DOCKETED	
Docket Number:	18-MISC-05
Project Title:	Disaggregated Demand Data Cleaning Workshop
TN #:	226400
Document Title:	Steve Uhler Comments MISC-18-05 Data minimization, aggregation and anonymization
Description:	N/A
Filer:	System
Organization:	Steve Uhler
Submitter Role:	Public
Submission Date:	1/28/2019 10:20:53 PM
Docketed Date:	1/28/2019

Comment Received From: Steve Uhler
Submitted On: 1/28/2019
Docket Number: 18-MISC-05

MISC-18-05 Data minimization, aggregation and anonymization

Additional submitted attachment is included below.

Disaggregated Demand Data Cleaning Workshop, CEC Docket Number 18-MISC-05

Data minimization, aggregation and anonymization

To minimize is to reduce to smallest amount possible.

To aggregate is to form a whole by combining several elements.

To anonymize is to make not identified by name.

The heart of the data is meter records. Trillions of these records are and will be waiting to be processed. These records can be easily minimized, aggregated, anonymized and processed into actionable information.

By aggregating anonymized meter records, timely actionable information will be found most efficiently while maintaining confidentiality of customer information.

Anonymized meter reference IDs can be foreign key linked to supporting information through decode tables.

Decode tables for each reporting entity where the true meter ID is linked to a randomized ID can be used to produce anonymized meter records. Each reporting entity would be supplied with enough non consecutive IDs to cover one and a half times the number of meters they expect to have. This will add to the difficulty in identifying who's meter data is being made available for study.

This method easily supports any unit of measure. Data structure, reports and charts can be reused for kWh, Peak kW, therms, etc. Bills Of Material (BOM) can be used to show meter to meter relationships.

Adhering to the requirements of 1342 are made simple by the anonymization process. Presenting sufficient detail to allow replication of the results by the Commission staff and by other experts in the field will be easy and secure. Keeping customer identifiable specific information secure as arcane knowledge on a need to know basis will have little effect on the quality of actionable information available for setting energy policy.

These records then can be linked to aggregation tables such as county or region. The aggregations if large enough can make it very difficult to identify the customer associated with meter data while providing actionable information. Access to the meter data can be of great benefit to local planners who can apply the knowledge in the real world and not just analog modeling.

As to PG&E's request for greater clarity regarding the timeline to apply the above methodology, many of the queries and reports already exist, I use them daily to study CAISO data. See this example of charts with links to source data not unlike meter data, Week 36, Thermal Production Dashboard for CAISO. The CAISO source data is a text file that has changed format over time, the method used easily adapts to such changes. I have produced decode tables that contain 100 million meter IDs, made sample meter data tables with 50 million records. Wrote queries to find events of interest such as large demand spikes, zero value patterns and synchronous events. Most of the queries are ones I have used for many years in many different places with much success.

Looking forward to getting started with real meter data.

Steve Uhler
sau@wwmpd.com

Links to meter data some studies on processing meter data:

https://aceee.org/files/proceedings/2012/data/papers/0193-000375.pdf

https://www.google.com/url?
sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=2ahUKEwiUgc_034zgAhVrja0KHc
LxAlgQFjADegQIARAC&url=https%3A%2F%2Fwww.ons.gov.uk%2Ffile%3Furi%3D%2Faboutus
%2Fwhatwedo%2Fprogrammesandprojects%2Ftheonsbigdataproject
%2Fmodellingsampledatafromsmarttypeelectricitymeterstoassesspotentialwithinofficial
statistics_tcm77-408756(1).pdf&usg=AOvVaw2n2Ts469EGmpkL-YpvbuYb