in the Smart Grid Working groups do not have a background in computer science and are not familiar with the technical possibilities for data protection including aggregation, anonymization and selective access.
3. *Diversity of need.* Some stakeholders are interested in general customer usage information, others are looking for 15-minute interval smart meter data, some would like better access to evaluation results and yet others are looking to access renewables supply information and grid reliability data.
4. *Difficult to plan for the future when data on current performance is inaccessible.* For local governments trying to plan programs, data on participation, program costs, energy savings, and other evaluation results are critical elements in deciding which programs to continue and which to end.
5. *The CA Solar Statistics website puts energy efficiency and demand response projects at a data disadvantage relative to solar*. One small energy storage company said that they felt that non-solar projects were at a disadvantage due to the excellent data available to the solar industry through the CPUC website California Solar Statistics.[6]

Energy Data User Survey

Methods

The energy data user survey was targeted to current and future users of energy data in California, and covered all research questions. Email participants received a message with an introduction to the survey and a link to take it online. The survey had 13 questions and most respondents spent 10-15 minutes using the online instrument. A total of 108 respondents took the survey. The survey was pretested in the last week of March and consisted of six user survey

[4] More information about Green Button can be found at http://energy.gov/data/green-button and http://greenbuttondata.org/.
[5] The Office of Rate Payer Advocates is an independent office within the California Public Utilities Commission whose job is to represent and advocate for ratepayer interests in regulatory proceedings. http://www.ora.ca.gov/default.aspx.
[6] http://www.californiasolarstatistics.ca.gov/.

responses and review by two survey design professionals. The resulting survey had fewer components and room for write-in answers for most questions.

Recruiting was performed over a three-week period, April 1-24, 2013 and sought to reach those who have tried to access utility data in the past or would be likely to do so in the future. Respondents were recruited via email, linked-in message, or in-person request. Respondents were contacted via the following means.

- Smart grid service list, R.08-12-009
- Direct email to 200+ primary contacts
- UC Berkeley Energy Community email
- LinkedIn messages
- Paper survey at energy events

A snowball sampling method[7] was used to maximize survey distribution and participation from individuals not currently participating in a CPUC or CEC rulemaking. Because accurately describing a population of future users is difficult, current user interests and needs were used as an indicator of the type of people who would likely want to access the data if were publicly available. The snowball sampling method was also particularly useful for reaching colleagues and business contacts of direct recipients of the survey email. Figure 1 shows the survey dispersion for one email sent to an energy consultant that reached four other connected respondents who took the survey. It's particularly important to understand the connections between respondents as results from all types of respondents are needed to understand the use and role that such data currently play in the existing climate and energy landscape.

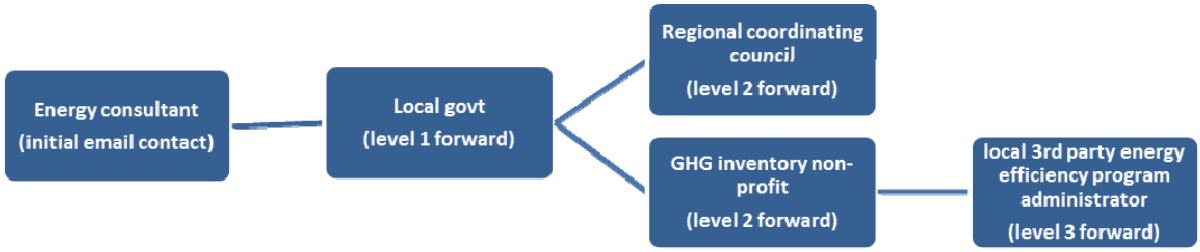

Figure 1. Snowball sampling survey dispersion--showing levels of forwarding for an initial email to an energy consultant.

Demographics

Respondents were asked to identify their employer type as shown in Figure 2 below. Among the 30% who said that they worked for the government, the majority were employed by cities and counties. Local government made up 62% of the government employers, and 22% of all respondents. One fifth worked at state government level for organizations like the CPUC or CEC. Federal government made up 13% of the government wedge, and employers were national laboratories and universities.

[7] Snowball sampling method is a non-probability sampling technique that groups typically used when the population is unknown or difficult to locate. It relies on previously identified group members to identify other members of the population, and as new members identify others, the sample grows like a snowball. (Henry, G. 1990. "Practical Sampling." *Applied Social Research Methods Series, Volume 21*: 21. Sage Publications.)

Respondents were also asked to identify by job type, but the responses for this category were less consistent so only demographics by employer are displayed.

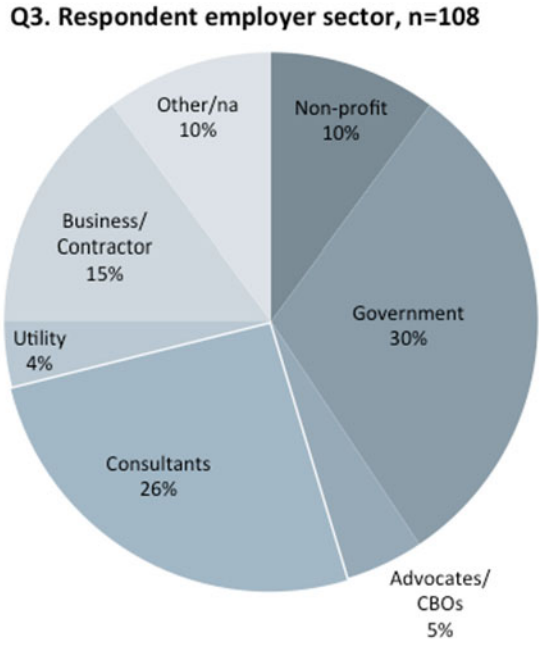

Figure 2. User survey respondent demographics by employer.

Findings

Benefits and Concerns about Open Access

To understand the range of attitudes and opinions towards open access, respondents were asked to rank benefits versus concerns about publicly releasing the data. The eight benefits and concerns were selected based on issues raised by parties in the smart grid proceeding and during the stakeholder interviews. Surprisingly, 36% of respondents (30/83) believed that promoting the indirect outcomes of open data was most important. Indirect outcomes were described as "programmatic efficiencies and cost savings, improving program offerings, reducing procurement costs, and leading to more equitable program delivery." Providing open access to data was ranked second most important (18/83), followed by maintaining customer privacy (16/83). Of least importance for first choice, was concern about government possession of data and reducing fraud from energy contractors who provide services to homeowners and renters. Figure 3 shows the results of the first choice rankings. We also analyzed second and third choice rankings but the results were not substantially different from the first choice rankings.

Q13. Most common 1st choice for importance	count	%
Promoting the indirect outcomes of open access to data such as programmatic efficiencies and cost savings, improving program offerings, reducing procurement costs, and leading to more equitable program delivery	30	36%
Providing open access to data	18	22%
Maintaining transparency of utilities	8	10%
Maintaining transparency of state, local govt, regulator	3	4%
Maintaining privacy of utility customers	16	19%
Keeping costs low for utility ratepayers	6	7%
Reducing fraud from energy contractors who provide services to homeowners and renters	2	2%
Concern about government possession of data	0	0%
Total	**83**	**100%**

Figure 3. Choices for most important (rank 1) benefit or concern about open data (results shown as a count, N=83).

Data of Interest

We asked respondents to select all types of data they were interested in accessing. Since most discussions and concerns in the smart grid proceeding had been focused on very detailed household usage data collected by smart meters, this question quantifies the amount of interest in various types of energy data that could be made available.

Four-fifths of respondents said they were interested in customer billing data while two-thirds said they would like to access energy efficiency information (Figure 4). CPUC evaluation results were also of interest even though these results are for work performed by the CPUC rather than utilities.

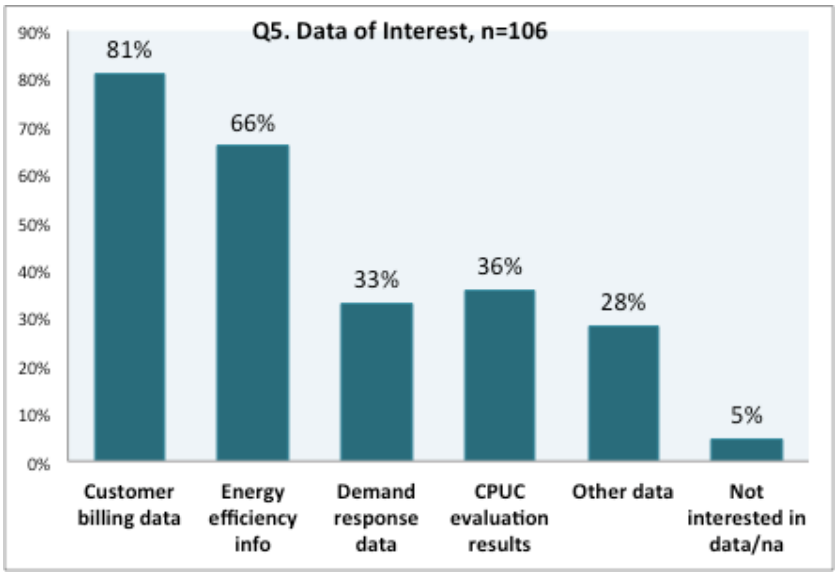

Figure 4. Data of interest for all respondents.

If respondents selected energy efficiency, demand response, CPUC evaluation results or other they were asked to specify the data of interest (Figure 5). For energy efficiency, respondents were most interested in program results and realized savings, followed by program

participation and usage data. Those interested in CPUC evaluation results were most interested in pre and post retrofit information and accessing evaluation reports. For demand response, load data was most desired, followed by interest in program participation.

Other data respondents would like to access include smart grid, distributed generation, 15 minute interval data, and city consumption by rate type.

Q5. Data of interest, please specify

Energy Efficieny (66%)	CPUC Evaluation Results (36%)	Demand Response (33%)	Other data (28%)
Program results/realized savings (16)	Pre and post retrofit (5)	Load (9)	Smart grid: transformer load, substation, system demand (4)
Program participation (past and current)(15)	CPUC reports (4)	Participation (4)	Distributed generation (3)
Usage data (10)	Program spending and cost-effectiveness by region (3)	Effectivenes (3)	15 minute interval data (or less) (3)
measure performance (9)	Behavior (2)	interval data (2)	Aggregate city wide consumption data by rate type (2)
Local govt savings (4)		Incentives (2)	
Ex ante/ex post (3)			
Completed projects in a region/city (3)			
Program costs (2)			
Savings for GHG reductions (2)			

Figure 5. Specific data of interest for energy efficiency, CPUC evaluation results, demand response, or other.

Desired Data Resolution

Echoing comments heard by parties during the Smart Grid Working Group meetings, most respondents would like access to data at all resolutions. For this multiple response question, Figure 6 shows participants had roughly equal interested in high level data, somewhat aggregated data, and raw or household level data. This suggests that releasing some data at a very high level would be valued by most respondents even while they continue to be interested in higher resolution data.

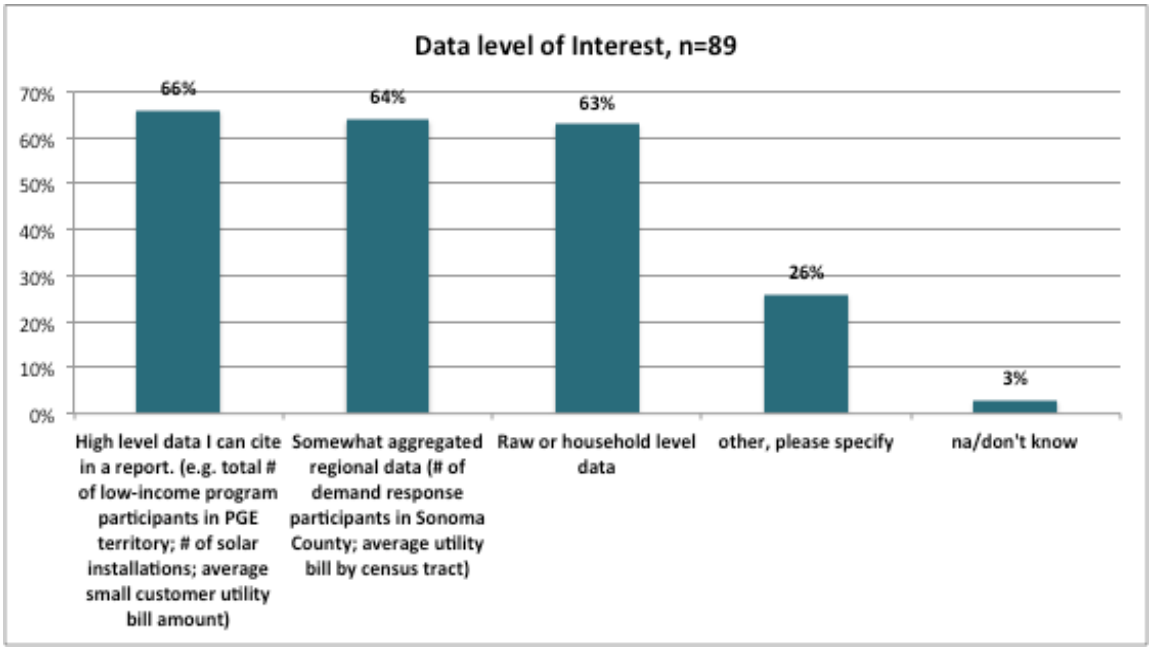

Figure 6. Interest in high level, somewhat aggregated, or raw/household level data.

When desired resolution level of interest was shown by user type, most groups were too small to discern any statistically significant differences.

Reasons for Requesting Data

We asked respondents to provide a description of the purposes for which they are interested in accessing utility energy data. Because participants entered their responses in a text field, many different uses arose and differed among respondents. Figure 7 shows the text responses. The most common reasons for accessing data were: to understand program performance; for local government climate action planning benchmark efforts; to inform the design implementation, outreach and evaluation of energy programs; to improve program targeting; and to forecast energy savings and GHG reductions. Among respondents who worked in government, the most important purpose was local government benchmarking and climate action plans. For non-profits, creating accountability and quantifying energy savings was seen as most valuable.

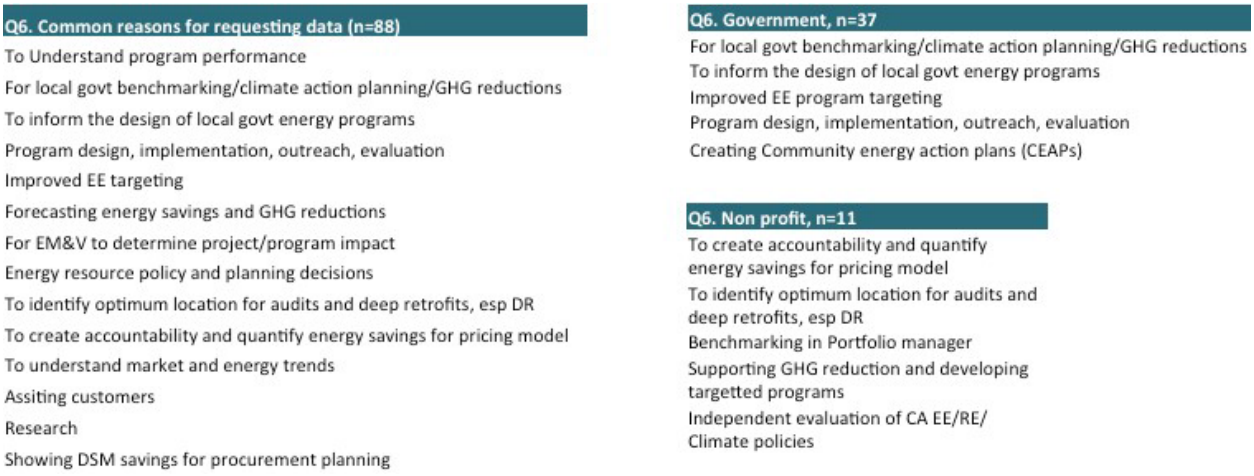

Figure 7. Reasons for requesting data (for all respondents, for government only, and non-profit only).

Other Data Sources

To receive a more complete picture of how energy data fits into the routines and habits of data users, respondents were asked what other data sources they regularly access. Figure 8 shows the most commonly used sources of energy information. The most frequently accessed source was the U.S. Census, followed by the U.S. Energy Information Agency (EIA), utility websites and reports, the California Energy Commission (CEC), and the Bureau of Labor Statics (BLS).

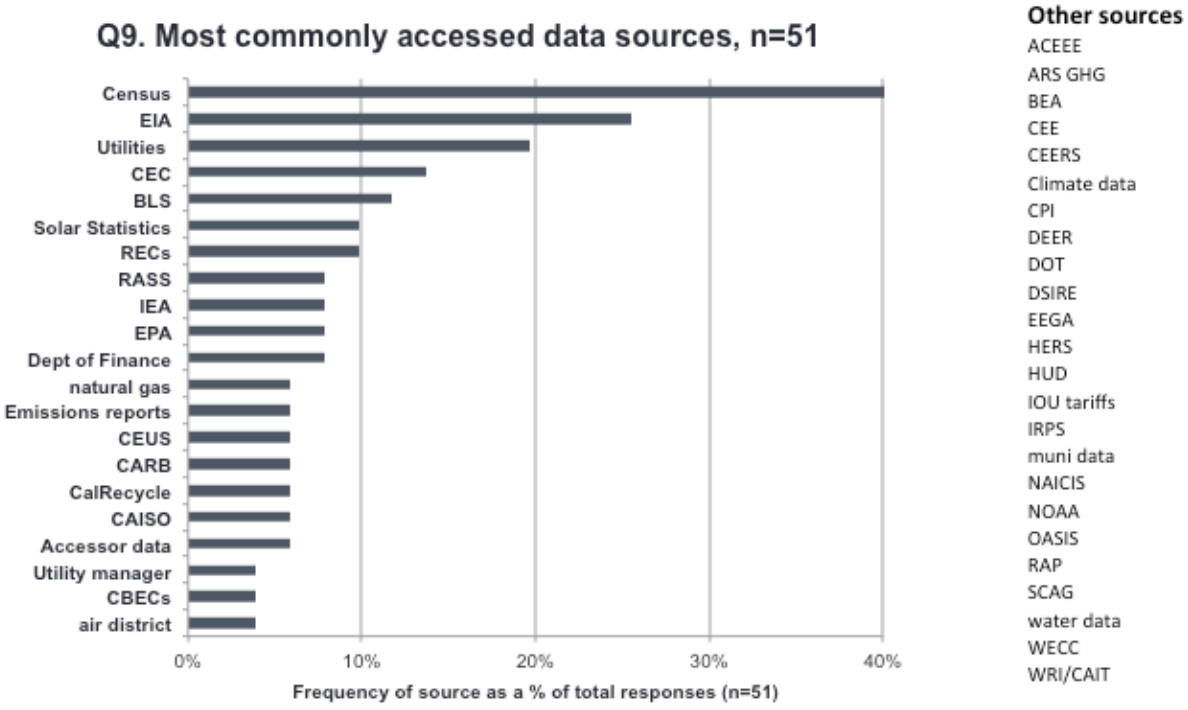

Figure 8. Other commonly accessed data sources.

Stakeholder Research Conclusions

The Current Process for Requesting IOU Data is Difficult and Inconsistent

One respondent wrote that accessing data is "harder than it should be and less consistent than it should be." Utilities need a better approach to data request response tracking and providing data of consistent quality. Multiple pain points exist, and data is delivered in different formats and of differing quality. Utilities also react differently to the same request: "One IOU is very responsive, one very protective, the rest are in between."

Not All Users Need Raw, Detailed Data

Many respondents indicated that they were equally interested in high-level (total number of installations, bill amount, number of participants, etc.) data and raw data (household level data or interval data directly from smart meters). This diversity of needs means that some users would be able to able to complete their analyses with high-level data only. While many users also expressed an interest in detailed data, the top three most common reasons for accessing energy information (to understand program performance; for local government benchmarking; and to inform the design of local government energy programs) can be achieved using aggregated energy and evaluation data. Local governments were the largest single respondent group, and would likely benefit most from high-level data since annual energy use is required to calculate GHG emissions for energy action plans and benchmarking.

Users are Most Interested in Customer Billing and Energy Efficiency Data

Four-fifths (81%) of respondents expressed an interest in customer billing data while two-thirds (66%) said they would like to access energy efficiency information. This suggests that future efforts should be focused on developing accessibility protocols for these two sources first.

Energy Data is One Source Among Many Used By Professionals

Professionals working in the energy field use many data sources including Census data, EIA, bureau of Labor Statistics and utility reports and data from the CEC. While energy data is considered useful or extremely useful by 95% of survey respondents, this data is part of a growing number of California and US datasets that can be utilized in analyses and are publicly available.

There is Cost to Delay

One survey respondent wrote: "our agency would have . . . better allocated funds to programs that were performing better than others." Stakeholders described projects held up by lack of data, missed deadlines, and ignorance about effectiveness of program strategies. One company complained about losing business due to the time involved in getting data in order to develop savings projections from installation of their product. Users acknowledged being able to access their own data but contractors have said it is often time consuming to access data on behalf of a customer—even with written consent.

Agreement that Solution Should Maximize Amount of Data Available Within Privacy Constraints

All stakeholders expressed a belief that energy data should be publicly available in some form. The majority of users see the indirect benefits of data as most important relative to concerns about government possession of data and government transparency.

"Energy Data of Interest" Includes CPUC Evaluation Results

In addition to IOU energy data, stakeholders are very interested in accessing CPUC evaluation results. Over one third (36%) of survey participants expressed an interest in accessing CPUC evaluation results. This issue also arose in several of the narrative responses which described the need to understand the effectiveness of various programs and frustration that all program information is not in the same place. One respondent wrote about the CPUC evaluation results "some figures were available on an annual basis, others seemed to be available only for the multi-year program cycle. It was difficult to identify the appropriate documents on the CPUC site."

Future of Energy Data Access

In May 2014, the CPUC issued its final decision in Phase 3 of the Smart Grid Proceeding. The decision[8] was approved unanimously and represents a big step forward in articulating the balance between data sharing and privacy protection. It requires the IOUs to post aggregated usage data, provide data to researchers, and develop a protocol for deciding who gets access and how to protect customer privacy.

The real value of energy data lies in the ability to combine it with other data sets and develop new insights. Increased researcher access will enable us to answer questions such as: Which neighborhoods consume the most electricity? How would the grid be impacted if the customers with diesel and gas water pumps switched to electric ones? What rate structures would most likely influence customers to change their behavior? And what are the characteristics of customers who consume less energy? Answers to these questions will help improve the targeting of energy efficiency programs, identify ways to reduce energy use in the least cost manner, and guide future policy planning decisions. And it is these improvements which will lead to reduced energy costs to customers, improved services, increased transparency, improved power grid reliability and management.

The CPUC decision will undoubtedly serve as precedent for other entities seeking to realize the benefits of their data while keeping an eye on protection of privacy and data security. Future discussions on energy data access in California and elsewhere will likely revolve around how to continue to establish the correct balance between reducing the risks of data sharing and the potential benefits of its use. This tradeoff is one that may be lessened as more technical solutions are developed to ensure the security and transfer of data, yet, the matter of to whom to grant access will likely continue to be debated.

[8] D.14-05016, approved May 1, 2014.