JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Volume 12, No. 1
April 2019
ISSN: 1946-1836

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is three issues a year. The first date of publication was December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of AITP-EDSIG who perform the editorial and review processes for JISAR.

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Editors

2019 JISAR Editorial Board

Standardizing Public Utility Data: A Case Study of a Rural Mid-Size Utility

Edgar Hassler
hassleree@appstate.edu

Joseph Cazier
cazierja@appstate.edu

Jamie Russell
russellja@appstate.edu

Thomas Mueller
muellerts@appstate.edu

Daniel Paprocki
paprockidj@appstate.edu

Center for Analytics Research and Education
Appalachian State University
Boone, NC 28608

Abstract

Energy is important to our daily lives. Energy data is important to utilities to meet operational goals and have better relationships with their customers. However, to get the most value out of their data they need to combine it with data from other secondary sources and apply advanced analytics techniques. This can only be done effectively and efficiently if utilities standardize their data in a way that allows the merger of their data with that of other sources to take place seamlessly. This article uses a case study to illustrate why small to midsize utilities should adopt a data standard, discusses some of the challenges in choosing and adopting a standard and concludes the process of moving to a data standard in our case. Results will be interesting to utilities and any industry contemplating taking advantage of the opportunities afforded by big data.

Keywords: Data Standardization, Energy Analytics, Sustainability, Big Data, Analytics

1. INTRODUCTION

Energy is ubiquitous to our modern lives. Consequently, energy data is also becoming ubiquitous in our modern digital era. Using this data to improve customer service, predict and prevent power outages, optimize power generation and persuade users to modify their energy consumption patterns can significantly improve the efficiency, profitability and sustainability of regional power providers. This is especially true when operational utility data is combined with other secondary data sources as described below.

During the 1990's, there was a wave of business innovation focused on what was branded *Business Intelligence* (BI). While definitions vary, the primary focus of BI was to gather and make sense of a company's *internal* data from their operations (Negash, 2014). This led to many operational improvements and efficiencies.

Today we are evolving into the next wave of data. Companies are integrating external or secondary data from outside of their core business to create even more value (McAfee & Brynjolfsson, 2012). This has come to be known as the era of *Big Data*, where data from multiple sources is combined, aggregated and analyzed in new and more advanced ways. Often by adding data from multiple sources there is a synergistic or exponential increase in the benefits that can accrue for the organization (Barton & Court, 2012).

Examples of secondary data in the utility industry might include merging property tax records. Such records can give the utility information such as the size of the home, heating source, tax value, age and other information that can be analyzed to help understand power consumption patterns. Likewise, weather data, a key driver of electricity consumption, can be merged into the data set to better understand how it impacts electricity use, used to predict future energy usage, and simulate what-if scenarios. Additional information like demographic data can be acquired and merged into the company operational data to better segment and reach out to consumers with individualized messaging via a consumer preferred media channel based on age, income, occupation, educational level and other factors. All of this can lead to a smarter, data driven utility.

To take advantage of the promise offered by big data and analytics, companies need to first be able to acquire the data by successfully merging their operational data with other relevant secondary data for analysis. To do this, there needs to be a way to connect their data with relevant data from outside the organization. This is usually done in the form of a primary key that uniquely identifies a record in each of the internal and external data sets.

However, the merge can only take place if the keys match. Since the most common identifier across most utility and external sources of data is the utility service address, this is generally the best key to use. However, in many data sets, addresses do not follow a set standard with many possible variations in how they are recorded and stored. Additionally, the data are fraught with errors as clerks often try to decipher hard to read handwriting or manage typos in a consistent manner. There are many "right" ways to write and store an address. This is the fundamental problem, as different formats (or lack of formats) won't match on a secondary data set merge.

To solve this problem, utilities can adopt and consistently follow a data standard. By having their data recorded in a consistent methodological way, they can merge their data with outside sources using the same format or convert the data to that format with a conversion code block. However, to do so efficiently, the data needs to be stored in a standard consistent format for transformation and merging of the data.

In this applied research paper, we utilize a case study to discuss some of the benefits data science can bring to the utility industry, illuminate the role of adopting, migrating to and enforcing a data standard in the process and share our experience with a rural public utility as they work to transform their data and adopt a data standard.

2. LITERATURE REVIEW

Public utilities makeup to 94.3% of all utilities in the U.S. and serve 68.3% of all energy consumers (Public Power, 2017). Public power utilities are governed by elected and appointed boards, which include mayors, city council members and citizens for the common good, rather than by large investors groups. The mission of a publicly utility is to optimize benefits for local customers, usually in the form affordable energy rates (California Energy Commission, 2017).

Some of the current and potential benefits derived from the application of analytics to a utility's big data merged with secondary data sources include: benefits for the utility, consumers and society as a whole. We will briefly discuss each of these in the subsections below.

For the Utility

As with all businesses, utilities provide a service for which consumers are willing to pay. That service is the reliable and affordable delivery of power on demand. At its most simplistic, the connections between the utility and the consumer are the power cables and the mail service by which the bill is delivered and the check is received. The records or data which have been typically recorded consist of a monthly meter reading. The widespread adoption of smart

meters and other monitoring equipment is rapidly changing this traditional model bringing with it great potential benefits and also some real challenges.

The adoption of smart meters and other monitoring equipment generates near real-time information which has the potential to dramatically improve utility operations in many areas including: successfully integrating intermittent renewable energy and other distributed energy resources, predicting and quickly respond to outages, increasing overall grid efficiency, improving load forecasting, providing rapid anomaly detection, improving demand side management, and, in the future, successfully integrating the growing demand for electric vehicle charging (Aria and Bae, 2016; Katz, 2018; Schuelke-Leech et al, 2015; Wen et al, 2018; Zhou et al, 2016). Another benefit of successfully interpreting this data is the ability to change the communication paradigm from a monthly bill mailing to real-time, individualized, interactive consumer messaging.

The benefits above come with the challenges of managing a huge increase in data volume compounded by integration of newer technologies into older existing infrastructure, and the fact that these increased data volumes are being overlaid upon legacy databases and other dated information systems (Katz, 2018; Schuelke-Leech et al, 2015; Zhou et al, 2016). Suggestions for managing these challenges include data compression, the creation of comprehensive data frameworks, and making sure that utility data systems are aligned with the best practices of modern data analytics and data science (Akhavan-Hejazi and Mohsenian-Rad, 2018; Munshi and Mohamed, 2017; Wen et al, 2018).

The potential benefits of big data for utilities are numerous. Even though the specific mechanisms for enacting those benefits are not fully developed, the informational underpinning for these systems will be based on the best practices of data analytics.

For the Consumer

Targeted, clear and well-presented messaging from power providers, to consumers, will build an advantageous relationship. Energy consumers are partial to information that allows them to make informed decisions on utilitarian, psychological and social implications within their communities (Hartmann & Apaolaza-Ibáñez, 2012). Social media has become an essential tool in promoting the corporate brand and communicating directly with the consumer. The purpose, many times, is to share a pro-environment attitude and to encourage customers to take action (Ballew, Omoto, & Winter, 2015). Social media consumers hold reduced risk aversion, sustain higher brand loyalty, and are overall more satisfied with the brand (power company) when social media information is available (Reisenwitz, 2013).

An essential component in energy consumption decision making is sustainable development. It has become a crucial topic for consumers. There is heightened interest in waste management, greenhouse gas emissions and renewable energy. Social media technologies have attracted attention as being a viable tool to encourage sustainable actions (Sogari, Pucci, Aquilani, & Zanni, 2017). Many times, consumers who engage social media are more aware of sustainability practices, then make buying and consumption decisions based on social media information. Studies indicate millennial consumers hold a strong affinity with social messaging and appreciate companies that help in acquiring information (Dabija, Bejan, & Tipi, 2018; Hartmann and Apaolaza-Ibáñez, 2012).

For Society

Big data and cutting-edge analytics have the opportunity to play an important role in improving the way the energy sector interacts with society as a whole. As the world becomes more focused on utilizing resources in an efficient manner, it is imperative that the energy sector continues to innovate to meet these societal demands. By using ever improving computing power and data analytics, the energy sector can begin to have a focused conversation with their clientele regarding energy consumption and conservation. Utility companies can use analytics to communicate with customers regarding upcoming high electric demand events (peak power) that put a strain on the environment and people's checkbook. This can help society understand when electricity generation is most expensive and most carbon intensive, and how they might conserve during those periods (Kenworthy, 2016).

This benefits everyone in a community, as the impact of power production on the environment is lessened, both the utility company and the clientele benefit from reduced peak power costs, and there is a general increased community exposure to these important issues (Feldman et. al., 2015). The educational benefits can cascade down into other sectors of one's life, changing the way one makes choices that may impact the planet. This not only includes energy, but also

sustainable consumption of water, and food. There is a great potential for society to benefit from our energy providers taking advantage of these exciting new technologies.

Next, we discuss some of the challenges to adopting big data techniques and how to overcome them.

Challenges to Adopting Big Data in Public Power

While these benefits are real, many challenges remain in place before they can be fully realized, especially among our small rural utilities and cooperative municipalities which make up the majority of utilities in the U.S. (Public Power, 2017). There is a lot of data being collected, especially as we start to enter the age of smart meters and appliances. However most of these public utilities use the data primarily for their billing system and internal operational efficiencies. While important, this leaves the true potential of big data for customer engagement, sustainability and total business optimization largely untapped.

Analytical tools continue to be developed and provide value for many firms in other sectors, or the better funded large private utilities, but many public utilities continue to be left behind. The primary reasons include:

- *Non-Standard Data* - Each public utility has evolved to keep data in its own way, with their own system targeted to their business environment and influenced by their history and focus. Some controlled data entry, like customer addresses, tightly. Most do not.

- *Legacy Rules and Regulations* - While rules and regulations differ depending on the utility size, ownership structure and location, they can be quite influential on a regulated public utility's use, or lack of use, of their data for activities beyond billing. A detailed analysis of rules and regulations is beyond the scope of this paper, but we acknowledge it here.

- *Lack of knowledge* - Most public utility managers and engineers are smart, well trained hard-working people. However, the discipline of Data Science did not exist as a discipline or course of study when most of them finished school, thus they need training and guidance to understand the benefits of applying analytics to their data to improve their business and customer relationships. While we give some examples of how this can be used in this paper, the primary purpose here is to discuss the importance of a data standard.

Data Standardization

Although all these reasons are important barriers to taking advantage of the promise of big data and analytics in this digital age, the primary focus of this paper is on the issue of data standardization. The rest of this paper explores how data standardization can create value for the consumer, public utilities (especial small to mid-size ones) and society; discusses some of the challenges to standardizing data; and shares a case study for a small rural utility.

3. MATERIALS AND METHOD

A Primer on Merging Data

Pieces of information about a customer (i.e. name, address, phone, etc.) are usually stored together as a unit called a record. As like records are collected into a set it becomes necessary to have a unique identifier for each record. This unique identifier is known as a primary key in database terms (Elmasri & Navathe, 2000). The primary key may be a simple account number or other identifier that is assigned to only one customer.

In order to relate different record sets to each other, the primary key from one record is included as a piece of information in other related records (Elmasri & Navathe, 2000). For example, a customer's name and contact details may be stored in one record set, while their meter reading with the date and time of the reading are stored in a different record set. This avoids duplication of customer details in each record of a reading and improves efficiency. To ensure that each reading is related to the correct customer, we simply include the associated customer identifier in each meter reading.

To recombine the details of records from two different record sets one simply matches the primary key from a record with the records containing the same key in the second set. This is known as a *join* in database terms (Elmasri & Navathe, 2000).

The ability to join two different record sets that do not share a key is thus challenging, but pivotal to linking data from different sources to create extended data records for each consumer.

Standardizing Public Utility Data

By standardizing the way that small utilities record addresses of their meters, the ability to merge secondary data, such as demographics

and property records, becomes much more achievable. If they have a public standard, such as using the postal record standard (USPS Pub 28) or a 911 standard, the new data can be transformed into this same format and merged with the existing data.

Challenges to Data Standardization

With all the potential benefits of using big data and analytics to improve operational efficiencies, reduce their environmental impact and build a more engaging relationships with their customers, it is important to address the barriers of data standardization. However, some of the technical and operational barriers to achieving this can be formidable, as illustrated below.

Resource Constraints

For most small organizations, few resources can be dedicated to cleaning data and performing analytics. Data analytics can be very costly in several ways, not just financially. To retroactively apply the standard to addresses already in the system is very time-consuming and can take hundreds, even thousands of hours. Many smaller utilities do not have the time, capital, or man-power to put towards retroactively cleaning addresses to conduct analysis. This presents a technical and resource issue that many companies do not have the ability to overcome on their own.

Customer Variability

Utility firms do not cater to only one type of customer, there are several different types to consider when determining a standard format for the input of addressing data. Energy meters are attached to many different types of structures, including residential, commercial, educational, governmental, and infrastructure. Each structural type represents a distinct set of shared characteristics and is charged a rate aligning with the purpose of the structure. Houses and certain apartment complexes are considered to be residential. Local private businesses are the commercial structures. The buildings located on a university campus are educational structures. Meters classified as infrastructure are structures that use power but are not buildings such as street lights, well pumps, or large signs covered with lighting. Each of these types of structures has its own unique identifiers that need to be clear and available in the addressing data. Additionally, there are subsets within some of the categories, such as single-family homes, and duplexes. It is non-trivial finding an addressing standard that accommodates all the different categories. Utilities need to start with a base addressing standard, and then expand to accommodate all possibilities, including residential, commercial, other building types, and infrastructure addresses.

Multiple Types of Merges

An additional challenge is that one address does not always mean one meter. For example, the property records received from county officials are optimized for tax collection. They generally contain one record for each tax structure, not utility meter. Thus, there are four common types of matches when merging the datasets:

1. One meter address to one property record (e.g. single family home)

2. Multiple meter addresses to one property record (e.g. apartment building)

3. One meter address to multiple property records (e.g. home with detached garage or barn)

4. Multiple meter addresses to multiple property records (e.g. shopping mall complex)

In cases 2 - 4 this creates additional challenges because one cannot discern exactly to which structure a meter is attached.

Standards of Available Data to Merge

Many towns and cities utilize the E911 addressing standards, which assigns a standardized address for emergency organizations to easily locate a building or structure. These standards are often self-determined but follow addressing standards that have been developed over years of trial and error. This standard is based primarily on the guidelines laid out in the United State postal code, which is universal throughout the nation. However, because each municipality has their own digressions from the absolutes of the standard outlined in the postal code, E911 addressing standards are not necessarily the same from one town to the next. This poses a problem because each municipality may impose a slightly different set of addressing standards than the set of standards that have been chosen for one application.

4. CASE STUDY FOR A RURAL UTILITY

We worked with a small rural utility in the Southeastern part us the U.S. with around 10,000 meters to help identify a data standard and transition their legacy data to that standard. The goal was to prepare them to integrate additional secondary data into their system, so they could segment their customer base and personalize their outreach to consumers around important

issues including peak power, pre-pay options, and energy assistance programs.

In the sections that follow we discuss some of the challenges faced and how we addressed them.

Merge Goal and Approach
As a first step, the utility wanted to merge publicly available government data from property tax records with their billing records to better understand power consumption patterns as they relate to home characteristics. The county exported a spreadsheet (Excel) from their database containing information such as address, year built, tax value, square feet and heating source. The task was to then merge this with 10 years of monthly energy readings for each residence. ~8,000 relevant meter ID's and addresses were selected for this analysis. A sample of the type of data from each original source is shown in Figure 1 (in Appendix A).

Recall from the previous discussion on databases that in order to merge (join) two different record sets there must be a shared key between them. The most immediate key to merge secondary data to the example utility company meter locations is the address. However, each meter location has two fields, each containing a portion of the information required to create a standardized address. To create a standard address, each field must be broken down into granular pieces and the pieces then recombined to form a complete address.

Constructing a Standard
Once information from the two address fields are broken down into distinct pieces, a standard for each field needs to be in place. As a starting point, USPS Publication 28 - which outlines preferred ways of representing all aspects of an address - was chosen. Because Publication 28 standards are meant to accurately represent a location for the purpose of parcel delivery, it provides a widely recognizable method of storing address information. Other possibilities include the E911 standard, geocodes or other positioning systems.

The final standard for address data consists of 12 individual fields. The fields were derived through an iterative process in which fields were added to the USPS Publication 28 base. A complete list of fields, along with their definitions, can be found in Appendix B.

Data Conversion
To illustrate how an address is broken up and cleaned, Figure 2 (see appendix A) represents the raw data. From this point, programming scripts are used to separate out all of the pertinent information.

There are several pieces of information located in each field of the raw data. The field titled Service Address contains a place name (Nathans Walk), Secondary Unit Indicator (A3), City (Local City), State (NC), and Zip (77777). The Line and Pole field contains a utility grid identifier (D16), Street Number (872), Predirectional (W), Street Name (EMPEROR), Address Suffix (ST), and Secondary Indicator (APT). Figure 3 (see Appendix A) shows the address fields properly segmented.

The individual fields may then be viewed as subcomponents of a larger, composite key that will be used to merge the datasets.

Data Filtering
After determining a standard, data management software was used to retroactively apply the standard to all addresses in the current dataset. The addresses were then filtered to uncover records missing critical components of the address such as street number, street name or street suffix. Records for which this information was unrecoverable were removed from the dataset as they cannot be matched. Essentially, they do not have a key.

Logical Steps of the Merge with Secondary Data
The following is a summary of the steps taken to prepare and merge the data, including an approximate percent of merged observations after each step.

Step 1: Upper case the "Service Address", "Line and Pole", and county address fields.

- This ensures that two fields with equivalent content are not prevented from matching due to a difference in casing.

- Nothing merges at this point, because there are no identically defined fields to match with in the meter address data set.

Step 2: Begin creation of distinct address fields by parsing (breaking up) the "Line and Pole" and "Service Address" fields.

- The "Line and Pole" field begins with a grid identification number, which is not part of an address. This is first segmented out into its own field and stored for future use but will not be useful for the merge with secondary data.

- After removing the grid identification number, the beginning of each string now contains a numeric value indicating the street number.

This numeric value is separated from the character values that follow and used to fill the address number field. The remaining characters were moved into a temporary address street column to be further broken apart.

- Nothing merges at this point, because the data is not fully segmented.

Step 3: Create Street Name and Suffix fields then standardize Street Suffix to Publication 28 addressing standards.

- A new field for Street Suffix is formed, using the example utility meter address data. First, to properly format the suffix field, the main dataset is segmented into smaller datasets based on patterns displayed in the line-and-pole address that indicate a proper suffix. For example, all variations of the suffix ROAD are removed to form a new dataset, standardized to RD, and then reintegrated back into the master list. A similar process is followed for all other major suffixes.

- This step is completed for both the utility and county datasets and the fields for address number, street name, and street suffix will be utilized as keys to merge the meter data with secondary data from the property records.

- After performing these operations, approximately 50% of the meter locations merged with a property record from the county tax records.

Step 4: Correct and clean data entry errors found in the Street Name fields.

- Many observations have small errors in the street name that prevent the two data sources from merging properly. Therefore, it is appropriate to look at the observations that do not merge, which isolates some of the errors. Then manually, one must sort through the street names and document errors. For example, a street name like "George Critcher", might be abbreviated as "GEO CRITCH." This type of error or misspellings should be isolated, and then changed with programming scripts.

- These operations increase the overall merge yield, while standardizing the data at the same.

- After performing these operations, approximately 83% of the meter locations merged with a county property record.

Step 5: Manually search for parcel outlines in the tax record website to match parcels to addresses.

- Each meter is located on a parcel of land, which is recorded in the county tax records. Therefore, searching each meter address in google maps shows the parcel outline - providing a method of visually matching a meter address with a parcel ID. The parcel id is manually recorded for each unmatched meter, and then the tax records for these remaining meter addresses are merged by that parcel ID.

- After performing these operations, 87.6% of the meter locations merged with a property record from the county tax records.

While this process does not completely merge all of the records, it does allow us to perform analytics on the majority of the meter locations. It was decided to have our analytics team stop at this point as merging additional records would require a person to physically visit each meter to confirm the address. The utility plans to have some of their meter readers complete this task in the near future and report back with the data.

Discussions with the utility database administrator has revealed that a project is underway to geolocate each meter. As shapefiles are available for each parcel, it is hoped that an additional step of matching meter locations inside a parcel-shape can be added to complete the merging and standardization of the datasets when geolocations become available.

Figure 4 provides a visually summary of the types of merges found in the final merged dataset.

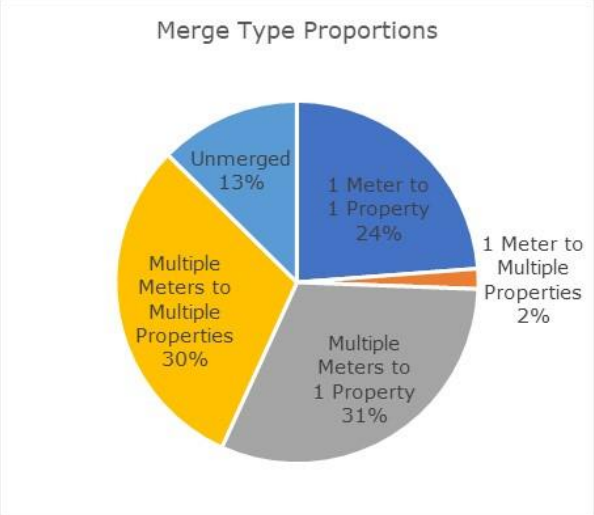

Figure 4. Merge types in final dataset

5. DISCUSSION

When merging enhancement data into an existing dataset, considerable work must be accomplished before analytics can begin - especially when relying on address data as a key for the merge. With address information, it is important to define a standard, convert all data to the standard through dissection and iterative cleaning, and filter the data for cases that must be handled manually.

In the course of our investigation, we found that using a more granular address standard required additional preparation and parsing work in the beginning, however it simplified many of the cleaning tasks and, more importantly, we have found it is now easier to reassemble the pieces in new formats to complete additional merges.

Despite only merging 88% of the records, the utility in this case was very satisfied with the results. The resultant dataset is sufficient to complete the next phase of the project, which is segmenting their customers to better understand power consumption patterns as they relate to home characteristics. Based on this knowledge, better communication and incentive programs can be designed and targeted where they will have the most impact on reducing peak demand.

There is also immediate value in standardizing data before analytics even begin. The utility has increased confidence in billing records, it facilitates standard operating procedures, and assists efforts regarding legal and regulatory compliance.

Smart meters are changing the landscape for utilities. Data standards becomes even more important as the volume, velocity, and variety of data increase. The increasing need for real-time analysis means there is reduced time for extended data cleanup efforts.

This means that it is important to maintain data within standards once they are implemented. To do so may include staff training programs to improve data entry and curation. It may also include improved checks and enforcement within processing systems.

6. CONCLUSION

Data standards are an important part of the business intelligence and data analytics framework. They make the merging of additional data sources possible and facilitate the transfer of data to modern tools, thus enhancing decision-making.

In future work, we are excited to expand the merging of data to additional sources such as demographics and studies regarding how and what to communicate with consumers to reduce peak demand. With additional utilities collaborating in the pooling and merging of data, big data, indeed clean big data, that can lead to peak demand reductions is within the grasp of even small rural utilities.

7. ACKNOWLEDGEMENTS

Special thanks to our graduate students Dalton Stout and Lucas Stinson for their help on this project.

8. REFERENCES

Akhavan-Hejazi, H., & Mohsenian-Rad, H. (2018). Power systems big data analytics: An assessment of paradigm shift barriers and prospects. *Energy Reports*, *4*, 91–100. https://doi.org/10.1016/j.egyr.2017.11.002

Arghandeh, R., & Zhou, Y. (2017). *Big Data Application in Power Systems*. Elsevier.

Arias, M. B., & Bae, S. (2016). Electric vehicle charging demand forecasting model based on big data technologies. *Applied Energy*, *183*, 327–339. https://doi.org/10.1016/j.apenergy.2016.08.080

Ballew, M. T., Omoto, A. M., & Winter, P. L. (2015). Using Web 2.0 and Social Media Technologies to Foster Proenvironmental Action. *Sustainability*, *7*(8), 10620–10648. https://doi.org/10.3390/su70810620

Barton, D., & Court, D. (2012). Making Advanced Analytics Work For You. *Harvard Busimess Review*, *90*(10), 78.

Dabija, D.-C., Bejan, B. M., & Tipi, N. (2018). Generation X versus Millennials communication behaviour on social media when purchasing food versus tourist services. https://doi.org/10.15240/tul/001/2018-1-013

Differences Between Publicly and Investor-Owned Utilities. (n.d.). Retrieved July 14, 2018, from http://www.energy.ca.gov/pou_reporting/background/difference_pou_iou.html

Electricity and the Environment - Energy Explained. (2014, November 22). Retrieved July 14, 2018, from https://www.eia.gov/energyexplained/index.php?page=electricity_environment

Elmasri, R., & Navathe, S. B. (2000). *Fundamentals of Database Systems* (3rd ed.). Reading, MA: Addison-Wesley.

Feldman, B., Tanner, M., & Rose, C. (2015). *PEAK DEMAND REDUCTION STRATEGY* (p. 59). Advanced Energy Economy. Retrieved from http://info.aee.net/hubfs/PDF/aee-peak-demand-reduction-strategy.pdf?t=1446657847375

Hartmann, P., & Apaolaza-Ibáñez, V. (2012). Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern. *Journal of Business Research*, *65*(9), 1254–1263. https://doi.org/10.1016/j.jbusres.2011.11.001

Kenworthy, B. (n.d.). Real-time Data Analytics. Retrieved July 14, 2018, from https://www.elp.com/articles/powergrid_international/print/volume-21/issue-3/features/real-time-data-analytics.html

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, *90*(10), 60–68.

Munshi, A. A., & Mohamed, Y. A.-R. I. (2017). Big data framework for analytics in smart grids. *Electric Power Systems Research*, *151*, 369–380. https://doi.org/10.1016/j.epsr.2017.06.006

Negash, S. (2004). BUSINESS INTELLIGENCE. *Communications of the Association for Information Systems*, *13*, 20.

Publication 28 - Postal Addressing Standards. (n.d.). U.S. Postal Service.

Reisenwitz, T. H. (2013). A comparison of the social media consumer and the non-social media consumer. *International Journal of Internet Marketing and Advertising*, *8*(1), 19–31. https://doi.org/10.1504/IJIMA.2013.056587

Schuelke-Leech, B.-A., Barry, B., Muratori, M., & Yurkovich, B. J. (2015). Big Data issues and opportunities for electric utilities. *Renewable and Sustainable Energy Reviews*, *52*, 937–947. https://doi.org/10.1016/j.rser.2015.07.128

Sogari, G., Pucci, T., Aquilani, B., & Zanni, L. (2017). Millennial Generation and Environmental Sustainability: The Role of Social Media in the Consumer Purchasing Behavior for Wine. *Sustainability*, *9*(10), 1911. https://doi.org/10.3390/su9101911

Stats and Facts. (n.d.). Retrieved July 14, 2018, from https://www.publicpower.org/public-power/stats-and-facts

Wen, L., Zhou, K., Yang, S., & Li, L. (2018). Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews*, *91*, 59–69. https://doi.org/10.1016/j.rser.2018.03.088

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, *56*, 215–225. https://doi.org/10.1016/j.rser.2015.11.050

Appendix A

Example Figures

Location Id	Service Address	Line And Pole
5662682	University Air Quality Study, Local City, NC 77777	G14 University Air Quality Study Bldg
5662687	327 Academy St New Dining Facility SRV1, Local City, NC 77777	G18 327 ACADEMY ST
5662688	103 North St TS, Local City, NC 77777	D17 103 North St TS
5662689	327 Academy St New Dining Facility SRV2, Local City, NC 77777	E08 327 ACADEMY ST
5662704	199 Jefferson Rd B, Local City, NC 77777	E24 199 JEFFERSON RD UNIT B
5662705	355 Hunting Hills Ln, Local City, NC 77777	K26 355 HUNTING HILLS LN
5662706	419 Meadowview Dr Apt, Local City, NC 77777	M23 419 MEADOWVIEW DR APT
5662707	Local City Pointe 205, Local City, NC 77777	G21 148 Hwy 105 Ext Unit 205

Figure 1. Sample Merge Data from Source Tax Records and Utility Address

Note in Figure 1 that the service address field has information regarding the city, state, zip code, and a place name (Local City Pointe), whereas, the Line and Pole field has the street number, street name, secondary unit, and the secondary unit indicator. The first step in the merger process is to split this address into the granular bits of information so they can be matched as a composite key where each sub-component is matched separately. After this standardization, additional information (i.e. property tax records) can be matched to the addresses.

Location Id	Service Address	Line And Pole
5665987	Nathans Walk A-3, Local City, NC 77777	D16 872 W EMPEROR ST APT A-3

Figure 2. Sample Non-Standard Addressing Data

Primary Address						
Address Number	Predirectional	Street Name	Address Suffix	City	State	Zip
872	W	EMPEROR	ST	LOCAL CITY	NC	77777

Secondary Address	
Secondary Indicator	Secondary Unit Indicator
APT	A3

Figure 3. Properly Segmented Address Fields

Appendix B

Final Addressing Standard

Street number - the number listed on the front door and/or mailbox of the property. This is positional descriptor along a street and is related to the cut in the street for the structures driveway. All addresses should contain a street number.

Predirectional Identifier - North, South, East, or West, shortened to only the first letter of the direction. This includes street such as "E King St" or "West King St." Not all addresses will have a predirectional identifier.

Street Name - the name of the road containing a curb cut for the site-in-question driveway. All addresses should contain a street name.

Street Suffix - the type of street the structure sits on, such as street, road, lane, etc. All the suffixes are abbreviated for entry into this field. All addresses should contain a street suffix.

Primary Street Suffix Name	Commonly Used Street Suffix or Abbreviation	Postal Service Standard Suffix Abbreviation
BOTTOM	BOT	**BTM**
	BTM	
	BOTTM	
	BOTTOM	
BOULEVARD	BLVD	**BLVD**
	BOUL	
	BOULEVARD	
	BOULV	
BRANCH	BR	**BR**
	BRNCH	
	BRANCH	
BRIDGE	BRDGE	**BRG**
	BRG	
	BRIDGE	

Figure B-1. Suffix Abbreviations (Publication 28)

Postdirectional Identifier - North, South, East, or West, shortened to only the first letter of the direction. This includes addresses such as "US Hwy 421 S" or "US Hwy 421 N". Not all addresses will have a postdirectional identifier.

Secondary Unit Indicator - these are identifiers that indicate the metered address is part of a larger structure. Secondary unit indicators include apartment (apt), suite (ste), and unit. Not all addresses will have a secondary unit indicator.

APARTMENT	**APT**
BUILDING	**BLDG**
FLOOR	**FL**
SUITE	**STE**
UNIT	**UNIT**
ROOM	**RM**
DEPARTMENT	**DEPT**

Figure B-2. Secondary Unit Designators and Abbreviations (Publication 28)

Secondary Number - this number identifies the apartment or unit number of the metered address. In some cases, the secondary number can be a letter or a combination of numbers and letters. Not all addresses will have a secondary number.

Infrastructure Identifier - this contains information indicating that the meter is not attached to a residence. A structural identifier can indicate the meter is for a sign, well, traffic signal, etc. Not all addresses will have a structural identifier. We decided that this requires a separate field because this type of address is not a structure like most addresses. These observations should not have associated tax record information, as they are not buildings. In the future, we may decide to separate these items out to analyze energy consumption for items other than commercial or residential properties.

Place Name - this identifies the title of the collective in cases where the meter is assigned to one parcel or unit of a larger structure. Place names can be titles of apartment complexes, shopping centers, malls, etc. Not all addresses will have a place name. This information is important to separate out, as it contains unique identifying information.

City - the city that each address can be found in. Almost all of the example utility company meters are located in the city limits aside from a few, which are located just outside city limits in unincorporated territory. Not all addresses will list a city.

State - this indicates the state that each meter is located in. All the example utility company meters are located in North Carolina (NC). All addresses will list a state.

Zip Code - the postal code assigned by the USPS to each address. All of the addresses in the example utility company data share the same zip code except for the few on-campus addresses which have the university-specific zip code. All addresses will list a zip code.