# Collecting Energy Usage Data for California's Proposition 39 K-12 Program: Lessons Learned, Best Practices Defined

***Diane S. Pepetone, L'Monte Information Services***
***Justin Regnier, California Energy Commission***

**ABSTRACT**

This paper describes the process of defining, collecting and validating energy usage and cost data for schools that received funds from California's K-12 Proposition 39 program. Critical to successful analysis of Proposition 39 funded projects was collecting consistent energy usage data from all energy utilities required to report. The objective was to collect high quality data to start with and avoid performing costly manual data cleansing of collected data on the back end. This paper describes how a data exchange specification was developed with built-in validation of data characteristics using a collaborative process with key stakeholders and how this resulted in successful collection of energy usage data. The authors share the lessons learned, results achieved, and the best practices that were applied to this topic. Additionally, it provides general recommendations for others embarking on similar efforts to collect energy usage data for evaluating implemented energy efficiency measures.

## Introduction and Background

The California Energy Commission has been in operation for over forty years and during that time its energy efficiency programs have grown in coverage and complexity. Starting in 1977 the Energy Commission implemented its first major program, the Title 24 Building Energy Efficiency Standards focused on new construction, which experienced steady growth for the next three decades. With the slowdown in new construction triggered by the 2008 recession, pre-Title 24 residential and commercial buildings became a new target for energy efficiency programming.

Concomitant with the expansion in programs has been an increase in the number and size of data stores generated from these programs. Over the years, adoption of new computerized technologies at the Energy Commission resulted in a wide range of software solutions, each specific to a particular program. Most organizations with data centric programs follow a similar path until they realize that they have unintentionally created multiple data silos, each with its own data language and programming tools. The potential for research and streamlining data collection offered by sharing data between these data silos is usually impossible without extensive and disruptive redesign. However, eventually most organizations arrive at a tipping point where the potential benefits of a common data language across programs outweighs the effort required.

### Energy Commission's Standards Data Dictionary

In 2007 the Buildings Standards and the Standards Implementation offices at the Energy Commission embarked on developing a shared controlled vocabulary for Title 24 Standards and compliance programs. The Energy Commission, utilizing an information systems contractor, embarked upon the process of defining data terms and data structures and vetting them with subject matter experts at the Energy Commission, the DOE labs and in the building and HVAC industries. The resulting Standards Data Dictionary (SDD) has data structures and properties for defining all components of a building that are evaluated in assessing its energy efficiency. For exchanging SDD based data, an Extensible Markup Language (XML) schema using the XML Schema Definition Language (XSDL) was developed to take advantage of its benefits. XSDL is a widely adopted, publicly available standard for data exchange (W3C 2012), developed by the World Wide Web Consortium whose 385 members include Apple, Microsoft, Dell, AT&T, NIST, and NASA (W3C 2016). XSDL has built-in capabilities for unambiguously expressing data types, ranges, required values and required order of reported data. Validating that XML data files are compliant to their XML schema specification only requires running an XSDL validating parser. This parser is included in all major XML editors and software development platforms.

Currently the SDD is stored in a SQL Server database accessible on the web through a secure SharePoint website. From the full SDD several different XML schemas can be code-generated for use in assessing energy efficiency in residential and non-residential buildings.

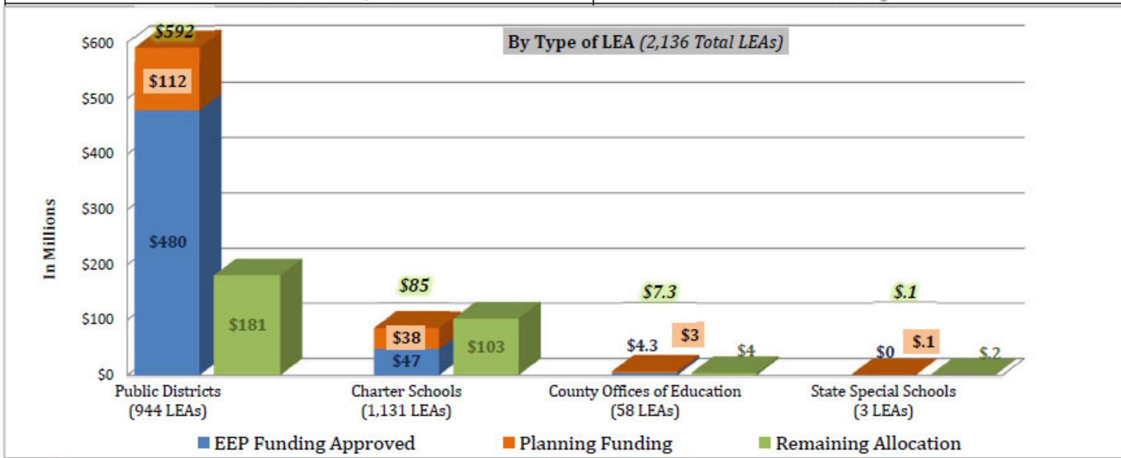
## California Clean Energy Jobs Act (Proposition 39 K-12 Program) Overview

"The California Clean Energy Jobs Act (Prop. 39) changed the corporate income tax code and allocated projected revenue to California's General Fund and the Clean Energy Job Creation Fund for five fiscal years, beginning with fiscal year 2013-14. Under the initiative, roughly up to $550 million annually is available for appropriation by the Legislature for eligible projects to improve energy efficiency and expand clean energy generation in schools." (California Energy Commission, 2016a).

The Proposition 39 K-12 program was one of the first pilots for AB 758, Comprehensive Energy Efficiency Program for Existing Buildings, and Governor Jerry Brown's Clean Energy Jobs Plan. Figure 1 shows the Proposition 39 funding allocations and projected energy savings since the program started in 2013. Proposition 39 also tracks project construction costs and job creation data to capture the whole picture of the implemented energy efficiency measures. The variety of external data sources highlighted the importance of an internal controlled vocabulary used to create a data exchange specification.

California Energy Commission (as of 02.16.2016)
California Clean Energy Jobs Act: Proposition 39 (K-12) Program *Snapshot*

| Funding Allocations *(in millions)* | Local Educational Agency (LEA) Award Program |
|---|---|
| FY 2013/14 Allocation - $381 | LEAs are required to submit an Energy Expenditure Plan (EEP) to the Energy Commission for review and approval. |
| FY 2014/15 Allocation - $279 | |
| FY 2015/16 Allocation - $313 | CA Department of Education is responsible for dispersing approved funding. |
| **Three Year Subtotal - $973** | |




| **Estimated Annual Energy Savings | **Estimated Annual Energy Cost Savings: $39,629,902 |
|---|---|
| Total kWh savings = 216,896,004kWh | **Estimated Greenhouse Gas (GHG) Emission Reductions: 165,859,723 lbs of carbon dioxide emissions |
| Total Therm savings = 1,390,685 therms | ***These numbers represent projected annual energy cost savings, energy savings, and GHG emission reductions when all of the Energy Commission approved projects are complete and final. LEAs are required to report final data 12-15 months after projects are complete and final. The numbers provided are estimated, not actual.* |
| Total Propane savings = 77,763 gallons | |
| Total Fuel Oil savings = 67,049 gallons | |

*(Source: Energy Commission)*

Figure 1. California Clean Energy Jobs Act: Proposition 39 (K-12) Program Snapshot as of 02.16.2016. *Source:* California Energy Commission 2016b

In 2013 the Local Assistance and Financing Office in the Efficiency Division began implementation of Proposition 39. When the staff began planning the energy usage data collection from the state's utilities, they recognized the need for a controlled vocabulary and a standardized data exchange specification. The makeup of California's electric utilities is a complicated picture. There are more than fifty electric utilities of which seven are investor owned, four are cooperatives and the majority are publicly owned. The six largest electric utilities combined represent 87 % of the electric usage in the state and 91% of all electric utility accounts. (California Energy Commission 2015a). Therefore, the staff focused their efforts on defining a data exchange specification that would cover all the variations across the four largest investor owned utilities Pacific Gas and Electric Company (PG&E), Southern California Edison (SCE), Southern California Gas Company (SCG), San Diego Gas and Electric Company (SDG&E) and the two largest publicly owned utilities, Sacramento Municipal Utility District (SMUD) and Los Angeles Department of Water and Power (LADWP). The SDD with its controlled vocabulary and XML schema capability was then adopted as the foundation for the Proposition 39 data architecture.

## Proposition 39 Data Collection Requirements

As the first effort at the Energy Commission to collect energy usage and cost data from utilities, the Proposition 39 program generated a lot of expectations for future benchmarking programs. A project chartering workshop was identified as an appropriate approach to capture and clarify these goals and objectives and get commitment from key stakeholders. Energy Commission stakeholders from the Efficiency Division, the Information Technology Services Branch, and advisors to Commissioner Andrew McAllister, and Commissioner Robert Weisenmiller participated in the project chartering workshop using the process described in *Requirements by Collaboration: Workshops for Defining Needs* (Gottesdiener, 2002.). The workshop resulted in the group defining a two phased approach. Phase 1, the Proposition 39 program was defined as a pilot project that would prototype processes and expand the SDD to generate a data exchange specification for collecting utility data. The work in Phase 1 would provide the foundation for Phase 2 development of a comprehensive existing building energy performance database for storing data from energy efficiency programs for existing buildings.

In addition to identifying stakeholder groups, high level scope and critical success factors, the workshop participants outlined several constraints for Phase 1. Because of limited budget and schedule, Phase 1 scope did not include developing a database for storing the collected Proposition 39 energy usage data. Therefore the data exchange specification had to provide built-in data typing and validation so that collected data could be validated for correct form and completeness in its delivered form. Building on the existing SDD which could be expressed in XML schema definition language would automatically provide this powerful yet simple validation capability. In Phase 2, the data exchange specification would be the model for developing database structures, so there would be minimal processing required to import the collected data for Proposition 39 projects into the future database.

**Developing a Collaborative Process with Utilities**

Use case models were developed next to model the interactions between the Proposition 39 system and its users and help to define data transfers in those interactions. As major stakeholders in the process, California's four largest investor owned utilities, PG&E, SCE, SCG and SDG&E were asked to participate. The California Public Utilities Commission (CPUC), with experience collecting data from investor owned utilities for evaluating their efficiency programs, also participated in the process. As with most new efforts the real progress in the early meetings was building trust among the stakeholders and identifying shared objectives. This required defusing the built-in tension between a regulator stakeholder and a regulated business stakeholder. Having a meeting facilitator helped especially to offer objective, constructive input when a statement or action from one party was misconstrued as obstructionist by the other. The collaborative process evolved into a combination of techniques including one on one sessions with each utility to address the particulars of their Customer Information System (CIS) database, and whole group sessions for technical issues that effected all of the utilities.

**Use Case Models Help Capture Challenges and Solutions**

Table 1 lists the three summary use cases that were further developed to identify all the data and workflow details involved in collecting energy usage data. For each summary use case, the Energy Commission, CPUC and utility stakeholders identified the use case actors, preconditions and likely scenarios.

Table 1. Proposition 39 Energy Usage Data Collection Summary Use Case Table

| Use Case ID | Primary Actor | Use Case Name |
|---|---|---|
| LEA1 | LEA | Submit Proposition 39 Utility Data Release Form to Utilities |
| CEC1 | CEC | Collect energy usage data for each LEA school from utility for two years prior to their Proposition 39 project approval until nine years after project approval. |
| Util1 | Utility | Send the Energy Commission energy usage data for each school in an LEA that received Proposition 39 funding covering the required time periods. |

The full set of possible data recorded for each use case is shown in Table 2 and based on *Writing Effective Use Cases* which provides an extensive discussion of use case formats and fields in Chapter 11(Cockburn, Alistair. 2002). Use case LEA1 was tackled first because no energy usage data could be sent to the Energy Commission until this release form was successfully processed. Developing LEA1 scenarios revealed the first major challenge to collecting energy usage for individual schools. Specifically, California K-12 public schools are not usually identified as customers in utility CIS databases because they rarely pay their own utility bills. The local educational agency (LEA), which is either a county office of education, school district, or direct-funded charter school, usually pays energy utility bills for all the schools under its jurisdiction (CDE. 2015). Prior to the Proposition 39 funding requirements most LEAs had no reason to track each school site's energy independently and their utility bills reflected this lack of granularity. Most utilities assign North American Industry Classification System (NAICS) codes to nonresidential customers but these proved inadequate for identifying LEA's and their schools.

Table 2. Use Case Record Fields

| Use Case Field | Purpose |
|---|---|
| Use Case ID | A string that uniquely identifies the use cases |
| Use Case Name | A brief statement of an action goal accomplished by the use case |
| Use Case Level | The level of functional granularity in the use case. Use case level is either Summary or End User |
| Actors | A list of agents that interact with other agents in accomplishing the goal of the use case. |
| Preconditions | A list of events including other use cases that must have taken place before this use case can execute. Each precondition is numbered |
| Success Scenarios | A list of steps required to accomplish the use case goal. Each step involves interaction between actors and the system |
| Failure Scenarios | Failure scenarios may be developed to identify issues to address. |
| Open Issues | A list of issues identified in the process of developing the use case. Each issue is numbered and paired with suggested solutions. |
| Decisions or Recommendations | Decisions or recommendations, made as a result of modelling the system with the use case, often related to an identified issue. |

A basic issue that was identified while developing failure scenarios was the problem of matching utility customer records to LEA's and their schools. The utilities provided examples of LEA customer records showing that the various names on the account did not match the official name of the LEA or the names of its schools. This led to the discussion of how to leverage the California Department of Education's fourteen digit County District School (CDS) code. The CDS code uniquely identifies the LEA with two numbers for the county, followed by five numbers for the school district and the last seven digits are all zeros. CDS codes for schools have the first seven digits of their LEA followed by a combination of seven numbers that uniquely identify it. Several possible solutions were entertained for mapping utility customer records and service agreements to the LEA and school CDS codes. Some utilities considered adding the CDS code to their CIS. Others planned to create a mapping solution external to their CIS.

Developing failure scenarios also elucidated workflow issues in processing of the Utility Data Release Form. Similar to the utilities' Customer Information Standard Request (CISR) form, this paper form completed by the LEA, authorized the utilities to share the LEA's customer utility data with the Energy Commission. It had places for the LEA to list customer accounts and agreements associated with the schools in their district. From interviews with Energy Commission staff, and LEA and utility stakeholders, a picture emerged of several problems with the current process. Being a paper form meant a manual and error-prone process that was especially onerous for large LEA's with more than one hundred schools. In a few instances the LEA just put "All accounts" instead of listing them. The form also did not require the school name or other key data needed to map school sites to their meters.

The final LEA1 use case record included a number of recommendations for improvements that were adopted. The Utility Data Release Form was revised to incorporate several automation features to make it easier to complete and less error prone. The utilities established new internal processes to ensure that the submitted forms were successfully routed to and processed by persons familiar with Proposition 39 requirements. The Energy Commission provided a list of utility email addresses for submitting the forms on their Proposition 39 webpage and revised the Proposition 39 Energy Expenditure Plan Handbook to reflect these changes and clarify LEA's submission workflow.

Work on the CEC1/Util1 use cases went hand in hand as they are opposite sides of the same process. CEC stakeholders identified energy usage data elements needed for analyzing the effectiveness of school Proposition 39 projects. Then in group and individual sessions with utilities a first draft of the data exchange specification was developed that would work with all of the CA IOU's different data names and structures. The final step was to evaluate existing systems that collect energy usage data to see what could be adopted or incorporated into the final data exchange specification.

## High Level Requirements Distilled from Requirements Models

After three months of collaborative work with Energy Commission project staff and utility stakeholders, the following high level requirements for collecting energy usage and cost data were distilled from the project chartering workshop and the use cases.

- The data exchange specification will be an expansion of the SDD.
- New utility data will incorporate elements from existing systems already supported by the utilities or the Energy Commission.

- The data exchange specification will be represented as an XML schema to be used by California's four largest investor owned utilities, PG&E, SCE, SDG&E, SCG, and two largest publicly owned utilities SMUD and LADWP. All other utilities who do not want to submit XML files will be required to complete either a smart spreadsheet or online form that collects the same data for each LEA and school.
- Each XML file, data file or online form will report energy usage and cost data for one type of energy for one site in a local education agency (LEA) that received Proposition 39 funding.
- Reporting period will be July 1st to June 30th and utility data files will initially be submitted annually at the end of the following calendar year, and not continuously.
- The data exchange specification will define core data required for every site from every utility and optional data that may not be available for all schools.
- The CDS code will be included as the unique identifier for the LEA and the school.
- The data exchange specification will include identification data for the LEA and the school including data for each customer agreement collecting energy usage for the school.
- The data exchange specification will include energy usage and cost data for every customer agreement in the identification section. The data will contain monthly billing data and interval level energy usage if available for the customer agreement.

**Requirement to Incorporate Elements from Existing Utility Data Collection Systems**

Several existing XML schemas and a database schema were evaluated for possible adoption in the SDD. Initial SDD development in 2007 went through a similar process of analyzing existing data models and incorporating key data definitions, structures and relationships. In this evaluation U.S. Department of Energy (DOE) Building Energy Data Exchange Specification (BEDES) (DOE 2016), Green Button XML schemas for Connect My Data application (NIST 2013) and California's Public Utility Commission (CPUC) Claims Access Database System (CPUC 2016), were evaluated based on criteria required for the Proposition 39 energy usage data exchange specification.

BEDES was selected because as its name indicates it was an effort to standardize building energy data specifically for data exchange purposes. The Green Button XML schemas were of interest because their main purpose was to provide a means for sharing customer energy usage with third parties. The Green Button XML schemas were based on the Energy Services Provider Interface (ESPI) data standard released by the North American Energy Standards Board (NAESB). Because of its particular focus on data collected from CA's IOU's for evaluating their efficiency programs the California's Public Utility Commission claims access database system was of particular interest. Analysis of the CPUC claims database was invaluable in providing a window into the available fields and nomenclature in the CA IOU CIS databases. It was also helpful to discuss the experiences the claim database team had in working with utilities on the system.

The candidates were evaluated on a number of criteria. A summary of the results of the evaluation are shown in Table 3. All candidates did well with the Coverage criteria, a simple indication of how well the candidate system covered the data needed for the Proposition 39 data exchange specification. The second criteria SDD naming conventions indicate how well the candidates follow the SDD naming conventions which were developed to avoid validation errors caused by complicated or inconsistent naming conventions. SDD naming conventions would be applied to any terms adopted from external sources and therefore alignment with them was important. The SDD naming conventions applied to all data names and string enumerations are the following:

- No punctuation except underscore
- No spaces
- Term or value is created from one or more descriptive words to embed meaning
- Capitalization follows Pascal casing which capitalizes the first letter of each word in a term, e.g. EnergyGenerationOnSite.

Terms in the Green Button schemas performed the best on following the same naming conventions. The XML schema format criteria indicates if the system can be expressed as an XML schema because that was the required format for the Proposition 39 data exchange specification and any system using the XSDL would be easier to adopt. In this Green Button was the only candidate that passed. The final criteria, XML schema parser validation indicates whether the candidate requires more than an XML schema parser for validating the correct form and completeness of the delivered data. Because of limited budget and tight schedule and the quantity of data involved, it was important to avoid systems that required effort to implement additional validation software or perform manual processing to validate correct form and completeness of the delivered data. None of the candidates satisfied this criterion. This was the main reason Green Button was not adopted, however many of its terms were incorporated into the SDD.

Table 3. Evaluation of Existing Utility Data Collection Systems for Adoption or Incorporation

| Evaluation Criteria | BEDES | Green Button | CPUC |
|---|---|---|---|
| 1. Coverage | Good coverage | Excellent coverage | Good coverage |
| 2. SDD Naming Conventions: Are terms compatible with SDD naming conventions? | No, terms are expressed in natural language with no consistent naming conventions. | Yes, except enumerated strings which are expressed as numbers and require mapping | Yes except some terms that do not convey the meaning by embedding words in term names. |
| 3. XML Schema Format: Is there an XML schema? | No, not with the current version of BEDES | Yes, schema files: espiDerived.xsd, RetailCustomer.xsd and atom.xsd | No, not currently |
| 4. Can delivered data be validated for correctness with an XML schema parser? | No, not in its current format | No, additional Green Button software has to be implemented | No, not in its current format |

## Proposition 39 Utility Data Exchange Specification

At the core of the Proposition 39 utility data exchange specification are the data elements required for analyzing the energy usage and cost of schools with Proposition 39 projects. How the data is represented in the schema follows the namespace architecture, design patterns and naming conventions of the SDD.

Namespace is an XSDL technique for organizing a large knowledge domain into usable containers of data. Each namespace contains all the names of terms that are allowed for defining a content area which can vary greatly in scope. Organizing terms by namespaces for specific content has many advantages. It helps in finding type definitions by limiting the search space. An analogy would be how books are arranged in a library. Similar to namespaces, books are usually arranged by categories which helps in finding the right book. Namespaces make maintenance of the XML schema easier because each type definition occurs just once in its namespace and then it can be used over and over in other schemas by importing the namespace, rather than having to create duplicate copies of the type definition when it is needed in different schemas. For example the SDD DataTypes namespace as its name suggests contains simple data types which are available in all other schemas by importing the DataTypes namespace. The SDD Building namespace has the broadest scope for modeling a building for compliance and it contains the SDD Envelope, HVAC and Lighting namespaces as well as the DataTypes namespace. With their flexibility of configuration, namespaces support reuse of data, and the ability to combine namespaces to represent different data models.

For every reporting purpose, a specific reporting schema is created that includes all model subschemas it needs to draw upon for required data types. This ensures that the same data collected across several different types of reports can be compared because the data has the same underlying definition. The reporting schema also specifies the order of the data, attributes that define if the data is required or optional, choices which allow different options in the data, and constraints on the data unique to the report such as allowing only specific values, or a limited range of values for a data type. Figure 2 depicts the data architecture built on the Standards Data Dictionary that is used in the Efficiency Division at the Energy Commission.

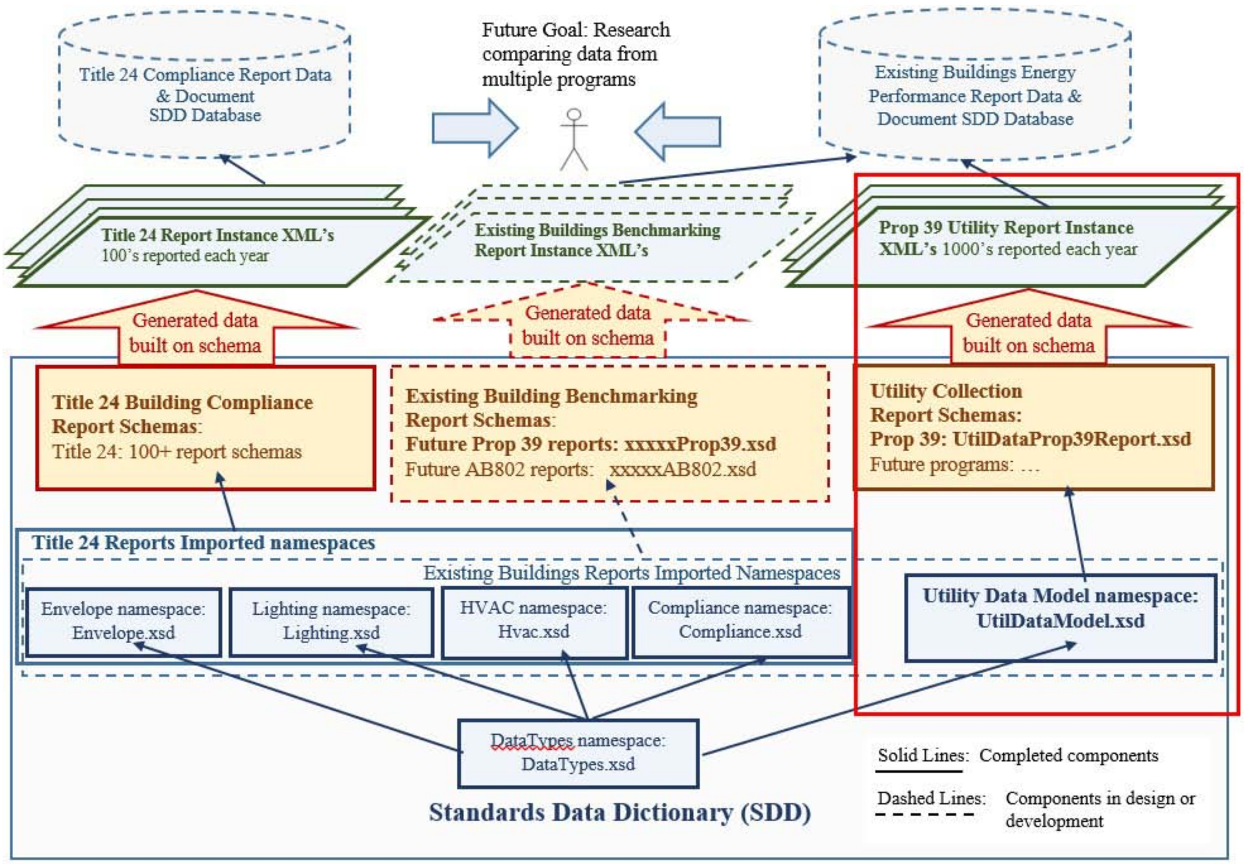



Figure 2. SDD-based Architecture from Namespaces to Databases. Proposition 39 architecture outlined in red.

## Proposition 39 Utility XML Schema Details

With the Proposition 39 data exchange specification a new SDD namespace was created to contain all data types and structures needed to define the utility data model. The new utility subschema, called UtilityDataModel includes the DataTypes namespace and contains many core utility data types adopted from the Green Button XML schemas. It also includes other data types to explicitly define relationships between a utility customer and their service agreements, service locations and meter. Following the SDD model-report design pattern, a report subschema was created that represents the Proposition 39 utility data exchange specification. This subschema, called UtilDataProp39Report, draws all of its data types from the UtilityDataModel namespace and defines the required order of data in submitted XML files, specifies any constraints of the data used for data validation and provides choices to support different configurations in the utilities and different data available for specific schools.

A graphical representation of the high level structure and organization of the UtilDataProp39Report schema for submitted utility XML files is shown in Figure 3. The UsageData element has the Choice symbol indicating that either ElectricityUsageData is supplied or GasUsageData but not both in the same XML file.

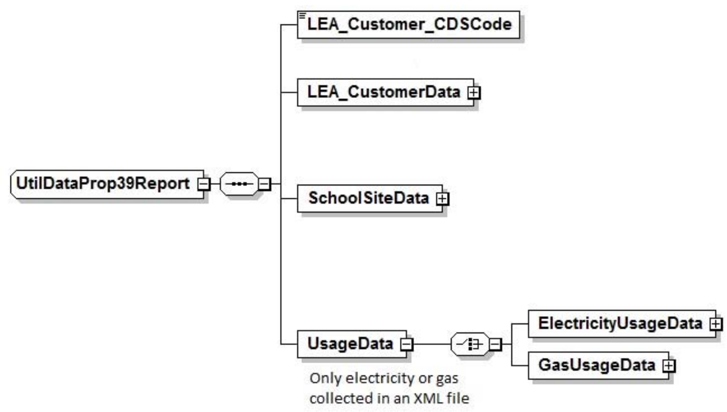



Figure 3. High level organization of UtilDataProp39Report.xsd *Source*: California Energy Commission, 2015b.

The ElectricityMonthlyBillDataPerAgreement in Figure 4 is the main data element in ElectricityUsageData (see Figure 3). ElectricityMonthlyBillDataPerAgreement defines all monthly billing data for a school CustomerAgreement. The symbol 1…∞ indicates that there may be one or more of these elements in the XML file.

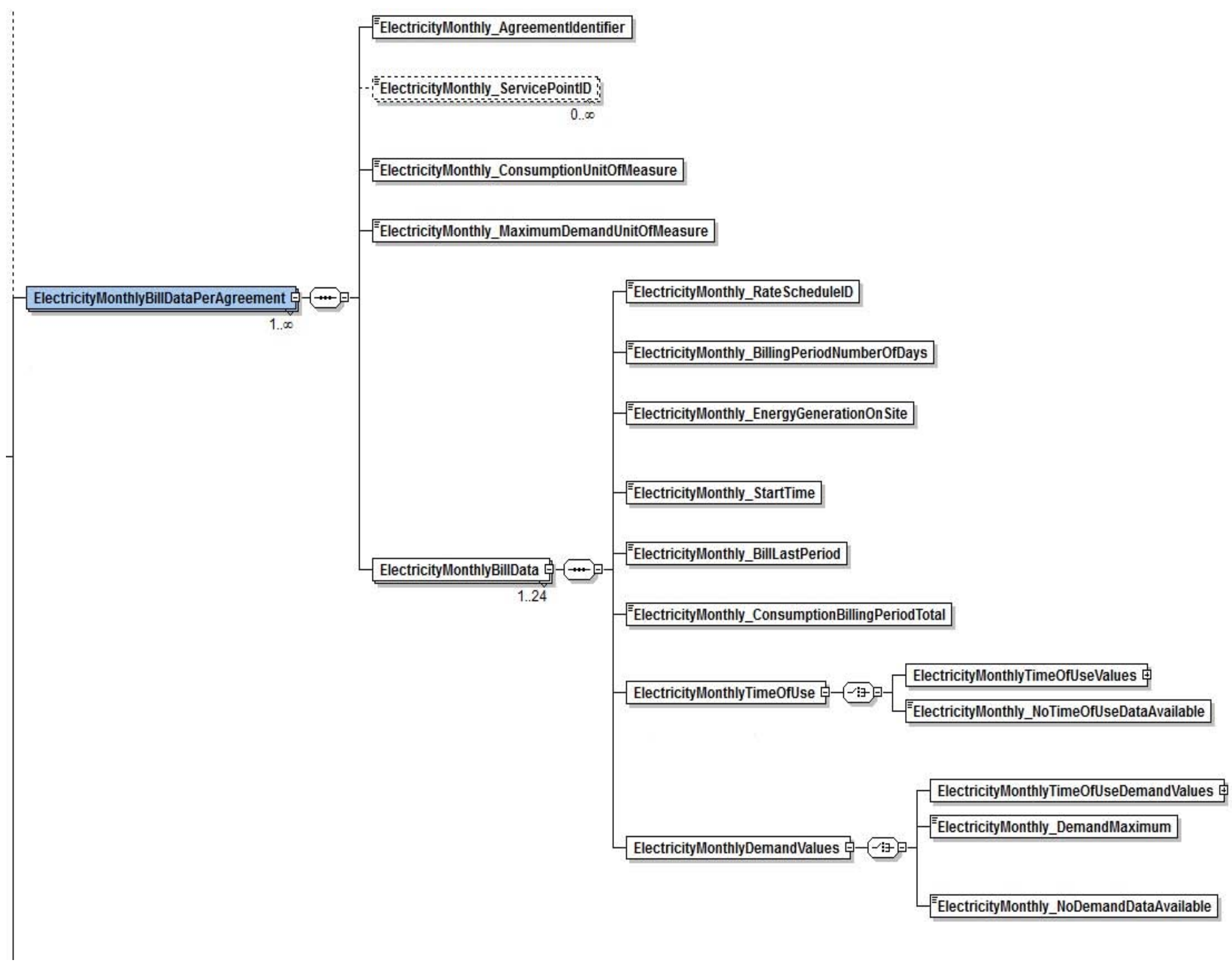



Figure 4. Organization and data elements of the ElectricityMonthlyBillDataPerAgreement element in the UtilDataProp39Report.xsd. *Source*: California Energy Commission, 2015b.

Finally Figure 5 shows the interval energy usage data collected for those School Customer Agreements that have interval data. Many of the data structures and element names in ElectricityIntervalData were adopted from the Green Button XML schema. Data elements that are represented with a dotted border indicate that the element is optional. For example the data element ElectricityInterval_ServicePointID is a virtual identifier for grouping meters at a ServiceLocation which some but not all utilities use to have a constant ID rather than depend on the ID of meters which change when the meter is replaced.

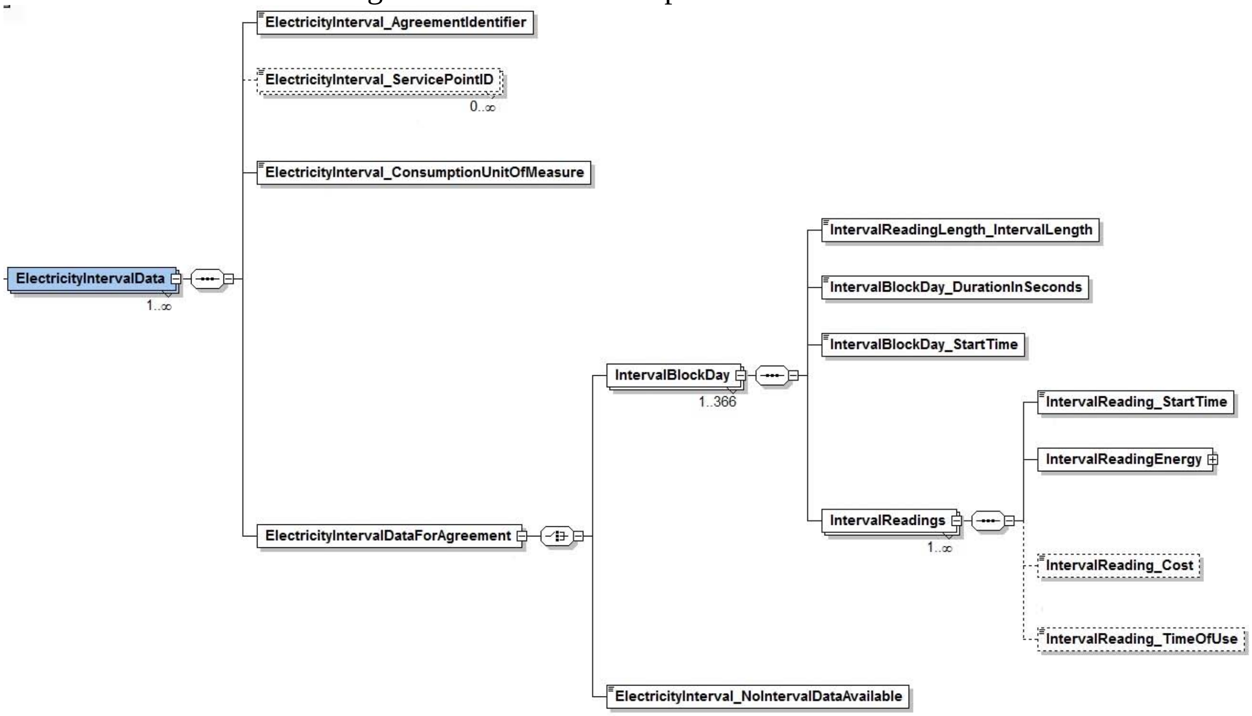



Figure 5. Organization and data elements of the ElectricityIntervalData element in the UtilDataProp39Report.xsd, *Source*: California Energy Commission, 2015b.

## Collecting the Proposition 39 Utility Data

The Proposition 39 utility data exchange specification consists of three files, UtilDataProp39Report.xsd, UtilityDataModel.xsd and DataTypes.xsd. The beta version was published and delivered to the California IOU's a week prior to a final whole group work session held in person at the CPUC on January 27, 2015. This meeting was followed by a period of several weeks when the utilities worked on implementing functionality to generate energy usage data that would validate against the XML schemas. After several iterations the official release version 1.0 was published and delivered to the utilities and then the Energy Commission began negotiating with the IOU's on when they would be expected to deliver the first year of energy usage data. During implementation several utilities ran into a problem with a commercial tool that didn't handle the XSDL Choice structure properly. This caused validation errors in the generated XML data. CEC project team worked with the effected utilities to create their own internal version of the XML schema that would work with their tool and still generate XML data valid with the official Proposition 39 XML schemas. Another area that required changes to the schema were the variations in rate structures and how each utility recorded peak demand and time of use values. Over several months the number of problems decreased and the XML schema has remained stable since October, 2015 with interim release 1.004b (California Energy Commission 2015b).

Once issues with the XML schema were resolved, the utilities focused on mapping the Proposition 39 project sites to their utility customer records. There were three major problems encountered during this process. As stated earlier the CDS code was used in the XML schema to uniquely identifying facilities in the LEA. At design time it wasn't clear that that this only applied to instructional facilities. During the collection process it was discovered that non-instructional facilities such as district offices or bus transport facilities have either the same CDS code as the LEA or no CDS code at all. The collected data for these facilities could not be correctly identified with just the CDS code as planned. One or more additional data fields had to be selected that could uniquely identify non-instructional facility data. The second problem was that some facility names and location data provided by the LEA did not match utility records. To solve these cases utilities used a combination of techniques including geospatial mapping of meters, referring to CDE records, and matching by hand using online mapping applications (i.e. Google, Bing). Despite these efforts, there were still some facilities that could not be matched successfully. Additional problems arose from the incorrect completion of required utility forms. In some cases, the release of liability form was not completed, or completed by persons with no jurisdiction over the facilities. In others, the account number information was incorrect or incomplete.

As of the writing of this paper the electrical interval energy usage data has been successfully collected for 79 % of all facilities in LEA's that participated in Proposition 39 funds the first year, and 82% of the facilities participating in the second. In both cases, about three-fourths of problems were due to an inability to match the facility to utility records and one quarter were due to a lack of adequate documentation submitted to the utility by the LEA. The full set of utility energy usage data collected to date (2.4 GB) is publically available as of June 10, 2016 at the following link: http://www.energy.ca.gov/efficiency/proposition39/data/.

## Lessons Learned and Best Practices Defined

There were many lessons learned while developing the Proposition 39 utility data exchange specification and collecting the first year of data. The most serious challenges in the Proposition 39 project were not the technical challenges posed by data transport, but rather resulting from the fact the Proposition 39 project was a pilot project attempting something previously not undertaken.

### Best Practices for Pilot Projects

The Proposition 39 program was a pilot program for future benchmarking programs at the Energy Commission. As with most pilot projects, the Proposition 39 program had to create from scratch an infrastructure of new methods, new protocols and new alliances. If it was successful this infrastructure would be used for future existing buildings programs. Changes to staff required to set up a pilot project can be disruptive with new staff positions and changes to existing staff responsibilities required to support the pilot project. It often means the staff for the pilot project have never worked together before and there may be no "old timers" in the group to rely on for guidance and advice. The formation of a successful working group for a pilot project takes time, especially without a clear statement of objectives and a plan. Holding a project chartering workshop using collaborative techniques was found to be a good initial technique for accomplishing this. During the workshop stakeholders starting with different goals and agendas were able to develop a shared vision and define high level scope. The project chartering workshop provided an opportunity for stakeholders to pool their knowledge and correct wrong assumptions and misunderstandings in the process. For example the Proposition 39 project chartering process helped the key Energy Commission stakeholders realize there was a confusion about the scope for the pilot project versus the scope for the future larger existing buildings program. Requirements modeling of use cases also provided a mechanism for strengthening the group process. The Proposition 39 utility data collection use cases were particularly useful in identifying missed areas of functionality and gaps in processing that were then corrected.

For pilot projects a best practice is to find and adopt existing successful solutions for parts of the problem space because it reduces the amount of experimentation and potential for failure. The Proposition 39 project leveraged previous work at the Energy Commission in developing a controlled vocabulary, the Standards Data Dictionary (SDD) where some best practices in data architecture and data processing were already put in place and had been working for several years. One example that helped minimize technical problems in the Proposition 39 project was the best practice of entering data definitions once and only once to a source from which all required products and documentation could be generated. That single data source for the SDD was an SQL Server database, accessed through a set of spreadsheet-type lists on a SharePoint website. The existing SDD XML schema process used code generation to create the XML schemas from the SharePoint lists. Using this technique meant that the instructions for creating valid XML schemas were also defined once and only once and then used over and over again to generate each XML schema. The beauty of the once and once only best practice is the savings in time, cost and frustration by avoiding numerous small often hard to find errors that inevitably occur when the same data has to be entered multiple times. The other powerful advantage the SDD brought to the project was the ability to validate correct form and completeness of thousands of delivered XML files by simply running an XML schema validating parser on the files. This also meant that feedback to utilities on validation failures was prompt and precise. The XML schema validation results provided the utility application developers with the location and nature of the errors in each failing data file, information that facilitated a faster turn-around on correcting their code and resending the files. The Proposition 39 effort benefited by building upon the existing SDD with its foundational data definitions, XML schema generation capability and automated data validation.

## Best Practices and Lessons Learned for Successful Collaboration

At the start it became apparent that each utility would favor the simplest short term solution for them, as it would require the least amount of their resources. The Energy Commission on the other hand was trying to create a robust utility data exchange specification that would provide complete, consistent, high quality data across multiple utilities for analysis purposes. After a few sessions with little progress, different techniques were employed to handle these conflicting agendas.

A best practice when there are critical stakeholders with conflicting agendas is to have a facilitator. An outside consultant with facilitation experience was able to ask clarifying questions when conflicting agendas clouded communications which helped maintain forward momentum during work sessions. Another technique that improved the flow of information was setting up one on one sessions between the Energy Commission and each utility. This mitigated the effects of the competitive relationships between the utilities that often blocked the sharing of information about their CIS data. It was particularly important in understanding the problems different utilities were facing such as working with legacy systems that were difficult to extend and vying with other internal groups for available resources. The whole group sessions were used for addressing less sensitive topics such as how the data was to be delivered to the Energy Commission. Working on shared problems is another technique for strengthening the collaborative energy and replacing the drive to resist the process. Engaging the utility participants in the requirements modeling process gave them an opportunity to focus on actual problems they had with the current workflow and share their ideas on how to resolve these problems.

One area where the collaborative process could have been improved was by ensuring adequate involvement from LEA stakeholders in the use case process. Several problems were identified by Energy Commission staff and utility stakeholders during the use case process and improvements were made to the forms and workflow, however, the LEAs were not part of that process. During data collection it became clear that some LEAs did not understand the process and failed to provide the data required to accurately identify each of their facilities on the modified release form. In hindsight having an LEA stakeholder from each utility territory participate in the use case development process would probably have benefitted the final collection process.

## Conclusion

The successful implementation of the Proposition 39 data exchange specification was due to many factors. Holding a project chartering workshop with Energy Commission stakeholders and an outside facilitator was an important first step for this project because the group process brought needed agreement and clarity on the goals and objectives of the Proposition 39 project. Building on the Energy Commission's SDD with its XML schema capabilities provided a technical solution with a proven track record that saved the project the cost and risk of developing a technical solution from scratch. The major non-technical factor in the project's success was the use of a collaborative process that engaged the regulatory and regulated stakeholders in developing use cases to understand the data and workflow needs of both groups and how to resolve the issues they identified in the process. One conclusion of this paper is that including LEA stakeholders in the development of use cases would have improved the results of the data collection process. During the development of the final data exchange XML schemas, it was important to have regular communications with utility technical staff to understand the unique complexities within each utility CIS that required customization. Selecting XML schema to represent the data exchange specification proved invaluable for several reasons. The XML schema supported precise data definitions, constraints on data and consistent organization that guaranteed a level of quality in the delivered data without added cost of implementing validating software or performing extensive data cleansing. Validation of thousands of delivered XML files simply required running an XML schema validation parser to show what files conformed to the data exchange specification. This in turn supported a rapid reporting and correcting of files that did have validation errors.

## References


CDE (California Department of Education), 2015. *Local Educational Agency Plan.* http://www.cde.ca.gov/nclb/sr/le/

California Energy Commission, 2015a. *California Electric Utility Service Areas California, 2015*. http://www.energy.ca.gov/maps/serviceareas/Electric_Utility_Service_Areas.html

California Energy Commission, 2015b. *Proposition 39 Utility Data Exchange Specification XML Schema files*, available upon request from the California Energy Commission at Prop39@energy.ca.gov

California Energy Commission, 2016a. *The California Clean Energy Jobs Act (Proposition 39 (K-12) Program) Overview*. http://www.energy.ca.gov/efficiency/proposition39/

California Energy Commission, 2016b. *California Clean Energy Jobs Act: Proposition 39 (K-12) Program Snapshot.* http://www.energy.ca.gov/efficiency/proposition39/documents/Prop_39_Snapshot.pdf

Cockburn, Alistair. 2002. *Writing Effective use Cases.* Boston, MA: Addison-Wesley.

CPUC (California Public Utilities Commission) 2016. *California Energy Efficiency Statistics.* http://eestats.cpuc.ca.gov/StandardTables/GuidanceDocument.aspx

DOE (U.S. Department of Energy) 2016. *BEDES Technical Documentation.* https://bedes.lbl.gov/technical-documentation

Gottesdiener, Ellen 2002. *Requirements by Collaboration.* Boston, MA: Addison-Wesley.

Green Button Alliance 2015. *Green Button Developers*. Accessed January, 2015. http://www.greenbuttondata.org/developers/

NIST (National Institute of Standards and Technology) 2013. *Green Button Initiative.* *http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/GreenButtonInitiative*

W3C (World Wide Web Consortium), 2012. W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. https://www.w3.org/TR/xmlschema11-1/

W3C (World Wide Web Consortium), 2016. About W3C. https://www.w3.org/Consortium/