Unleashing the Power of Big Data on Efficiency? Not So Fast

Anne McKibbin, Elevate Energy

ABSTRACT

As Advanced Metering Infrastructure (AMI) is deployed across the US, utilities, energy efficiency program administrators, researchers, and competitive electric suppliers are clamoring to protect, obtain, and use customer energy use data. At the same time, the public is increasingly aware of privacy issues and routine consumer data collection. As a result, customer energy use data is becoming a hot topic among regulators and legislators, who must strike a balance between consumer data privacy and the desire to unleash the power of big data on energy efficiency efforts.

This paper outlines how program administrators and researchers use customer energy use data to improve building energy efficiency programs. It also surveys data access regulations established by state regulators and legislators across the U.S. It discusses the types of restrictions that are placed on access to and use of customer energy use data, and their effect on program design and administration and research possibilities. Finally, the paper makes recommendations for engaging policymakers to ensure a productive balance between protecting consumer privacy and meeting research and program administration data needs that move energy efficiency forward.

Introduction

AMI meters are being deployed across the U.S., and the rich trove of energy use data they provide promises to unleash significant innovation in the design and implementation of energy efficiency and dynamic pricing programs. As energy efficiency programs become more sophisticated and aim for deeper savings, utilities and program administrators need access to this energy use information to identify continuing sources of energy waste and ways to conserve. As a result, utilities, energy efficiency program administrators, program contractors, researchers, and competitive electric suppliers are seeking to obtain access to this data. Regulators, however, are increasingly wary of releasing customer data, as current events such as the Target data theft have increased exponentially the public's interest in consumer data privacy.

The granularity of AMI data and the meters' ability to communicate with the customer in near real-time holds particular promise for customer education and automated response. Monthly customer energy use data from traditional meters, however, remains very useful in designing and improving energy efficiency programs. This paper will discuss access to and restrictions already in place on customer energy use data from both types of meters, but will focus on the urgency created by AMI meters for 'big data'- what could be a powerful insight into customer energy consumption patterns. It's estimated that 46 million AMI electric meters had been installed by July 2013, reaching 40% of US households. (IEE 2012) With more deployments certain by 2015, the smart meter landscape could soon look like Figure 1, below, taken from the Edison Foundation's 2013 IEE Report titled *Utility-Scale Smart Meter Deployments: A Foundation for Expanded Grid Benefits.* This expansion of smart grid is creating even more urgency around the questions of who can access customer energy use data, for what purpose, and under what conditions.

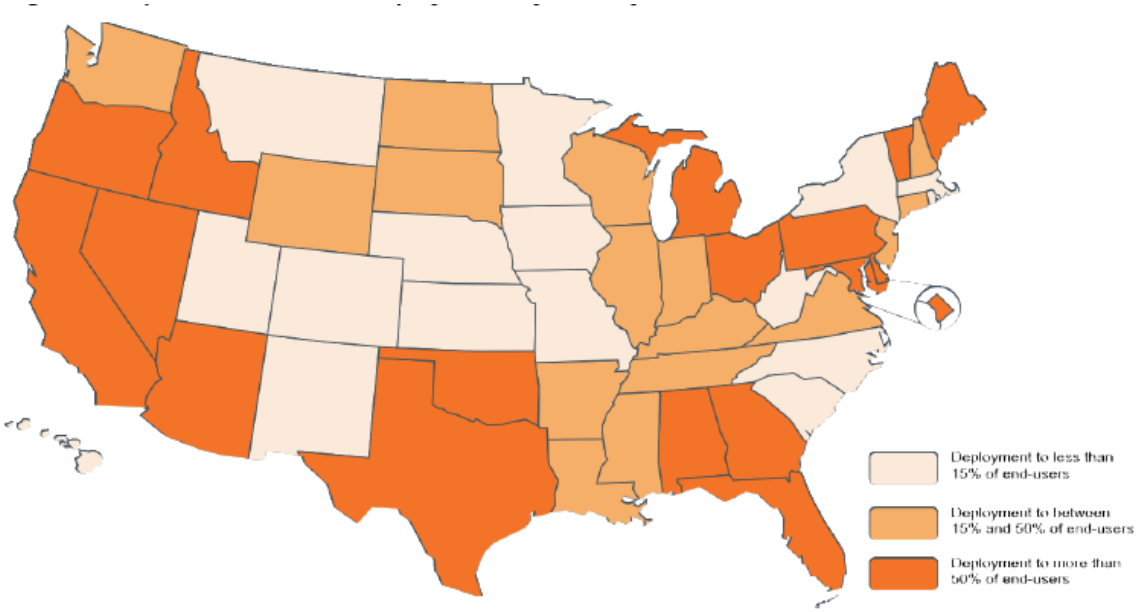

Figure 1: Map of expected AMI deployments by 2014. *Source*: IEE 2012.

Utilities hold immense amounts of personal customer information, including name and account numbers, billing and premises addresses, meter numbers, and the all-important energy use data. Before undergoing extensive AMI meter installations, utilities typically submit a plan to their regulators, which may include information on the types of data to be collected and its storage. In approving and overseeing these installations, regulators must balance privacy concerns with the need to create customer benefits from energy efficiency and other programs enabled by AMI interval data and monthly customer energy use data, and are under increasing pressure to tilt that balance in favor of protecting privacy. Up to this point, the existing datasets have not been used to their fullest potential. This will likely change as more granular data becomes available and the industry seeks new ways to use it. So far, this regulatory balancing act has created a confusing patchwork of regulations that both enables and restricts how customer energy use data can be used to improve energy efficiency programs.

This paper will discuss several ways that program administrators and researchers can use customer energy use data to improve building energy efficiency programs. While the paper does not present an exhaustive list of the research methods that could benefit these programs, the chosen studies highlight the various ways that data access restrictions can limit our understanding of our building stock, the consumption patterns of customers, and their energy efficiency needs. This paper also discusses these limitations and finally, makes recommendations that would help to ensure a productive balance between protecting consumer data privacy and meeting research and program administration data needs that move energy efficiency forward.

Using Customer Energy Use Data to Improve Energy Efficiency Programs

Customer energy use data can be used to design better energy efficiency programs in a number of ways. AMI meters can be used to inform customers of their energy use on a near real-time basis, creating opportunities for customer education and automated responses. In addition, both AMI and traditional meter data can be used to inform building owners of their whole-building energy use performance and to compare the local building stock's energy use with national averages. When combined with additional data, such as Census tract-level income and demographic information or housing characteristics data from a local property assessor's office, energy use datasets can become a particularly powerful tool to seek out and target energy waste in particular building types and communities. Each of these uses is discussed in Table 1, and in more detail below, along with the types and granularity of data required for each activity to proceed.

Table 1. Uses for customer energy use data and data fields needed and granularity required

Uses for Customer Energy Use Data	Data Fields Needed and Granularity Required
Informing Customer's Own Choices	Customer's own energy use data; near real-time
Benchmarking Building Energy Use and Post-Retrofit QA/QC	Customer's own energy use data for common areas and aggregated (summed) tenant energy use data; monthly
Calibration Studies	Energy use data for all buildings in a geographic area; identifying data is needed only before merger with building characteristics data; monthly
Housing Characterization Studies	Energy use data for all buildings in a geographic area; identifying data is needed only before merger with building characteristics data; monthly
Program Targeting	Energy use data for all buildings in a geographic area; identifying data is needed throughout the process; monthly

Informing Customer Choices

Customers who have access to their own energy use data, on a near-real-time basis, can use that information to inform their energy use choices and to control programmable in-home devices or other applications. As discussed further in the next section, there is little regulatory controversy surrounding customers' access to their own data or their ability to give a third party authorization to access that data on their behalf.

Benchmarking Building Energy Use and Post-Retrofit QA/QC

Monthly, or more granular, customer energy use data can be used to benchmark building energy use and to perform post-retrofit quality assurance checks. Building energy use benchmarking helps commercial and multifamily building owners understand and manage their buildings' energy use by providing a consistent, comparable measure of energy use throughout the entire building. Benchmarking is a useful tool to guide investments in energy improvements

that benefit building tenants, and is necessary to earn a US EPA Energy Star Buildings certification and comply with a number of recently passed benchmarking and disclosure ordinances in US cities. (Krukowski and Burr 2012)

Some energy efficiency retrofit programs, including the Energy Savers multifamily retrofit program administered by Elevate Energy, rely in part on post-retrofit analysis of actual building energy use data monitor the effectiveness of HVAC, insulation, and air sealing improvements by verifying that actual savings are in line with expectations. (CNT Energy 2013)

To benchmark their buildings, or perform whole-building post-retrofit quality checks, owners, managers, program administrators, and their consultants need to obtain aggregated (summed) tenant energy use data and enter it into a benchmarking tool, such as US EPA's Portfolio Manager. (Krukowsi and Burr 2012) Building owners often need to obtain this information without the express authorization of each individual tenant, which would delay benchmarking or make it impractical. (CNT Energy 2013)

Calibration Studies

Energy efficiency program administrators make program design decisions based on the building energy use data available to them. Residential energy analysis tools, such as energy assessment software should be calibrated to local building energy use conditions. However, without access to actual building energy use data, these tools and processes can only use the sample data provided by national surveys such as the US Energy Information Agency's Residential Energy Consumption Survey (RECS) and Commercial Buildings Energy Consumption Survey (CBECS) as inputs.

Local building energy use conditions, however, may vary dramatically from RECS and CBECS averages for several reasons, including small RECS and CBECS sample sizes, differences between the age of the local and survey sample housing stock, variations in climate across the survey region, and differences in the level of homogeneity of the housing stock in the local area and across the survey region. For all of these reasons, decisions based on the RECS and CBECS surveys, without additional information provided by a local analysis of actual energy use data, can be consistently biased and inaccurate within a particular geography. (CNT Energy 2013)

To understand the relationship between these surveys and local conditions, program administrators need access to residential or commercial energy use data, by building type, for a geographic area. Sensitive information such as address or account or meter numbers are needed to construct datasets that accurately reflect building energy use in buildings with multiple meters or other complexities. After an accurate dataset is complete, however, these sensitive variables could be replaced by random proxy variables (for example, replacing addresses with 'House A' or 'Small Commercial Building B') to protect privacy for the remainder of the analysis.

Housing Characterization Studies

A housing characterization study identifies and categorizes the most common housing types in a geographic area by their building characteristics and energy use patterns. These studies then identify common characteristics for each building type and describe the related energy efficiency opportunities. Housing characterization studies can improve the effectiveness of efficiency programs by helping program administrators understand which building types present the greatest efficiency opportunities, and by improving the efficiency of the energy assessment

process. (CNT Energy 2013) For example, a study might find that masonry structures built before 1945 have the highest energy use per square foot in a particular metropolitan area, allowing program administrators to focus on finding new and better ways to reduce energy use in that type of structure.

Housing characterization studies require a dataset of energy use information for residential customers in the targeted geographic area. This energy use dataset must be merged with a housing characteristics dataset that includes information such as each home's age, its size and other available information, such as architectural style or framing material. These housing characteristics are sometimes available from property assessor offices. To merge the two datasets, both datasets must contain a unique variable that identifies each individual building, such as an address or meter number. After the merger is made, the resulting dataset will tell the researcher each home's age, size, and energy use. At this point in the process, it may be possible to strip the dataset of any personal information such as address or account number, leaving only minimal geographic information that does not identify the individual home. However, sensitive information that identifies each individual building is crucial to performing the necessary merger step, without which the study cannot be completed. (CNT Energy 2013)

Program Targeting

As energy efficiency programs become more sophisticated and aim for deeper savings, utilities and program administrators need access to energy use information to identify continuing sources of energy waste. This information will help design programs that target this waste and ensure availability to the customers who would benefit most. In addition, program administrators and regulators may want to target programs to lower income residents or areas with an abundance of a particular housing type.

To identify sources of energy waste and target programs effectively, program administrators and researchers need datasets that merge building- or neighborhood-level energy use, building characteristics, and demographic information, such as that available at the Census tract level. Creating such a rich dataset requires the merger of information from multiple datasets. The accurate merger of these datasets will almost always require a customer-identifying variable such as address or meter number. In addition, target marketing the resulting programs requires a dataset that includes some means of identifying the individual or classes of customers who would most benefit from a program. (CNT Energy 2013)

Data Access Restrictions and Their Effects

Below, this paper classifies several types of data restrictions that are currently in use in the US. This section also discusses the effect that each of these types of restrictions have on program administrators' ability to perform each of the activities discussed above. These effects are summarized in Table 1, at the end of the section.

"Unrestricted" Data and Authorization Requirements

There is general consensus among regulators nationwide that a customer's energy use data should be freely available in two circumstances: (1) where the customer wishes to access and use his or her own energy use information, and (2) where the customer has given a third party authorization to access his or her energy use information. (SEE Action 2012)

Green Button is an effort to allow customers to download their own energy use data from utilities in a common technical standard and share this information with third-party applications. Since its launch in January 2012, over three dozen utilities have agreed to make data available using the standard, and over half of these utilities have implemented the program. (Green Button) While easy access to customers' own information is not yet ubiquitous, these initiatives have created a pathway to widespread access by customers in the future.

Currently, situations where customers give a third party authorization to access their energy use information are more complicated, because there is no universally-accepted standard for customer authorization. Acceptable authorization could range from a 'wet' signature, which would require the exchange of documents and a physical signature from the customer, to the mere possession of an account number, which presumes that the account number was acquired legitimately and is being used for its intended purpose. (SEE Action 2012) In between these two extremes are faxed or scanned copies of signatures; electronic signatures, which can include password/PIN combinations or clicking a designated box on a website; and recorded voice authorizations. (Cortez and Green)

Each of these methods of authorization provides a different balance between customer convenience and privacy. Each authorization type requires different procedures to acquire authorization, with the wet signature being the most onerous for energy efficiency program providers, and the account number or phone authorization being perhaps the easiest to obtain. And, as a practical matter, utilities have not all established standard procedures for accepting authorization. Without standard procedures and clear authority for utility staff to accept authorization, an efficiency provider may find its access to data cut off when the utility staffer who provided data to authorized users changes jobs.

Anonymity Restrictions

There are several basic types of restrictions on access to customer energy use data without customer authorization: anonymity restrictions, use restrictions, and recipient restrictions.

Anonymity restrictions protect customer privacy by ensuring that the recipient of energy use data cannot match that information to a specific customer. Typically, as in Colorado and Oklahoma, data subject to these restrictions can be released after treated with a two-step process. First, information that directly identifies an individual customer, such as address or account number, is removed from the dataset. The resulting data may then be grouped or summed by sector or customer class (i.e., residential, commercial, or industrial) and geographically, so long as the group is not so small that an individual customer can be identified. (SEE Action 2012) Anonymity restrictions may apply to individual customer energy use data or to the aggregated, or summed, tenant energy use data used to benchmark building energy use.

This second step hinges on the size of the group of data points that is released. Vermont, for example, allows individual customer data to be grouped only at the municipal level. (SEE Action 2012) Alternatively, several states have adopted the 15/15 rule, which requires that customer data be released in groups of no less than 15 customers, so long as no one customer makes up more than 15% of the group's total energy use. (SEE Action 2012).

Anonymity restrictions can make it impossible to complete the calibration studies, housing characterization studies, and program targeting discussed in the previous section. If identifying data is stripped from the dataset before it is merged with housing, Census, or other

necessary datasets, then these studies simply cannot be completed. Potential solutions to this problem are discussed in the Recommendations section below.

The choice of threshold data group size, if set too low, can also make it impossible to benchmark the energy use of small and medium sized commercial and multifamily buildings. Several cities with benchmarking ordinances, for example, require benchmarking for buildings over 50,000 square feet. A residential building of this size could reasonably hold only 25 large apartments. So, where these ordinances apply to residential buildings, a threshold that allows the building owner to easily obtain aggregated (summed) tenant energy use information only if the building has 30 or more tenants would make it virtually impossible for the building to comply with the local benchmarking ordinance. To do so, the building owner would have to obtain authorization from each tenant, a situation that would be time consuming at best and likely impractical and impossible. (CNT Energy 2013b)

A too-low threshold may also prevent program administrators from performing quality control checks on smaller buildings, where those checks are partly dependent on comparing post-retrofit energy use information to similar buildings or pre-retrofit usage. (CNT Energy 2013b)

Use Restrictions

Use restrictions address how the recipients of data may use the information. Some states that allow the use of customer energy use data for the provision of utility-provided energy efficiency programs, for example, require customer authorization before the utility can give a third party this data for some other purpose. (SEE Action 2012) Use restrictions could become a barrier to targeting energy efficiency programs. Use restrictions are less likely to become a barrier to determining sources of energy waste than anonymity restrictions. However, they could create a significant barrier to applying these findings if they prevent efforts to identify and market to customers who would benefit most from energy efficiency programs.

Recipient Restrictions

Recipient restrictions address the types of organizations that may receive customer energy use data from a utility without customer authorization. The provision of this data to some third parties may create significant customer benefits. Some efficiency program providers, for example, operate without direct contracts with the utility. University-based researchers and energy efficiency advocates may be able to point out programmatic gaps or hidden efficiencies that programs can capture. And, municipalities may wish to use this type of information to inform energy planning processes.

Existing regulations typically apply to utilities and their ability to share data with third parties under contract to perform energy efficiency-related services. Some states, such as Colorado, have exercised jurisdiction over the transfer of consumer energy use data by placing strict restrictions on the parties that can obtain this data. (SEE Action 2012) Other states, such as California, require that third parties who receive customer energy use data be registered with the regulatory commission.

Recipient restrictions such as registration pose little barrier to the beneficial use of customer energy use data. Parties that are using the data in good faith should have little objection to registering, so long as the process is not an administrative burden. Restrictions that prevent third parties from obtaining the data altogether, though, would prevent those parties from

engaging in any calibration studies, housing characterization studies, or program targeting. These restrictions would inhibit the customer benefits that can arise from these activities.

The effects of data access restrictions on energy efficiency programs, and the studies and data-related processes that would benefit these programs are summarized in Table 1:

Table 2. Effect of data access restrictions on energy efficiency programs

Data Use	Restriction that Creates Most Significant Barrier	Effect on EE Programs
Informing Customer's Own Choices	Prescribed method of authorization	Method of authorization could cause delay
Benchmarking Building Energy Use and Post-Retrofit QA/QC	Threshold on number of tenants to obtain aggregated (summed) tenant data	Small buildings prevented from benchmarking; QA/QC analysis incomplete
Calibration Studies	Identifying data removed	If identifying data is removed before merger with housing data, it is not possible to complete study
Housing Characterization Studies	Identifying data removed	If identifying data is removed before merger with building type, it is not possible to complete study
Program Targeting	Identifying data removed	If identifying data is removed before merger with Census and building type data, the study will yield little useable information. Removing identifying data after the merger will prevent individual customer targeting, but some targeting by geographic area may still be possible.

Recommendations

If utilities and advocates are to meet their goals over the long term, the customer energy use data created by AMI and traditional meters must help design and improve energy efficiency programs. In doing so, it is critical that the industry and regulators retain customers trust by respecting their privacy. To balance these objectives, regulators, consumer advocates and the industry must:

- Educate themselves on the methods of using customer energy use data to design and improve energy efficiency programs;
- Create consistency by using common definitions and terms;
- Recognize that different uses for customer data require different regulatory responses;
- Learn from other industries that deal with sensitive information;

- Encourage the construction of rich datasets while maintaining privacy; and
- Create regulatory certainty.

This section discusses each of these recommendations further.

Policymaker and Stakeholder Education

The regulatory urgency created by the swift adoption of AMI meters and current events raises the risk that regulations will be created without a full understanding of their consequences on our ability to design and improve energy efficiency programs. Policymakers and the industry must educate themselves on the ways that customer energy use data can be used to benefit customers and how data access restrictions affect the industry's understanding of the building stock and customer energy use. As with many areas of regulation, rules around customer energy use data must be carefully written to avoid unintended negative consequences. They must account for the need to merge energy data with other data sources and for how the actual data will be manipulated in the process of creating useful studies. While this area of inquiry is exceedingly detailed and technical, the potential customer benefits from the responsible use of customer energy use data are significant.

Common Definitions and Terms

Policymakers and the industry can facilitate efforts to make customer energy use data anonymous, while remaining useful, by applying common definitions of data fields and data manipulation techniques. As mentioned above, this area is exceedingly technical. To avoid misunderstandings, confusions, and errors, we must all use the same language. Terms such as 'aggregated' have very specific meanings to those who work with building energy data every day, but those meanings may not be shared with colleagues who work in policy or regulatory areas. Without common terms, we run the risk of not being able to effectively negotiate, argue, or implement regulatory orders because of a lack of clarity about what they mean.

Recognize that Different Uses for Customer Data Require Different Regulatory Responses

No single blanket rule will ever protect customer privacy while also allowing energy use data to be studied in a way that creates customer benefits. An excellent example is the threshold data group size discussed under the Anonymity Restrictions section above. There, customer energy use data is made anonymous by removing identifying information. Then, to prevent anyone from using the remaining geographic information to deduce the identity of a particular customer, the data is only released in customer groups over a particular threshold size. There is, however, no single 'correct' group size. Instead, the minimum size of a group that maintains privacy will depend on how the information is to be used. It is considerably more difficult to 'reverse engineer' a single customer's energy use information out of a summed total of multiple customers than from a grouped list of individual, though anonymous, customers' information. Consequently, privacy is protected by a lower threshold number of tenants when dealing with summed data. This is just one example of the kind of care that is required to protect customer data while also allowing its use to create customer benefits.

Learn from Other Industries

There are a number of other industries that deal with sensitive information daily, and we must learn from them. The consumer credit industry has been cited by Philip Henderson from NRDC as an exemplar for sharing highly sensitive customer information in a manner that is automated, fast, and easy for the customer. (Henderson 2012)

Public health provides another example, having arrived at far lower threshold grouping levels than those widely used for customer energy use data. Guidelines from the public health field are instructive, as disease data is very sensitive information, but its release, when handled correctly, can help public health officials prevent and treat widespread health problems. Consequently, public health officials have developed extensive guidelines on when to suppress information in their public reports because it would compromise individual privacy. The National Birth Defects Prevention Network's *Guidelines for Conducting Birth Defects Surveillance*, for example, cite the data suppression rules for various disease surveillance systems in Table 2 (NBDPN 2008):

Table 3. Threshold levels to ensure privacy in public health data systems

Data System	Suppression Criteria
HIV/AIDS Surveillance System	< 4 cases
National Notifiable Diseases Surveillance System	Race and Hispanic Origin if < 4 cases
STD Surveillance System	County: < 4 cases; State: < 6 cases; National: None

National Birth Defects Prevention Network, "Guidelines for Conducting Birth Defects Surveillance: Appendix 11.1 Data Suppression," 2008. Available at: http://www.nbdpn.org/docs/Appendix11-1.pdf

Such information should be considered as a matter of course as we develop regulations for the use of customer energy use data.

Encourage the Construction of Rich Datasets that Maintain Privacy

So far in the evolution of customer data access rules, regulators have focused on finding the balance between barring the release of certain data and using it for beneficial purposes. However, we may be able to have our cake and eat it too. Some of the studies listed above require datasets that can only be created by using sensitive information to merge energy use data with housing characteristics or Census datasets. But, for calibration and housing characterization studies, sensitive information such as name, address, or account number is only needed to perform the data merger, not for the analysis itself. Consequently, it may be acceptable to release these datasets to third parties after the merger is performed and sensitive information is stripped from the dataset. This type of data manipulation can be costly, however. So, regulators and the industry should consider arrangements that allow this work to go forward, but covers the cost in some way.

The industry may also wish to consider more advanced methods of manipulating data to ensure privacy, like private key encryption and proxy variables that could be used to join multiple sensitive datasets before stripping the sensitive information and leaving more general, but still useful, data. (CNT Energy 2013c)

Create Regulatory Certainty and Consistency

No matter which methods are used to join datasets or make them anonymous, regulators must create certainty so that customers will feel secure that their privacy is being respected, and utilities will feel secure in sharing appropriate data with third parties. In addition, consistency across states and utility jurisdictions will reduce the cost and administrative effort needed to protect, share, and use data appropriately. Without this certainty and consistency, research and other activities that could improve energy efficiency programs will be slow to develop.

References

CNT Energy. 2013. *CNT Energy's Initial Verified Comments*. Illinois Commerce Commission Docket No. 13-0506. http://www.icc.illinois.gov/docket/files.aspx?no=13-0506&docId=204298. CNT Energy changed its name to Elevate Energy on Feb. 4, 2014.

———. 2013b. *CNT Energy's Verified Reply Comments*. Illinois Commerce Commission Docket No. 13-0506. http://www.icc.illinois.gov/docket/files.aspx?no=13-0506&docId=205289. CNT Energy changed its name to Elevate Energy on Feb. 4, 2014.

———. 2013c. *CNT Energy's Surreply Verified Comments*. Illinois Commerce Commission Docket No. 13-0506. http://www.icc.illinois.gov/docket/files.aspx?no=13-0506&docId=205911. CNT Energy changed its name to Elevate Energy on Feb. 4, 2014.

Cortez, E. and Green, L. *Digital vs. "Wet" Signatures.* Office of the Los Angeles County Counsel. Accessed March 3, 2014. http://www.caceo58.org/assets/documents/Digital.pdf

Green Button. Accessed March 4, 2014. http://greenbuttondata.org

Henderson, P. 2012. "Making Better Use of Energy Usage Information: Lessons from the Credit Reporting Industry." In *Proceedings of the ACEEE 2012 Summer Study on Energy Efficiency in Buildings*. Washington, DC: ACEEE. http://www.aceee.org/files/proceedings/2012/start.htm

IEE. 2012. Utility-Scale Smart Meter Deployments: A Foundation for Expanded Grid Benefits. Washington, DC: Edison Foundation. file:///C:/Users/Anne's%20Labtp/Downloads/IEE_SmartMeterUpdate_0813.pdf

Krukowski, A. and Burr, A. 2012. Energy Transparency in the Multifamily Housing Sector. Washington, DC: Institute for Market Transformation. http://www.imt.org/uploads/resources/files/Energy_Trans_MFSector_IMT_Final.pdf

NBDPN (National Birth Defect Prevention Network). 2008. "Guidelines for Conducting Birth Defect Surveillance: Appendix 11.1 Data Suppression." Atlanta, GA: National Birth Defect Prevention Network. http://www.nbdpn.org/docs/Appendix11-1.pdf

SEE Action. 2012. A Regulator's Privacy Guide to Third-Party data Access for Energy Efficiency. Washington, DC: SEE Action Network. http://www1.eere.energy.gov/seeaction/pdfs/cib_regulator_privacy_guide.pdf