

## **Minneapolis comments on risk mitigation strategies for the release of CEUD to third parties**

4/18/14

Per the request of Judge Pust via her email dated April 10, 2014, Minneapolis staff submits these written comments for the April 18<sup>th</sup> CEUD workgroup meeting regarding risk mitigation alternatives for the release of CEUD to third parties. These comments do not necessarily reflect the official position of the City of Minneapolis or its elected officials, but are the opinion of staff assigned to the workgroup.

**Finding a balance between pursuing state energy goals and protecting individual customer data from re-identification is an important charge of the workgroup.** The Commission understood the connection between state energy goals and access to energy usage data when they directed the formation of the CEUD workgroup. Discussions in the workgroup, as well as comments made to the workgroup from outside parties, demonstrate the need for data at various scales to undertake efforts which support state energy goals. The need for the protection of individually-identifiable energy usage data has also been identified during the workgroup process. We believe that the Commission can adopt risk mitigation measures to be followed by utilities which include a reasonable mitigation of risk, and are appropriate to the level of risk presented by the re-identification of energy usage data.

**Further statistical and background research would benefit the workgroup and the Commission.** The workgroup does not currently have the capacity to analyze aggregation thresholds through a statistical analysis to understand the mathematical probability of re-identification of a customer using a particular risk mitigation strategy such as aggregation or anonymization. If the Commission chooses not to conduct a study on Minnesota data, the completion of a similar analysis currently being conducted by DOE's Pacific Northwest National Labs (PNNL) may be informative. This study will likely be released in late 2014. However, this study only reviews re-identification probabilities associated with whole-building data, and does not review use cases that include data aggregated at a neighborhood or community scale.

**Utilities need guidance on the release of data to third parties.** Even if a separate Minnesota study on risk mitigation measures is conducted in the near future, utilities will continue to receive many requests for data and will need to make decisions about what data to release, and how to do so. Communities and building owners are seeking data to comply with local, state and federal laws and initiatives that are currently ongoing. We suggest that the Commission adopt reasonable risk mitigation guidance for utilities, and consider revisiting the issue if additional useful analysis is developed. The Commission could also reconvene a data access workgroup on an annual or semi-annual basis to review data requests received by utilities, identify any issues associated with those requests, and report back to the Commission.

**The scale of risk mitigation measures should be proportional to the risk of re-identification based upon a reasonable assessment.** The discussion about the risk present from the re-identification of a customer from energy usage data at the workgroup has been almost completely theoretical to date. Potential risks that have been identified include threats to an individual company's business interests from unique knowledge about energy use, the ability to use energy data to target an individual

customer for crime, and the potential for law enforcement to use energy usage data to overreach their authority.

Some of the risks identified do not apply to any of the use cases currently before the workgroup. For example, targeting an individual household for a crime based on usage data would likely require a finer grain of detail than what has been identified in use cases to date (monthly or annual data only). Risks that do not apply to the scope of data presented in the use cases should not be used to identify mitigation measures.

The analysis of risk must also take into account what access to individual customer data already exists, and what results (positive and negative) that access has produced. For example, access to individual and aggregations of small groups of industrial and commercial customer data has long been available to third parties through certain data sets collected by public agencies, like the Energy Information Administration (EIA) (see “Minneapolis Use Case Comments”, March 2014). Given the long history of public access to this data, and the fact that no real-world examples of malicious use of this data have been presented to the workgroup, we have some reason to believe that access by third-parties to even less specific data sets - like monthly or annual usage data aggregated with multiple customers - may present a low level of risk.

**Examples of risk mitigation measures exist in the energy industry and other industries which could guide the Commission, even without a complete statistical analysis of risk mitigation measures.**

Realizing that the workgroup and the Commission will likely move forward before any statistical analysis of risk mitigation measures, like aggregation, has been completed at a national or state level, we believe the Commission can look to other fields for guidance on what constitutes a reasonable level of risk mitigation in the absence of perfect information.

For example, the Census Bureau has been responsible for disseminating demographic and economic data since 1903. The Census Bureau provides very fine-grained levels of data to third parties that include topics which would likely be considered much more sensitive than energy usage data, such as household size and composition, income, race and ethnic background, and in the case of businesses, payrolls and counts of employees. Discussions with staff at the Bureau’s Center for Disclosure Avoidance Research revealed that in most cases an aggregation threshold of three individuals, households or businesses provides a reasonable level of protection against re-identification. In some cases, the Bureau also uses a “percentage of total” method to scrub data (e.g. if an individual business’s payroll is above X% of the aggregated total for a given geography, it is excluded from the data set).

While they could not reveal the exact percentage the Bureau uses for screening, they confirmed it was above 50% of the total. Census data is provided at many geographies, the smallest being a census block (roughly equivalent to a traditional city block).

There are also a number of utilities that have adopted policies, some governed by PUC guidance, others not, regarding the provision of whole-building CEUD to third parties like building owners or managers (see the document “Summary of MPLS aggregation levels identified 3 17 14”). These policies for building-level data generally rely on aggregation of customer accounts as one risk mitigation measure.

These aggregation thresholds range from zero to 15. A threshold of 15 has been adopted by Xcel Energy, and forms the basis for the Commission’s action to form a CEUD workgroup. The next highest aggregation threshold after Xcel’s is 5 customers. Many of these utilities employ additional risk mitigation measures, such as requiring the requestor to verify or state that they are the building owner, or requiring them to identify the meter numbers within the building.

**For data requests beyond the building scale, producing one set of data on an annual basis could reduce risk and reduce expenditure of public and utility resources.** In contrast with requests to utilities by third parties, access to Census data is not subject to repeated, overlapping requests, a source of potential risk identified by workgroup participants. Data is processed once, “scrubbed” using pre-identified risk mitigation procedures (e.g. aggregation), and published for third party access. A similar model could be adopted for community-scale energy usage data in Minnesota. Utilities, the Department of Commerce, or another entity could serve as the data “cleaner” and/or repository and publisher. This means utilities would not be subjected to continuing requests from many different entities. Local governments, non-profits, researchers and other entities would have a central location to locate energy usage data on a specified schedule, streamlining their processes. Risk mitigation measures could be standardized across utilities, and applied uniformly. This type of process would also eliminate the issue of service territories not aligning perfectly with geopolitical boundaries, an issue identified by some utilities during the workgroup process.

**Different risk mitigation approaches should be applied to building-scale and community-scale use cases.** While further statistical analysis of risk mitigation strategies would benefit the Commission, guidance from the Commission to utilities and their community partners is needed in the short term to move forward on advancing Minnesota’s energy goals. Therefore, Minneapolis proposes the following framework, shown in the table below, for the distribution of CEUD to third parties for two main use case types, which apply to many of the specific use cases identified by workgroup members (though not all). Given the availability of other aggregated and non-aggregated data inside and outside the utility industry, we feel that this framework should provide a reasonable level of protection against the re-identification of individual customers in these use case types.

For data beyond the building scale, we suggest that the Commission explore the concept, reviewed above, of collecting data on an annual basis, from all utilities, and distributing it as a single data set(s). Theoretically, this can reduce risk and minimize expenditure of utility and public resources.

<b>Use Case Type</b>	<b>Requirements for availability</b>	<b>Other risk mitigation measures</b>
Building-scale use cases (monthly or annual whole-building usage data, from multiple tenants, provided to a building owner/manager or other entity)	Minimum of 4 customers, aggregated	Standardized requestor verification form and other measures to validate ownership/management interest in the building (meter numbers, etc).
Community-scale uses cases (monthly, quarterly or annual usage or program participation data at the block group level for each of the following customer types: commercial, residential and industrial)	Minimum of 4 customers of any one type (residential, commercial, industrial), with no one customer using more than 80% of the total usage for that customer type in that geography. Geographies can be combined until thresholds are met.	Standardized requestor information form, and agreement to terms of use, submitted to publishing entity (utility, DOC, etc).