Turning Today's Data into Tomorrow's Reality

Melissa Culbertson and Alek Antczak, CLEAResult Consulting, Inc.
Kerry Meade, EMI Consulting

ABSTRACT

We've all heard the buzz, big data is going to transform our world, but as demand side management (DSM) program implementers and designers, we are already there. Each day we generate an enormous volume of important market data, from calculating measure savings, to processing incentive payments, to analyzing home and commercial facility audits that characterize the conditions found in structures across the United States. But countless differences in vocabulary, building classifications, program data frames and other jurisdiction specific difference render use of the data a daunting endeavor. How can we leverage the data we generate now, before "big data" gets rolling? What should we be doing to prepare our data for analysis? What does our world look like in the future?

In this paper, we will answer these questions by examining two very different applications of data: a simple use of Geographic Information Systems to help characterize the commercial facilities within a utility service area for more targeted program outreach and design, and a recent award-winning application of residential customer energy use data to design a mobile phone application that interacts with the user about personal energy use. We'll explain how each project applied data, what type and how accessible that data is, and then explore how each example points to a more sophisticated, broader effort to leverage data generated in current ratepayer funded energy efficiency programs. Starting to use what you already know is the starting point for identifying, with big data, what you seek to understand.

Introduction

Green Tech Media (GTM) research has pegged the value of the global utility data analytics market at a cumulative $20 billion between 2013 and 2020, growing from an annual spend of $1.1 billion in 2014 to nearly $4 billion by decade's end. (St. John 2013) In 2011 the cumulative spend on DSM in North America topped $9 billion. (Laitner 2013)

Utilities tend to begin their foray into big data through business operations where the big data is: smart metering or automated metering infrastructure (AMI). Over the past few years, we've seen AMI-centered projects emerge from pilot deployments to full-scale commercial offerings[1], whether in the form of embedded IT systems at large investor-owned utilities, or as a cloud-based software-as-a-service (SaaS) model. Similarly, in the DSM industry, we have seen the automation of the program delivery process with the deployment of measure calculation and field tools, to opportunity tracking systems, all the way up to integrated portfolio management systems.

But the move from one-off, proof-of-concept deployments to broader adoption of analytics is still a work in progress for the utility industry. A host of reports out this year prove

[1] For example, Baltimore Gas & Electric recently deployed C3 Energy Smart Grid Analytics in its service area. The initial phase will employ C3's 'revenue protection' to identify load profile anomalies and outliers and C3 Operations to verify the integrity of the AMI Network (Energy Business Review).

that many utilities haven't even captured AMI data's value to make sure their core meter-to-cash and communications systems are working at optimum value, let alone begun adding more advanced functions to the mix. Based on our experience as a leader in the industry, program implementers are still struggling with the unique challenges that our industry faces, and have yet to realize the full value of the data we are collecting already.

Buzz around Big Data Entering the DSM Space, and the Paralysis Generated by the Buzz

Utilities have been collecting data for a while now, but the amount of data and the speed at which they can collect it has been steadily increasing. The first smart meters were rolled out almost ten years ago, utilities are continually developing advanced metering infrastructure and the use of meter communication infrastructure (MCI) analysis platforms. The question is, what is there to do with all of the data that is now accessible?

There is considerable buzz around new firms entering the utility space, in particular with DSM. Start-ups and more established data and analytics firms are jockeying for position even as their entry into the industry is pushing the boundaries of how to define DSM. New entrants are bringing behavioral approaches and controllable technologies that bridge the energy efficiency and demand response gap, which add to the traditional DSM offerings. And with that comes the ever increasing accumulation of data.

One of the most interesting aspects of the recent upturn in the collection of data is that most of the efforts have been one-off proof of concept type exercises, deployments designed to fill a particular need at a particular time. Evaluation, Measurement and Verification (EM&V) protocols require that a significant amount of data be collected in nearly all jurisdictions. However, what implementers are required to collect is really only what's relevant to the goals of the utility that hired them, or the responsible utility commission in its jurisdiction. This is compounded by the fact that collecting and analyzing data that does not directly connect the program efforts to the goals of the program is an extra cost not typically factored into program budgets; given the thin budgets often associated with program implementation, this leaves utilities and implementers challenged to find time in their data collection efforts for data points that may be of use to the utility in program planning down the road but won't be directly used in the current implementation effort. The lack of consistency and coordination has made it difficult to use data to prove out concepts from one jurisdiction to another. CLEAResult finds this to be true sometimes where they run nearly identical programs in two different jurisdictions. This struggle combined with the buzz surrounding "Big Data" and the muddying of the DSM waters with the expansion of DSM beyond technology retrofits, has led to considerable inaction on the part of many implementation firms in improving their capabilities in collecting and using data in a useful, or even consistent, manner. Lastly, EM&V firms may have data needs that they are aware of, but do not have the capability of amending the tracking systems within utilities and implementation firms.

Who is Entering into the DSM Industry?

Over the past several years, we've seen startups like AutoGrid, C3 Energy, Trove, Opower and Verdeeco, and IT giants like Oracle, IBM, SAS, Teradata, EMC and SAP, and grid giants including General Electric, Siemens/eMeter, ABB/Ventyx, Schneider Electric/Telvent, Toshiba/Landis+Gyr and more, leap into the big data fray in the DSM space. (Leeds 2012)

What Does Their Entry Signify?

There are a number of implications to the sudden entry of a variety of new players in the DSM space.

Data driven businesses are becoming more commonplace. Energy consumption data is a very powerful behavioral metric, which can be sold to companies that want to know about consumer behavior, such advertisers.

A market has recently arisen for utility customers that want a window into their consumption patterns. Given the tougher economic times, many home and business owners are looking to trim their expenses as much as possible, and energy consumption is directly controllable. Giving them knowledge about their volume and cost of their consumption allows them to make more informed behavior and purchase decisions.

Many of the firms that are entering the DSM space and collecting data are coming first as vendors of new widgets, such as intelligent thermostats, mobile apps, commercial lighting monitoring and control systems, and building energy management systems. These widgets are a transaction between the customer and a third party vendor, and are outside of the regulated space. The data that is collected does not have the same rules and restrictions governing usage, and in many cases the user agrees (through a terms and conditions agreement) to countless, open-ended ways the vendor may use their data. This is considerably different from utilities who operate in a strictly regulated environment wherein use of customer data is carefully monitored. While this is a topic for another paper, it's important to note the distinction—if data analytics is the way of the future, what does this competitive imbalance mean for utilities?

There is an opportunity to exploit the surge in popularity of social media; which we refer to as the microphone effect. Customers like the ability to integrate their energy consumption, particularly when it makes them look good, into the rest of the information they regularly broadcast about themselves through social media. This is being addressed by companies like OPower and Google through massive investment in social media integrated dashboards. There is the potential to generate an overwhelming amount of data through this purpose. (Lacey 2013)

In line with the other implications, there is the desire to consolidate and assimilate as many disparate sources of information as we can, to simplify our lives in the face of ever increasing complexity. How we receive and consume energy is already exceedingly complicated and difficult to understand. Widgets and platforms that allow us to control our home energy use from our mobile phones, or monitor the performance of the HVAC system in any one of several dozen store locations, greatly simplifies and incorporates the energy component.

While these scenarios are still in early stages of development, it is important to remember that we currently collect a large quantity of usable data, and despite the challenges to interfacing with the data, there are creative applications to fill in the gaps. We can use the data we do have to develop capabilities to plot a smoother course, improve the quality of our program delivery, and reduce the cost in the interim.

What are the Barriers that Exist for Current Implementers to Collect and use Energy Efficiency Data?

Implementers do not often own the program data that is collected, or at the very least, there are layers of privacy protection that make it difficult to share outside of program implementation. This can often interfere with sharing best practices, benchmarking program performance against a baseline control group, or comparing the data to a group of customers in a different jurisdiction. There are many times when implementers look to industry standards and best practices to guide market adoption or implementation cost estimates. When pushing into a new market, or trying to shift a program from early start-up to a more established market transformation phase, it can be difficult to gauge how customers will react to a certain strategy without territory specific data to model performance. When a jurisdiction has very specific goals or a limited budget, it is important to be able to maximize the effect of program resources. When there is limited market intelligence, a more generalized and typically more costly approach is often required.

Related to the previous point, there are complications to making "apples to apples" comparisons, even when ample data is available. With limited budgets, the emphasis is on attaining program savings with the available expenditures. Most program funds are diverted to incentives and implementation related costs, as well as the required EM&V assessment. The upfront investment in data tracking is difficult to justify if it doesn't directly produce savings.

What has been more common is the development of a field tool, or a tracking database, which can be more reasonably justified in an annual or multi-year program budget, especially where multiple utilities are working collaboratively. Even those efforts rarely consider how the data being collected might be used beyond ensuring that the program met its goals. The extra cost of tracking fields that could benchmark the program or portfolio against current program implementation efforts in other jurisdictions is rarely considered. Future changes to program designs are also difficult to incorporate if the tools aren't built with flexibility and scalability in mind, which is also more expensive and therefore often foregone.

Additionally, the typical approach to regulated energy efficiency quantification is to collect a year's worth of program data and deliver it to an EM&V firm, most of the time in form of a spreadsheet. A few months later a report is published that specifies evaluated savings, realization rates, net-to-gross ratios and costs. This stage in the program lifecycle creates a significant barrier for the analysis and application of historical data points tracked within utility and implementation firm databases. EM&V changes the data frequently and after the fact, and it is often determined that the reconciliation of the program tracking systems is cost prohibitive. As one would imagine, the complexity of creating a system that could easily make these reconciliations is significant, especially considering the lack of consistency in reporting EM&V results across different jurisdictions. Further integrating that system into a program or portfolio tracking system only adds to said complexity, as well as upfront investment. This can become overwhelming very quickly and usually results in implementers and evaluators remaining in their own silos. (Kleinman 2014)

What is Collected Currently to Support the DSM Industry?

Implementers today accumulate enormous amounts of program related data, but still need to turn information into business value. As an implementation firm, CLEAResult runs over 200 DSM programs throughout the US and Canada, which only represents a small percentage of the total regulated DSM market. CLEAResult's central tracking system, which collects data on leads, trade allies, applications, goals, and all of our individual measure and project savings, currently houses over 17,000 discrete data fields and receives over 300,000 entry transactions per year. (Lewis 2014) We have been able to combine our data with other external data sources, both publicly available and through third party vendors, to produce some very powerful analytics, including;

- Targeting strategies for new program and pilot offerings for new and existing clients
- Helping existing programs re-focus their strategies to target more specific market segments or individual customers bases
- Managing sales pipelines to mitigate the year end “hockey stick” effect
- Developing market analytics to inform internal business decisions

Case Studies

The types of benefits that implementation firms can expect from internally collected program data, publicly available external data sources, and third party purchased data is best explained through examples. The following is a highlight several recent examples of the types of analysis described in the previous section. As these examples demonstrate, we can already benefit from big data in energy efficiency, even if there are existing barriers that still need to be overcome.

Demand Response Market Assessment

Overview

In 2013, CLEAResult provided analysis and recommendations to support a multi-state utility in finalizing a program design and approach to the non-managed account market for demand response (DR) programs, currently only available to managed accounts. The objective of the project was to provide data-driven market analytics on the potential for the deployment of DR program offerings to eligible, non-managed C&I customers by assessing the potential for specific market segments to meet eligibility requirements through load shedding. The key questions we sought to answer included:

- Which customer segments should be targeted?
- What are the characteristics of each of those segments and what is the propensity of customers in each to participate in demand response programs?
- What are the crucial channels to employ to successfully reach customers in each segment?
- What are the potential barriers to customer participation in each segment and where are the opportunities?

Types of Data

The analysis used customer billing data, publicly available 8,760 load shapes, 3rd party GIS data, and a CLEAResult proprietary peak analysis tool. The quantitative results were supplemented with qualitative information gleaned from secondary research and phone interviews with randomly selected customers from market segments identified as primary targets by the peak analysis tool.

How the Data was Applied

The first step in the DR program assessment process was to analyze customer monthly billing data and identify groups of customers that represented good candidates for participation in a DR program based on their load profiles. This analysis was performed to identify the segments of high program potential within the market and narrow the pool of customers with whom to solicit in-depth interviews for the second stage of the project.

We prepared the data for analysis using the following process:

- Preliminary data assessment and organization for development of general metrics such as maximum connected kW and average kWh/kW ratio (proxy for run hours) on all C&I customers to make sure there were no errors in the data
- Classification of customers by 3-digit SIC code, and development of more specific metrics for each industry segment. SIC code was the only available industry identifier in the client's customer data set. Rather than convert to NAICS and potentially introduce error, we chose to work with SIC codes
- Disaggregation of monthly meter level billing data into hourly end-use level billing data. This was accomplished using building specific load curves, keyed off of SIC codes

We reviewed best practices to identify the most common end-uses curtailed in DR programs and then removed data for end-uses that are not typically available for curtailment through DR. We then ran the data through our proprietary peak analysis tool to identify those customers with sufficiently curtailable peak load on likely DR end-uses. The tool isolated hourly consumption for the peak period and arranged it so that the magnitude of variability can be compared incrementally through time.

We grouped the monthly billing data according to SIC code and then developed a key that links each SIC code to a corresponding building type load shape file. Using a custom key we applied each of the proportional end-use load shapes to the monthly kW billing data to disaggregate it into hourly interval load by end-use for each utility customer.

We then removed all of the off-peak hours, leaving only the on-peak hours for the next part of the analysis. The proprietary tool we use to analyze load data for potential DR tracks each two hour increment of peak load across time, and compares it to the other increments. If the delta between the lowest increment and the highest increment crosses a preset threshold, it flags that customer as a potential participant in a demand response program. The underlying theory behind this methodology is that when the delta exceeds the threshold the end-use equipment has enough variability in usage that it could be temporarily ramped down or the usage could be shifted. The caveat here is that we had no information on customers with on-site generation.

We acknowledge that a customer with on-site generation could have a flat load profile, but still represent a good candidate for DR participation. Because the utility client does not track which customers have on-site generation capacity, we used GIS analysis to identify as many of those accounts that fit recommended industry profiles for on-site generation (e.g. specific production processes or systems, facility size, connected voltage, etc.) and scrubbed as many as we could from the total list. While not ideal, given the customer information available from the utility, we felt that along with the conservatism built into our other assumptions, this was the best methodology. However, this is a noted entry point for error into the analysis. Figure 1 below shows how a process load demonstrates the manner in which the tool plots for a process load.

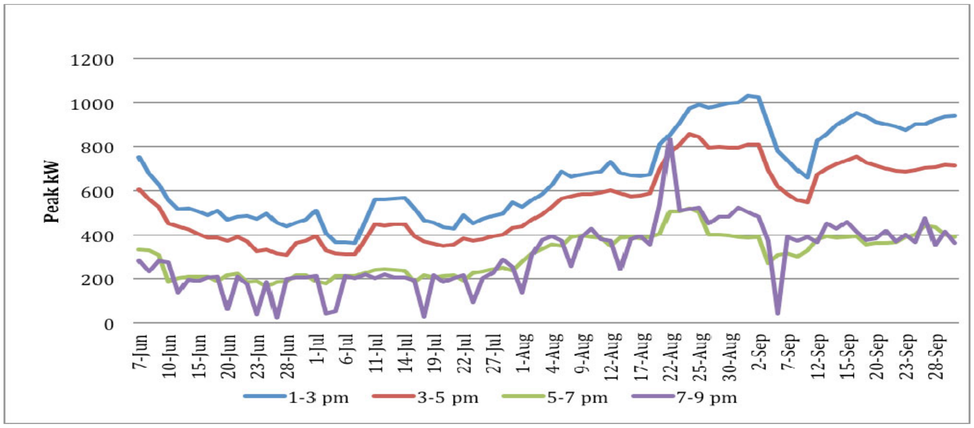

Figure 1. Peak kW Profile - Customer "X".

The tool also runs a check on the monthly demand trend to make sure that monthly average vs. peak indicates similar results. If the hourly load identifies the customer as a potential participant, but the monthly average demand does not indicate the same result in that month, it flags the customer for data review. Our goal in doing so is to identify potential outliers that could be driving the hourly increment data to flag the customer. Figure 2 below shows the average monthly demand trend for the same customers' process load.

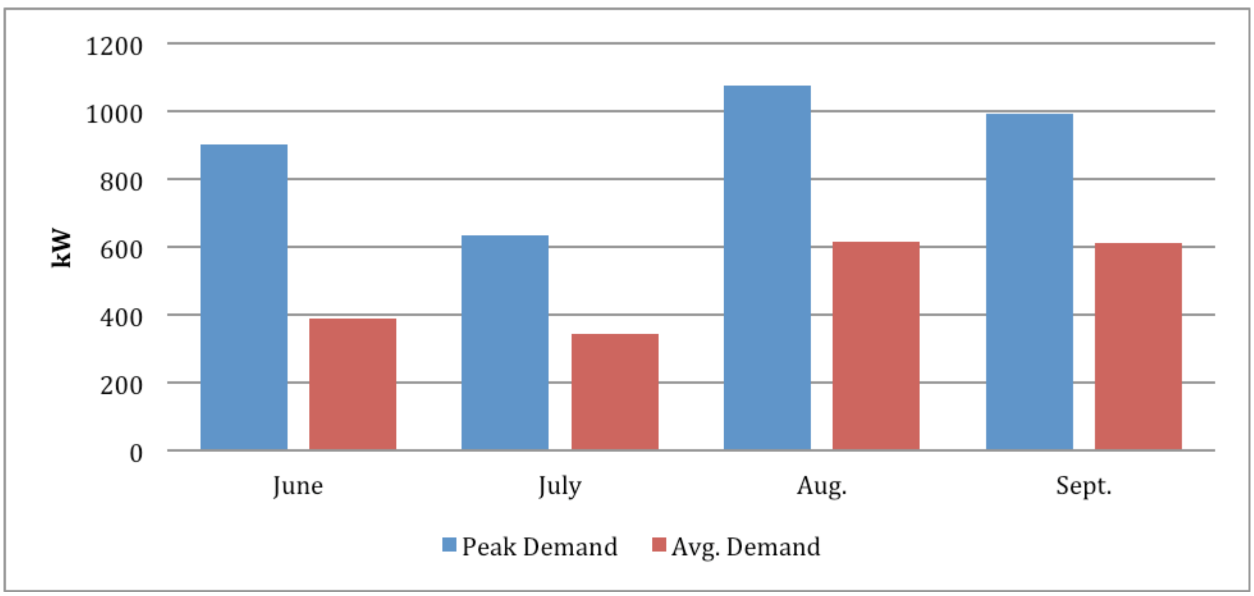

Figure 2. Process Demand Trend - Customer "X".

Because of the use of generic load shapes, we were conservative in setting our thresholds. Best practices suggest a threshold of 33% of the program minimum allowable connected kW. Because the program recommended minimum total connected kW of 200 is low for this market, we increased the lower limit to 300 kW. In Figure 2 above, the delta between the max peak kW and the minimum peak kW is well over 100 kW (33% of 300 kW) for nearly all the days in the summer peak period, and the demand trend data shows similar results on a monthly basis. This customer would then be flagged as a potential candidate for participation in a DR program.

The Availability of the Data

We used a patchwork approach of utility provided data, in-house CLEAResult data, publically available secondary source data and 3rd-party GIS data to complete the analysis. While customer usage data was provided by the utility, assumptions were required to disaggregate monthly usage data into hourly. In addition, the customer facility characteristics were analyzed through a 3rd party commercial facilities database, which required the use of GIS to match the 3rd party data to the utility account information.

Project Results

CLEAResult determined there to be 70 MW of achievable DR potential available via 825 program participants. We further developed segmentation and program delivery strategies with segment specific marketing collateral and approaches. The utility client has begun rolling out the DR program to the clients identified by the analysis. At the time of writing, the program has only been in operation for 3 months, and the DR peaking season has not yet arrived. Program results will be available next year, and we are excited to see how effective the analysis was in helping the client target the program to the most able and willing customers.

2013 SXSW Eco Hackathon

Overview

At the 2013 SXSW Eco Hackathon presented by Verizon Foundation and Pike Powers Labs, teams of volunteer hackers worked to transform vast disaggregated interval data from the Pecan Street Project (a University of Texas-based research organization) into usable tools within just 24 hours. The purpose was to provide a unique opportunity to create useful energy tools that transforms the way people make energy decisions. This app helps homeowners track regular energy usage and opportunities for energy efficiency, and it provides better utility customer interactions and aids in user behavior changes.

One note of importance is that with only 24 hours to develop the product from concept to app, the team was unable to fully develop a final product. The case study described below reviews the concept and idea behind the app, as if it were to have been fully completed.

Types of Data

The primary data that was used to create the app was a curated "wide table" set of participant data generated through the Pecan Street Project. The data we were provided was manipulated for the purpose of the Hackathon, to protect the privacy of the participants. It comprised of residential energy consumption and solar photovoltaic generation data. Each record in this dataset contains power readings taken on a per-minute level from eGauge monitoring systems located in the homes' breaker panels.

The supplemental datasets that were used to develop the complimentary functionality was California Energy + Environmental Economics E3 Energy Planning Tool load shapes, location-based weather data from National Oceanic and Atmospheric Administration (NOAA), and Austin Energy's Time of Use (TOU) five-tier rate tariff and "Value of Solar Tariff" (VOST).

How the Data was Applied

The first step in the development of the app was to analyze the disaggregated customer end use data and align that usage with an algorithm based on Austin Energy's residential electric rates, along with an offset for renewable generation. This analysis was performed to identify both the cost of the energy usage by end use and estimate the incremental cost of that usage. The algorithm was then applied to the generic load shapes to develop benchmarks for both national and local customers.

Next, using Ruby on Rails we developed the architecture for the app. The front end consisted of an input field for the users utility account number. After this point the app directed the user to a high level survey instrument, as well as the real time energy and price benchmarking screen.

Finally, after initial calibration, the app would analyze energy usage and make recommendations on better energy practices, as well as provide links to local and specific energy efficiency programs.

The Availability of the Data

The Pecan Street Project data that was provided to the contestants in the Hackathon was proprietary and only available through purchase. The complementary datasets were all publicly available, with the exception of the algorithm which was developed on site in excel.

Project Results

The app developed for the SXSW Eco Hackathon won the jury prize. The app was able to provide the following features:

- Real time energy use alerts when spikes accrue, capture the alert so the program learns behavior or predicts, and provide users with push/pull notifications
- Billing alerts with recommendations:
    - Provides information and awareness of tiered rate structure
    - Enables customers to set alerts for notification of approaching usage tier and shows better time of use
- Compares your data with data aggregated from similar homes in your geographic area to let you know how your consumption compares to others locally, as well as a national benchmark.
- Supplies customers with energy saving tips and leads user to participate in applicable DSM programs
- Monthly and minute cost of energy, kWh/Square foot and visualization of benchmark performance
- End use specific recommendations

The app is currently being developed for commercial application.

Conclusion

These are just two recent examples of the myriad ways that CLEAResult, and the rest of the DSM industry, are innovating to maintain relevance and improve the quality of their products and programs. We acknowledge that there are still tremendous challenges posed by data gaps and inconsistencies, and competition from "big data" entering the industry. We also feel that while it is important to be able to make the most of the data we are already collecting, the move towards having more data, from increasingly varied sources, and at a faster pace is inevitable. It does not bode well for the traditional implementation firms that stand idle and wait for the big data revolution to provide their program solutions for them. Those firms that continue to develop creative solutions to integrate and analyze the data they are already collecting will be that much better positioned to take advantage of the influx of data as more of it becomes available.

Acknowledgements

The authors of this paper would like to thank all the following people for their guidance and assistance in the development and review of this work (in alphabetical order): Steven Cummings, Annemarie Diaz, Jonathan Kleinman, David Lewis, Sean Nunes, Jeffrey Pape, Brian Smith, and Anthony Thompson.

References

Energy Business Review. May 15, 2014. *Baltimore Gas & Electric Deploys C3 Energy Smart Grid Analytics.* Accessed May 15, 2014 http://utilitiesretail.energy-business-review.com/news/baltimore-gas-and-electric-deploys-c3-energy-smart-grid-analytics-150514-4267151

Kleinman, Jonathan, interview by Melissa Culbertson. 2014. *VP, CLEAResult* (February 25).

Lacey, Stephen. 2013. "Nest and Opower TalStrategy for Residential Demand Response." *Green Tech Media.* July 13. Accessed January 16, 2014. http://www.greentechmedia.com/articles/read/nest-and-opower-make-more-big-moves-in-residential-demand-response.

Laitner, John A. "Skip". 2013. *Calculating the Nation's Annual Energy Efficiency Investments.* Report, Washington, D.C.: ACEEE.

Leeds, David J. 2012. *The Soft Grid 2013-2020: Big Data & Utility Analytics for Smart Grid.* Market Research, New York City: GTM Research.

Lewis, David, interview by Alek Antczak. 2014. *Director, CLEAResult* (February 22).

St. John, Jeff. 2013. "Big Data on the Smart Grid: 2013 in Review and 2014 Outlook." *Green Tech Media.* December 16. Accessed February 10, 2014. http://www.greentechmedia.com/articles/read/Big-Datas-5-Big-Steps-to-Smart-Grid-Growth-in-2014.

2013. *Utility Scale Smart Meter Deployments.* IEE Report, Washington, D.C.: Edison Foundation.