

Utility Customer Energy Usage Data Accessibility and Risk Mitigation Study Scope

Summary

The proposed study would provide contextual information to assist Minnesota utility regulators in establishing guidelines for utilities to make customer energy usage data (CEUD) and customer program participation data (CPPD) public or provide it to parties not involved with providing regulated utility service. The study would include:

1. Identification, summary, and analysis of existing publicly available and third-party access to CEUD/CPPD across the U.S., to include a discussion of the legal structure within which the data is made available;
2. An analysis or literature review of potential risks from the re-identification of utility customers from public or utility-provided CEUD/CPPD;
3. A review of privacy protection techniques currently in use to prevent re-identification within the utility industry;
4. A review of privacy protection techniques currently in use within other industries or with other types of data, such as census data, health care data, and “big data” computer science techniques ; and **one** of the following:
 - 5a. A statistical analysis of CEUD/PPD from Minnesota utilities that assesses re-identification risk given different data types, use cases, and differing privacy protection and risk mitigation techniques such as data aggregation and data anonymization; or
 - 5b. A statistical analysis of CEUD/CPPD that determines the validity of existing studies by examining their statistical methodologies, assumptions, and conclusions. Analysis must also include the supporting legal/regulatory framework as it compares to Minnesota, and a discussion of potential implementation strategies.

Study purpose & need

The Minnesota Public Utilities Commission has charged a working group (the CEUD workgroup) to examine issues surrounding utility release of customer energy usage data for purposes of furthering state energy goals while balancing customer privacy concerns. The workgroup is composed of utilities, state agency staffs, environmental and customer advocacy organizations, local government, and other interested parties. While the workgroup has engaged in rich conversations about different approaches to providing parties not involved in providing regulated utility service access to CEUD/CPPD, a robust analysis of privacy risk mitigation measures is beyond the expertise of the workgroup.

In order to provide fact-based recommendations to the Commission, it is necessary to engage experts in the fields of statistics, energy policy and privacy protection to conduct an analysis of practical risk mitigation approaches that can be applied to the “use cases” the workgroup has developed. The intent

of the analysis is to assess the magnitude and likelihood of re-identification of individual data based on particular use cases and mitigation techniques.

While several data aggregation standards for release or provision of utility customer usage data to entities that are not associated with the provision of regulated utility service exist across the country, there is no broad agreement or acceptance of a particular risk mitigation standard or method. Public Utility Commissions in California, Colorado and Minnesota are all currently engaged in conversations about how to balance individual privacy concerns with facilitating greater access to energy data for purposes of advancing public policy objectives. This study, therefore, has the potential to significantly advance energy policy in Minnesota and across the nation.

Study components

- 1. Identification, summary, and analysis of existing publicly available and third-party access to CEUD/CPD across the U.S., to include a discussion of the legal structure within which the data is made available.** Access to customer energy usage data by third parties is currently available in varying forms through utilities, regulators, government agencies, research organizations and other entities. To better understand the existing landscape of access to data, as well as real and perceived risk of greater CEUD/CPD availability, this study will analyze existing data access and availability, the purpose for the access/availability, the basis of the access/availability, including the underlying/supporting legal framework within which data is available, and the documented or perceived risks associated with that access. Examples include data access tools provided by utilities for building- or neighborhood-level data, data shared or published by third-party energy service providers, and data maintained by state and federal government agencies.
- 2. An analysis or literature review of potential risks from the re-identification of utility customers from public or utility-provided CEUD/CPD.** To balance the policy objectives of access to data with privacy considerations, regulators must understand the existing and potential risks that exist from the re-identification of CEUD/CPD. This study should review the landscape of third-party utility data access, and document cases of re-identification that have resulted, the basis/how the data was re-identified, and the consequences that resulted from the re-identification. The study should also consider changing technology, review potential risks that may be realized by increasing access to data, such as layering requests/datasets for the purpose/intent of re-identification. The spectrum of risk should be analyzed based on the categories of use cases presented during the Workgroup process.
- 3. A review of privacy protection techniques currently within the utility industry.** Numerous utilities and some states have studied this issue and established or proposed aggregations thresholds or other methodologies to apply to utility customer data, for it to be provided publicly or to third parties. The study must include a summary of these aggregation methods/thresholds, including the parameters and conditions of its release or publication (customer consent, contractual, etc.), the parties to whom it is/can be released, and the methods employed for access/release/publication of the data. The study must also include an analysis of the aggregation methods/thresholds with

respect to the protection of individual utility customer privacy, the purpose and basis upon which the standards/thresholds were formed, and any supporting legal or regulatory framework and how it compares to Minnesota's legal and regulatory framework.

- 4. A review of privacy protection techniques currently in use within other industries.** Other industries and fields of study, such as census data, health care data, and masking or other “big data” techniques used in the computer science field. These fields already employ sophisticated techniques for maintaining individual privacy and reducing the risk of re-identification of individual data, while providing third-party access to usable data that supports public policy objectives. Examples include demographic research and surveys (US Census) and health care information. The study should look at “big data” masking techniques and specific practices employed in these areas for the purpose of identifying methods that may be appropriate for the utility customer data use cases identified by the workgroup. A comparison of the magnitude and likelihood of risks associated with data types (e.g. health data vs. utility data) should be included when reviewing risk mitigation approaches.

- 5. A statistical analysis of CEUD/CPPD from Minnesota utilities that assesses re-identification risk given different data types, use cases, and differing privacy protection and risk mitigation techniques such as data aggregation and data anonymization.** This study should assess the likelihood of re-identification of individual data given different risk mitigation measures, such as data aggregation and data anonymization. Using Minnesota utility customer data and other publicly-available data, the likelihood of re-identification of individual customers should be assessed through a statistical analysis approach. Possible risk mitigation strategies should assess re-identification risk associated with various aggregation thresholds, anonymization, masking, temporal or spatial averaging or other statistical summarization, and other methods identified by the study team. Risk should be assessed for data interval and granularity identified in workgroup use cases (actual or average monthly by rate class, for example), and data type (kWh, therms, KW, CPPD, etc.). Data groupings analyzed must include whole building (residential and mixed-use), neighborhood or multi-building geographic unit (census block, zip code, etc.), and any differences between rural and urban utility populations.

Study Team

The workgroup assumes that this study will require a multi-disciplinary team that includes expertise in the areas of statistics, demographic analysis, data privacy, and energy policy and law. Additional areas of expertise the workgroup believes may be needed include computer science, health policy and utility regulation.

Schedule

The duration of the study should not exceed 1 year.

The study schedule will include multiple opportunities for feedback by members of the CEUD workgroup prior to the completion of the final report. Study authors should be prepared to meet with members of

the working group at least twice to collect feedback on work products before making a final presentation of the study to the workgroup. In addition, the Study authors are expected to present study findings to the Commission and provide general support to the Commission as it considers the implications of any new policies and procedures associated with utility release of CEUD/CPPD. The workgroup believes the Commission may initiate a procedural comments process following delivery of the final report, which would potentially subject the study to broad public input and scrutiny.

Deliverables

Deliverables will include a final report covering items 1 through 5, one scoping meeting with the workgroup as well as two feedback sessions during the report drafting, and final presentations to the workgroup and the Commission. The final report will address feedback provided by workgroup members, and separately identify how the authors responded to the feedback.