

MEMORANDUM

6/6/14

TO: CEUD Workgroup
FR: Brendon Slotterback, City of Minneapolis
RE: Additional details on the development of an Energy Data Center for the processing and distribution of CEUD

City of Minneapolis staff submits these comments in response to the discussion at the 5-16-14 CEUD workgroup meeting regarding the desire for more detail about the proposal for developing a central point for the collection, processing and distribution of CEUD for specific use cases. For the purposes of this memo, this central point is referred to as an Energy Data Center. This memo will present a framework for the functioning of the Energy Data Center, identify what use cases it might satisfy, and briefly discuss approaches for use cases in which the Data Center is not an appropriate approach for granting access to CEUD.

Use cases will be identified by number, consistent with the numbering found in the "Use Cases and Utility Data Availability" matrix dated 5-14-14.

Summary of proposed CEUD access methods by use case

Use Case(s)	Description of Use Cases	Proposed Method to Access CEUD	Notes
1-7	Neighborhood, City, County, Utility Service territory aggregated data	Energy Data Center	Utilities provide energy usage data to Center, where it is processed using appropriate disclosure avoidance techniques. Public data sets are published once per year. No special requests, no overlapping requests.
8-13, 31, 32	Whole-building data from buildings with multiple tenants	Utility-run building aggregation/benchmarking system	New tools developed by utilities (see DOE Data Accelerator). Accessed by owner/manager/third party service provider. May require verification of requestor's identity.
14-30	Single customer, groups of single customers, special requests	Request to utility for manual processing OR access via online customer billing system (Green Button)	Usually requires consent form(s). Govt subpoena, research, and unique requests may have special processes (NDAs, etc).

How the Energy Data Center would work

Collect data from utilities

Once per year, all regulated utilities (or other interested non-regulated utilities) would send individual customer energy usage and program participation data to the Energy Data Center. This data would be considered non-public and protected. It would include usage, applicable program participation information, premise location (address or other more accurate location information such as geocoded point), and class (R/C/I).

Process data using disclosure avoidance techniques

The Energy Data Center would process the data using appropriate disclosure avoidance techniques to develop data sets that could be made public that would present a low risk of an individual customer's usage being estimated by a third party. Techniques would likely include aggregation, anonymization, and rules to protect very large users, like the n-k rule (see Minneapolis comments dated 5-5-14 from more on n-k and other disclosure avoidance techniques in use by federal agencies).

GIS technology would be used to develop tabular and map data that met standards for disclosure avoidance. This should enable the Data Center to provide more fine-grained data than utilities can with their existing infrastructure (some utilities may be limited to zip codes, according to CEUD workgroup discussion).

Publish public data sets

The Energy Data Center would publish annually specific data sets in tabular and map form that could be accessed electronically. As noted in the above table, these data sets could satisfy CEUD requests for use cases 1-7. Published data would be machine-readable. Suggested data sets include:

- Aggregated class usage (total commercial, total industrial, total residential usage) by census block group or census tract for all regulated utility service territory. Could be monthly and/or annual.
- Aggregated class usage (total commercial, total industrial, total residential usage) by political boundaries (city and county) for all regulated utility service territory. Could be monthly and/or annual.
- Aggregated class usage (total commercial, total industrial, total residential usage) by utility service territory (city and county) for all regulated utility service territory. Could be monthly and/or annual.
- Aggregated program participation (total commercial, total industrial, total residential participation) by census block group or census tract for all regulated utility service territory. Could be monthly and/or annual.
- Aggregated program participation (total commercial, total industrial, total residential participation) by political boundaries (city and county) for all regulated utility service territory. Could be monthly and/or annual.

- Aggregated program participation (total commercial, total industrial, total residential usage) by utility service territory (city and county) for all regulated utility service territory. Could be monthly and/or annual.

Neighborhood or community groups and local governments could access these data sets to achieve their goals. Generally, census block group or tract-level data should be sufficient to satisfy all requestors identified during the CEUD workgroup process that are interested in data beyond the building scale. It should be noted that zip code-level data would very likely not be sufficient to meet the needs of these requestors, given their size. Changes in the geographic boundaries of zip codes are also not documented like those of census block groups and tracts, making comparisons over time more difficult.

These data sets would be available through an online map and in tabular format for download. Before publication, all data sets would be screened, as noted above, using disclosure avoidance techniques to ensure that either individually or in combination their publication would present a low risk for the estimation of an individual customer’s usage.

No custom data requests, overlapping requests, or requests to individual utilities

The Energy Data Center would publish specific sets of data on an annual basis for the previous year. Risks associated with multiple, overlapping requests made to utilities would be mitigated, since such requests for community-scale data would not be possible.

This approach would also limit the need for each utility to process requests that fit these use cases, each adopting their own techniques, staff and technical resources. This could be a significant cost savings to rate payers.

An example of tabular data

The table below is a sample of what a table of published CEUD might look like at the Census block group level. This table is for illustrative purposes only, values are not based on real data.

Block Group	Census Place	Count of Residential Customers	Count of Commercial Customers	Count of Industrial Customers	Total residential kWh	Total commercial kWh	Total industrial kWh	Total kWh
2892	Minneapolis	1,000	7	0	4,200,000	58,800	0	4,258,800
564	Minneapolis	1,010	10	6	4,242,000	60,400	100,800	4,403,200
3911	Minneapolis	900	0	3	3,780,000	0	N	N

In Census block group 3911, the “Total industrial kWh” cell is marked “N” because data has been suppressed. Data was suppressed because in that geography, the minimum aggregation threshold for industrial customers was not reached. In addition, “Total kWh” for block group 3911 was suppressed because knowing that total would allow a third party to determine the aggregate total for industrial kWh.

About the Energy Data Center

This proposal assumes the Energy Data Center is operated by an entity that can adequately protect non-public data. This could be a state agency or other public entity, or a non-profit. Examples of entities

that gather private, sensitive data and process it for publishing include the Census Bureau, the Bureau of Labor Statistics, the Energy Information Agency, Minnesota Community Measurement (for health care data) and many others. Approaches used by the Data Center could model these entities, even though the proposed scope of the Center is much more limited.

Funding the Center could be approached similarly to individual utility methods of funding CEUD processing infrastructure.

Summary of May 21st Conference Call

Based on the discussion at the May 16th CUED Workgroup meeting, Xcel Energy organized a conference call to further discuss the details of an “Energy Data Center” as proposed by the City of Minneapolis. The call included at least representatives from Xcel Energy, Centerpoint Energy, Dakota Electric, Minnesota Power, the Large Industrial Group, City of Minneapolis, Fresh Energy, and the Center for Energy and Environment.

The City of Minneapolis reviewed the details of what an Energy Data Center might look like if implemented in Minnesota. Call attendees expressed some cautious interest in having non-building data collected centrally by the state or other entity, recognizing that this may be a cost-saving approach. There was still concern about the organization, costs and other details of the Energy Center, which are beyond the scope of the Workgroup to develop.

In addition, some utilities suggested the data made public should be limited to zip code-level aggregations, rather than anything smaller. This concern did not seem to be based on any aggregation standard rationale, but rather a concern that utilities would be responsible for producing geographic aggregations (rather than the Data Center), which they feel is beyond their current capacity. The point was also made that the conversation about the data center did not resolve any of the outstanding issues about the appropriate aggregation threshold for customer data to be considered safe from re-identification. Xcel Energy also expressed concern that the workgroup had not adequately discussed program participation information as part of the use cases and that there was some concern about that being a data set published by the Data Center without more discussion.