TABLE OF CONTENTS

INTRODUCTION

During 2019, the Virginia General Assembly approved HB 2332, which required the Virginia State Corporation Commission (SCC) to convene a stakeholder process to obtain recommendations related to electricity data access and privacy and provide a report to the General Assembly by April 1, 2020.

This paper collects perspectives from various stakeholders across the United States in order to inform the SCC on viable options for data aggregation. After conducting a literature review to understand and define the concept of data aggregation, the Community Climate Collaborative (C3) began exploring the relationship between data anonymization and aggregation, as well as identifying their associated costs. Then, C3 analyzed various thresholds for aggregation based on privacy and usefulness. From the perspective of privacy, thresholds were identified that did not disclose private, consumer-specific information. From the perspective of usefulness, thresholds that provided relevant information for studies performed by academic, civil, and governmental organizations were prioritized.

After exploring various thresholds, C3 discussed the topic of timing, or when data would be aggregated. Then, several data security measures were described, as storing data in a safe manner is essential for privacy protection. C3 also detailed the best ways to access data in a standardized, functional manner. Finally, this report offers recommendations regarding the proper threshold of data aggregation for the Commonwealth of Virginia.

LITERATURE REVIEW

When energy data users such as municipal governments seek data from energy utilities, they can experience a range of utility and regulatory barriers. For example, state rules intended to protect customer privacy may be ambiguous or overly restrictive. Providing requested data may be at the discretion of the utility company, which often tend to err on the side of confidentiality, arguing that this is in the best interest of the customer. This is particularly unfortunate given that third parties can use energy data to create programs, policies, and services that benefit utilities' ratepayers and the utilities themselves (Crandall 2019).

When utilities do provide data to third parties, they typically process the data using two methods to avoid identifying specific customers: anonymization and aggregation.

ANONYMIZATION

Anonymization removes unique personal identifiers like name, address, or account number from a customer's individual data, such that the customer's utility use data could be released without it being attributable to a particular individual (Crandall 2019). Personal information could either be deleted or generalized; for example, the consumer's date of birth could be deleted completely or generalized to a year (Feinauer, et al. 2016).

This process of anonymization could make data available at the single-home level, creating the potential for granular data analysis, while still preventing third parties from assigning that data to an actual customer or location. However, anonymizing data may not sufficiently prevent re-identification and thus anonymization methods should be only a part of the solution to ensure data privacy (Kalkbrenner and Unger 2018).

AGGREGATION

Aggregation combines annual energy use data of a particular demographic to create a larger total, diluting individual-level records (Crandall 2019). Aggregating across groups could occur at the block, neighborhood, census tract, town, or even regional level. Aggregating across time could combine statistics for a single person or entity across sections of time, such as over a month, week, or day (Feinauer, et al. 2016).

By aggregating data across group or across time, the data becomes less granular and more generalized (Kalkbrenner and Unger 2018). However, aggregated data does address privacy concerns; for example, if energy-consumption data associated with 100 homes in a particular area is aggregated, one could still observe the average energy consumption per home in that area or overall energy consumption patterns across time, without seeing the energy consumption for any particular household or entity (Feinauer, et al. 2016).

PRIVACY, USEFULNESS, AND COSTS

The most valuable data is unaltered, granular information about individual energy consumption, which provides third parties with the best insights. Any reductions in the granularity of the data reduces the usefulness of the information; hence, a larger aggregated grouping will be less useful than a small grouping (Feinauer, et al. 2016).

On the other hand, sharing highly granular data also contains the largest privacy risks. Aggregation and anonymization are approaches for reducing the granularity of the data—and thereby reducing threats to privacy—while still providing useful information (Feinauer, et al. 2016). Among other goals, C3 aims with this report to inform legislators about the proper level of aggregation necessary to balance privacy concerns with usefulness of data, widely accepted by different stakeholders around the United States.

The extent to which data is anonymized and aggregated also affects the computational and staff costs necessary to prepare the requested datasets and/or providing the specific energy data reports. The more the information is anonymized and aggregated, the more time and effort will be required to determine how to treat the solicited energy data in order to protect privacy and still provide useful information. Furthermore, there will be more computational time and costs to modify the data. Nonetheless, the more that data is aggregated, the smaller the resulting data set may be, which may reduce data storage and transfer costs (Feinauer, et al. 2016).

THRESHOLDS FOR AGGREGATION

Low levels of anonymization or aggregation provide data capable to tracking closely to the individual decisions faced by customers, allowing third parties to target problems on a granular level or develop models with individual data points. Mid-level aggregation and anonymization on street-wide, neighborhood-wide, or census tract level practices provides useful data; however, the ability to target single home usage practices becomes significantly more difficult. At high levels of aggregation and anonymization—such as at the ZIP code level—third parties can use the data for more general modeling purposes (Feinauer, et al. 2016).

15/15 RULE

A growing number of public utility commissions have been adopting a practice called "15/15" in order to manage requests for aggregated energy data. Under this 15/15 standard, data can only be released if there are at least 15 customers and no one customer makes up more than 15 percent of the data (Crandall 2019).

However, 15/15 may not be the best threshold, since a significant amount of energy data could be withheld from publication. The American Statistical Association Committee on Privacy and Confidentiality has called the 15/15 standard "overly restrictive". Similarly, a study conducted by the New York Public Service Commission found that data from 35 to 80 percent of geographic areas would not have been released under the 15/15 rule for small commercial customers; that number jumps to 80 to 100 percent for large industrial or transportation customers. Additionally, some municipalities find that customers near the 15 percent threshold can be removed or re-added year-to-year depending on the rest of the community's energy usage (Crandall 2019).

There are ways to modify 15/15 to avoid the problem of too much data being withheld. For example, several municipalities in Colorado—including Denver—petitioned for a far less aggregated requirement such as a 4/80 rule during Colorado's data access rulemaking. Colorado's Public Utilities Commission ruled, however, that the more stringent 15/15 rule was necessary to protect consumer privacy (Feinauer, et al. 2016). Colorado ultimately enacted 15/15 along with the condition that data will be rolled up into larger categories, rather than removed, if it violates the 15/15 standard. This means a municipality may receive data for a single category of "business" customers, for instance, where commercial and industrial customers could not be separately provided (Crandall 2019).

OTHER THRESHOLDS

There are many less restrictive thresholds that still protect consumer privacy. For example, the U.S. Department of Energy's Better Building Energy Data Accelerator determined several best practices for data aggregation (see

Table 1) (DOE 2016). These thresholds depend on the number of individual meters included in an energy data request and can range from only two to five meters, rather than the minimum amount of 15 meters requested by 15/15 rule.

Table 1. Summary of Utility Aggregation Thresholds.
The first number represents the minimum number of meters at a property that must be available for aggregation. The percentage figure (where noted) represents the maximum percentage contribution of any single meter to the aggregated energy consumption total. If the meter threshold is not met, or if the percentage threshold is exceeded, the utility will not provide whole-building data access. These thresholds are intended to safeguard customer confidentiality. [1] [2]

Utility Company (Service Territory)	Aggregation Thresholds
Austin Energy (Texas)	4/80%
Baltimore Gas & Electric (Maryland)	5
Clark Public Utilities (Washington)	2
Commonwealth Edison (Illinois)	4
Consolidated Edison (New York City)	2
Eversource (Boston & Cambridge, MA)	4/50%
National Grid (Boston, MA)	4/50%
National Grid (New York City)	4/50%
Pacific Power (Oregon)	5
Peoples Gas (Illinois)	5
Pepco (District of Columbia)	5
PSEG Long Island (New York City)	2
Puget Sound Energy (Washington)	5
Rocky Mountain Power (Utah)	5
Seattle City Light (Washington)	2
Xcel Energy (Minnesota, Colorado)	4/50%

Furthermore, a 2014 study by the Pacific Northwest National Laboratory (PNNL) explored the statistical likelihood that individual energy consumption data could be estimated from aggregated data. PNNL found that the greatest improvements in privacy protection take place as aggregation thresholds increase from two to six meters; if an aggregation threshold reaches six meters or higher, the incremental increase in privacy protection is small compared to the loss of eligible properties (DOE 2016).

TIMING

The timing of energy data collection directly impacts consumer privacy and data usefulness, so the timing intervals for data aggregation must be specified. For example, a larger interval, such as data aggregated at the month level, results in more private consumer information but less useful data (see Figure 1) (DOE 2016).

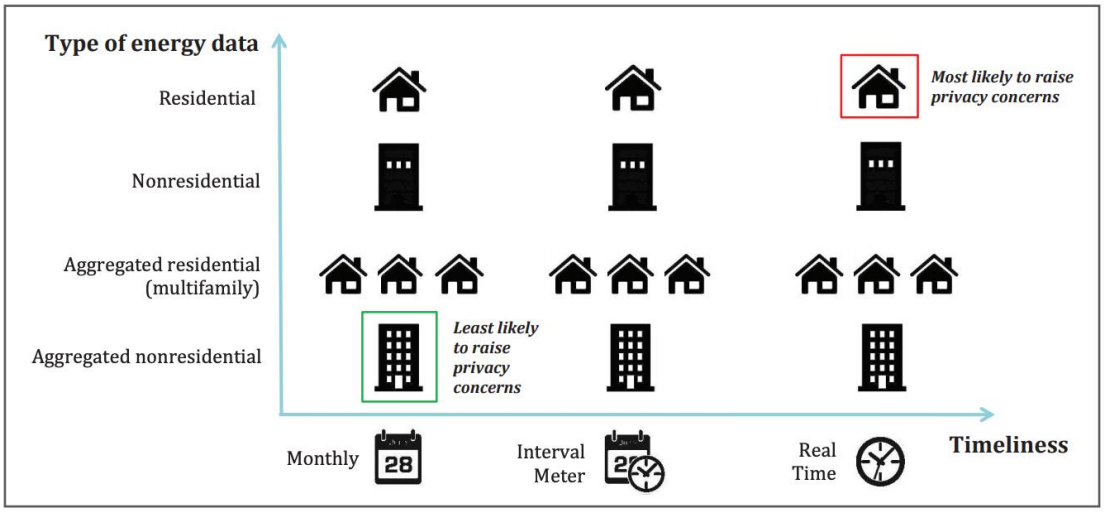

Figure 1. Utility Data Sensitivity Chart.

One method of aggregation would be to combine data from similar slices of time on a weekly basis. For instance, this would require the utility company to average the energy consumed in a household over two-hour intervals and then to aggregate those two-hour slices across the entire week, which provides protection against third parties seeing the consumers' moment-by-moment use of energy. The resulting data set would then show the weekly average amount of energy used during each two-hour-window, allowing third parties to see where to target energy-efficiency measures related to time-of-day use without receiving more detail than necessary (Feinauer, et al. 2016).

STORAGE/SECURITY

Policymakers will have to consider how to ensure that utilities transfer aggregated energy data to third parties securely and that third parties protect the aggregated data adequately. A law or standard requiring that utility companies and data recipients implement security measures for the specific type of aggregated energy-data would be beneficial (Feinauer, et al. 2016).

First, a law could require each of the regulated entities to develop, document, and update specific procedures to meet a reasonable standard. However, leaving this determination up to individual companies may result in company-friendly procedures that vary from firm to firm and potentially go unpublished; some consumer groups may perceive this lack of transparency as threatening the accountability of utility companies. Ultimately, both utility companies and data recipients may want more specific guidance on what constitutes reasonable security measures to help shield them from liability (Feinauer, et al. 2016).

Second, a rule could obligate utility companies and data recipients to implement specific security measures. This would allow the interests of multiple stakeholders to be taken into account, but would be more time-consuming to determine what the proper security measures are. Nonetheless, several national organizations have released relevant data-security guidance documents that would be helpful to review (Feinauer, et al. 2016).

Third, utility companies and data recipients could be mandated to adopt industry standards. For example, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), both independent non-governmental organizations, have developed industry standards for security management systems. This approach would ultimately lead to greater uniformity in the procedures that utility companies and data recipients employ (Feinauer, et al. 2016).

ACCESS

Accurate and easily-accessible building energy data is crucial to measuring, managing, and tracking the energy consumption of buildings. Since collecting the data and/or obtaining written consent from individual tenants is time-consuming and burdensome for building owners, utilities are beginning to offer access to aggregated whole-building data, which combines the consumption of all tenant and common area spaces and avoids privacy concerns for individual tenants (Friedman, et al. 2016).

The best of these programs have (Friedman, et al. 2016):

- Clear, user-friendly instructions for accessing the data online;
- An aggregation threshold (such as four units and above) where individual tenant consent is not required, and standard electronic

forms to be used when specific tenant authorization is still needed (such as in buildings with three units or fewer);

- A standard data format;
- Automated transfer of whole building data directly into benchmarking tools, such as ENERGY STAR Portfolio Manager, via web services, significantly reducing administrative burdens for both utilities and building owners;
- Continual access (e.g. no need to resubmit forms yearly).

The DOE's "Green Button" has been one successful way to access information by providing a clickable button on the utility's website for downloading data. Similarly, Portfolio Manager Web Services Data Exchange allows for utility data to automatically be uploaded into a building owner's ENERGY STAR Portfolio Manager account and can provide automated data transfer on an ongoing basis. These allow building owners or operators easier and quicker benchmarking—which can lead to better energy consumption measurement without compromising access to private billing information. (Friedman, et al. 2016)

C3'S RECOMMENDATIONS

Without a clear standard for fulfilling energy data requests, access to energy data falls under the discretion of utility companies, which often keep data confidential to protect consumer privacy. However, this report has demonstrated that setting a standardized threshold for data aggregation can allow for broader access to energy data without compromising privacy. With publicized data on aggregated energy consumption, municipal governments, nonprofits, and other stakeholders can create more effective programs and prioritize actions that benefit underserved locations. Therefore, a clear rule regarding energy data access is essential to promoting energy-efficiency efforts throughout the Commonwealth of Virginia.

Relevant literature has revealed a dichotomy between consumer privacy and data usefulness, so the chosen aggregation threshold must be able to balance these concerns. A low threshold for aggregation will provide granular data but may infringe on customer privacy; a high threshold will ensure privacy is protected but may not provide useful insights. Therefore, a mid-level threshold for aggregation is best to provide useful, generalized data. When aggregating across groups, this may mean aggregation at the neighborhood or census tract level. Aggregation across time may be best at a specified interval, such as monthly average energy consumption. A mid-level threshold for aggregation would also reduce costs of data storage and the effort required to aggregate the data itself.

As a result of the literature review presented in this report, C3 recommends that the Commonwealth of Virginia adopts the 15/15 rule for data aggregation, which will allow for privacy protection and useful information for third parties. No matter the specific threshold, energy data access must be prioritized to provide proper benchmarking for data-driven policy making. Ultimately, a clear standard regarding energy data publication will greatly benefit the Commonwealth and its residents.

References

Crandall, Kelly. 2019. *Rethinking Energy Data Access: Conquering Barriers to Achieve Local Climate Goals.* Institute for Market Transformation; Urban Sustainability Directors Networl (USDN).

DOE, US. 2016. *Guide to Data Access and Utility Customer Confidentiality Energy Data Accelerator.* US Department of Energy.

Feinauer, Evan, Sean Fernandes, Cole Francis, Alex Gross, Molly Jardine, Nick Oliver, Anna Sims, and Mark Templeton. 2016. *Freeing Energy Data A guide for regulators to reduce one barrier to residential energy efficiency.* Chicago: Abrams Environmental Law Clinic University of Chicago Law School.

Friedman, Julia, Charlie Taylor, Ashley Fournier, Erik Fowler, and Christine Brinker. 2016. *Multifamily Energy Efficiency Retrofits: Barriers and Opportunities for Deep Energy Savings.* Regional Energy Efficiency Organizations.

Kalkbrenner, Astrid, and Jason Unger. 2018. *Energy Consumption Data and Rights to Privacy: Climate change mitigation policy, privacy and the "internet of things" in Alberta.* Environmental Law Centre (Alberta).