FILED
07-07-09
11:50 AM

Attachment C

Energy Division Straw Proposal on EM&V Issues
for June 17th EM&V Workshop

F. (Updated 6-23-09) Data Availability, Data Quality Improvement, and Reporting

The Energy Division believes that the quality of the program tracking data[1] and information provided to the Energy Division and their EM&V contractors to support the evaluation efforts needs considerable improvement. The program tracking databases contain detailed information on program participants, specific energy efficiency projects, and specific energy efficiency measures rebated by the program. In order to measure and verify energy impacts, the Energy Division and their EM&V contractors depend on the program tracking data to design sampling plans and as a baseline reference for updating IOU claims with ex-post EM&V results. Many problems discovered with using the program tracking data records as a basis for verifying program impacts are documented in the Energy Division's Interim Performance Basis Report[2] and 2006-2007 Verification Report.[3] Energy Division also solicited responses from the evaluation contractors regarding they key issues they faced with data quality and availability and are cross referenced here to highlight some specific issues. Key issues with data availability, data quality and reporting are summarized below.

- In general, the data and information provided by the IOUs is useful for conducting evaluation activities, but in many cases repeated follow-up data requests are necessary as critical data are sometimes initially missing. (Examples of these issues are highlighted in the Small Commercial , Commercial Facilities ; and Major Commercial memos.)

- While utility staff appear to make a good faith effort to fulfill EM&V data requests, there are some critical delays that have a cascading impact on the timing of EM&V field work and Energy Division's reporting schedule. (Examples can be found in the Commercial Facilities memo; Small Commercial memo)

- The IOU program tracking systems have many limitations which make their use in evaluation time-consuming and labor-intensive: (The issues below are presented in more detail in the Local Government Programs memo, and the Major Commercial / RCx memo as well as several of the other memos)

[1] The terms "program tracking data", "program tracking databases" and "program tracking system" are generically used here to refer to the elementary underlying information on projects and measures installed and rebated through the utility energy efficiency programs. Each utility maintains different systems and procedures for managing program related data.
[2] Available at http://www.energydataweb.com/cpuc/
[3] http://www.cpuc.ca.gov/PUC/energy/Energy+Efficiency/EM+and+V/081117_Verification+Report.htm

- Lack of consistent measure descriptions and naming conventions substantially increases the time and expense required to segregate measures into common measure groupings.
- Program tracking data lacks unique key fields to enable tracking of a line item or project throughout the program cycle and identification of records that have changed from one data submission to the next.
- Program tracking data parameter estimates for some line items are incrementally updated, sometimes spanning several quarters or program years, and are often not accompanied with notification, a clear rationale, or updated workpaper. This means the quarter and year of each portion of the claimed savings cannot be tracked or easily understood, which is highly problematic when developing samples.
- Program tracking systems are not consistent across the utilities in terms of content, format, and quality, making it difficult to perform cross-IOU data management and analysis.
- Significant time must be spent conducting reviews, quality control, data cleaning and consolidation each time a new program tracking dataset is delivered.
- Project baseline and measure base case conditions are rarely included in the program tracking systems. Assumptions must therefore be made regarding pre-existing conditions for most project sites. The precise type and conditions of existing equipment should be recorded as a program requirement before new installations are made.
- The lack of complete location data makes assignment of appropriate climate zones and locating sample cases time-consuming and difficult.
- Measure units (e.g. lamp, fixture, refrigerator, BTU, ft^2 of floor space, etc.) are poorly reported across all IOUs. Units are commonly left blank, or are reported on a "kWh saved" basis or simply as "Unit". This greatly hampers analysis, as there is little insight into how the line item savings are calculated.
- References to DEER measures are inconsistent across all IOUs, or are absent. Sometimes only generic DEER measure IDs are given, sometimes DEER 2005 Run IDs rather than DEER 2008 IDs are given. In many cases, the savings value given for a measure does not match the DEER savings value for the Run ID that is referenced.
- Many key data are not collected or entered into the program tracking systems, or are poorly specified, such as program delivery mechanism building type
- Sometimes measure quantity is equal to "0", or it is the equivalent of the net or gross savings. In either of these cases, we gain no knowledge of the actual quantity that was installed.
- Frequently, SEMPRA and PG&E report zero savings for a line item; however, installed quantities and rebate payments are reported.
- Key upstream measures, such as CFLs, are tracked by quantity shipped rather than quantity sold or installed.

- Matching E3 calculator data with program tracking data remains a significant challenge. A number of data elements required by the E3 calculator are only presented in the E3 line items and not in the program tracking system, making reproducing the cost-effectiveness and savings results difficult. (This issue is presented in more detail in the Major Commercial/RC X memo, and noted in the Small Commercial memo.)

- The existence of multiple different reports and databases (monthly reports, quarterly reports, cost-effectiveness spreadsheets, utility program tracking databases, subcontractor tracking databases and spreadsheets, annual reports, etc.) has substantially increased the data processing and interpretation work for

EM&V contractors, Energy Division staff and IOU staff and has sometimes led to multiple parties using dissimilar assumptions regarding program participants and portfolio costs and impacts. (This issue is presented in more detail in the Major Commercial/RC X memo.)

- In some cases, large custom project details needed by the evaluation are stored as paper files only, or do not have supporting electronic materials. In other cases, electronic files are made available only as PDF image files that are not text searchable. (This issue is presented in more detail in the Commercial Facilities, and PG&E Ag memos)

- Non-DEER measure workpaper quality and availability varies. (This issue is presented in more detail in the Local Government Programs Major memo.)

- For programs targeted upstream or midstream, the program tracking database does not always capture information about the participating upstream or midstream market actors. In addition, installation site addresses and contact information is not always captured in program tracking databases. This information is critical for verification and evaluation of upstream and midstream program activities. (Examples of missing data can be found in the Small Commercial memo)

Initial Recommendations

The IOUs, Energy Division, and EM&V contractors will collaborate as soon as possible to examine, improve, and streamline the management and sharing of program tracking data to facilitate the earliest practical transition to a combined accomplishment reporting and EM&V tracking quarterly IOU reporting submission.

Upgrades to the program tracking systems shall be required and shall not be done without Energy Division review and approval, and shall include EM&V contractor input. Upgrades to the program tracking systems shall be uniform across the utilities and shall be designed to contain all of the data needed for all EM&V and reporting purposes and required to use cross-IOU standardized fields, standardized naming conventions, standardized data elements and data validation rules consistent with DEER. The IOUs shall only deliver program tracking data that conforms to these standards.

Data tracking requirements for all upstream programs will require special attention to make sure that the kinds of data that are required for verification and evaluation are tracked.

The IOUs will be required to document their data quality control procedures and work with the Energy Division and EM&V contractors if improvement to those procedures are needed.

EM&V data request response times are a key factor in keeping the Energy Division's EM&V performance reporting work on schedule.

The IOUs shall develop a system for making all energy efficiency project documentation readily available to the Energy Division and their EM&V contractors in electronic format wherever possible. Quarterly reporting for custom measures shall include this documentation for those measures determined by Energy Division, during the non-DEER measure review and approval process, as requiring such documentation.

The Commission should adopt just one form of data to be conveyed on a regular basis from the IOUs to Energy Division. This should be the most disaggregate form, i.e., program tracking data and detailed expenditure data. The Energy Division recommends that program tracking and financial data be submitted to the Energy Division on a quarterly basis along with outputs from the new Energy Division managed CE Tool and simplified, but highly specified, narrative reports. Additional reporting detail may be required when the new RRIM and the 2009-2011 portfolios are adopted.

Appendix D

Data quality memos prepared for the Energy Division by EM&V contractors

----------Commercial Facilities Data Memo----------

Date: May 12, 2009
From: David Diebel
To: Kay Hardy
Cc: Don Dohrmann
Subject: **Problems Obtaining Program Data from Utilities**

EEGA requests for information on sampled projects submitted by ADM Associates, Inc., included the following language: "Please provide a copy of the contents of each project file listed in the attached Excel file, including but not limited to documentation of measures installed, cut sheets, correspondence between the IOU and the participant, engineering calculations, energy simulations (with inputs/outputs), etc." While we did receive paper files for projects, the spreadsheets and other electronic documents used to generate tables and other information in the paper files were not provided as requested. In some cases, they were provided at a later date upon further request by ADM personnel.

Project files were often mailed later than the EEGA request due date. This was due to IOU resource constraints in dedicating staff to photocopy files available only in paper format. Sometimes there was notification from the IOU that the data would be delivered late; at other times, there was not. Additionally, some project files were not delivered in a timely fashion because they needed to be "requested from file storage."

----------PG&E Agriculture and Food Processing Data Memo---------

To: Kay Hardy Date: May 11, 2009

From: Fred Coito

Subject: Data Availability and Quality Review for 06-08

Following is a brief summary of data availability and quality as they pertain to the PG&E Ag-Food evaluation. We have organized the discussion according to the topics outlined in the memo that was issued on April 24, 2009.

AVAILABILITY OF DATA

1. Were the right types of information available from the program or utility to answer the research questions at hand? (Please note variations by utility, and/or specific requests.)

For PG&E, our main source of data, the project files that we have been working with generally provide a good starting point for our analyses. However, they do not always provide a complete documentation of the project.

Since files are hardcopy, or scanned versions of hardcopy, it is not possible to do electronic searches of the files, which slows things down.

2. What was missing (if anything) from data requests (either not available or provided only partially) or general information that was needed to conduct a thorough evaluation?

Our main data requests of PG&E have been for project files and associated materials. We ask for printed materials and all electronic materials associated with each application. The following language is typically used:

> "Please find attached the list of applications that are included in the evaluation sample. As part of our evaluation process we are asking for copies of:
>
> - Any and all paperwork that are relevant to these applications
> - Any and all electronic materials that are relevant to these applications"

Thus far, we have only been getting hardcopies (or sometimes scanned copies) of the project files (1st bullet) and none of the electronic files (second bullet).

We almost always would like the engineering analysis files such as those completed by the review contractors and applicants, but we have not received these during the first round of a data request. We have a strong preference for the XLS files but at least printouts of them or SPC results.

Since we don't get these files from the initial request, we have to go to back to the appropriate parties (as identified in the hardcopy files) directly for this information or we have to submit additional for formal data requests to PG&E. These additional efforts that are required to get at supplement information costs us billing time and calendar time.

3. How did this missing data/information affect your ability to conduct the evaluation?

Missing data has mainly slowed us down, as we need to follow up with the appropriate parties to get this information.

4. What recommendations can you provide that would improve the type of data available in the future.

We suggest that all relevant electronic data (spreadsheets, models, etc.) that are referred to in a project files be stored in a central location, similar to how the hardcopy files are stored, but electronically. Then when a data request comes in, PG&E can readily supply all the hardcopy and electronic files.

Also, it would help a lot if the files came in text PDFs instead of hard copy or PDF images, so we could search for words and numbers. This would involve the IOUs converting to an electronic-based (versus paper-based) filing system.

Finally, it would be good to see a one-page or two-page summary sheet in each project file that summarizes the measures, final IOU savings, best contacts, and key documents that are relevant to the project analysis.

QUALITY OF DATA

1. Was the data complete and accurate (to the best of your knowledge)?

Data has been fairly complete and accurate, with the exception of several projects that required follow-up requests for information.

2. Was the quality of the data consistent?

Yes. With a few exceptions, the project file data provided by PG&E has been fairly consistent with respect to quality.

3. What types of problems did you have with the quality of the data that was available for your evaluation?

For PG&E, we have run into some minor changes in the tracking data between what was supplied in March 2008 and what was supplied in March 2009. These changes include:

- Several projects that have been moved from PGE2001 to PGE2007

- Several projects that have been reclassified from Industrial New Construction to Commercial New Construction (which causes a shifting from the Ag-Food evaluation to the Commercial New Construction evaluation).
- Several project whose savings values have changed from the 2008 data to the 2009 data.

4. How did it affect your evaluation?

Data quality issues have mainly caused us to use additional time to investigate the potential problems.

5. What recommendations can you provide that would improve the type of data available in the future?

No recommendations at this time.

OTHER ISSUES WITH DATA

1. Timeliness of the data responses (ED staff will do a review of EEGA data request turn around so comments can be qualitative assessments.)

For PG&E, data responses have been fairly timely. While not always on time, the requests have usually been less than two weeks late.

For Sempra (SDG&E, SCG), we have only submitted one data request. The response was 3-4 weeks late.

2. Was the data understandable, i.e. were tools needed to interpret the data (such as keys) provided?

All data received to date have been understandable.

3. Availability/willingness of IOU staff to answer questions about how the data was developed/collected/aggregated.

Thus far, the IOUs have shown willingness to work with us on data-related questions.

----------Small Commercial Data Memo----------

Itron

Memo

Electric / Gas / Water
Information collection, analysis and application

11236 El Camino Real
San Diego, CA 92130-2650
858.724.2620 Tel
858.724.2690 Fax
www.itron.com

Date: July 7, 2009

To: Jeorge Tagnipes, CPUC

From: Elsia Galawish, Itron, et. al.

Re: Data Availability and Quality Review for 06-08
Projects: 10718/10720/10721

Availability of Data or Other Necessary Information/Quality of Data or Other Necessary Information

Generally the data and other info had sufficient information to help generate a research plan and support most sampling activities. Overall, some critical data were not as comprehensive as we would have desired.
Attached is an Appendix that includes a detailed review of the IOU Program Tracking Databases through 2007 Q4. Many of these items are no longer applicable, but should be reviewed as background for issues encountered during these projects.

10718 Small Commercial

Sempra:

- Missing account number, phone numbers, and NAICs codes for some records in the tracking database. They returned most of them for 3012, 3020, and 3507 but it is still not complete.
- Mailing address is not filled in at all

PGE:

- The tracking database was split by year due to the size. This makes it very difficult to look at the raw data for a single program.
- PGE2080 does not line up to E3. We have not checked the latest version of the tracking database. We think it is a result of 2008. The differences are across measures.

SCE:

- Missing rebate information for 2511.

- They provide us with one or more databases per program which is fine but the field names across databases are not consistent and each program has slightly different information.
- The measure names in the tracking database are different than the measure names in the E3 calculator. We had to manually create a lookup table to get from one to the other.
- Some target sectors in the E3 show up in the Comments field to match to the tracking. (This is more of an E3 issue than a tracking issue…)
- Savings values for 2511 and 2517 match E3

Recommendations

- Consistent field names and information given across utilities and programs - certainly at least across programs within a utility.
- Split databases by programs instead of by years if they are too large (Res vs NonRes or by contract group) but keep all variable names and data consistent throughout each database.
- Put some checks in place for zip codes (many "service zips" are out of state or nonexistent zip codes)
- Some sort of unique identifier to match to E3 would greatly help with lining up tracking to E3. Sempra already does this which makes it very straightforward. PGE and SCE require a different combination of fields to map them to each other. For example, one measure for SCE might be a match based on measure name and climate zone and another is a match based on measure name, climate zone, target sector and comments.

10720 PG&E Industrial

The extracts are missing a description of the measure installed:

- There is a measure description field, but the values are high-level and inconsistently applied across contributing programs. For example: pump-off controllers are coded "PROCESS OTHER" in the SPC program, "PROCESS ENERGY EFFICIENT MOTOR" in the Nonres Retrofit program, "PROCESS (CUSTOMIZED)" in the NRNC program, and "ROD BEAM PUMP-OFF CONTROLLER" in the Global Energy Partners third-party program.
- The core PG&E program extract does include a project description field, but in many cases that field contains the name of the customer rather than a description of the project. If project details exist, they are very general (e.g. "POC for new oil well - 75hp")
- The extracts are also missing a description of the measure "units" and the number of units installed: For example, in a lighting project one would expect the tracking database to list the type and number of lamps installed. But the "HIGH

EFFICIENCY LIGHTING" measures recorded in the NRNC program have neither (a typical project description entry would read "expansion project" and the number of units would be recorded as 1).

Recommendations:

These two issues can be overcome in the engineering work, where a paperwork request is necessary and the information is eventually available. But the net-to-gross work is difficult, because the tracking system does not contain specific enough information that can be used in a customer survey (e.g. "…we are calling about the 150 T8 lamps that you installed at X date, and for which you received a rebate from Y program" becomes "we are calling about the lighting expansion project".)
Ideally, PG&E would add a measure description field to their database in which they would record specific information about the measure(s) installed in each project.

Project 10721 SCE Industrial and Ag – SCE's Pump Testing program (SCE2510Pump).

- Phone number is missing
- No data dictionary provided
- No variable was provided to tell us whether or not the pump test resulted in a recommendation for repair. Moreover, in the Q4 database "Grs_kwh_sav" and "Grs_kw_reduc" were not populated at all (but were populated in Q3). We were using these values on the previous quarterly reports to figure out whether or not the pump test resulted in a recommendation for repair (but unsure if this is an accurate way to do this).
- There is no easy way to merge this database to the Pump repair database (SCE2510Ag) to see if someone went on to do a repair on the pump that was tested. We can merge by account number to see if one of the pumps on an account was repaired, but we can't merge by the specific pump. The pump description variables don't match in the two databases (Plantname vs. Sitename)
- Consistency across extracts: SCE program specs for program SCE2510 Pump Testing changed right before the Q4, 2008 extract. SCE is now more conservative as far as impacts claimed (impacts are claimed only once for customers with repeat pump tests within the same program cycle.) However, in Q4, 2008 this caused thousands of line items to disappear from the tracking extract for each of the three program years (2006, 2007, 2008.) Fortunately, the evaluation of this program is based on customer surveys only, and the sample design was based on the Q4, 2008 extract. Had Itron sampled customers prior to this final data extract, some of our surveys might have been unusable – with significant time and cost . Ideally, SCE would determine program specs prior to the first data extract delivered to CPUC, and then not change the specs until the following program cycle.

- Consistency between program extracts and E3 accomplishments: this affects mostly program SCE2510 Ag, where the tracking extract and the E3 accomplishments did not match. This was caused by the migration to a new tracking system in January 2008 and was finally resolved for the Q4, 2008 extract.

Missing Data Effects:

- A great deal of time to clean and prepare for research activities.
- Samples for onsite recruiting and surveying could not be pulled in a timely manner.

Other issues with data or other necessary information

IOU Staff support/timeliness of data extracts:

Generally, Itron received most data request in a timely manner and IOU staff was always available and willing to answer questions about how the data were developed/collected and aggregated. Invariably when all contract groups request data/info at the same time, it is understandable that some requests may be slow or extensions needed.

Understanding the data, i.e. were tools needed to interpret the data (such as keys) provided?

Along with the provided documentation and IOU staff willingness to answer our questions, Itron did not have too many difficulties understanding the data.

-----Small Commercial Detailed Q4 Data Inventory Memo----------

Utility Tracking Data Assessment for Small Commercial Contract Group

This report summarizes efforts to inventory the program tracking data for the fourth quarter (Q4) provided by Pacific Gas & Electric (PG&E), Southern California Edison (SCE), and the two Sempra utilities (San Diego Gas & Electric (SDG&E) and Southern California Gas (SCG)) in support of the Small Commercial Contract Group (SCCG) evaluation efforts. The primary objectives of the inventory are to:

- Validate the data with respect to their consistency with the monthly reports posted on the Energy Efficiency Groupware Application (EEGA) Web site
- Assess the data for their conformity with the original data request

It is important to note that we recognize the inherent difficulty in collecting and storing data of this nature. Many of the issues raised in this report are only observations and cannot necessarily be resolved. Furthermore, the findings in this report are the result of analysis of the data over a compressed time frame based on our interpretation of the supporting documentation. As such we do not intend to claim any statements as facts, as there will likely be cases where the data's providers will be able to explain where a misunderstanding or error in data processing has led to an erroneous conclusion. Feedback is welcome, and to that end we have attempted to provide as much information on the native field and table names as possible.

General Inventory

The data provided by the utilities reflect significant differences in the how data are collected and stored among the utilities. For PG&E, the data reside in an Access data base (**EMV_24months.mdb**) with a dozen different tables. For the small commercial programs, the relevant data are contained in two of those tables. All of the site-specific installation data (measure descriptions, kWh and therm savings, etc.) are stored in the table "*pge_frozen_data_030108*." Customer contact information is found in the other tables, though primarily in the table called "*pge_customer_data_030108*." These two tables are easily joined on the "*sa_id*" variable (service account ID, or account number). An Excel workbook (**EnergyEfficiency2006-2008-Programs_DR_EMV-ED_009-Q01Atch02 - updated 111307.xls**) defines the various fields and a Word document (**UserGuide_102507_Part II.doc**) describes some example queries to join to the two tables. The structure of the data is consistent with earlier extracts provided for Q2 and Q3.

While the Q2 data for SDG&E and SCG were provided in a single Access data base, the Q4 data are stored in separate Access files (**Installation Database SDG&E 2006 - 2007.mdb** and **Installation Database SoCalGas 2006 - 2007.mdb**). The structure of the SDG&E data is identical to the original, joint file, but the new SCG data does have some small differences. The most significant issue with both data bases is that they require the complex join of up to nine tables to assemble all the required variables. An Excel workbook (**Data_Dictionary_02_07_2008.xls**) accompanied the Q4 data with more detailed descriptions of the available fields.

The data provided by SCE are provided in separate flat Access files for each program. One immediate issue is the absence of data for SCE2559 (Lighting Energy Efficiency PAR 38/30 CFL Program) and SCE2563 (Plug Load Reduction Program). While the savings are small for these programs, our understanding is that there should be at least some tracking data as of Q4. Only the data for SCE2511 (**SCE2511 - Nonresidential Direct Installation.mdb**) were analyzed thoroughly due to their importance for the immediate research activities

Comparison with EEGA Reports

One of the first steps in assessing the Q4 data is to compare the total savings by program with the official utility EEGA monthly reports for December 2007. This comparison serves two purposes. First, any discrepancies might reveal an error in the processing of the data, which is particularly relevant for the SDG&E and SCE. Second, the comparison ensures that the tracking data are internally consistent with official numbers, so that the site-specific information on installations, savings, etc., in the tracking data can reliably tied to the utilities' reports.

The results of the comparison of tracking data savings with the official utility reports are presented in

Table 1. The programs in the table represent all those associated with the SCCG. Rows with "N/A" in the Unique Account Numbers indicate programs that have no records in either the tracking data or the EEGA monthly report. Overall, the tracking data summaries match the EEGA reports very well. Most of the large programs match either perfectly or within one half of a percent. There are a few programs where the tracking data report savings up to 27% lower than the EEGA reports, which will require further investigation to determine the source of the discrepancy[4].

Table 1. Comparison of Program-Level Tracking Data Savings with EEGA December 2007

[4] While the discrepancies are likely due to processing of the data, it appears that the proper treatment of the data requires knowledge of undocumented assumptions. It would be valuable for SDG&E and SCG to provide far more detailed documentation on the structure of the tables and their relationships. The queries in the Access files that reproduce the monthly report figures are complex and come with no documentation to facilitate the use of the data.

EEGA Code	Program Name	Unique Account Numbers in Tracking Data	Net kWh - Tracking Data	Net kWh - Monthly Report	Tracking kWh as % of Monthly Report	Net Therms - Tracking Data	Net Therms - Monthly Report	Tracking Therms as % of Monthly Report
PGE2003	PG&E Retail	136	14,633,759	14,633,759	100.0%	-33,949	-33,949	100.0%
PGE2032	PGE Sonoma	50	1,248,708	1,248,708	100.0%	19,575	19,575	100.0%
PGE2047	PGE Laundry Coin-Op	303	166,469	166,469	100.0%	64,561	64,561	100.0%
PGE2048	PGE Water Conserv	1,591	662,969	662,969	100.0%	94,780	94,780	100.0%
PGE2051	PGE RightLights	1,655	19,859,121	19,859,121	100.0%	170,820	170,820	100.0%
PGE2054	PGE Energy Fitness	784	11,498,314	11,477,886	100.2%	0	0	N/A
PGE2060	PGE Cool Control Plus	129	9,417,200	9,274,071	101.5%	0	0	N/A
PGE2074	PGE SBEA	2,423	6,683,973	6,683,973	100.0%	0	0	N/A
PGE2080	PG&E Com MM	16,107	828,467,184	822,379,488	100.7%	5,589,248	5,589,844	100.0%
PGE2085	Cool and Light Program	N/A	N/A	0	N/A	N/A	0	N/A
PGE2089	California Preschool Energy Efficiency Program (CPEEP)	N/A	N/A	0	N/A	N/A	0	N/A
SCE2511	SCE Nonres Direct Install	39,262	205,010,349	204,734,989	100.1%	N/A	0	N/A
SCG3507	SCG Express Efficiency	2,342	0	0	N/A	14,335,020	14,269,402	100.5%
SCG3514	SCG On-Bill Financing	N/A	N/A	0	N/A	N/A	0	N/A
SCG3525	SCG Energy Coalition Direct Install	N/A	N/A	0	N/A	N/A	0	N/A
SCG3526	SCG Water Conservation	1	0	0	N/A	643,290	643,290	100.0%
SCG3538	SCG Gas Cooling	4	0	0	N/A	13,142	13,142	100.0%
SCG3540	SCG Laundry Coin-op	N/A	N/A	0	N/A	N/A	0	N/A
SDGE3012	SDGE Express Efficiency	517	38,024,869	38,024,869	100.0%	507,298	507,306	100.0%
SDGE3019	SDGE On-Bill Financing	N/A	N/A	0	N/A	N/A	0	N/A
SDGE3020	SDGE Small Bus	11,093	144,307,192	144,296,628	100.0%	618,665	616,823	100.3%
SDGE3030	SDGE CPEEP	N/A	N/A	0	N/A	N/A	0	N/A
SDGE3039	SDGE MobileEnergy	1	1,581,168	1,581,168	100.0%	60,824	82,418	73.8%
SDGE3042	SDGE Laundry Coin-OP	212	221,326	221,326	100.0%	65,591	66,271	99.0%

Conformity with Data Request

This section summarizes how well the IOU tracking data conform to the items in the initial data request and identifies any key missing items that will require more immediate attention from the IOUs. The items of interest for this inventory are as follows:

i. *Unique and permanent record identifiers to allow records for an application/measure to be merged across quarterly submissions of the data.*
ii. *Application identifier needed to aggregate measure associated with a customer application.*
iii. *CPUC Program ID on each measure record.*
iv. *Identifying fields that can be used subsequently to link implemented measures to customer electric and/or gas billing records.*
v. *Data and business rule(s) needed to determine whether a measure was paid and when payment occurred, i.e., paid date/install date/completion date/post inspection date, or other fields used to document project status.*
vi. *Identification and contact information for the entity paid.*
vii. *Identification and contact information for the site where measure was implemented.*
viii. *Identification and contact information for third parties that the program used to identify/develop/implement the measure.*
ix. *Data describing measures recommended by program sponsored audits, system assessments or equipment testing activities and implementation status of these recommendations.*
x. *Measure code/description, unit counts, and savings (kWh, kW, Therms as applicable).*
xi. *Ex ante cost effectiveness inputs, including but not limited to total project cost (labor and material), incentive paid, Effective Useful Life, Net-to-Gross Ratio.*

Account and Application Information

Items i, ii, iii, and iv, which deal with specific account and application information, are based on different combinations of fields depending on the specific application and utility. This data requirement has been largely satisfied. For PG&E, a participant account number and an application ID clearly identify most unique participant applications. For both SDG&E and SCG, an account number in conjunction with a site number and a project ID serves the same purpose. For SCE, the account number can be used along with the invoice information to identify different account applications.

The account number, which is a key field associated with this data requirement, merits some additional discussion because it is ultimately used to develop an identifier for unique participants and will be used to associate billing data with individual participants.

For PG&E, the account number is the "*sa_id*" column from both the "*pge_frozen_data_030108*" and "*pge_customer_data_030108*" tables. For the SCCG programs, there are 22,908 unique values for "*sa_id*" associated with 83,939 rows of detailed measure information in the frozen data table. Of these unique account numbers, 22,897 had a matching record from the customer contact table. This small discrepancy between the two numbers is due to a number of IDs that are invalid or unidentified.

Although there are only a few invalid account numbers, they represent a significant portion of the total records, primarily for the upstream programs.

For SDG&E, the account number comes from "*acct_nbr*" field in the "*cs1tm04_accounts*" table. This account number is joined with the measure data in "*cs1tm10_msrs*" table by the "*proj_id*" and "*site_nbr*" fields. For the small commercial programs, there are 36,890 observations in the measure table representing 11,814 unique account numbers. Of the total observations in the measure tables, 3,175 have no account (almost nine percent)[5].

For SCG, the account ID also comes from "*acct_nbr*" field in the cs1tm04_accounts table. This account number is joined with the measure data in "cs1tm10_msrs" table by the "*proj_id*" and "*site_nbr*' fields. There are 5,355 records of tracking data associated with the small commercial programs representing 2,346 unique account numbers. Of these records, we were not able to associate an account number with 2,127 records (around 40%)[6].

For SCE, the account ID comes from the "*servAcctNum*" field. The 155,001 observations of program data for SCE1511 consist of 39,262 unique accounts. There are no missing values for this field.

One of the primary purposes of these data is to develop a sampling frame to support a number of different research activities. For these purposes, the data need to have a unique participant ID that represents the unique sites that can be used for the selection of samples without pulling a single program participant more than once.

While the individual utility account numbers discussed above might seem suitable for this purpose, there are a number of reasons why they are not sufficient. One reason is there is some overlap in service territories so that a participant in both Edison and SoCal Gas would have two participant IDs using only the account numbers. A more significant issue, however, are the instances of intra-utility account duplicates, which come in several different forms.

The most straightforward example is PG&E's use of separate account numbers for gas and electric, as in the following example:

Service Account ID	Account Fuel	Customer Name	Address	City	State and Zip	Telephone
DELETED - CONFIDENTIAL CUSTOMER INFORMATION						

[5] The missing accounts are associated with SDGE3039 (3,163 observations) and SDGE3042 (12 observations).
[6] The observations with missing accounts are all associated with SCG3526. While this is a longer-term concern, it does not affect the activities for the verification research.

There are more problematic instances of accounts that need to be collapsed. Take the following case where two different account numbers clearly represent the same site. If this were treated as separate participants, the research activities would run the risk of selecting the same site twice during the sample selection process.

Service Account ID	Account Fuel	Customer Name	Address	City	State and Zip	Telephone
DELETED - CONFIDENTIAL CUSTOMER INFORMATION						

There are a number of different ways in which duplicate sites or customers appear in the tracking data and all four IOUs have instances of each. Without going into the details of how these cases are treated, the end result is that the original number of unique account numbers discussed above is reduced to a set of unique participant IDs intended to represent individual participant sites. These reductions are as follows:

PG&E: 22,908 account numbers reduced to 18,469 participant IDs
SDG&E: 11,814 account numbers reduced to 10,595 participant IDs
SCG: 2,346 accounts reduced to 2,322 participant IDs
SCE: 39,262 unique account numbers reduced to 32,773 participant IDs

Given that the data cover two years of time and the numerous parties involved in data collection, we understand that it is not reasonable to expect perfectly clean data where a single account number will be associated with each project. Nevertheless, it would be helpful in terms of processing these data to get an understanding of what contributes to cases where single sites have multiple account numbers.

Measure Information

Item v deals primarily with information on measure installation and payment dates, which can all be represented by a number of different fields. Table 2 reduces this requirement to three key fields and shows whether each IOU's tracking data provide a field to satisfy the request. SDG&E and SCG both include an installation and pay date field ("*instal_dt*" and "*ap_vouch_dt*" from the "*cs1tm10_msrs*" and "*cs1tm13_ap_vouch*" tables, respectively). These fields are not entirely populated, however. SCG, for example, is missing every value of the payment date.

Table 2. Conformity with Data Request Item v

Project Status Field	Utility			
	PG&E	SCE	SDG&E	SCG
Payment Date	Yes	Yes	Partial	No
Install Date	No	Yes	Partial	Yes
Project Completion Date / Inspection Date	Field exists but mostly missing	Yes	No	No

Reliable contact information for customers and implementers is the requirement for items vi, vii and vii. To varying degrees, each utility has one or more fields to satisfy this

requirement for the contact name[7]. One significant issue, however, is the lack of a telephone number for a number of sites. Late processing of the data for SCG revealed that nearly all of the participants were missing telephone numbers, so a near-term action item is for SCG to explain how to process the data to find this field or to provide a table with a phone number for each SCG account number. For the sake of documentation, the phone number is coming from the "*cntct_ph*" from the "*cs1tm01_sites*" table, which is joined with the "*cs1tm10_msrs*" on the "*proj_id*" and "*site_nbr*" fields.

Although it was not a specific item in the original data request, the fields used to establish the building type present some concern. While PG&E has mapped descriptive names to this field, more than 50% of the observations are labeled "Both Residential and Commercial" or "No Building Type Available," which has no value for most research activities. SCE's field for this variable is also sufficiently descriptive, though 32% are "Misc._Commercial," which has limited value. For SCG and SDG&E around 90% of the observations (from the "*bldg_type*" from the "*cs1tm01_sites*" table) have the value "C," which we interpret as "Commercial." SDG&E does have some observations with a code (1 through 10) for this field, but there was no documentation on its meaning.

Items ix, x, and xi deal with measure-specific assumptions and data sources. While all the tracking systems have complete information on savings (both kWh and therms) and unit counts, the documentation on the sources for these figures could be more complete. Table 3 shows the desirable fields and each utility's compliance with the data request. PG&E satisfied in about 90% of the cases by either by the "*deer_measure_id*" or "comments" fields, which has information on the work papers. The "*deer_number*" field from "*cs1tl08_pgm_msrs*" is missing from 72% and 99% of the observations for SDG&E and SCG, respectively. The "*msr_cmnt*" field has some information on the measure source for SCG, but it is sparsely populated.

Table 3. Conformity with Items ix, x, and xi

Project Status Field	Utility			
	PG&E	SCE	SDG&E	SCG
Gross and Net kWh and/or Therm Savings	Yes	Yes	Yes	Yes
Units Installed	Yes	Yes	Yes	Yes
Net-to-Gross	Yes	Yes	Yes	Yes
EUL	Yes	Yes	Yes	Yes
Measure and Project Cost Variables	Rebate (p_rebate), Incremental Cost (p_incr_cost), Project Cost (p_project_cost) for about 8% of observations	Measure Cost (rmsrPricePerUnit)	Rebate (msr_incntv), Incremental Cost (cust_cst) for about 10% of observations	Rebate (msr_incntv)
DEER ID or Workpaper Info	Mostly Complete	Incomplete	Incomplete	Incomplete

[7] One shortcoming is PG&E's contact data is it only contains one field for the customer name and in many cases it is just the name of the business and not an actual customer contact name. While it would preferable to have both fields, it is not a major impediment to research activities.

Summary

As a general statement, the overall quality of the data is good. Each utility's data have their own strengths and weaknesses and they require a significant amount of work on the part of the contractor to clean and prepare for analysis, but with the exception of a few significant items, most of the information required is present. Based on this inventory, the following items need to be addressed with the current data:

- SDG&E and SCG need to provide missing account numbers, if possible, or provide documentation on how to obtain account numbers for all of the individual site numbers and projects. An attached Excel workbook has the site number, project ID, and associated program code for the records that are missing the account number for both utilities
- SDG&E and SCG need to provide the missing phone numbers, if possible, or provide documentation on how to associate a phone number with all of the individual accounts. The attached workbook has the account number, site number, project ID, and associated program for the records that are missing phone numbers.
- Verify with whether there should tracking data for the following programs:
    - PGE2085
    - SCE2559
    - SCE2563

In addition, the following items should be considered prior to assembling the next set of tracking data for delivery:

- SDG&E and SCG need to provide better documentation on the table relationships of the data
- If possible, all utilities should have fewer building type codes that fall into general categories
- More complete documentation of the sources for measure assumptions, either a DEER ID or the appropriate workpapers.
- Though not a significant issue should be cleaned of duplicate observations
    - Example #1: SDG&E's table "*cs1tm04_accounts*" has multiple account numbers for the same project ID and site number so that when joined with the measure data, the data are expanded.
    - Example #2: PG&E's customer contact table has duplicate values for the "*sa_id*" variable, with only phone numbers to differentiate them
- Limit the presence of field with entirely null values

----------Local Government Partnerships Data Memo----------

To: Energy Division Evaluation Staff

From: Floyd Keneipp

CC: Kevin Cooney, Bryan Ward, Eric Merkt

Date: May 11, 2009

RE: Request for Data Availability and Quality Review for 06-08 Local Government Partnership (LGP) Programs

---

This memo is in response to a request submitted by the California Public Utilities Commission Energy Division Evaluation Staff to the Energy Division Evaluation Contractors for a review of the availability and quality of the data provided by the utilities to support evaluation efforts of the 2006-2008 energy efficiency programs. The data related to resource and non-resource activities are discussed separately. Additionally, since quality and availability of IOU data varies widely from IOU to IOU, IOU-specific feedback is given when appropriate.

Resource Activities

The discussion that follows is focused on the characterization of availability and quality of project-level data provided by the IOUs and its impact on LGP program evaluations. Much of this characterization results from the VRT/PEB process, as an in-depth knowledge of IOU program tracking data was necessary in the creation of the SBC Interim Database that feeds the VRTs.

Availability of data or other necessary information

In general, the IOU program-level tracking data systems are antiquated and poorly designed. The systems do not have primary keys that would allow a line item or project to be tracked from its infancy to the end. Additionally, the systems are not built to a common data specification that would permit cross-IOU data management and/or analysis. In the case of SCE, there is not a common data specification within the utility; the result is a different stand-alone database, each with its own unique format, for every SCE program.

Inter-IOU data analysis is necessary for several activities: HIM study design and analysis, program-level evaluations for programs sponsored by more than one IOU, VRT and ERT reporting, DEER impact mapping, etc. Significant time must be spent vetting and bringing these datasets together each time a new quarter of program tracking data is delivered.

In terms of the impact on individual program evaluations, Summit Blue is the prime contractor on the LGP contract group, and this contract is evaluating 4 programs[8]. Of these only 1 program, the SCE/SCG Palm Desert Partnership, is significantly dependant

[8] These 4 programs actually include 12 distinct EEGA program numbers. UC/CSU and California Community Colleges actually include 4 programs each (1 per IOU), while both the Palm Desert and LA County programs each have an SCE and SCG component. .

on IOU data and we have found this data sufficient for the sampling purposes necessary at this stage of the evaluation. The other programs are closely managed either subcontractors or staff at the organizations[9] tend to be implementing larger, ESCO type projects. We have relied directly on these entities for the detailed project necessary for our current evaluation activities and have found this data to be adequate. At the present time we do not have an opinion on IOU data for these programs.

Recommendations on resource activity data

IOU program tracking databases need modernization; however, if this modernization were accomplished uniformly across all IOUs, with a single, highly specific data specification driving the modernization, then future program design and analysis could be much simpler.

Part of this data specification should require the IOUs to use DEER standards for reporting certain data elements when possible, including fields like building type, climate zone, vintage, measure name, and delivery mechanism.

Additionally, IOUs should be required to have primary keys built into the design of the database. In some cases, IOUs cannot specify which year the savings from a particular measure were claimed. There is limited ability to track a line item in an IOU database through to the end of the evaluation, or to see how a line item changes over time. Primary keys must be inherent in the IOU databases to provide traceability of projects that is currently not available.

SBC has worked with ED in identifying additional program tracking data fields and will continue to work with ED in determining the specifics of these fields and overall database design as needed.

Quality of data or other necessary information

The quality of data varies widely by IOU and program. In the case of SCE, some programs provide high levels of data fidelity, while others provide very little measure level information.

Specific data quality issues:

Building type data: Generally, building types are poorly specified across all the IOUs. In particular, SEMPRA provides very generic program tracking building types that provide little to no information about the actual building. SEMPRA does provide NAICS codes to clarify; however, the NAICS codes are unreliable and, even when they can be trusted, they are very difficult to map to DEER building types. Other IOUs have similar issues where building type specificity is low.

Measure descriptions: Measure descriptions are also generally poorly specified across the IOUs. There is little commonality of measure names, within IOUs and across IOUs. For example, SCE names a 13W CFL 13 different ways. Another issue is IOUs reporting ranges of measures; for example "13-26W CFL". In many cases, measure names are left blank or give little insight into what was actually installed. Make and model numbers, as well as vintage, are rarely given for measures. Finally, measure efficiency is rarely provided, except in the case when specific CFL wattage is provided. Measure efficiency would include information such as the unit demand of CFL (60W) or the unit demand of a chiller.

[9] UC/CSU and California Community Colleges are implemented with the support of Newcomb Anderson. Implementation and documentation of the LA County Program is being managed primarily by County staff.

Workpapers: Workpaper quality and availability varies from IOU to IOU. SCE generally provides very well-documented workpapers; however, workpapers are largely "under development" for a significant number of programs and measures. In the case of PG&E, in order to determine which workpaper is cited for a line item in the database, one must cross-reference the E3 calculator "comments" section. This can actually be impossible in certain cases, as there is not a one-to-one correlation between program tracking and E3 for all of PG&E's tracking data line items. Also, many of PG&E's workpapers refer to "Calc Workpapers", which provide absolutely no detail on the measure. If "Calc Workpapers" is cited, then the project files must be requested and review, which is highly time-consuming and requires shipment of large quantities of paper documents. This results in having little insight into the savings calculation methods for a large number of measures across all the IOUs.
Savings reporting: Frequently, SEMPRA and PG&E report zero savings (kW/kWh/Therms) for a line item; however, installed quantities and rebate payments are reported. Additionally, savings are being incrementally updated for a single line item, sometimes spanning several quarters or program years. This means the quarter and year of each portion of the claimed savings cannot be tracked, which is highly problematic when developing samples.
Baseline descriptions: Baseline case descriptions are rarely provided by the IOUs; therefore, assumptions must be made regarding baselines.
Units: Measure units (e.g. lamp, fixture, refrigerator, BTU, etc) are poorly reported across all IOUs. Units are commonly left blank, or are reported on a "kWh saved" basis or simply say "Unit". This greatly hampers analysis, as there is little insight into how the measures are being reported.
Delivery Mechanism: This is an essential field that is frequently not specified in the tracking data. Ideally, the IOU program tracking delivery mechanisms would map directly to DEER delivery mechanisms.
DEER references: References to DEER are made inconsistently across all IOUs. Sometimes only generic DEER measure IDs are given, sometimes DEER 2005 Run IDs are given. In many cases, the savings value given for a measure does not match the DEER savings value for the Run ID that is referenced.
Customer Data: The PG&E Q4 2008 program tracking dataset did not merge customer data with the "Frozen" program tracking data table. This will hamper field work and sample design.
Quantity: Frequently, quantity is equal to "0", or it is the equivalent of the net or gross savings. In either of these cases, we gain no knowledge of the actual quantity that was installed.
E3/Program Tracking Matching: This is a heavily discussed topic. In particular, the LGP Palm Desert program, SCG3543, had a significant disagreement (an order of magnitude) between what was stated in the Q4 08 narrative summary and what was stated in the Q4 08 E3 calculator.

Other issues with data or other necessary information

All IOU staff and partnership staff interactions were always very positive, professional, and supportive. Generally, responses to data requests were given in a timely manner and answered the appropriate questions. SCE and PGE were particularly responsive and highly available for phone and email conversations on specific data topics. SEMPRA

generally took longer to respond to requests which, in many cases, resulted in moot responses that were no longer applicable because of new datasets, etc.
Data responses were generally understandable, with appropriate keys provided. However, in many cases, keys were not unique, which resulted in multiple possible answers for a unique combination of keys. An example is a lookup table provided by PG&E that defined whether the program tracking gross savings was estimated using a "deemed" approach or a "custom" approach. The keys provided did not uniquely define this field for a significant percentage of the tracking line items.
Because portions of IOU data systems are significantly antiquated, sometimes very simple questions require inordinate lengths of time to answer. For example, IOUs generally do not track which implementation contractor is responsible for specific subprograms. It is generally known that data quality depends heavily on the implementation contractor in charge of a specific subprogram.

Non-Resource Activities

The discussion that follows is focused on the programs evaluated under the Government Partnerships contract and, more specifically, the evaluation of the non-resource efforts of those programs, with a primary objective of identifying near-term energy savings indirect impacts of those non-resource efforts.

Availability of data or other necessary information

On January 8th, 2008 the Summit Blue Local Government Partnership Evaluation team submitted a memo reviewing the data provided October 31, 2007 for Local Government Partnership (LGP) Programs in response to the CPUC master data request. This memo is included in Appendix A. The memo attempted to provide an indication of the suitability of the data provided for each program based on a ranking from 0-5 with 0 indicating no data was provided and 5 indicating the data was complete and in a format usable for evaluation. PG&E scored a 3 or better on 7 of 21 of their programs, SCE for 2 of 17 of their programs and Sempra had not provided any data for any of their 18 programs. The following prvides a few retrospective observations on the January 2008 memo.

1. The memo did not exclude any programs for which there were no meaningful non-resource activities. However, at the time, based on reviews of program implementation plans, quarterly narratives and interviews with program managers, all programs were at least indicating some efforts for non-resource activities (which included community outreach but excluded marketing efforts by our definition).

2. Much of the data provided by PG&E were sign-up sheets from community outreach events where visitors to a booth/table at an event would sign up to receive; a CFL, a chance to win something, or additional information. While SCE and Sempra were holding these same types of events they did not attempt to capture information from booth visitors. Ultimately, the community outreach events were not prioritized for our evaluation efforts as it was deemed unlikely to drive near-term energy savings. However, if the evaluation objective for such events were to identify their effectiveness in increasing awareness of the value of efficiency and increasing awareness and participation in energy efficiency related opportunities in their

communities, capturing information on visitors to the booth through a sign-up sheet would certainly be important for supporting this type of evaluation.

3. Local government partnership programs are a relatively new mechanism for delivering energy efficiency, as such their tracking systems tended to be relatively less developed than more experienced entities. And, their efforts tended to be more focused on developing systems for tracking resource efforts for which they had specific goals

4. Some of the non-resource activities specified in the PIP were abandoned to focus on resource activities, presumably under direction from the utilities, who also had incentives tied to meeting those resource goals.

5. Many of these programs were sponsored by multiple utilities and there was some coordination issues regarding which of the sponsoring utilities was responsible for managing the tracking of data and responding to data requests, I think especially true in the case of Sempra's Southern California Gas.

6. Information identifying activities was often incomplete, e.g. activities which are mentioned in a quarterly narrative, not listed in requests for lists of activities. Where information was provided, it was sometimes incomplete, e.g. limited or no participant and/or contact information, or missing contact information for a significant proportion of participants. Some participant information is very helpful in understanding who the participants are e.g. in a training specification of organization and title can provide valuable insight into who actually participated in the training (which can often differ from the specified target audience).

In order to conduct an evaluation of non-resource activities, you first need a clear picture of what activities are being pursued and at what levels. Information about this comes from multiple sources, e.g. what they plan to do or what is specified in the PIP, program managers what has gone on in the past and what is planned to happen in the future on a qualitative level, and what actually occurs. Understandably, what actually happens often differs from the best laid plans. Therefore, it is critical for evaluators trying to prioritize non-resource activities for evaluation to have good information on what activities are planned and what has actually occurred as well as information documenting the activities that occurred i.e. participation lists, training materials for training courses, recommendations made in audit reports.

Recommendations on non resource activity data

Tracking of non-resource activities was clearly a secondary priority. It is the evaluation teams understanding that goals for non-resource activities were only required for programs with no resource goals, which was only 7 of the 56 LGP programs. Only one of them actually specified any non-resource goals. I think that any non-resource activities the program deems as important element of their program logic model or to which significant resources are being expended, be required to have goals specified for those activities in the PIP, and that there are clear specifications for what information is tracked for reporting progress toward those goals. And what the expectations are for what will be submitted in response to data requests for information on those activities.

We developed a non-resource activity tracking template in MS Excel which was based on a template that PG&E had used to capture and submit some of the data for their non-resource activities. The template tried be provide some clarity about the types of information that we were seeking and how it differed by type of activity. This template is far from being the ideal system, but could be informative in development of requirements for tracking and reporting on non-resource activities. As indicated above it may not be practical to specify that level of tracking for all non-resource activities, but for any which are highlighted as being important to partnership logic model and/or where significant resources are being devoted.

Quality of data or other necessary information

Data completeness was very often an issue; however, the accuracy of data when provided was rarely an issue. There was significant variation in the consistency of the data across utilities and even across partnerships sponsored by utilities.

Other issues with data or other necessary information

For the majority of data requests, responses were timely and within the expected timeframe. In some cases, extensions were requested and granted, and the extended deadline was met. In a few cases the data requests were never met or were met, but were basically non-responsive. These were generally associated with requests where it was determined the data was just not available, or would be too onerous to compile, as such, so some requests were retracted.

Generally, in characterizing how understandable the data is, the better the tracking system, the better the documentation, or if documentation was not provided were largely self explanatory. Data that was provided in PDF format e.g. sign in sheets for training events were generally understandable, but could be very time consuming to sort out and put into a usable format. Data provided in spreadsheets that were from ad-hoc tracking systems or adapted to the NRAT template were often less understandable.

In cases where 3rd party contractors were providing services to multiple partnerships, associations with programs were often unclear. A spreadsheet would contain hundreds of records of participants with a program identifier, but it would not be at all consistent with the partnership names used by the IOU associated with the partnership ID, and almost never was the partnership ID used.

IOU staffs were very accessible and were, in general, responsive to data requests and providing information to the level their knowledge would allow. Their knowledge varied often by their level of involvement in the day-to-day activities of the programs. For example evaluation managers could often only respond in generalities based on their expectations of the programs and sometimes specifically about high level activities, IOU program managers were generally more knowledgeable about the details for specific programs, but often had to defer to partnership program managers for more detailed information or information about less high-profile activities, which non-resource activities often were. Our experience interacting with all IOU staff and partnership staff was always very positive, professional, and supportive. Although, they were also very concerned about the level of effort that might be required to accommodate data requests, especially when there were no electronic systems in place for tracking the requested data.

----------RCx and Major Commercial Data Memo----------

FROM: Michael Baker

TO: Carmen Best and Jeorge Tagnipes

DATE: May 11, 2009

RE: Request for comments on data availability and quality for 06-08

CC: Marc Schuldt, SBW

SBW is the prime contractor and has been responsible for processing IOU supplied data for both the Major Commercial and Retro-Commissioning contract groups. The comments below reflect our experiences in processing data for approximately 30 programs across all four IOUs that contributed data relevant to our HIM and program evaluation responsibilities. In addition, we have played a significant role in the development of the Interim Database and the VRT (Verification Reporting Template). In that role we have processed all of the E3s filed by the IOUs and have reconciled Program Tracking claims against all E3s.

1) Availability of data or other necessary information

a) Were the right types of information available from the program or utility to answer the research questions at hand? (Please note variations by utility, and/or specific requests.)

See next item.

b) What was missing (if anything) from data requests (either not available or provided only partially) or general information that was needed to conduct a thorough evaluation?

All IOU program tracking data lacked persistent and unique primary keys, i.e. a way to track over time each record of claim from the time that the IOU make a commitment to proceed with installation / payment to the time that the savings, costs and incentives are finalized and claimed. Also, missing for some programs and some time periods was the data needed to associate Program Tracking records with E3 line items. Missing throughout is a consistent coding of measure description. Other data elements were missing for some programs and time periods, such as measure location street address and ZIP code.

c) How did this missing data/information affect your ability to conduct the evaluation?

The lack of unique keys made it very difficult to determine what records had been changed, added or removed between successive rounds of data filings. A number of data elements required by the E3 calculator were only presented in the E3 line items, so difficulties in matching to Program Tracking translated into difficulties in reproducing the cost-effectiveness results. The lack of consistent measure descriptions substantially increased the time required to segregate measures into high impact groups. Lack of complete location data makes assignment of appropriate climate zones and locating sample cases more difficult.

d) What recommendations can you provide that would improve the type of data available in the future.

One general conclusion from our work is that the filing of four different reports (Monthly, Quarterly, E3 and Program Tracking) has substantially increased the data processing and interpretation work for all parties, contractors, ED staff and IOU staff. We would strongly urge that only one form of the data be conveyed on a regular basis from the IOU to the ED. This should be the most disaggregate form, i.e., Program Tracking, once this form has been upgrade such that it contains all of the data needed for all evaluation purposes. This program tracking data should also come with consistent measure descriptions and complete location data.

2) Quality of data or other necessary information

a) Was the data complete and accurate (to the best of your knowledge)?

As noted above, primary keys and consistent measures descriptions were missing. In addition, data elements were given different names across utilities, and for SCE, across programs. Also, the definition of the data elements; e.g., were the savings normalized per unit, changed over time.

b) Was the quality of the data consistent?

Location data was missing to varying degrees across our programs, and when present was not necessarily valid, i.e., measure location ZIP codes that actually are found within particular IOUs service area.

c) What types of problems did you have with the quality of the data that was available for your evaluation?

See previous comments.

d) How did it affect your evaluation?

Data quality issues increased the cost of handling the data, preparing samples, and complying with the requirements of the VRT. Variations in data definition also increased the chances of mistakes being made in the evaluation work.

e) What recommendations can you provide that would improve the type of data available in the future?

Standardized data elements should be established by the ED, both the name of the data element and rules that constrain the data values. The IOUs should only deliver Program Tracking records that conform to these standards.

3) Other issues with data or other necessary information

a) Timeliness of the data responses (ED staff will do a review of EEGA data request turn around so comments can be qualitative assessments.)

The amount of time to respond has varied.

b) Was the data understandable, i.e. were tools needed to interpret the data (such as keys) provided?

See previous comments.

c) *Availability/willingness of IOU staff to answer questions about how the data was developed/collected/aggregated.*

There was considerable variability in availability and willingness. It changed over time and between IOUs. In many instances, the time required by both sides would have been less if we had been able to speak more readily with the IOU staff that had direct knowledge of a programs data or a specific database. Often quick resolution of a data problem requires back and forth discussion, which is difficult to achieve in the EEGA request system.