

February 26, 2010

**Re: Data Request From Energy Division Dated February 10, 2010, Regarding All Distributed Generators and Transmission and Distribution System Information**

Dear Ms. Baker and Mr. Franz:

We are in receipt of your data request dated February 10, which asked for many categories of information about PG&E's transmission and distribution system, and about all "distributed generation technologies ... 20 MW or less that are currently operating and interconnected at any voltage." As you know, PG&E provides large amounts of information to Energy Division on a regular basis, and will continue to do so. However, the burden of assembling and producing the information requested in this particular request could be enormous, and we do not understand how it could be useful to Energy Division. There are also a number of other issues with this request. For these reasons, we would like to discuss this request with you in more detail. As you requested, the following is a discussion of some of the issues associated with this request.

**General Discussion and Objections.** We have the following issues with and objections to this data request:

**Confidentiality and Security Issues.** Your request asks for large amounts of information about customer equipment and payments, including names, addresses, and account numbers. We understand that you may intend to turn this data over to Black and Veatch (B&V) or possibly others. It is essential that customer confidentiality be protected and preserved. If any of this data is produced, adequate contractual protections must be negotiated and signed providing that B&V or any other contractor will preserve and protect customer confidentiality. Similarly, the request seeks a variety of maps and other system information that could pose risks to the security of essential systems if disclosed. Adequate protections must be negotiated assuring protection against disclosure of this data. Moreover, the agreement must address what commercial and other use the contractor may make of this data and when it will be returned or destroyed.

**Ambiguity.** The Commission has held that it does not consider facilities interconnected at transmission voltage as Distributed Generation (DG). D.09-08-026, p. 18. However, in your request, you appear to view any project smaller than 20 MW as DG. What is your definition of DG? In addition, in your cover note, you only ask about projects now operating, but in the attached spreadsheet you ask about projects in the queue. Are you asking for data on all generators, even if they were interconnected many decades ago?

**Burdens of Assembling The Data.** PG&E has over 38,000 small generators interconnected to its system, and adds nearly 800 new projects a month. We also have a vast amount of transmission and distribution system information. This data associated with these topics is spread among multiple departments, and is stored in multiple formats. Some records are on proprietary databases, some are on simple spreadsheets, and some data exists only in paper records. Some of the generators have been on line for over 30 years. Trying to even find all the requested data would be an enormous job, and for some categories it does not exist. Inputting the data into the requested spreadsheet would be even worse. For example, question 1 alone has 50 different fields for each of 38,000 projects, **for a total of 1.9 million requested fields.**

**Duplication Of Other Energy Division Data Requests.** In recent months, we have provided vast amounts of data to Energy Division concerning net metering issues, the California Solar Initiative, the Self Generation Incentive Program, and other related data. This request appears to overlap with and duplicate portions of those requests. These are discussed in more detail below.

**Relevance and Value of This Data.** We don't understand what Energy Division plans to do with this data, or what useful results can be achieved from it. For example, what use will be made of the customer name or address? Nor do we think the information would permit a meaningful analysis of places to site distributed generation. Among the many issues related to such an analysis is the fact that transmission and distribution systems are not static. New lines and system loads are added on a regular basis, and electrical loads can be switched among various transmission and distribution lines as system needs dictate. A line that might have excess capacity one day may be near capacity the next, depending on real time system operations. Moreover, as explained above, new DG systems are being added on a daily basis, and the generation from larger plants (not the subject of this request) is also being added and studied on a regular basis.

**Need For Further Discussions.** We would be happy to discuss this request with you and try to understand your goals, and how we can help achieve those goals.

Additional discussion of the three specific questions in your request follows below.

### **Question 1**

This request will require PG&E to pull together data for over 38,000 customers that operate generation. PG&E's various data sources are not integrated, as PG&E has had no specific business need to do so. For example, we have an interconnection database administered by the Generation Interconnection Services Department. Separately, a different department has detailed records concerning requests for Self Generation Incentive Funds or California Solar Initiate Funds. We also have separate administration of various RPS power purchase agreements, as well as records on standby service for generators. Most of these different data sources in different departments are not linked. Moreover, PG&E has not managed all the incentive programs and so does not have

complete incentive data. An example is the incentives administered by the California Energy Commission.

The data concerning generator interconnections is not consolidated and is in several different data sets. The distribution and transmission system data are also in different data sets. Actual interconnection and study costs are stored in the accounting database by site and would require manual retrieval. We do not store the distance of generators from substations. We have not yet added queue information for wholesale projects to our database. Queue information for transmission projects is kept in the ISO database. As noted above, some of the 20 MW and under generators are linked to our Transmission system.

Moreover, within the past year or two, PG&E has provided a large quantity of related data in response to various data requests from the Energy Division to support a variety of DG-related efforts, including Measurement and Evaluation (M&E) of the Self-Generation Incentive Program (SGIP), M&E of the California Solar Initiative (CSI), the Net Energy Metering (NEM) Cost-Benefit analysis, periodic reporting to the CPUC, and others. Examples of data that have been provided for those purposes, and which may overlap with the types of data requested in this particular data request efforts, include:

- Interval (15-min) generation data for PV systems receiving Performance-Based Incentives under the CSI program
- Interval (15-min) generation data for DG systems that received incentives from the SGIP
- Connectivity data and hourly loading and for several feeders
- Load research data for all NEM customers
- Billing data for all NEM customers
- Billing data for select SGIP participants
- Customer information data associated with the above facilities (protected by detailed confidentiality agreements).
- All NEM and non-NEM interconnections projects (includes NEMBio, NEMMT, NEMW, VNEM, NEMFC)

In addition, until the project is officially interconnected we cannot be sure that the proposed generator system information will be final. Also, is this a request to list all proposed applications including ones with incomplete information?

## **Question 2**

Like question 1, responding to question 2 will require a significant amount of work. To start, PG&E will need to collect data from various data sources (circuit connectivity data bases, load-growth projections, SCADA systems, etc.) that are not linked together. To give an idea of the volume of data, one of the sources would be a data-dump involving

approximately 3,000 line items in an excel spreadsheet (PG&E has about 3,000 distribution circuits). Another source would be approximately 250 excel files (with multiple worksheets) containing other data like peak loads and circuit ratings. Attached is the spreadsheet for question 2 which PG&E has annotated to indicate areas where data is more or less readily available as compared to other areas.

One potential way of compiling this information would be for PG&E to provide the data from the various sources to Energy Division or B&V personnel so they can compile the information into one spreadsheet. Someone will be required to take the data, circuit-by-circuit (all 3,000 of them), from each of the 250 excel files and align it with the 3,000 line item data-dump previously mentioned. Data from a SCADA historian would also need to be pulled for approximately one to two-thousand of banks and circuits. As discussed above, we are not sure what budget Energy Division or B&V have for this work, or what it is to be used for.

It is also worth noting that some of the excel spreadsheets previously mentioned contain customer specific information. PG&E would need to either redact the information or provide it under adequate confidentiality protections. Manually redacting the information will take time. Also, If PG&E needs to print out this information and redact customer information it will take us significantly longer.

The information for 2009 minimums and peak load is extremely burdensome. The data is in a data historian and the process to get the data is tedious and will require a significant amount of effort to retrieve for all banks and circuits where data is available (which is roughly estimated at one to two-thousand of banks and circuits).

An additional element to consider is the dynamic nature of electric distribution systems. Electric distribution systems change as a result of adding new capacity, connecting new customers and performing other work requested by third-parties, replacing aging assets, connecting new distributed generation facilities, reconfiguring existing circuits by switching and so on. Clearly, distribution systems change over time and the data PG&E provides in response to this request will only represent a “snap-shot” in time.

**Question 3:**

Responding to the question 3, which basically requests information regarding distribution system maps, does not appear to difficult. However, PG&E would like to discuss this question with Energy Division and Black & Veatch before responding in order to ensure common understanding of the terms used in the questions. Moreover, if the step that follows is a request for maps, then a further discussion will be needed, as issues of burden, relevance, security, and confidentiality will follow.

Distribution Data

Number	Substation Name	Substation ID	Address	City	County	Zip	Latitude	Longitude	Transformer ID	Design Capacity	Transformer Primary Voltage Rating	Transformer Secondary Voltage Rating	Short Circuit Current at Transformer Secondary	Feeder ID	Feeder Type (Over Head, Under Ground)	Length of Feeder	Design Capacity	2009 Peak Load	2009 Min Load	2009 Peak Load	2009 Min Load	2009 Peak Load	2009 Min Load	2009 Peak Load	2009 Min Load	2009 Peak Load	2009 Min Load	Peak Load - Type of Reading	Peak Load - Sampling Rate	Minimum Load - Type of Reading	Minimum Load - Sampling Rate
		(If different than Substation name)					Decimal format			(MVA)	(KV)	(KV)	(KA)	Include Spares	(OH, UG)	(ft)	(MW)	All Hours	June 21 to Sep 20 (5 am - 7 pm) (MW)	Sep 21 to Dec 20 (6:30 am - 5 pm) (MW)	Dec 21 to March 20 (7 am - 5 pm) (MW)	March 21 to June 20 (5 am - 7 pm) (MW)	(SCADA, Analog, Human)	(i.e. 10 sec. annual)	(SCADA, Analog, Human)	(i.e. 10 sec. annual)					
Bank	Y	Y	Y	Subject to confidentiality provisions					Y (PG&E interprets as bank number)	Y (PG&E interprets as transformer nameplate rating)	Y	Y	Y (different data source)	NA for Subst	NA	NA	NA	Y (different data source, manual process)	Y (Partial -- all banks do not have SCADA, also different data source, potentially very burdensome to sort through data historian)								Y	Y	Y	Y (if min load data available)	
Circuit	Y	Y	Not applicable (NA) for a circuit	NA					NA	NA	NA	NA	NA	Y	Y (PG&E can provide total cct miles of OH and UG)	Y	Y (PG&E interprets as circuit outlet rating as limiting factor)	Y (different data source, manual process)	Y (Partial -- all feeders do not have SCADA (less than 40-50%), also different data source, potentially very burdensome to sort through data historian)								Y	Y	Y	Y (if min load data available)	

Data is more or less readily available without too much difficulty

Data is available, but: 1) may not be complete; 2) will need to come from different data sources (manual process); 3) there are potential confidentiality concerns; or 4) some combination of these factors