

DOCKETED

Docket Number:	18-MISC-05
Project Title:	Disaggregated Demand Data Cleaning Workshop
TN #:	226274
Document Title:	Pacific Gas and Electric Company Comments - Disaggregated Data Comments
Description:	N/A
Filer:	System
Organization:	Pacific Gas and Electric Company
Submitter Role:	Public
Submission Date:	1/7/2019 3:45:52 PM
Docketed Date:	1/7/2019

*Comment Received From: Pacific Gas and Electric Company
Submitted On: 1/7/2019
Docket Number: 18-MISC-05*

Disaggregated Data Comments

Additional submitted attachment is included below.

January 7, 2019

**POSTED ELECTRONICALLY TO
DOCKET 18-MISC-05**

California Energy Commission
Dockets Office, MS-4
Docket No. 18-MISC-05
1516 Ninth Street
Sacramento, CA 95814-5512

Re: Docket 18-MISC-05: Pacific Gas and Electric Company Comments on the Disaggregated Demand Data Cleaning Workshop

Pacific Gas and Electric Company (PG&E) appreciates the opportunity to provide comments on the California Energy Commission's (CEC) Disaggregated Demand Data Cleaning Workshop. PG&E appreciates the continued efforts of CEC Staff to discuss and refine components of the Title 20, Section 1353, Disaggregated Demand Data effort.

PG&E has four general comments regarding the Disaggregated Demand Data Cleaning Workshop in addition to more detailed requests for technical clarifications (Appendix 1) and possible technical design adjustments (Appendix 2):

- I. Third Party Access to Data: PG&E understands that the current emphasis is on finalized data requirements and working out data transfer procedures to transmit Section 1353 data into Amazon Web Services (AWS) for the CEC's use. However, during the workshop, numerous parties, in their comments, implied that they would be accessing this raw data. PG&E notes that the adopted regulation specifically provides for the confidentiality of the customer interval meter data and that, while the CEC will be holding additional discussions on how the data could be aggregated, third parties will not have access to the customer private information. As a result, PG&E asks the CEC to clarify the following:
 - a. Expectations management for third parties – More guidance regarding data access for all parties other than the CEC should be provided. Data access by third parties must be limited to data which has been minimized, aggregated, and anonymized, which is consistent with customer privacy protections, including privacy requirements under the Consumer Privacy Act, the California Information Practices Act, and Title 20 itself.
 - b. Data access – More clarification regarding the process for determining who at the CEC will have access to the data, how those personnel will be qualified to receive data access, and how that data will be accessed is needed. PG&E notes that any data access granted to parties other than CEC should be consistent with CEC's Title 20 purposes, Institutional Review Boards (i.e., IRB approval), data security, and

study/analysis publication. PG&E would like greater clarity regarding the timeline and methodology of this process.

- c. Data sharing standards – More clarity on the process for determining the data minimization, aggregation, and anonymization standards. A better understanding of the timeline and methodology of this process is needed.
- II. Project Timeline: CEC noted in its workshop on December 11, 2018, that the timeline for data delivery may be pushed past the current February 15, 2019 deadline. PG&E also has concerns about the current Section 1353 data timeline, especially as the data requirements and the data transfer method have not yet been finalized. PG&E notes the following specific concerns which may affect its timeline for delivering Section 1353 data:
- a. Data transfer procedures – PG&E would like more guidance with respect to the timeline for constructing and verifying data transfer procedures, including the transfer of test files and the possible use of interim data transfer procedures while the AWS environment is being constructed and validated.
 - b. AWS data transfer – PG&E would like to have a better understanding of the options for moving data into the AWS environment and how PG&E’s existing data transfer tools might be used.
 - c. Data schema requirements – PG&E would like more guidance on the timeline for finalizing the data requirements and formats, including the verification of test data files for content and format.
 - d. Formalizing timeline – PG&E requests that the CEC formally issue a letter extending the due date for provision of the Section 1353 data.
- III. PG&E’s population of Section 1353 data will be “as is” in PG&E’s data reporting systems at the time of extraction. Based on the data elements outlined in Section 1353, not all of these will be available due to the nature of PG&E’s data reporting systems. Data are stored in various systems, and there may be minor discrepancies between the systems depending on how the data are stored and reported. Because of the quantity and variety of data involved, there will be a very small percentage of cases where the data may not precisely align between different tables. This is not expected to have any significant impact on any data analysis projects.
- IV. In the December workshop, it was suggested that the three tables regarding energy efficiency measures could be reduced to one simpler table that will allow the CEC to link to the California Energy Data and Reporting System (CEDARS). This would be a simpler and better approach, although PG&E is also able to support the current design. PG&E suggests a table consisting of the existing “ServiceAccountId” field and a matching “ClaimNumber”. The “ClaimNumber” field could then be matched to CEDARS.
- V. Finally, there are also some elements of the CEC’s detailed technical approach for Disaggregated Demand Data where clarifications or adjustments may be required. Clarifications are listed in Appendix 1 and other issues which may require adjustments are listed in Appendix 2. Please refer all comments and questions regarding the technical clarifications to the PG&E technical point of contact, Robert Lucadello (Robert.Lucadello@pge.com).

PG&E appreciates the opportunity to provide comments on the Disaggregated Demand Data Cleaning Workshop and looks forward to providing input and working with the CEC on the establishment of their Disaggregated Demand Data repository.

Sincerely,

/s/

Valerie J. Winn

Appendix 1 Requests for Technical Clarifications

CEC Table	CEC Field Name	Clarification Desired
Multiple	Multiple	Please provide examples of date/time combinations in the desired UTC format.
Multiple	Multiple	Please verify that the three-character time zone identifier is either PDT or PST.
Multiple	Multiple	<p>When a utility identifier has to added to the data in a field:</p> <ul style="list-style-type: none"> - Please verify that "PGE" is an acceptable utility identifier - Please clarify whether the identifier should be prepended or appended - FYI: PG&E is agnostic about whether to add utility identifiers to specific fields or to add a field identifying the utility in table rows
Multiple	Multiple	For Boolean fields where an affirmative response = "True", is null assumed to mean "False" or should we populate "False"?
B02 BillingConsumptionElec, BillingConsumptionGas	NaicsCode	Should PG&E handle NAICS for residential customers the same way it handles non-residential (by expanding out to 6 digits by adding zeros)?
Multiple	ServicePointId	The directions are to "Concatenate with PremiseId to form a unique key within each utility's data" but our SPID are unique in our system already. Given this, does PG&E still have to concatenate with its Premise ID?

CEC Table	CEC Field Name	Clarification Desired
G08 EnergyEfficiencyParticipation	EndDate	Does the data definition imply that this can be blank as long as the claim date is populated?
G08 EnergyEfficiencyParticipation	StartDate	Does the data definition imply that this can be blank as long as the claim date is populated?
G08 EnergyEfficiencyParticipation	ClaimDate	Does the data definition imply that this can be blank as long as start date or end data is populated?

Appendix 2 Technical Issues

CEC Table	CEC Field Name	Tech Issue
Multiple	CommunityChoiceAggCode	Suggest adding a utility identifier to avoid potential duplicate key issues in the CommunityChoiceAgg table.
Multiple	DirectAccessCode	Suggest adding a utility identifier to avoid potential duplicate key issues in the DirectAccess table.
Multiple	RateScheduleCode	<ul style="list-style-type: none"> - Suggest adding a utility identifier to avoid potential duplicate key issues in the RateSchedule table. - In standard reporting work, PG&E treats CARE and FERA as separate flags rather than trying to parse them out of rate codes. These are potentially important flags for analysis work and different utilities may use different methods to incorporate them into rate schedules, so we suggest considering splitting them out as flags.
Multiple	EeProgramCode	Suggest adding a utility identifier to avoid potential duplicate key issues in the EEProgramCode table.
Multiple	EeMeasureCode	Suggest adding a utility identifier to avoid potential duplicate key issues in the EEMeasureCode table.