| DOCKETED | |
|---|---|
| **Docket Number:** | 22-MISC-03 |
| **Project Title:** | Energy Data Modernization |
| **TN #:** | 248588 |
| **Document Title:** | Lumen Comments on Energy Data Modernization and Analytics |
| **Description:** | N/A |
| **Filer:** | System |
| **Organization:** | Lumen |
| **Submitter Role:** | Public |
| **Submission Date:** | 1/27/2023 4:27:33 PM |
| **Docketed Date:** | 1/27/2023 |

*Comment Received From: Mariko Geronimo Aydin*
*Submitted On: 1/27/2023*
*Docket Number: 22-MISC-03*

## Lumen Comments on Energy Data Modernization and Analytics

*Additional submitted attachment is included below.*

January 27, 2023

**Vice Chair Gunda, Commissioner McAllister, Jason Harville, and Maggie Deng**
California Energy Commission (CEC)
Letter submitted via CEC e-commenting page:
https://efiling.energy.ca.gov/EComment/EComment.aspx?docketnumber=22-MISC-03

RE: January 13, 2023 Commissioner Workshop on Energy Data Modernization and Analytics, under CEC docket number 22-MISC-003.

Dear Vice Chair Gunda, Commissioner McAllister, Jason Harville, and Maggie Deng,

Thank you for this opportunity to provide comments on behalf of Lumen Energy Strategy, LLC (Lumen, we), in response to the following questions you posed in the above-referenced workshop (workshop):

- What else should the CEC be doing with the data to advance the state's goals of a clean energy future for all?
- What data and analytics can the CEC provide that will have the most value?

I understand from the workshop these questions are in the context of three CEC pursuits:

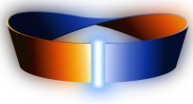
- The CEC's intention to publicly disseminate some form of the customer-level meter data it collects under Title 20 Section 1353 of the California Code of Regulations (meter data)—these data are collected to support forecast and assessment of energy loads and resources.
- The CEC's migration to Amazon Web Services, Snowflake, Tableau, other modern data management tools and services, and, in parallel, "big data" management and analytics.
- The CEC's ongoing efforts to address intra- and inter-agency data management-related barriers to the quality and efficiency of energy assessments and forecasts. I infer these barriers are amplified by explosive growth in the volumes and complexities of energy-related data.

Lumen shares the CEC's desire to enable state and local agencies, resource planners, researchers, and others in California's energy ecosystem to have access to the best available data to support robust and efficient grid planning and operations.

The Lumen team is currently working under a CEC research grant to advance analysis of climate resilience in the state's resource planning models which we're calling the *WARP to Resilience* study. We also are just completing the CPUC's inaugural Energy Storage Procurement Study pursuant to AB 2514 and required under the CPUC's landmark energy storage decision 13-10-040. Consequently, the comments contained herein are from a statewide grid planning and climate resilience perspective.

## Summary of comments

In direct response to meter data-related questions the CEC posed during the workshop my comments can be summarized as follows:



- We strongly encourage the CEC to release some form of aggregated data with limited analytics by the end of March 2023.
- We suggest publishing anonymized and aggregated meter data with some meter-level quality control, aggregated by zip code and sub-sector clusters, also aggregated by hour for interval-level data, and in monthly .csv or .txt flat files, as described in more detail below.

In support of extracting the best value from the meter data, advancing the state's clean energy goals, my additional comments are:

- We suggest the CEC's energy data modernization team review the Energy Storage Procurement Study data management recommendations to the CPUC and consider how they may synergize with the CEC's efforts.
- We strongly encourage the CEC to work with the CPUC to collect and disseminate customer-sited energy storage operating data, which will be essential to understanding the (Title 20 Section 1353) meter data and its applications to the demand forecast.
- We strongly encourage the CEC to consolidate efforts with the CPUC to build a centralized energy storage database of existing and planned installations which, among its many benefits, can help to improve the quality of grid planning models, and can help improve knowledge exchange with local agencies for more effective resilience safety risk management.

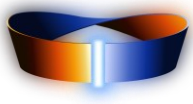
## Meter data

**Urgency of data release.** Our first suggestion is process-related regarding release of the meter data. We strongly encourage the CEC to release some form of aggregated data by the end of March 2023.

This timing is imperative to provide stakeholders enough time to learn from the data and support the 2023 Integrated Energy Policy Report (IEPR) demand forecast with meaningful insights. As the CEC is well aware, the in-depth bi-annual IEPR demand forecast is foundational to all of the state's electricity resource planning efforts. The 2023 IEPR demand forecast, specifically, will feed into resource planning models for the next two to three years including, but not limited to: the state agencies' next round of SB 100 planning process, the CPUC's 2024/2025 long-term planning process, load-serving entities' 2024 integrated resource plans, and the CAISO's 2024-25 and 2025–26 transmission planning processes cycles.

In order for those planning models to more accurately capture climate resilience-related vulnerabilities of electricity service to Californians it is essential that the demand forecast reflect the best available information on weather and climate change trends and volatilities, customer behaviors in response to heat and other environmental factors, sensitivities of distributed energy resources (DERs) or load-modifying resources to heat and other environmental factors. Much can be learned from the granular time profile and from the rich cross section of spatially-granular and customer-granular consumption behaviors the meter data offer.

In order to be useful to the 2023 IEPR demand forecast the meter data need to be fully digested and analyzed in the context of new climate data and demand forecast models this year. That is a daunting and complex work flow. The CEC should not have to bear that burden by itself; it can lean on its experienced community stakeholders. But that opportunity is lost if the CEC postpones release of the data.



I understand from the workshop that several stakeholders have the desire and capabilities to derive meaningful analyses from meter-level data, and that data privacy concerns present stakeholder frictions and potentially legal challenges that are not likely to be resolved any time soon.

Time is of the essence and it is better to publish any form of aggregated data now than to wait until meter-level data or advanced analytics can be released—although these options should not be mutually exclusive. We urge the CEC to release aggregated data with limited analytics now, and more detailed data and analytics later when available.

**Suggested "rapid release" meter data.** Recognizing both the urgency of data release and the need to derive meaningful insights on the temporal and spatial patterns to weather-sensitive demand we suggest publishing the meter data with the following limited analytics, data granularity and aggregations, data fields, and data format.

*Quality control.* Some quality control will be needed at the meter level that cannot be done by stakeholders once the data are aggregated. To avoid major issues with using the aggregated we suggest the CEC performs some quality control analytics such as (but not limited to):

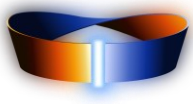
- Inspection of zero values and whether they should alternatively be coded and treated as missing (e.g., affects meter counts in aggregated data)
- A simple screening analysis of the reliability of a particular meter's data, e.g. measured by continuity of data, anomalies in absolute or order of magnitude of units recorded like many values at "9999" or "-1"
- High-level checks of totals against other sources, e.g., by utility, do the profiles add up to annual retail sales independently reported by each utility?

We suspect the CEC has already inspected the data in this way, and likely even more thoroughly, but wanted to highlight its importance.

*Granularity and aggregations.* We suggest the following level of data granularity and aggregations:

- Hourly data, aggregated by zip code and sector-cluster
- 5-digit zip codes (about 2,000 zip codes in California), to provide some geographic granularity that is relatively easy to merge with other public datasets (e.g., the Self-Generation Incentive Program reports solar and storage installations by zip code)
- Clusters within each sector (see explanation below)
- If a cluster analysis cannot be performed, data by rate schedule and, within each rate schedule, by presence of onsite solar PV and/or storage presence (in four groups: no solar or storage, solar plus storage, solar only, and storage only)

*Cluster analysis.* Data by sector do not have sufficient granularity to understand customer and DER weather-sensitive behaviors. Generally, we prefer the data to tell us what electricity use behaviors are present rather than imposing behavioral categories by sector or rate schedule.



A cluster analysis is a statistical method for aggregating large volumes of data into distinct groups that each contain shared behaviors or patterns. For example, suppose two neighbors each have onsite standalone storage but one tends to charge during the day and the other at night. The neighbors are in the same sector and on the same rate schedule, but they have very different behaviors that are relevant to demand analysis. A cluster analysis would put these two neighbors in different groups.

We used a cluster analysis in the Energy Storage Procurement Study to analyze the interval meter-level operations of several hundred individual customer-sited installations. We found this analysis to be an efficient and robust method to reveal distinct storage operating profiles and overall it yielded good insights. We believe a cluster analysis would be a particularly well-suited method for aggregating the meter data.

We encourage the CEC to perform a cluster analysis within each sector, following a framework and methodology developed and demonstrated by a team of researchers led by Lawrence Berkeley National Lab (LBNL). This work is summarized in LBNL's 2018 "Uses for Smart Meter Data Webinar Series" which includes publicly-available slides and webinar videos.[1] The researchers published additional details on their clustering techniques which we encourage the CEC to review and implement.[2]

*Data fields.* We suggest the following data fields:

- GMT hour ending and local clock time hour ending
- Fields corresponding to granularity and aggregations; specific fields depend on long versus wide format of data discussed in the data format section below, but would reflect zip code and sector-cluster (or zip code, rate schedule, and solar/storage category)
- Total hourly volume of metered kWh
- A count of the number of meters in that hour-zip-sector-cluster (e.g. to account for rate switching)

*Data format.* We suggest the following:

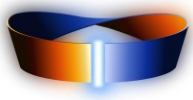
- All in .csv, .txt, or similar standard 2-dimensional (rows and columns), i.e. flat file format
- Divided into monthly files—we would expect each monthly file to have around 175 million rows and 10 columns (if in long format) or 1.5 million rows and 128 columns (if in a widened format but zip codes kept long).

## Insights from the CPUC/Lumen Energy Storage Procurement Study

During the workshop the CEC and panelists raised broader questions around public and cross-agency data access, how to make best use of the meter data, and how to engage with local agencies. We

[1] Please see: https://emp.lbl.gov/publications/uses-smart-meter-data-webinar-series

[2] For example, please see "Load Shape Clustering Using Residential Smart Meter Data: a Technical Memorandum" (2016): https://eta-publications.lbl.gov/sites/default/files/jin_-_loadshape_paper.pdf



believe findings and recommendations in our Energy Storage Procurement Study may be useful in providing a few more pieces to the larger data management puzzle. More specifically, we highlight findings and recommendations around (1) customer-sited energy storage operating data and (2) a cross-grid domain energy storage database that would importantly include two-way information exchange with local agencies

**Recommendations on data management.** The Energy Storage Procurement Study included as part of its core scope an analysis of actual energy storage operations in the 5-year period 2017–2021. To do so, we compiled operating data across all grid domains (customer-sited, distribution-connected, transmission-connected) and use cases (e.g., time shift of renewable generation versus microgrid). We also assessed the overall trajectory of energy storage towards large-scale cost-effective deployment to support state goals and policies. For this, we compiled an energy storage database—again, across all grid domains and use cases.

We were surprised at how difficult it was to collect the most basic information on installed and planned storage resources for a comprehensive view of the portfolio. We faced many of the challenges mentioned by panelists in the workshop. Data management became one of six recommendation themes we presented to the CPUC. We believe the recommendations are in alignment with the CEC's data modernization efforts, and that CEC has an important role in the CPUC's ability to address those recommendations.

We suggest the CEC review and consider these recommendations. The draft report is available at www.lumenenergystrategy.com\energystorage with the final report forthcoming.

**Customer-sited energy storage operating data.** Customer-sited energy storage is distinct from many other DERs in that it is a dispatchable resource and highly modular and flexible in how it is operated. Operations cannot be inferred from other information, like how solar PV profiles can be estimated from solar installed capacity, configuration, and weather conditions.

At the same time, energy storage installations have ramped up considerably over the past 10 years. Customer-sited stationary energy storage capacity grew from 61 MW at the start of 2017 to at least 582 MW by the end of 2021, largely driven by 468 MW of SGIP-funded installations. Recent SGIP reports indicate that energy storage installations outside of SGIP are beginning to scale up. Data on customer-sited energy storage operations is quickly becoming a critical ingredient to understanding the (Title 20 Section 1353) meter data.

Data on ratepayer-funded customer-sited energy storage operations were nearly impossible to get access to, even with the weight of the CPUC behind us. We were eventually able to access data for resources receiving *performance-based incentives* and required to report operations under SGIP. For other SGIP installations and resources outside of SGIP, operating data are only accessible through private arrangements with energy storage developers and installers.

We strongly encourage the CEC to work with the CPUC to (a) develop mandatory data collection, retention, quality control, and reporting of interval-level operations for all ratepayer-funded energy

storage resources, (b) elevate this effort for customer-sited installations, and (c) build channels for accessing these data across study groups, agencies, or even publicly.

**Cross-domain energy storage database.** Finally, the Energy Storage Procurement Study identifies the need for a centralized and comprehensive energy storage database that is broadly accessible and useful throughout the state agencies and its stakeholders. This has important linkages to the state agencies' ability to track resource development and how the meter data fits with downstream resource planning processes. Furthermore, electrochemical energy storage carries serious but manageable safety vulnerabilities distinct from other types of resources on the grid. Safety was also part of the core scope of the Energy Storage Procurement Study and we dedicated a report attachment to the issue (Attachment F). The CEC, CPUC, and a variety of their stakeholders, including local agencies and local emergency responders, need to know exactly where and what type of energy storage resources are built and planned (regardless of grid domain) in order to manage those risks effectively.

We strongly encourage the CEC to merge minds and efforts with the CPUC to develop a centralized and more comprehensive cross-domain storage database that is broadly accessible and useful for the state agencies and its stakeholders.

Please contact me if I can provide any further clarification of our comments and suggestions. Thank you, and the Lumen team looks forward to studying and learning from the meter data.

Written with high regard for the CEC's work,

*/s/ Mariko Geronimo*

**Mariko Geronimo**

Co-founder and Chief Energy Economist
Lumen Energy Strategy, LLC
mariko@lumenenergystrategy.com

Oakland, CA
www.lumenenergystrategy.com
*WBE, MBE, and SBE certified*