R33 Observations & Recommendations from CT Residential Program Database Interviews

FINAL

1/26/2016

SUBMITTED TO:
Connecticut Energy Efficiency Board

SUBMITTED BY:
NMR Group, Inc.

Table of Contents

NMR
Group, Inc.

CT Residential Program Database Interviews

The purpose of this project is to document strategies that can improve the efficiencies of working with evaluation data from the Connecticut Companies, memorializing suggestions for both the Companies and for the evaluators. NMR conducted this project in three phases. First, NMR conducted an internal review of challenges we have experienced working with Connecticut data from Eversource and United Illuminating ("the Companies") based on evaluation project experience over the last three years. Second, NMR identified and proposed energy efficiency program administrators to interview for this study and compare against the Companies' practices. The Team proposed organizations for interviews based on a combination of its experience with program data from other jurisdictions and a search for papers and studies addressing program databases and tracking in recent IEPEC proceedings and the CALMAC website. Third, NMR conducted interviews with Eversource and UI staff who are responsible for the Companies' residential databases and for responding to our requests for data[1] to discuss interim findings and gather additional context and options. The Company interviews made it clear that the Companies are not in a position to establish completely new customer tracking systems, and as we explain in more detail below, many of the database issues the Team identified could be traced to difficulties of communication. Given this, it made little sense to interview program administrators about proprietary customer information databases that the Companies were not in a position to adopt. Since the California Public Utilities Commission (CPUC) has addressed the issue of inconsistency among billing databases of the state's investor-owned utilities (IOUs) by establishing a statewide IOU customer database, and this seemed like an appropriate analogy to the Connecticut situation, with approval of the EEB Evaluation Consultant, the Team interviewed just one other organization in addition to the Companies, a representative of the CPUC's statewide IOU customer database.

In this report, we summarize key observations from the interviews and recommend solutions in light of the information gathered.

KEY OBSERVATIONS AND RECOMMENDATIONS

Clarifying data requests through the use of data dictionaries. As we have noted previously, the team found the variable names in the UI data to be unclear; that is for, some variables, the names did not clearly indicate to someone outside the program the type of data contained in field and the meanings of values, particularly ones that denote missing or imputed information. Eversource has a data dictionary that the team has found useful for developing data requests. UI does not currently have a data dictionary and voiced concern about the time it would take to develop a data dictionary for the hundreds of variables in its database versus the likely return. UI has offered to develop a template that evaluators can use to submit requests and that could improve the likelihood of data requests yielding the information needed the first time. This might be less time-consuming than developing a data dictionary. It would be helpful for UI to review Eversource's data dictionary in order to understand what a data dictionary entails and explore the possibility of developing a UI data dictionary modeled on Eversource's dictionary.

---

[1] We interviewed two systems administrator staff for the Eversource HES and HES-IE programs. We interviewed a system developer, a senior systems analyst, and two managers with United Illuminating HES and HES-IE program responsibilities.

Recommendation #1:
- *We recommend that the Evaluation Team work with the Connecticut Energy Efficiency Board (EEB) Evaluation Consultants and appropriate staff of both Companies to develop*
  - *Lists and descriptions of the information that are most commonly requested for (1) process evaluation and (2) impact evaluation. The lists should include the variable names under which each Company stores the information. The lists should also note what values are used to denote missing data for each variable and what special values might be found in each data field that could affect analysis. (For example, the information that an ID number of an Eversource HES or Multifamily participant that ends in -2 supercedes an ID number that ends in -1 but is otherwise identical.)*
  - *Company-specific data request templates. The templates would be built on the lists of information and variable names described above. The purpose of the template would be to standardize data collection requests as much as possible.*
- *We recommend that Eversource consider sharing its data dictionary with select UI staff to help UI staff in planning for a UI data dictionary.*
- *Developing the UI-specific data request template should take UI much of the way toward putting together a data dictionary. We recommend that UI staff review Eversource's data dictionary to assess what additional work would be needed to complete a UI data dictionary.*

Improving the tracking of measure-specific inputs and providing details regarding calculations. Previously we have noted instances of missing measure-specific inputs and lack of details regarding calculations. In our conversations with the Companies about these instances, we found that the measure-specific inputs and calculations the evaluators thought were missing actually *do* exist in the databases. Either these inputs and calculations have only recently been added (the Companies are continually adding new elements to their databases) or the evaluators asked for the information in a manner that was not immediately understood by the Companies' database managers. Had we communicated about the issues directly with the Companies' staff who are responsible for the program databases, we most likely would have come to understand that the data were available and thus we would have been able to request and obtain this information in a manner more readily understood by the database managers.

In previous communications about data issues the team has suggested a need for data quality assurance checks. In their interviews the Companies described quality control processes to reduce data entry errors. For example, Eversource verifies account numbers with its customer information system (CIS) and automatically populates the program record with demographic information from the CIS in order to avoid errors. Eversource uses system reporting features to identify projects that are out of variance with pre-determined parameters and reviews these on a quarterly basis. UI also described a system of automated checks to avoid data entry errors. These include checks for internal data consistency and data that are out of range that happen as system users input data, and a final review by the program manager when closing out projects. To help reduce data entry errors, at both Companies HES field techs enter data on a handheld device. While HES-IE techs at both Companies still record data on paper, data entry by handheld device will soon be available for HES-IE as well.

The Companies also described their processes and quotas for inspecting projects undertaken in single-family and multifamily homes through HES, HES-IE, and the multifamily initiative, as well as how this information is recorded for each program in the database. Both UI and Eversource inspect

a sample of program homes after measure installation. Errors in work or recording of data found in the inspections are corrected in the program database. How these are recorded varies by program and utility.

It appears to the team that the Companies have instituted some of the quality control measures that the team has suggested since 2011. Interviewees indicated that such improvements occur on a regular basis based on requests from program managers. The team also believes that a lack of communication about the variation in the databases has led to misunderstandings regarding quality control.

Recommendation #2:
- *Third-party evaluation staff, the EEB Evaluation Consultant, and Companies establish an expectation that each evaluation will include at least two formal meetings about data requests: (1) A meeting at the beginning of each evaluation for third party evaluation staff to communicate directly with designated Company program database staff. The purpose of this meeting would be for evaluators to learn in an efficient and timely fashion what relevant data are available for a study and provide them with the information they need to develop complete and clear data requests for the Companies. (2) A "data request kick-off meeting" promptly after the third-party evaluator delivers the data request for a project. The purpose of the data request kick-off meeting is to encourage detailed discussion of the intent of the data request, data format, and data terminology. Both meetings would include the EEB Evaluation Consultant.*
- *Oftentimes third-party evaluation staff have new questions once they begin cleaning or analyzing the data. These questions are typically time-sensitive. Once third-party evaluation staff and Company program database staff have had the data request kick-off meeting, the EEB consider allowing third-party evaluators and Company database staff to ask each other data-specific questions and provide data-related clarification as the need arises over the course of a study by phone and email without waiting for the EEB Evaluation Consultant to be available for these ad hoc communications.*
- *During the evaluation planning stage, even before an evaluation one-page description is approved, the EEB consider allowing third-party evaluation staff and Company database staff to communicate about data in the presence of the EEC Consultant, as part of formal or informal assessments of the evaluability of particular questions or programs. Assessing a study's evaluability—including the data available that are relevant to the study—before approving work plans would help EEB spend evaluation funds more effectively. The EEB should set aside budget for these evaluability assessments to ensure that evaluators are paid for the exploratory work on projects ultimately deemed "not evaluable."*

Consistency between utility tracking systems for programs and measures. As we have documented previously, team members have found what appear to be errors and inconsistencies within utility databases, and inconsistent data formats and terminologies between utility databases.

In the interviews the Companies noted that they already align units and terminology in their respective program databases with those that appear in the Program Savings Document (PSD), and thus with each other. During our interview we discussed issues of consistency both within each Company's database and between the two Companies' databases.

Within a Company, in some cases, the Companies may be able to make specific changes to a program database in response to a need identified by evaluators. For example, in recent years Eversource added fields to track the equipment recommended to a participant as a result of an audit. Such changes would need to be addressed on a case-by-case basis because they can be time consuming for the Companies.

In the course of our conversations, the team found that some instances of what appeared to be inconsistency within or across data fields in one or the other Company's database were not inconsistencies—but the key to understanding them was not included with the data. For example, in the Eversource program database, changes resulting from quality control inspections lead to additional ID numbers being added to the database that duplicate original numbers, but end in -1 or -2. Data are associated with each ID number. The ID number that should be used for analysis in these cases—the original number or the one ending in -1 or -2—varies depending on the program. During the interviews, we did not have time to delve into each of the inconsistencies we had identified previously, but the example illustrates a source of confusion regarding which inputs to use in our studies—a source that nevertheless serves a very real and important purpose to the Companies. Having the ability to communicate more readily with Company database staff about data-related questions as they arise in data analysis would help avoid future data misunderstandings.

Looking across Companies, the Companies noted that aligning other terminology—such as field names and codes for missing data, etc.—between the Companies' databases would be a difficult and expensive undertaking requiring management approval. This is especially true for billing data for those customers with both natural gas and electric service and the information associated with it. Even something as simple as the way in which an address is tracked reflects a legacy at each Company. Using the "|" as the element delimiting fields,[2] one Company may list a home as Jane Doe | 123 Main Street #3| Anywhere | 06000, while another lists John Doe | 123 | Main St.| Apt. 3 | Anywhere | CT| 06000. The simple differences—listing different household members as the contact and the structure of the address—can make it extremely difficult for evaluators to link electric and gas accounts and program information. Yet, it is also the case that altering the fields tracked in the databases is not a simple matter for the Companies, requiring complex and expensive reprogramming of their systems.

California may point to a solution. In California, the CPUC has addressed the issue of inconsistency among the billing databases of Pacific Gas & Electric, San Diego Gas & Electric, Southern California Edison, and Southern California Gas ("the IOUs") by establishing a statewide IOU customer database. The California IOUs are mandated to provide the CPUC's third-party evaluator with customer billing data. The CPUC's evaluator aligns the different data names and file formats, consolidates the data into one format, and makes the data available on an as-needed basis to qualified users to conduct approved studies. This system does not require the IOUs to make any changes in file format, naming convention, etc., but it does require that the CPUC hire a third-party data management company on a continued basis. Massachusetts is moving toward a customer database model that is similar to California's. In the section California's Statewide Customer Database, we describe the origins of the California database and how it works.

---

[2] For example, "|" signals a new column in a spreadsheet.

Recommendation #3:

The EEB and Companies may wish to explore establishing a statewide residential electric and gas customer billing and participation database similar to California's, to be managed by a third-party firm. This database would contain customer electric and gas use and program participation information. (For more details about California's database, see the section "California's Statewide Residential Customer Database.")

Tracking of project data for multifamily buildings with consistent unit-level reporting.
Previously the team identified issues of inconsistency unit-level data in multifamily buildings, primarily those in UI data. UI is aware of the difficulties of matching electric and gas meters and accounts for the same building. In their interview UI staff noted that their service territory includes many mixed-use multifamily homes with commercial space on the ground floor and dwelling units above, and matching electric and gas meters with individual buildings will be particularly difficult for this type of situation. This is because the residential units often have individually metered electric heat, and there is one large boiler in the basement to heat the commercial space. From the electric perspective there are multiple residential customers, but from the gas perspective there is one commercial customer. Residential units with tenants who choose to participate in HES or HES-IE each get their own project number, while the gas-heated space would be a separate commercial project with its own project number, since a different customer would pay the gas bill. UI is exploring ways to match up all the units and commercial spaces within a single multifamily building, regardless of fuel type. Matching meters serving a particular building should facilitate grouping project numbers associated with particular buildings.

UI has already begun work to address this issue. Because this is a complex issue and it could easily take well over a year to implement the recommendation, allowing evaluators and Company program database staff to communicate more readily should help evaluators deal with this problem until it is resolved.

Accurate tracking of both electric and gas account numbers. Previously the team identified issues in the UI data with incomplete or inaccurate unit number and address information. At the time, the UI data management system did not have unique project identifiers to facilitate mapping of energy-efficiency projects at the customer level to both electric and gas billing data. It appears that since then, UI has addressed the issues, and there does not seem to be need for further action. In the interview with UI staff, we learned that they have been addressing this issue in a variety of ways. First, UI as been working with auditors to improve the quality of the gas account information they collect. Second, in January 2015, UI began to assign a unique project ID to each project and use this to cross-reference electric and gas information. Third, in March 2015, UI began requiring vendors to use the Eversource project number with participants served by Connecticut Natural Gas and Southern Connecticut Gas, both of which are Eversource utilities. For these customers, the Eversource project number is now being used as the mechanism for uploading project data into UI's system. This is expected to alleviate difficulties matching electric and gas account data on projects served by these utilities.

CALIFORNIA'S STATEWIDE RESIDENTIAL CUSTOMER DATABASE

Overview

California's statewide customer database was established in 2006, but was only recently named the California Energy Data and Reporting System (CEDARS). The database is a CPUC effort.[3] It is funded by ratepayers through the Public Purpose Programs charge.

Two research contractors manage the data for the CPUC. One, the "C&I customer & program tracking data" contractor, is responsible for housing non-residential customer data and program tracking data. The other, the "residential customer data" contractor, is responsible for aggregating and housing residential customer billing data and for quality assurance of residential, commercial, and industrial program tracking data that pertain to the evaluations conducted by this contractor.

The IOUs supply all of the billing data, and most of the program data, that are housed in the database. Program data that are not supplied by the IOUs comes from Regional Energy Networks (RENs)[4]. Both the IOUs and RENs supply data for the database directly to the California Public Utilities Commission. The CPUC turns it over to the data contractor for consolidation and cleaning.

The residential version of CEDARS is very large and has the following three elements:

1. Residential customer contact and billing data for all four IOUs. This totals more than 12 million accounts and includes data for master-metered buildings. With these data it is possible to obtain all energy use for the same premise, even if the premise was occupied by different customers over time, or follow a customer from address to address.
2. Energy-efficiency program tracking data for all four IOUs. In addition to program tracking data from the IOUs and RENs, this includes data reported by IOUs that have been corrected with data imputed by Energy Division contractors, and additional calculations undertaken by the Energy Division contractors.
3. Interval data from smart meters: These data are supplied by IOUs on an as-needed basis, only for specific programs that will use the data in evaluation. The CPUC can request interval data from the IOUs with two weeks' notice.

Each year, the contractor responsible for the residential billing data determines what customer data will be needed for planned studies and develops the customer data requests for the IOUs. The IOUs deliver the data to the CPUC, which provides it to the residential customer data contractor.

Currently, IOUs are supplying residential data on the following schedule:

- Residential customer data: once a year
- C&I customer data: twice a year
- Program tracking data: four times year
- Custom ad hoc data requests: as needed, with a two-week turnaround

---

[3] The system has been in place, in one form or another, since 2006, but only recently took on the California Energy Data and Reporting System and CEDARS acronym.
[4] Regional Energy Networks enable local government entities to plan and administer energy efficiency programs independent of the IOUs, working together on a regional basis. RENs are selected by the CPUC.

Purpose

The CPUC chose to establish a statewide database because it wants California evaluations to be as coordinated as possible, in order both to save money and to provide consistency to allow the CPUC to compare results from one study to the next.

CEDARS has a dual purpose: (1) to provide a central source of billing data, and (2) to provide a source of data with which to develop aggregated public information for the CPUC website. The dual purposes mean that CEDARS has internal and external faces. The external, public-facing information is available here: http://eestats.cpuc.ca.gov/Views/EEDataLandingPage.aspx

CEDARS also serves as a statewide residential sample frame for evaluation contractors. In the vast majority of cases, the residential customer data contractor samples customers without replacement. (That is, once the contractor has pulled a sample for a study in a particular year, those customers will not be part of any other study in the year.)

Contents

Residential customer contact and billing data include:

1. Customer name and customer account number(s). (This information makes it possible to track a single customer with multiple accounts at multiple addresses—for example, a landlord.)

2. Billing address

3. Service address

4. Premise identifier. This refers to an individual dwelling unit (e.g. an apartment unit or a single-family home).

5. Service account number. This tracks the individual combinations of customer(s) at a premise, including rate changes. Single-family homes have just one service account number; multi-family buildings have multiple service account numbers.

6. Actual electric and gas use and claimed energy savings.

7. Rate information, including presence of Solar or other DG (mandatory), and EVs if the customers' choose an optional EV rate.

8. In the near future, utilities' water savings claims based on program measures—but *not* customers' actual water use data.

Energy-efficiency program tracking data includes information about customer participation in programs for which the IOUs claim energy savings. For downstream programs, the service account or premise number links program tracking data to billing data. However, for certain programs, such as those that target upstream or midstream market actors, the program tracking information does *not* match a residential customer's service account or premise. For example, the service account number may be for a store at which the sales force received mid-stream incentives for qualified equipment or products sold. In all cases for which the IOUs have information about the end-use location, such as from a rebate form, the information is linked to the customer's service account number. For equipment or measures installed by a contactor and so likely to stay with the home, the premise ID is also linked. If a customer participated in a program that required a contractor, this is tracked as well.

The data the IOUs provide for the CPUC data set are a substantial subset of what is in each IOU's Customer Information System. The customer data come directly from the IOU billing systems and only include what the residential customer data contractor requests. (For example, the data exclude

credit ratings or other payment information because they are not necessary for the EM&V functions conducted by the evaluation contractors on behalf of the CPUC.)

Data Format

Utilities are required to submit *program tracking* data in a common format, but only for the programs for which they claim savings. Each utility submits *customer billing* data in whatever format is easiest for the utility (e.g. Access, SAS, etc.). The residential customer data contractor reviews the data to make sure they are complete, and then puts them into a common SAS format.

Because the IOUs do not claim savings from Low-Income programs, they are not required to supply Low-Income program data in any particular format. Any Low-Income program data that may be supplied for evaluation is supplied in whatever format the IOUs choose.

Data Matching

The CPUC is about to issue a contract to get help linking meters in multifamily buildings with units in the buildings. Making this connection is a challenge with large buildings that have multiple addresses (for example, a building that straddles a city block could have an address on more than one street) or with units that have their own addresses within a building. (Even having the geo-tracking coordinates of each meter has not resolved this problem.)

In the case of utilities that supply both electric and gas service, the electric and gas data are delivered separately but can be matched to a single customer. When a customer is serviced by separate gas and electric utilities, the residential customer data contractor will match the electric and gas data by address. (This matching is sometimes problematic, but in general it works well.)

Availability of Data

The aggregated billing data that are provided for public use are summarized by utility and state. Regulations that were enacted recently require utilities to provide data at the zip code level. Users who wish to drill deeper into the data—for example, by census tract—need special permission to access the more detailed data.

The database does *not* give IOUs access to each other's detailed billing data. When each IOU supplies customer data to the residential customer data contractor for the database, the data are under the control of the CPUC. The CPUC makes the specific data needed for each evaluation available to approved third parties but *not* to the IOUs.

Data are made available to evaluators carrying out CPUC-mandated evaluations and to other organizations whose formal request for the data has been approved by the CPUC. The request must explain why the organization wants to use the data and how it will use them. For the CPUC to grant permission to use the data, both the reason given for why the data are needed and the plans for using the data must meet the CPUC's criteria. When the CPUC agrees to share data with an organization, the CPUC has the residential customer data contractor prepare the data and deliver it to the CPUC. The CPUC then provides the data to the requesting organization. Organizations authorized to perform an evaluation with the residential data request the data directly from the residential customer data contractor, which prepares the data per the request.

Data Privacy & Security

Most data transfers are performed through Accelion, an ultra-secure cloud data transfer system. The data must be encrypted when in transit from one party to the next.

The data are housed on a central server. The residential customer data contractor's data set resides on an SAS server behind locked doors with fire protection. Only authorized individuals can enter the room with the server. The data are backed up regularly.

At the residential contractor's organization, billing data sets must stay on the SAS server unless an exception is granted to put small data subsets on a laptop or tablet. (In general, the data must not be used on a portable device, but laptops are sometimes required for field work.)

Consumer protection laws did not have to be changed in order for this database to be established. However, there have been changes in California consumer protection laws that have to do with avoiding the data being hacked. A particular concern is loss of data through use with portable devices.