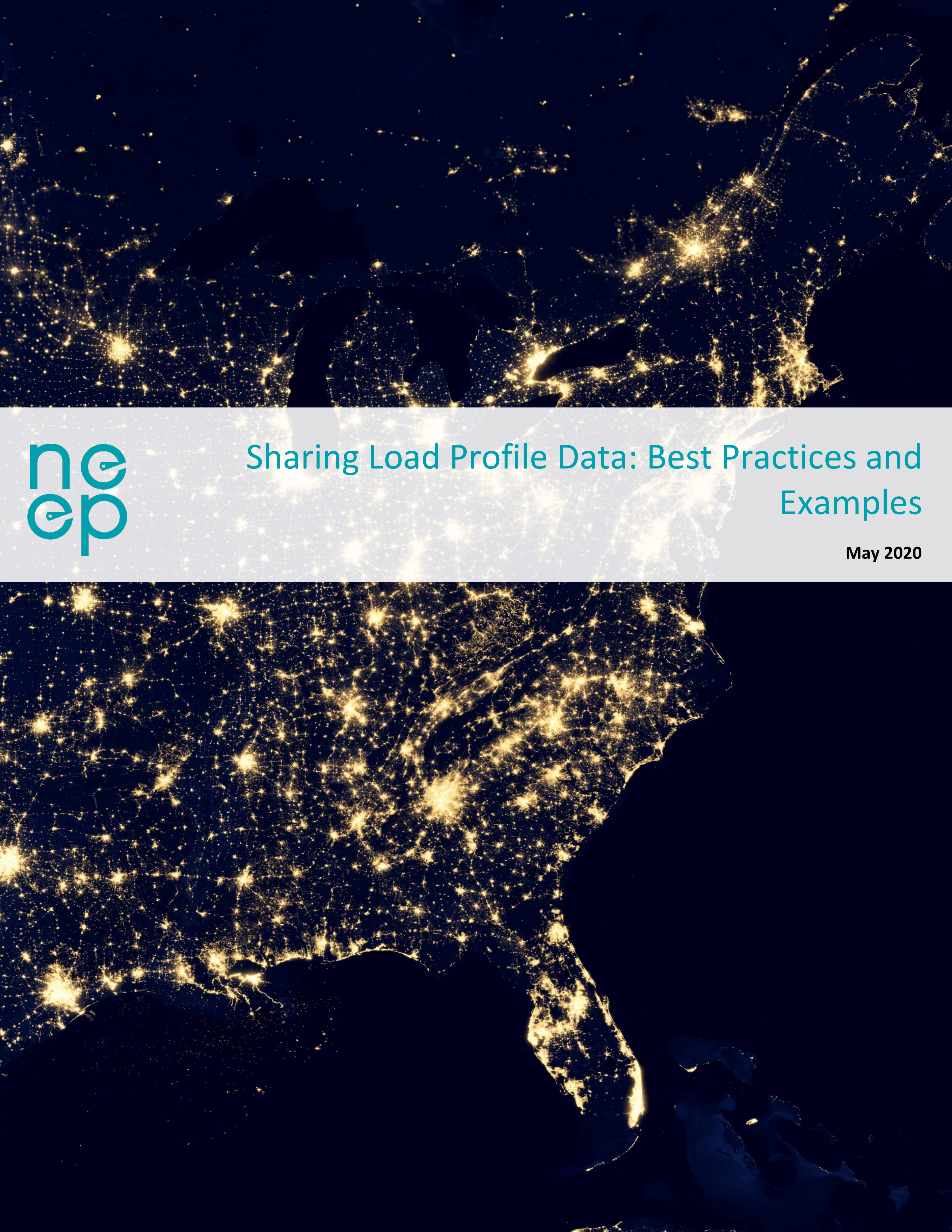


Sharing Load Profile Data: Best Practices and Examples

May 2020

Table of Contents

Executive Summary 4

Purpose and Scope 7

Background and Definitions 7

- Energy Consumption Data 7
- Load Profile 7
- Data Generation and Sharing 8

Overcoming Data Sharing Barriers 9

- Applicability 9
- Access 10
- Emerging Regulatory Support for Data Access and Sharing 10

Uses of End Use Load Profile Data 12

Best Practice Guidelines: End Use Load Profile Data Sharing Projects in EE and Related Applications.... 14

Gaining Access: Examples of Individual Site Whole Building Load Profile Applications 16

Best Practice Guidelines: Individual Site Whole Building Load Profile Data Sharing Projects 17

Data Sharing Case Studies 19

- Case Study #1: EM&V Forum Load Shape Research and Data Catalog Project 19
- Case Study #2: Home Energy Labeling Information eXchange (HELIX) 24
- Case Study #3: Silicon Valley Clean Energy Meter Data Project 29

Appendix: Published Sources - Data Access Guidance 34

Acknowledgements

This report was prepared as part of NEEP's End Use Load Profiles Residential and Commercial Buildings Project, with support from the Massachusetts Clean Energy Council (MassCEC) and done in collaboration with MassCEC, the New York State Energy Research and Development Authority (NYSERDA) and the U.S. Department of Energy (U.S. DOE).

This report reflects the invaluable contributions of multiple individuals. We would like to recognize the report's lead author, Stephen Waite. Independent energy consultant. Several NEEP staff served key roles in the development of the report including Elizabeth Titus, Senior Advisor for Evaluation and Research, and Cecily McChalicher, Research and Analysis Manager. Samantha Caputo, Senior Policy & Research Associate, and Bryan Evans, Residential Program Associate, provided valuable information for the HELIX case study. Formatting and edits were provided by Lisa Cascio, Director of Partner Engagement, and Victoria Salvatore, Marketing Associate.

NEEP would like to thank other contributors to this report, including Jeremy Roberts, Green Button Alliance, Aimee Bailey, Silicon Valley Clean Energy, Devin Hampton, UtilityAPI, Michael Murray, Mission:data Coalition, Joe Buonannnata, Isabelle Hazlewood, and Selya Price, Connecticut Green Bank.

NEEP would also like to recognize and thank members of its End Use Load Shape Regional Advisory Group for their participation in reviewing this report and providing input into the creation of this document. These individuals include Joseph Borowiec, NYSERDA; Ziggy Majunder, NYSERDA; Meg Howard, MA Clean Energy Center; Peter McPhee, MA Clean Energy Center; Adam Hinge, Sustainable Energy Partnerships; Will Lange, Water Furnace Renewable Energy; Barry Coflan, Independent consultant; Ralph Prahl, Independent consultant; Michael DeSocio, NY ISO; Charles Alonge, NY ISO; Danielle Bond, CPower Energy Management; Ben Griffiths, MA Attorney General's Office; Greg Thomas, PSD Consulting; David Jacobson, Jacobson Energy Consulting; Henry Yoshimura, ISO-NE; Puja Vohra, Slipstream; Claire Cowan, Slipstream; Chris Ashley, Energy Hub; Rodney Sobin, NASEO; Pasi Miettenen, Sagewell; and Claire Miziolek, MA 80 X 50.

About NEEP

NEEP was founded in 1996 as a non-profit whose mission is to serve the Northeast and Mid-Atlantic to accelerate regional collaboration to promote advanced energy efficiency and related solutions in home, buildings, industry, and communities. Our vision is that the region's homes, buildings, and communities are transformed into efficient, affordable, low-carbon resilient places to live, work, and play.

Executive Summary

Energy consumption data are used in diverse applications, ranging from research to energy efficiency program impacts to business and real estate investments. Useful energy data are not widely available due to the cost of data collection, restrictions on authorized data access, limited sources of publicly-available data, and other barriers. Data sharing makes existing data available to users who do not have the resources to undertake their own data collection activities. Those who cannot undertake primary data collection must prioritize their information requirements and seek to identify opportunities to leverage existing data that can serve as a reasonable alternative or supplement to primary data collection.

This brief provides guidance to facilitate effective data sharing among energy data users. It presents background information and definitions important to data sharing. It outlines data sharing barriers and best practice guidelines to help overcome those barriers for several applications in which data sharing is feasible and potentially beneficial. Three data sharing case studies are presented. They demonstrate the best practice guidelines in actual projects. While the primary focus of the brief is sharing electric load profile data, it is relevant to other energy-related uses, as illustrated among the case studies.

A load profile is a time series of energy consumption values that characterizes the variation in electric demand measured or estimated at the same site during a calendar year. **Whole building load profile** data quantify total site energy consumption whereas **end use load profile** data quantify energy consumption of specific equipment or equipment systems that serve a designated end use.

There are several barriers to data sharing for both end use load profile and whole building load profile data. The principal barrier to end use load profile data sharing is limited *applicability* of existing data to diverse applications and end-user populations. The central barrier encountered by users of whole building load profile data is gaining *access* to the required source data. Another barrier common to all types of data sharing is the lack of *transparency* of the process employed to generate the source data.

Best practice guidelines for end use load profile data sharing projects are outlined in this brief as a series of seven elements.

1. Select the measure or end use categories.
2. Define the required load profile parameters and compliance standards.
3. Define the measurement boundary.
4. Specify the normalization variables.
5. Specify the level of site aggregation and segmentation.
6. Create a flexible user interface.
7. Explore opportunities to leverage secondary data.

Best practice guidelines for individual site whole building load profile data sharing are outlined in this brief as a series of 10 elements.

- Customers should be informed by the service provider regarding its privacy and data security policies and practices.
- Customers should have complete access to their own data.

- Customers should have the ability to grant and revoke third party access to their data.
- Prior customer consent for data disclosure to a third party should be required.
- Customer data should be secured against unauthorized access.
- Service providers should adopt a standardized data exchange protocol for the authorized transfer of customer energy usage data to a third party.
- Third parties seeking access to customer data should be required to register with the service provider and execute a contractual agreement to binding terms of use.
- Third party access to customer data should be available without prior consent if the shared data set cannot be used to reveal the identity of an individual.
- The methodology employed to prevent customer identification should conform to accepted standards of data aggregation or anonymization.
- Enforcement of security policies should be practiced.

The elements draw upon documents published by public and private stakeholders that address regulatory policy objectives and options, policy recommendations, proposed principles of conduct and other issues related to energy data sharing practices.[1] The guidelines are intended to represent the core elements of a comprehensive approach to data sharing policy. It is important to note that the elements presented above for both end use load profile and whole building load profile data sharing are situational; some may not apply to certain data sharing applications.

Each case study is an example of recent or ongoing successful energy data sharing projects, illustrating how some of the best practice elements are being implemented.

Case Study 1: The EM&V Forum Load Shape Project demonstrates the feasibility and scope of effective end use load profile data sharing, given careful attention to critical elements of applicability in source data generation. A number of useful insights can be gained from this example of best practice:

> A well-designed, regional approach to primary data collection and analysis can leverage limited program administrator resources and budgets for data acquisition. Pooling program administrator resources and existing data is an effective strategy when budgets for EE program planning, evaluation, and research are insufficient to fund independent primary data collection activities to develop end use load profile data for each program administrator.
>
> The barriers to end use load profile data sharing vary depending on the measure category. Different study designs are required to promote applicability and minimize the cost of primary data collection. Each measure category, therefore, warrants a separate study to minimize inefficiencies in data collection and analysis.
>
> Developing a flexible user interface that allows for user-specified end use load profile parameter definitions and population segment weights enhances the applicability of the source data. The user is enabled to efficiently adapt the shared data to a unique target population and use case and also update parameter

[1] See Appendix A: Published Sources - Data Access Guidance

definitions and segment weights at a later time to reflect changes in the target populations, wholesale electricity market rules, peak/off-peak periods, etc.

Case Study 2: Home Energy Labeling Information eXchange (HELIX) is a secure, cloud-based, open-source data platform that was developed to enable the transfer of home energy labels and certifications to real estate Multiple Listing Services (MLS) in order to help promote transparency (and more accurate market valuation) of home energy attributes, as approved by the seller. It illustrates best practice learnings about data sharing because its purpose and design make energy data more publicly accessible in the real estate market, and construction of HELIX required overcoming the data access barrier. HELIX demonstrates the feasibility and scope of effective data sharing for a primary purpose and could be used for multiple other purposes. Insights from this example of best practice include:

A well-designed, regional approach to data collection and analysis can leverage limited resources and available sources of information. Use of home energy labels for real estate communications can encourage market transformation and enhance the value of the investment that states, utilities, and residential customers made in energy efficiency.

The barriers to data sharing vary depending on the organization providing the data and the terms they require. Each data source, therefore, warrants a separate negotiation.

Standardization of data fields where possible can ensure accuracy and add efficiency to the process of data sharing.

Case Study 3: Silicon Valley Clean Energy (SVCE) Data Hive project is a led by SVCE, a community-choice aggregator serving thirteen communities in Santa Clara County, California. Clean energy providers and customers interact with SVCE via the Data Hive, a data platform provided to SVCE by UtilityAPI. The SVCE Data Hive's ability to share data between a customer and a vendor or prospective provider to develop behind-the-meter clean energy projects (e.g. installing solar photovoltaics, battery storage, electric vehicle chargers, etc.) is essential to achieving SVCE's mission to achieve community-wide deep decarbonization. The platform and associated data sharing features demonstrate the feasibility and scope of effective sharing of individual customer metering data sharing. Useful insights from this example include:

A well-designed, efficient and secure system for primary data sharing can leverage resources and minimize burdens on customers, utilities, aggregators and prospective service providers by streamlining and standardizing processes. It can also level the playing field for market participants.

Developing a flexible user interface that provides the option of user-specified selections enhances the applicability of the source data.

Purpose and Scope

The purpose of this brief is to provide guidance to facilitate effective data sharing among energy data users. The brief first presents background information and definitions important to data sharing. It then outlines data sharing barriers and best practice guidelines to help overcome those barriers for several applications in which data sharing is feasible and potentially beneficial. This is followed by a presentation of selected data sharing case studies that demonstrate the best practice guidelines in actual projects. While the primary focus of the brief is sharing load profile data, other energy-related uses will be included among the selected case studies.

Background and Definitions

This section provides background information and key terms important to understanding data sharing processes and applications.

Energy Consumption Data

Energy consumption data are used in diverse applications with specialized requirements pertaining to the source and attributes of the required information. Useful energy data is not widely available due to the cost of data collection and restrictions on authorized data access. The primary source of available site-specific energy consumption information is **utility meter data.** Electric utilities routinely collect this consumption data to determine the amount of revenue to bill customers for energy supply and delivery.

Utility meter data can be used in certain applications that require *whole building* annual, monthly, daily, hourly, or more frequent kWh consumption data over a specified period of time. Such applications include building system energy management, building simulation calibration, time of use and interruptible rate design, baseline studies, demand response analysis, distributed energy resource (DER) analysis, building energy benchmarking, utility resource planning, and analyzing energy consumption impacts on power plant emissions.

Other applications require *end-use* energy consumption data that are not routinely collected to support electric utility operations. However, since the 1980s, energy efficiency (EE) program administrators have conducted site measurement and verification (M&V) of program participant energy savings and specialized end-use research to develop equipment load profiles required to analyze programs and measure peak demand reduction and the seasonal peak and off-peak period distribution of annual energy savings.

Load Profile

A load profile is a time series of energy consumption values that characterizes the variation in electric demand, measured or estimated, at the same site over the course of time (e.g. a calendar year). Load profile data represent the typical load "shape" of annual energy consumption, i.e., the breakdown of annual consumption into consecutive discrete time intervals. A load profile can be expressed in absolute values (e.g., kWh per time interval) or normalized values (e.g., kWh per time interval/connected kW). Load profiles that map to a specific time (e.g. a given calendar year), can serve as inputs to models built on reference year assumptions, such as resource planning. For example, load profiles adjusted to reflect normalized weather years are used to describe usage under average conditions and can support program planning or qualifying resources for wholesale markets. The term **load shape** is often applied to normalized load profiles. In EE applications, end use load profiles are often normalized on the basis of connected kW or nominal equipment capacity such as tons of cooling or controlled horsepower of a motor control.

Whole building load profile data quantify total site energy consumption whereas **end use load profile** data quantify energy consumption of specific equipment or equipment systems that serve a designated end use. In EE M&V applications, collecting end use load profile data may be restricted to the operation of the equipment (measure) that is affected by an efficiency upgrade.

The scope and scale of collecting end use energy consumption data is often constrained by budget and time limitations. Data sharing makes existing data available to users who cannot undertake their own data collection activities. Those who cannot undertake primary data collection must prioritize their information requirements and seek to identify opportunities to leverage existing data that can serve as a reasonable alternative or supplement to primary data collection.

Data Generation and Sharing

The following terms are used to describe the process of data generation and sharing:

- **Site data**: information collected at a specific site of energy consumption.
- **Source data:** information derived from site data for a specific use.
- **Data sharing:** the process whereby one or more potential users are given access to site or source data, where access and use are subject to statutory, regulatory, contractual and code of conduct restrictions.
- **Data access:** the authorized transfer of source data.

Table 1 illustrates the derivation of source data from site data for two different data applications.

Table 1. Site Data and Source Data

Site Data	Source Data	Application
Whole Building Interval Data	Monthly Energy Consumption and Peak Demand	Customer Billing
End-Use Interval Data	EE Measure Seasonal On/Off Peak Energy and Coincident Peak Demand Savings	EE Program Benefit-Cost Analysis

The distinction between site and source data is fundamental to understanding the data sharing process. The microdata collected at each site are the raw inputs used to derive the source data which contain the information required in a specific application. In this context, the term **interval data** refers to any time series of site measurements such as energy consumption, operational status, thermostat set point, space occupancy, ambient temperature, etc. In the two applications presented in Table 1, the site data must be sufficient to derive monthly billing determinants for every customer or the determinants of avoided energy and capacity costs for EE measures installed by program participants.

Advanced metering infrastructure (AMI) has dramatically expanded the production of high-resolution time-differentiated energy consumption data that can be used to derive whole building load profile source data for any site where a smart meter has been installed. Initiatives such as Green Button Connect (GBC) enable utility customers to gain access to their usage data and to grant access to selected third parties. Furthermore, several states have authorized access by third parties to whole building load profile source data that is sufficiently aggregated to preserve individual customer anonymity.

These developments, however, have not created a corresponding increase in opportunities for data sharing in applications that require end use load profile data. EE program planning and evaluation depend on *separate* quantification of time-differentiated energy consumption by the specific types of equipment that are targeted as program measures. Other applications, notably load forecasting and integrated resource planning, utilize methods that aggregate energy consumption and hourly load projections for different building sectors and end-use categories to project system energy and capacity requirements. Because end use load profile data remain a scarce resource, the demand for innovative approaches to the cost-effective generation and sharing of source data that can serve diverse applications will continue into the foreseeable future.

Overcoming Data Sharing Barriers

There are several barriers to data sharing for both end use load profile and whole building load profile data. The principal barrier to end use load profile data sharing is limited *applicability* of existing data to diverse applications and end-user populations. The central barrier to whole building load profile data sharing is gaining *access* to the required source data. Another barrier common to all types of data sharing is the lack of *transparency* of the process employed to generate the source data. Proper transparency requires adequate documentation of provenance and metadata.

The best practice guidelines presented later in this brief provide the foundation for overcoming the applicability and access barriers to data sharing. The following sections discuss these barriers in more detail.

Applicability

The first barrier to data sharing is when the data available for sharing is not **applicable** to the data application of interest. In these cases, the attributes of the source data are not aligned with the attributes of the information required by the user in a specific application. Aligning source data attributes and user needs is a threshold condition for data sharing.

Taking the Table 1 applications as an example, source data consisting of monthly accumulated energy consumption and peak demand values derived from 30-minute whole building interval site data would be applicable to customer billing, but would not be applicable to EE program benefit-cost analysis because the load profile must be able to differentiate between alternating peak and off-peak periods of consumption during each day and week of the month.

Access

The second barrier to data sharing is difficulty gaining **access** to the applicable source data. Access to customer-specific energy usage data is generally restricted, especially if the source data includes personally identifiable information, i.e. customer names, addresses and other data that can be used to determine customer identity. Limitations on data access are motivated by the recognition that energy consumers have a legitimate privacy interest in the exercise of control over the disclosure and use of their energy consumption data. Releasing data to third parties raises security and privacy concerns regarding potential breach of confidentiality, unauthorized disclosure, and use beyond the intended purpose of data sharing. For example, interval energy usage data from smart meters could reveal the approximate location of the customer and patterns of household occupancy. This data may pose a risk that a residence could be targeted for criminal or other unauthorized purposes.

Barriers to access are encountered in applications that require energy consumption or energy-related data associated with an individual customer site. Utility customer AMI data provide a rich source of site-specific whole building load profile and energy cost (billing) information that has a wide range of potential applications. Utilities, however, are subject to statutes, regulations, and commission orders that govern third-party access to customer data. Moreover, utilities typically have established their own internal policies that restrict access to customer usage data. In addition to the potential risk of legal exposure for unauthorized data sharing, utilities have concerns regarding the additional costs associated with developing and implementing systems to enable sharing large volumes of historical usage data as a routine business operation. Access to data may also be subject to limitations imposed by the cost of data retrieval, consolidation and transfer to a community of potential users.

The same issues confronted as barriers to access to customer meter data also apply to end-use data. Increasingly connected device manufacturers have this data, but sharing it with third parties can be problematic.

Emerging Regulatory Support for Data Access and Sharing

Despite the challenges, recent developments at the national and state levels are sending the message that data sharing is growing in importance, enough to warrant the effort and innovation required to make it happen.

On the national level, several sources of guidance on how to protect data privacy and security are now publicly available. Formal standards and principles addressing data privacy help demonstrate that providing data access and protecting privacy and security are not mutually exclusive. The North American Energy Standards Board (NAESB), which is a membership organization, recently developed NAESB Retail Data Privacy Standard REQ.22, *Third Party Access to Smart Meter-based Model Business Practices*. NAESB developed this national standard for data privacy to satisfy distribution company and regulatory needs and interests. It also includes a testing and certification program for third party users of data.

DataGuard is a free resource developed by the U.S. Department of Energy in 2013. Like NAESB REQ.22, it is a set of voluntary principles related to aggregating and anonymizing data and other security practices. It also includes guidance on interactions with customers. While it is less specific than the NAESB standard, it was designed with unregulated third party needs in mind. Entities that self-certify that they have adopted DataGuard principles can use the DataGuard brand to provide some level of assurance to customers and the market that they use data responsibly and protect customer privacy.

Mission:data is a non-profit coalition of companies including smart technology manufacturers, energy management firms, and software companies that empowers consumers with access to their own energy usage data. Mission:data supports customer-friendly permission-based data access policies throughout the country. This organization began at the same time AMI meters became more prevalent in utility service territories. Many utilities have terms and conditions addressing customer permission for data sharing, but some of these processes and requirements can be barriers to implementing data sharing, especially at scale. Mission:data advocates for two clear principles: 1) customer-level data should be electronically accessible, and 2) data sharing should be standardized across the country, for example through state utility commission adoption of Green Button Connect. Mission:data's 2017 report, *Energy Data: Unlocking Innovation with Smart Policy*, synthesizes data sharing policies of some leading states and the District of Columbia into a comprehensive guide. It discusses policy elements relevant to state regulatory oversight, including to privacy, consumer protections, technical standards, and enforcement issues.

Pecan Street is a 10-year-old non-profit organization with a mission that includes providing "access to the world's best data on consumer energy and water consumption behavior, testing, and verification of technology solutions and commercialization services" with a goal of accelerating the transition to clean energy.[2] It was born at the University of Texas at Austin. A global network of researchers and partnerships with Stanford University, UC Berkeley, and Cornell University helped Pecan Street become a key source of data and analysis for technology startups and established corporations. Pecan Street specializes in modeling, testing, and verifying new smart home technology, electricity pricing, electric vehicle infrastructure, solar energy, and energy storage.

Dataport is Pecan Street's data platform, housing a unique and robust collection of disaggregated residential electricity data. Pecan Street measures circuit-level electricity use and generation from a "test bed" of nearly 1,000 volunteer homes across its research network in Texas, California, and New York. Those measurements are anonymized in accordance with its privacy policy and made available for free to university researchers or to non-university users via commercial licenses, both individual and organizational. Datasets specific to electric vehicle charging, rooftop solar generation, energy storage and energy use down to individual household circuits, including heating and cooling systems are available. In addition to data measured and collected by Pecan Street, Dataport also includes curated datasets from other sources, including the U.S. government, weather institutions, utility industry institutions, and others that aid researchers conducting innovative analyses.

On the state level, California has been on the leading edge of developing energy data sharing and access policies. In 2011, the California Public Utilities Commission (CPUC) adopted rules protecting the privacy and security of customers' electric usage data generated by smart meters and transmitted by the smart grid. The CPUC's decision adopted 1) reporting and audit requirements regarding the utilities' data privacy and security practices, third-party access to customer usage information, and security breaches of customer usage information, and 2) policies to govern access to customer usage data by customers and by authorized third parties.

In July 2019, New Hampshire enacted SB 284, which requires utilities to create and operate a multi-use secure online data platform that provides access to information about personal energy usage to consumers and

---

[2] See https://www.pecanstreet.org/about/

stakeholders. The NH utilities can charge third parties for data access through the multi-use, online data platform and retrieve costs from customers in a manner approved by the commission. The platform's features include:

- A common foundation of energy data
- Adherence to standards regarding data accuracy, retention, availability, privacy, and security
- User-friendly interface
- Opt-in requirements for third party access
- Voluntary involvement of municipal and rural co-ops

In early 2020, the New York Public Service Commission (NY PSC) initiated a proceeding to address the strategic use of energy-related data. NY PSC staff will file two white papers for public comment. The first paper will address creating an integrated energy data resource that would provide a platform for access to customer and system data, and the second will focus on developing a data access policy framework that standardizes the necessary privacy and cybersecurity requirements for access to energy related data. More information about the 2020 proceeding may be obtained from www.dps.ny.gov (Case Number 20-M-0082).

This builds on the NY PSC's 2018 order (CASE 18-M-0084) and Governor Cuomo's Green New Deal. The order established guiding principles to serve as foundational elements for developing policies that appropriately balance privacy concerns with the rapidly changing energy marketplace. These include:

- Increase customer familiarity with and consent to appropriate data sharing;
- Movement towards improved access by third-party service providers to customer energy usage data, consistent with consent;
- Linking customer energy usage data with other sources of building data, energy use drivers, and energy systems data to enable enhanced identification of EE/DER opportunities;
- Providing that the mechanisms for appropriate access to customer energy usage data are implemented in a useful, timely, and quality-assured manner.

In addition to these national and state-level efforts, NEEP recently prepared guidance on data aggregation best practices and a report on Utility Data Access for the Public Sector related to building benchmarking.

Uses of End Use Load Profile Data

Several critical attributes of existing source data must be carefully considered when evaluating applicability to a given use. These attributes should guide the development of new source data for users with similar information requirements. The information requirements of various applications can be broadly classified according to two principal source data attributes, level of *aggregati*on and *measurement boundary*.

The leading applications of end use load profile data require a certain level of site aggregation. This is because the end use load profiles are being used to characterize the diversified hourly demand of specific types of equipment being operated at sites within different building sectors of a designated population of end users. The load profile of the aggregated demand determines the energy and capacity resource requirements associated with a particular end use and end user segment of the population.

Several leading applications of end use load profile data are outlined below: EE program planning and evaluation, demand response, integrated resource planning, and electrification.

Energy Efficiency Program Planning and Evaluation entails diverse activities that require end use load profile data to quantify the impact of specific EE measures on hourly system electric loads throughout the calendar year. End use load profiles are used in benefit-cost analysis to determine the amount of annual measure savings that occurs each hour of the year in order to quantify annual avoided energy and capacity costs, which are the primary benefits of efficiency upgrades by program participants. EE potential studies and program design depend upon assessments of measure cost-effectiveness and the magnitude of achievable cost-effective energy savings in different residential, commercial and industrial market segments. Projections of EE program impact on utility and regional hourly load forecasts require energy savings inputs and end use load profile data that are conformable to the sector and end-use structure of forecasting models. M&V of annual energy savings achieved by measures installed by program participants often rely upon end use load profiles to quantify the coincident peak load demand reduction associated with different measures and customer categories.

Demand Response Planning and Evaluation (DR) provides an opportunity for consumers to play a significant role in the operation of the electric grid by reducing or shifting their electricity usage during peak periods in response to time-based rates or other forms of financial incentives. Demand response programs are being used by some electric system planners and operators as resource options for balancing supply and demand. As with energy efficiency planning and evaluation, end use load profiles meet diverse needs for demand response programs. The profiles identify the timing and magnitude of peak consumption, both of which are important considerations when designing strategies and programs for reducing or managing peak consumption as part of demand response program planning and design.

Integrated Resource Planning (IRP) relies on long-term forecasting of aggregate energy demand by a large and diverse population of individual energy consumers. The demand for energy is a derived demand, so called because the quantity of consumed energy is directly dependent on the demand for a specific end use (e.g. lighting, space conditioning, etc.). An accurate multi-year projection of aggregate energy demand therefore requires a disaggregated forecast of end-use service demand by major segments of the consumer population within a utility service territory, or a larger region, depending on the resource planning context. The structure of the energy demand forecasting model determines the level of site and end-use aggregation of load profile data.[3]

Electrification refers to programs, policies, and market forces that increase the demand for electricity. There is a growing movement to convert end-use technologies that consume fossil fuel to electric-powered alternatives as part of a comprehensive building decarbonization strategy. System planners utilize end use load profiles to determine the impact of projected conversion of heating, water heating, process loads, and motor vehicles to electric technologies on future generation, transmission, and distribution resource requirements. Heat pumps and electric vehicles are key electric technologies that will influence the demand for electricity under electrification.

[3] For example, the California Energy Commission electricity demand forecasting model process uses the Hourly Electricity Load Model to convert annual energy use forecasts to hourly demand forecasts by application of whole-building, end-use, and EE load shapes. There are two sectors (single family and multifamily) and twenty-four end-uses in the residential model. The commercial sector consists of twelve different building types, with ten end-uses.

Resource planners and EE program administrators have limited resources available for data acquisition. This is problematic given the scope and scale of end use load profile data requirements for the applications described above. Because ongoing primary data collection is essential to developing accurate end use load profiles, users with similar information requirements must adopt a collaborative approach to data generation and combine their resources. Joint funding of primary data collection and analysis can generate source data more cost-effectively than independent data acquisition projects that are focused exclusively on data applicability to each planner's or program administrator's end-user and program participant population. In most cases, even a large multi-state utility or regional planning organization would not have sufficient resources to independently fund the generation of the source data necessary to comply with minimum standards of data accuracy and validity for a given application.

Taking EE program planning and evaluation as an example, program administrators located within the same region can jointly fund primary data collection projects which are designed to generate end use load profile source data. This data can be used not only to serve the needs of every sponsor, but also potentially the requirements of non-sponsors in the same region, or even those of program administrators in other regions. Identifying the key elements of project design that can achieve this objective establishes the basis for the following best practice guidance.

Best Practice Guidelines for End Use Load Profile Data Sharing Projects in EE and Related Applications

The following guidelines for end use load profile data sharing projects, like those described in the previous section, are intended to represent the core elements of a comprehensive approach to data sharing policy. They are outlined below as seven elements. It is important to note that these elements are situational; all of the steps may not apply for every project.

1: Select the measure or end use categories.

The first element is to decide which measures or end-use equipment category load profiles to produce. Collaborative projects involve certain compromises in order to achieve consensus on the scope of work. In multi-state, regional, collaborative research, program administrators often have divergent data priorities due to differences in state regulatory policies and energy market conditions that affect program and portfolio design.

2: Define the required load profile parameters and compliance standards.

A load profile parameter is a measurable characteristic of the 8760 annual hourly load profile values that is a critical determinant of the impact of efficiency measure installation (adoption) on the resource cost of end-use service. It can typically be expressed as the sum or average of a defined subset of the annual values, e.g. seasonal peak load coincidence factor (CF), annual equivalent full load hours (EFLH), seasonal on/off-peak energy period per cent of EFLH, etc.

Different subsets of users of the end use load profile data are likely to require different definitions of a given load profile parameter depending on the benefit-cost analysis framework and models employed within a state or jurisdiction and the energy and capacity market conditions. The definition of seasonal coincidence factors and on/off-peak energy period hours are likely to vary within a multi-state region.

Variations in the definition of a given load profile parameter have important practical implications for effective data sharing. EE program administrators are subject to requirements that data collection and analysis in certain applications conform to standards of statistical precision that must be addressed in project sample design and data analysis. Precision criteria may differ among program administrators, and different parameter specifications require larger samples to attain the same level of precision. These considerations underscore the importance of the creation of a flexible user interface that enables efficient calculation of user-specified load profile parameters and associated relative precisions.

3: Define the measurement boundary.

End use load profile data are, by definition, derived from separate measurements of the operation of a subset of the energy-consuming equipment at each site, as opposed to the total metered energy consumption of the facility. Again, it is important to accommodate the different source data needs of program administrators. Programmatic differences in measure categorization and corresponding differences in prescribed savings calculations (as set out in Technical Reference Manuals) have implications for the data collection protocols employed at each site.

4: Specify the normalization variables.

Normalization of end use load profile data is a critical attribute of source data that enables general applicability to a range of use cases. As noted above, the term load profile may refer to a time series of energy consumption values expressed in physical units (kWh) or as normalized values (kWh/connected kW), where the normalization variable is congruent with the units of the tracking variable used to quantify the magnitude of energy demand for every site in the target population. In EE applications, end use load profiles are often normalized on the basis of connected kW or nominal equipment capacity such as tons of cooling or controlled horsepower of a motor control.

5: Specify the level of site aggregation and segmentation.

The cost savings afforded by data sharing depends on the assumption that sample data can be pooled across different program administrator customer populations to form an aggregate estimate that is applicable to each constituent population. Because program administrator customer populations are heterogeneous due to geographic diversity in demographic, economic and climatic factors, as well as differences in program eligibility criteria, aggregate load profile estimates derived from sites selected from different program administrator populations may not be representative of any one program administrator population.

Site segmentation of the source data can mitigate the bias resulting from site aggregation of the program administrator populations. This is accomplished by identifying the principal components of end use load profile variation for a designated end use. For example, equipment operating schedules are determined by building occupancy and use. While certain equipment (e.g. refrigeration) must be in continuous operation during all hours of the year, HVAC equipment operation varies according to weather conditions as well as building occupancy schedule.

In commercial sector EE programs, for example, the space use and occupancy schedules of sites within the same building type will be relatively homogeneous compared to variation between building types. Therefore, assuming building type is the primary source of variation in the end use load profile, a pooled estimate within each segment will be relatively unbiased. The bias in the aggregate estimate derives from variation in the distribution of building types among program administrator populations, not from geographic diversity within each segment. An aggregate end use load profile estimate for each program administrator population can be calculated as the weighted average of the segment estimates, utilizing the respective population distribution weights. Aggregate estimates formed in this way mitigate the bias created by unsegmented aggregation and afford the benefit of increased precision from pooling across populations within each segment.

6: Create a flexible user interface for the end use load profile data.

Each end use load profile contains all of the consumption data required to calculate any load profile parameter that is required for a specific application. Users, however, typically do not have resources for the data analysis necessary to derive various parameter values and associated confidence intervals from hourly data. A user interface is required to automatically perform the appropriate calculations for a set of default parameters. It is also important to provide the flexibility to calculate user-defined parameters required by another application. The interface must also enable the user to specify the appropriate population weights to customer segments in order to minimize aggregation bias as discussed above.

***7: Explore opportunities to leverage secondary data**.*

Program administrators must conduct primary data collection to comply with accepted protocols and standards of practice adopted for M&V of EE measure savings. M&V end-use monitoring data collected as part of program impact evaluations are a potential source of usable site data that can supplement, or in some cases defer, new primary data collection activities.

Gaining Access: Examples of Individual Site Whole Building Load Profile Applications

Access to customer-specific end use data can facilitate market transactions that have a significant impact on building EE and site energy consumption. The demand for site-specific energy usage data has been largely stimulated by the potential benefits from effective utilization of AMI data. There are many potential applications of whole building load profile data. This section highlights several examples of those potential applications.

Customer and contractor access to whole building load profile data can facilitate deployment of distributed energy resources (DERs). Hourly usage data enable a more accurate analysis to optimize the size and cost-effectiveness of solar PV or energy storage installations and also assist in targeting DER opportunities within a geographic area.

Demand response (DR) service providers (aggregators) require access to customer AMI data to enable customer participation in wholesale energy and capacity markets.

Benchmarking provides building owners and managers with information they need to make informed decisions about building system optimization and investments in energy efficient building upgrades. Benchmarking employs building energy performance metrics such as energy use intensity (EUI) and ENERGY STAR score, which is a normative measure comparing a building to similar buildings nationwide. Calculation of the building EUI requires access to historical energy usage data, including monthly utility metered consumption. Building energy benchmarking is designed to increase energy efficiency and inform potential purchasers of a building's current level of energy efficiency and eventual energy costs.

Home energy management systems (HEMs) encompass a wide range of products that enable varying degrees of monitoring and control of home energy consumption. HEMs can access and utilize high resolution interval data to identify opportunities to reduce or shift load to optimize and integrate the operation of energy consuming equipment, energy storage devices and distributed on-site generation (e.g., solar PV). A significant benefit of this functionality is the reduction in carbon emissions that can be achieved by shifting load to be more coincident with maximum renewable DER output. HEMs can also enable remote control of devices via DR protocols established by utilities or grid operators.

Access to interval energy consumption data can assist customers and energy service providers evaluate the costs of different utility time of use pricing options and energy management measures to minimize site energy cost.

An important application of whole building load profile data is in building simulation calibration. Building energy modeling is utilized to adjust model input assumptions to achieve a closer approximation of the modeled whole building load profile to the metered data. Building energy simulation modeling applications include calculation of individual site end use load profiles, building energy benchmarking, and calculation of site-specific EE measure savings.

The scope of potential applications of individual site whole building load profile data presents a compelling argument that a concerted effort to enhance access to customer energy use data can accelerate the attainment of energy savings, renewable energy and carbon reduction goals.

Best Practice Guidelines for Individual Site Whole Building Load Profile Data Sharing Projects

The following best practice guidelines draw upon documents published by public and private stakeholders that address regulatory policy objectives and options, policy recommendations, proposed principles of conduct and other issues related to energy data sharing practices.[4] The guidelines are intended to represent the core elements of a comprehensive approach to data sharing policy.

The guidelines are presented as 10 elements of best practice for whole building load profile data sharing projects. They also provide a context within which to evaluate alternative proposals to achieve the dual objectives of enhanced data access and protection of customer privacy.

1: Customers should be informed by the service provider regarding its privacy and data security policies and practices.

---

[4] See Appendix A: Published Sources - Data Access Guidance

The service provider should present customers with an explanation of the procedures in place to enable access to their usage data and to authorize disclosure of the data to a third party. The service provider should also describe its data security policy and practice.

2: Customers should have complete access to their own data.
The service provider should establish a convenient procedure for customers to gain access to all of their own usage data.

3: Customers should have the ability to grant and revoke third party access to their data.
The service provider should establish a process whereby customers can authorize disclosure of their data to a designated third party. The process should enable the customer to set limits on the data elements, the purpose and the term of use of the shared data. The customer should also be able to rescind prior authorization for disclosure to a third party.

4: Prior customer consent for data disclosure to a third party should be required.
The data sharing authorization process should require affirmative customer consent for data disclosure to a third party. The process should be secured against disclosures based on fraudulent consent.

5: Customer data should be secured against unauthorized access.
Service providers should implement a process to protect customer data from unauthorized access, use, dissemination and corruption throughout its lifecycle.

6: Service providers should adopt a standardized data exchange protocol for the authorized transfer of customer energy usage data to a third party.
Service providers, typically utilities, should utilize a standardized protocol for the authorization of third party access to customer usage data and a standardized interface to facilitate automatic data transfer to authorized third parties. A standardized data format and interface for the exchange of customer usage information will remove barriers to the development of innovative products. This allows developers to create software that does not have to be adapted to each utility's proprietary system.

7: Third parties seeking access to customer data should be required to register with the service provider and execute a contractual agreement to binding terms of use.
Third party registration should require the execution of a contract with the service provider. The terms of agreement should address limitations on the use of the shared customer data and the third party's privacy policy that will be in effect to protect data confidentiality.

8: Third party access to customer data should be enabled without prior consent only if the shared data set cannot be used to reveal the identity of an individual customer.
Service providers should make it possible for third parties to access customer data without customer consent if the customer's identity cannot reasonably be determined from the information content of the data.

9: The methodology employed to prevent customer identification should conform to accepted standards of data aggregation or anonymization.

All identifying information should be removed from the shared data. Load profiles unique to particular customers should not be disclosed without their consent. Data aggregation standards should be sufficient to reasonably prevent customer identification but not overly restrictive. Aggregation standards should provide the flexibility to account for the geographic scope of site aggregation, the degree of heterogeneity of the aggregated load profiles and the temporal resolution of the usage data.

10: Data sharing policies should be enforced.

Regulators should encourage third parties to adopt a publicly endorsed privacy policy to which it would be held accountable.

Data Sharing Case Studies

The following three case studies are examples of recent or ongoing successful energy data sharing projects. Each case study begins with an overview of the project, followed by a discussion of how the project overcame either the accessibility or access data sharing barrier by following many of the data sharing best practice guidelines outlined in this report.

The first case study describes the how the EM&V Forum Load Shape Research and Data Catalog project overcame the data applicability barrier, while the second case study describes how the Home Energy Labeling Information eXchange (HELIX) project overcame the data access barrier. The third case study, the Silicon Valley Clean Energy Data Hive project, overcame both the data access and applicability barriers. All three case studies show the importance of transparency in data sharing projects.

Case Study #1: EM&V Forum Load Shape Research and Data Catalog Project

Northeast Energy Efficiency Partnerships' (NEEP) Regional Evaluation Measurement and Verification Forum (EM&V Forum) served the Northeast and Mid-Atlantic region through various cosponsored projects from 2009 to 2016. It supported use and transparency of current best practices in evaluation, measurement, verification, and reporting of energy and demand savings, costs, avoided emissions, and other impacts of energy efficiency. Several EM&V Forum projects developed end use load profiles or made related parameters from preexisting studies publicly available. This facilitated consistency where appropriate, and it leveraged these resources, since meter-based evaluations and load profile studies are time consuming and expensive. Here is a collection of EM&V Forum end use load profile projects presented as a case study illustrating successful examples of the data sharing best practice guidelines related to applicability.

The first of these projects took place in 2009 when the EM&V Forum joined with the Northwest Power and Conservation Council Regional Technical Forum (RTF) to commission Phase 1 of the Load Shape Research and Data Catalog project. The purpose of the study was to:

- Research and inventory existing load shape data;
- Determine what load shape data is necessary to meet utility energy efficiency program, ISO New England and PJM Capacity Markets, and air quality regulatory needs;

- Identify weaknesses in the existing data for use in EE programs, capacity markets, and air quality regulations;
- Evaluate the transferability and applicability of existing load shape data to the regions;
- Provide a road map for meeting future short term and long term end-use metering needs.

The results of these activities included:

- Overview of the approach to identifying usable end use load datasets and the results of the data collection effort;
- Summary of needs of users of end-use load shape data by EE planners, capacity market operators, and air quality analysts;
- Metrics for determining usability of end-use load datasets and review of issues related to transferring data between regions;
- Synopses of studies that were classified as “usable” and review of gaps in available data;
- A high level prioritization scheme for load shape improvements for each region;
- Road map for coordinating future end-use data efforts;
- Recommendations for near-, mid- and long-term load shape improvement efforts.

In 2010, the EM&V Forum, representing utility sponsors in the Northeast and mid-Atlantic regions, initiated Phase 2 of the project, which consisted of a series of studies to develop end use load profile data for C&I EE applications by the EM&V Forum sponsors. Four separate studies were subsequently completed for different commercial program measure categories. The following reports document the methodology and results of the research:

- Lighting Load Shape Project Final Report (2011)
- Unitary HVAC Load Shape Project Final Report (2011)
- Variable Speed Drive Load Shape Project August (2014)
- Refrigeration Load Shape Project (2015)

These studies were all designed to provide source data for the following applications:

Forward Capacity Market Demand Resource: Quantification of EE measure demand reduction during peak load hours as defined by ISO New England and PJM market rules governing participation in wholesale forward capacity markets.

EE Program Benefit-Cost Analysis: Quantification of 1) EE measure energy savings during summer/winter, on-peak/off-peak energy costing periods, and 2) EE measure demand reduction during peak load hours for the purpose of calculating avoided energy and capacity cost benefits of program implementation.

Emissions Analysis: Quantification of efficient lighting savings during High Electric Demand Days (HEDD) for the purpose of air-quality regulation modeling.

The Load Shape source data requirements to support these applications were specified in the project scope of work:

- Weighted aggregate 8760 hour load shape (normalized end use load profile);
- Disaggregation of the load shape by facility type, size and weather zone in order to provide for the calculation of aggregate load shapes that reflect the facility composition of different program administrator customer populations;
- Statistical accuracy (precision and bias) in compliance with FCM M&V standards;
- Spreadsheet Tool to calculate user-specified end use load profile parameters:
    - Coincident Demand Reduction (kW)
    - Coincident Energy Savings (kWh)
    - Maximum Demand Reduction (kW)
    - Full Load Equivalent Hours
    - Annual Demand Reduction (kW)
    - Annual Energy Savings (kWh)
    - Period kWh savings / Annual kWh savings
    - Number of period hours

Table 2 summarizes the data generation process employed in each study.

Table 2. EM&V Forum Load Shape Study Data Generation Process

Data Generation	Lighting	Unitary HVAC	Variable Speed Drive	Refrigeration
Measures	Fluorescent, HID, LED	Unitary HVAC	VSD retrofit of HVAC SF/RF/CWP/HWP	ECM retrofit/control of EF/ASDH
Measurement Boundary	Site-level aggregation	HVAC unit	Equipment type	Equipment circuit
Normalization	Connected kW	Connected kW	Controlled motor HP	HP/door
Site Aggregation	All sites	Weather region	Equipment type	Equipment type
Site Segmentation	Facility type Customer size	Cooling capacity	Weather region	Weather region
Data Collection	Secondary	Primary/secondary	Primary	Primary/secondary

As a companion to the Phase 2 studies, the EM&V Forum and its sponsors commissioned a set of data collection protocols. The purpose of the protocols was to identify the site specific metadata that is important to record and save along with any metering data collected from field studies so that applicability of the data for various purposes can be evaluated and data can be leveraged appropriately.

Many aspects of the EM&V Forum studies illustrate best practices relating to ensuring applicability of data. Highlights are summarized in Table 3 below.

Table 3. Data Applicability Best Practices and EM&V Forum Load Shape Study Examples

Data Applicability Best Practice Elements	EM&V Forum Study Experience/Example of Best Practice
Select the measure or end use categories.	Informed by the Phase 1 findings and recommendations, the EM&V Forum sponsors reached consensus on the four measure categories listed above as priorities for end use load profile data development in Phase 2 of the project after considerable deliberation.
Define the required load profile parameters and compliance standards.	Program administrators located in New England and the Mid-Atlantic states must comply with the M&V standards established by ISO New England and PJM for demand resource offers into the respective forward capacity markets (FCM). The site selection in each study was therefore designed to yield estimates of the aggregate Coincident Demand Reduction parameters as defined by the FCM rules in conformity with the standards of precision adopted in the M&V standards. The end use load profile users interviewed during the Phase 1 assessment concluded that compliance with wholesale capacity market standards of precision and accuracy for measurement and verification of demand resources would suffice to comply with other regulatory requirements for states' reporting energy efficiency program impacts.
Define the measurement boundary.	In every study, each type of equipment was separately metered. The refrigeration study measured the load on circuits that in some cases were connected to several units of the same type. The lighting study aggregated the lighting logger data for all fixtures at the site, weighted by connected kW, because 93% of the loggers were installed on the same fixture type (linear fluorescent).
Specify the normalization variables.	As noted earlier, the term load profile may refer to a time series of energy consumption values expressed in physical units (kWh) or as normalized values (kWh/connected kW), where the normalization variable is congruent with the units of the auxiliary tracking variable used to quantify the magnitude of energy demand for every site in the target population. The EM&V Forum End Use Load Profile studies employed normalization variables specified to be congruent with the

	program tracking data recorded for the program administrator participant populations.
Specify the level of site aggregation and segmentation.	With the exception of the Unitary HVAC study, aggregate estimates of measure category end use load profiles were derived from the entire sample of available sites with installed measures. The level of aggregation of Unitary HVAC sites was one of six weather regions that encompassed all of the program administrator service territories located in New England, New York and the Mid-Atlantic states. Thus, data from multiple program administrator sites were aggregated within each weather region. The Lighting End Use Load Profile study sites were segmented by CBECS facility type and large or small customer size. The Unitary HVAC sites were segmented by unit cooling capacity (1-11.25 tons/11.5-100 tons). These site segmentations were employed to allow for differences in the composition of program administrator participant populations, which enables the formation of separate aggregate estimates to mitigate bias.
Create a flexible user interface.	One of the products for each of the EM&V Forum end use load profile projects was a spreadsheet tool which included normalized hourly load profiles for the measures being studied, but the tool was designed to meet flexible user needs. It is a MS Excel workbook that is designed to serve as a user interface that can generate end use load profile source data configured to represent a specific target end-user population weather region and distribution across different customer segments. The normalized load profiles can be scaled by user input of the appropriate normalizing variable. In addition to the automatic calculation of the end use load profile parameters defined above, it allows the user to specify the hours of the year used to calculate the values of parameters that are different than the default definitions.
Explore opportunities to leverage secondary data.	Based on the findings and recommendations of the Phase 1 review and analysis of completed studies and informal feedback from Forum members, the project subcommittee determined that available existing data were sufficient in quality and quantity to derive reasonable estimates of C&I lighting load shapes for the designated applications. Therefore, the study scope of work was limited to the compilation and analysis of existing measured data available at the time of project initiation. The data sources consisted entirely of interval lighting meter data previously collected by the project sponsoring utilities in their respective impact evaluations of lighting measures. The Phase 1 review determined that primary data collection was required to generate the source data required for the Unitary HVAC, Variable Speed Drive and Refrigeration data requirements. However, the Forum sponsors were able to leverage some existing end use metered site data to reduce the primary data collection costs of the Unitary HVAC and Refrigeration studies.

General Observations

The EM&V Forum Load Shape Project demonstrates the feasibility and scope of effective end use load profile data sharing, given careful attention to critical elements of applicability in source data generation. A number of useful insights can be gained from this example of best practice:

- A well-designed, regional approach to primary data collection and analysis can leverage limited program administrator resources and budgets for data acquisition. Pooling program administrator resources and existing data is an effective strategy when budgets for EE program planning, evaluation and research are insufficient to fund independent primary data collection activities to develop end use load profile data for each program administrator.
- The barriers to end use load profile data sharing vary depending on the measure category. Different study designs are required to promote applicability and minimize the cost of primary data collection. Each measure category, therefore, warrants a separate study to minimize inefficiencies in data collection and analysis.
- Developing a flexible user interface that allows for user-specified end use load profile parameter definitions and population segment weights enhances the applicability of the source data. The user is enabled to efficiently adapt the shared data to a unique target population and use case and also update parameter definitions and segment weights at a later time to reflect changes in the target populations, FCM market rules, peak/off-peak periods, etc.

Case Study #2: Home Energy Labeling Information eXchange (HELIX)

HELIX is a secure, cloud-based, open-source data platform that was developed to enable the transfer of home energy labels and certifications to real estate Multiple Listing Services (MLS) in order to help promote transparency (and more accurate market valuation) of home energy attributes, as approved by the seller. The HELIX platform is managed by NEEP with financial support from the U.S. Department of Energy (U.S. DOE) and states that currently subscribe to HELIX. While HELIX does not involve collecting end use metering data, its inputs pertain to energy data for individual residential properties. Across the region, states have different confidential information laws regarding access to customer utility information due to privacy concerns. The inability to gain access to this data, namely customer addresses, has created a barrier to realizing the many benefits of accessible home energy information. One of HELIX's goals is to help overcome this barrier.

This HELIX case study illustrates best practice learnings about data sharing because its purpose and design make energy data more publicly accessible in the real estate market, and construction of HELIX required overcoming the data access barrier. HELIX also demonstrates many of the best practice guidelines related to data access.

HELIX has been designed to populate MLS with several types of home energy labels. It is currently being populated for 13 states in the Northeast and will expand to up to 20 states by the end of 2021. It brings together home energy labels, including U.S. DOE Home Energy Score, RESNET Home Energy Rating System (HERS), and various green certifications, as well as data on photovoltaics in homes, into one central repository. HELIX is capable of being further customized to serve multiple state and local needs for access to credible sources of residential energy data. The HELIX platform contains:

- **Data**: The *Data* tab is where all accepted datasets (i.e. HERS, LEED, solar PV, etc.) are loaded.
- **Inventory**: The *Inventory* tab lists the properties with critical data points from the compiled datasets above.
- **Certifications**: Includes a list of certifications that have been accepted into HELIX for a subscriber's state.
- **Organizations**: Lists all of the organizations and sub-organizations associated to the subscriber's account.

Data security and privacy protections are highly important due to the nature of the information and type of data (i.e. home addresses) being shared between organizations. HELIX takes all legal measures and technical processes to prevent any sharing breach. This helps to assure states, data providers, real estate professionals, home buyers, and renters that the information is secure. HELIX's structure allows data to be stored by the unique source because states often have varying data providers, especially if they supply solar PV information. This structure protects the raw data of each individual provider since data cannot be shared between sub-organizations without the expressed permission of the data provider. For example, the Vermont Department of Public Service is a primary organization, and it has given permission to a suborganization, VEIC, to store data for HELIX.

Input Data

As shown in Table 4, populating HELIX requires access to home energy labeling results from many associated sources. Within any given state in the NEEP region, HELIX is populated with six default labels and certifications and has the ability to add new certification types specific to a state, such as the Efficiency Vermont Residential New Construction label. HELIX accommodates many different fields, including the building label or rating metric, the source of the rating, the year the rating was performed, the version of the building rating, and the building address.

Table 4. HELIX Certification Data Labels

Northeast	National Data Certifications & Labels in HELIX	Certification & Label Data Provider	Uploaded into HELIX
Maine	EnergyStar Certified	Environmental Protection Agency (EPA)	✓
	DOE ZERH	Department of Energy (DOE)	✓
	HERS	Residential Energy Services	✓

		Network (RESNET)	
	NGBS	Home Innovation Research Labs	✓
	LEED	United States Green Building Council (USGBC)	✓
Vermont	EnergyStar Certified	Environmental Protection Agency (EPA)	✓
	DOE ZERH	Department of Energy (DOE)	✓
	HERS	Residential Energy Services Network (RESNET)	✓
	HES	Department of Energy (DOE)	
	NGBS	Home Innovation Research Labs	✓
	Residential New Construction (RNC)	Efficiency Vermont	✓

HELIX also has access to photovoltaics (PV) data. The requested data points include the physical address, system size (kW), system capacity (typically nameplate capacity rating in kW-DC), installation date, and type of ownership. This information is obtained state-by-state through various providers including utilities, state energy offices, and public utility commissions that have different data sharing laws, as outlined in Table 5.

Table 5. HELIX Photovoltaics (PV) Data

PV Data Source	Requested PV Data Points	Uploaded in HELIX	PV Data Sharing Process
Vermont Energy Atlas	Address, city, zip code, capacity (kW), type of mount, system specs, installation date, electricity type (net metered)	✓	Publicly available with no agreement required through the Vermont Energy Atlas
New Hampshire Public Utilities	Total system capacity (kW-DC), address, date	✓	Right to Know law request to the PUC

Commission (PUC)	rebate paid, type of mount, ownership		
Massachusetts Clean Energy Center	The Production Tracking System (PTS) data fields requested include PTS ID, system name, system capacity (DC), date in service, facility type, address, city, zip code, 3rd party owner, estimated annual production	✓	Publicly available via a Freedom of Information Act (FOIA) request
Connecticut Green Bank	Requested fields include address, completion date, interconnection date, nameplate capacity (kW-DC), ownership	✓	Solar PV data is available through a FOIA request and NDA to restrict the use of the data
Vermont Energy Atlas	Address, city, zip code, capacity (kW), type of mount, system specs, installation data, electricity type (net metered)	✓	Publicly available with no agreement required through the Vermont Energy Atlas
New Hampshire Public Utilities	Total system capacity (kW-DC), address, date	✓	Right to Know law request to the PUC

States have various options to integrate solar PV data into HELIX for their use. HELIX can access data through pathways including a public database, a Freedom of Information Act (FOIA) request through utilities and state agencies, or through a non-disclosure agreement (NDA). The HELIX team works with a state entity to access solar PV information.

> Connecticut Green Bank is a quasi-public agency that administers clean energy financing and incentive programs; currently this includes incentivizing homeowners' investments in solar energy and is one source of input data on PV for HELIX.

MLS Integration

MLS integration can be done through three paths: 1) from HELIX to the MLS' data aggregator to an MLS, 2) from HELIX straight to the MLS via an API, and 3) from HELIX straight to the MLS via deep linking. Integrating new systems and data can be a slow-moving process. Privacy protections are implemented at the certification/label level in HELIX to ensure homeowners who opt-out of information sharing will not have their information shared with MLS and other parties. In order for the real estate industry to effectively utilize home energy data, the receiving MLS (and likely the MLS's tax data aggregator) must implement the Real Estate Standards Organization (RESO) Green Verification Fields, which provide a dictionary of standardized field definitions (standardization in field naming, field structure, and construct within different databases and across various platforms helps ensure that data is sent, received, and viewed consistently). For more information about RESO, see The Green [Data] Evolution by Meg Garabrant.

Table 6 lists the best practice elements related to data access and outlines how HELIX follows many of them.

Table 6. Data Access Best Practices and HELIX Examples

Data Access Best Practice Elements	HELIX Experience/Example of Best Practice
Customers should be informed by the service provider regarding its privacy and data security policies and practices.	HELIX does not interact with homeowners, rather with the home labeling organization or third-party provider of PV data.
Customers should have complete access to their own data.	Property owners with energy labels have access to their home energy label via the specific certifier, not through HELIX. Their label would also be available in the MLS listing through the Green Verification fields.
Customers should have the ability to control third party access to their data.	Customers can opt-in and opt-out to share information and scores. It is standard practice within the real estate transaction process that the homeowner has to approve the listing created by the selling agent before it is made public. HELIX leverages this process.
Prior customer consent for data disclosure to a third party should be required.	As stated above, customers have the ability to opt-in or opt-out to share their information with an MLS. No information can be shared without the permission of the homeowner.
Customer data should be secured against unauthorized access.	Terms of an NDA with HELIX outline that HELIX does not intend to sell this information, as its only use is to properly market homes with efficiency and renewable attributes. In addition, under the current database access structure, no other state energy office or utility outside of the designated data provider would be able to see the data housed in HELIX.
Service providers should adopt a standardized data exchange protocol for the authorized transfer of customer energy usage data to a third party.	In order for energy efficiency data to be effectively utilized by the real estate industry, the receiving MLS and likely the MLS's tax data aggregator must have implemented the RESO Green Verification fields. This ensures that data is sent, received, and viewed in the proper format as it was originally intended.
Third parties seeking access to customer data should be required to register with the service provider and execute a contractual agreement to binding terms of use.	States have various options to integrate data into HELIX for their use. HELIX can access data through pathways including access via a public database, via a public FOIA request through utilities and state agencies, or through an NDA. The HELIX team works with a state entity to access solar PV information. Information will not be sold or marketed for any other reason outside of real estate purposes to properly valuate and market

	homes with efficiency and renewable attributes.
Third party access to customer data should be enabled without prior consent if the shared data set cannot be used to reveal the identity of an individual customer.	Addresses are a key component in HELIX to match records and map out the data. Addresses are considered confidential information in select states, therefore customer consent is needed to share with an MLS.
The methodology employed to prevent customer identification should conform to accepted standards of data aggregation or anonymization.	This is not applicable for HELIX.
Enforcement	NEEP has contacted various states regarding confidential information laws to respect the privacy requirements that protect customer data from unauthorized access, use, dissemination and corruption and HELIX complies with these requirements. HELIX can establish an NDA, submit a FOIA request or public information request that clearly outlines the intended use for the information.

General Observations

HELIX demonstrates the feasibility and scope of effective data sharing for a primary purpose and could be used for multiple other purposes. Insights from this example of best practice include:

- A well-designed, regional approach to data collection and analysis can leverage limited resources and available sources of information. Use of home energy labels for real estate communications can encourage market transformation and enhance the value of the investment that states, utilities, and residential customers made in home energy labels.
- The barriers to data sharing vary depending on the organization providing the data and the terms they require. Each data source, therefore, warrants a separate request.
- Standardization of data fields where possible can ensure accuracy and add efficiency to the process of data sharing.

Case Study #3: Silicon Valley Clean Energy Data Hive Project

Silicon Valley Clean Energy (SVCE) is a community-choice aggregator serving 13 communities in Santa Clara County, California. It provides residents and businesses with renewable and carbon-free electricity and innovative programs to meet its goal of cutting energy-related emissions in half by 2030 from the 2015 baseline. SVCE works in partnership with Pacific Gas & Electric (PG&E). While PG&E delivers electricity over existing infrastructure and handles customer service and billing, SVCE buys clean energy directly from the source on behalf of its communities. Additionally, dozens of clean energy providers are active in the SVCE service territory offering products and services directly to SVCE customers. They range from well-known solar and battery vendors to mom-and-pop solar industry providers, as well as EV installers, and EE and demand response providers. Clean energy providers and customers interact with SVCE via the Data Hive, a data platform provided to SVCE by UtilityAPI.

The SVCE Data Hive case study showcases the ability to share data between customers and vendors is essential to the success of SVCE's operations. The platform and associated data sharing features illustrate many best practices and strategies to overcome barriers to data access, transparency and applicability.

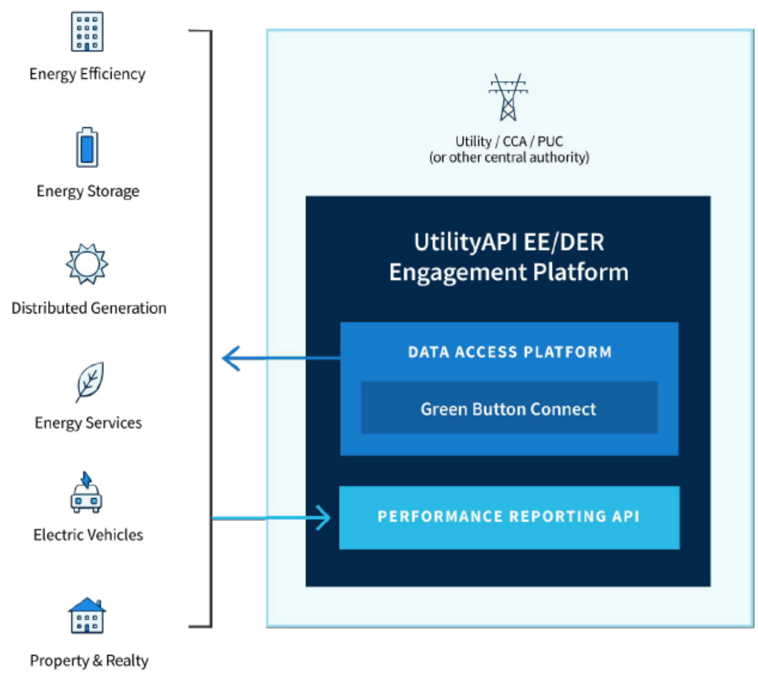
PG&E and California's other investor-owned utilities deployed their smart meter networks more than a decade ago. Revenue-grade meter data is of great value to energy services companies, energy efficiency providers and other parties that need to continuously monitor and maintain investments in building HVAC, lighting, and other major electrical loads. This data can also be invaluable for behind-the-meter batteries, smart thermostats, electric vehicle (EV) charging systems and other distributed energy resources.

Utility customers have a right to access this data and share it with whomever they want under California state law. SVCE has full access to the data from the PG&E smart meters that connect its customers to the grid. SVCE enables authorized, secure, streamlined access to applicable meter data from utility customers to registered energy vendors via the Data Hive. SVCE launched the Data Hive platform in 2020. Its features include:

- **Data standardization:** Standardization allows third parties to have easy access while empowering customers with the ability to securely transfer their data.
- **Consistency with nationally recognized standard format**: The Green Button *Connect My Data* (CMD) is the nationally recognized energy-industry standard – formally the *NAESB REQ.21 - Energy Services Provider Interface Model Business Practices* standard for enabling easy access to, and secure sharing of, utility-customer energy and water usage data. It establishes a consistent format allowing utilities to report customer electricity, natural gas, or water-usage data in any interval they choose including one-minute, 15-minute, hourly, daily, or monthly, as long as the data interval can be provided by the utility's meters. The Data Hive complies with and builds on Green Button CMD**.**
- **Easy-to-use interface:** The IT infrastructure that enable interactions between SVCE's customers and approved businesses resembles one-click services available with online retail and banking. The platform was designed to meet needs of diverse types of users, from small-scale residential solar installers to large-scale energy services providers. The Data Hive dashboard reduces the burden of access for small businesses. This was a major motivating factor in SVCE's choice of a tool for delivery of its program.
- **Streamlined authorization procedure:** Customers must authorize data sharing. Data Hive automates the process, including generation of the authorization forms which provides documentation for customers and reduces time and burden on the customers and the data recipients.
- **Accommodation of many data sources**: The platform can share sources ranging from utility smart meter databases to back-office billing systems and other stores of data, some of which may be characterized as unstructured.
- **Flexibility in applications:** Data that is shared ranges from providing historical snapshots of home energy usage to a customer to authorizing historical consumption data or ongoing streams of data to be made available to service providers.

- **Operation at scale:** As the market evolves, many businesses are likely to need to share this data. For example, in a world with demand-response providers or thermostat companies or large-scale EV deployments influencing the grid, scalability, and automation are key.

Figure 1. Data Hive Platform



Highlights of how SVCE illustrates best practices relating to ensuring applicability of data and data access are summarized in Tables 7 and 8 below.

Table 7. Data Applicability Best Practices and SVCE Examples

Data Applicability Best Practice Elements	SVCE Experience / Example of Best Practice
Select the measure or end use categories.	Billing data (costs), rates, up to 30 data points are available, including weather zone.
Define the required load profile parameters and compliance standards.	N/A
Define the measurement boundary.	SVCE has site level metering data. PV, car charging, and other measurements, and possibly also metadata describing features of a site will be available in the future.

Specify the normalization variables.	N/A
Specify the level of site aggregation and segmentation.	N/A (SVCE does not have a need to aggregate across customers)
Create a flexible user interface.	Different output formats are available to users. The user interface design is intentionally flexible; it includes a vendor dashboard and customer-friendly interface.
Explore opportunities to leverage secondary data.	The data originates from Pacific Gas and Electric. As a retailer, SVCE has to have access to the interval and billing data to carry out its core business functions, such as rate making, load forecasting, calculating the generation charges for the bill.

Table 8: Data Access Best Practices and SVCE Examples

Data Access Best Practice Elements	SVCE Experience/Example of Best Practice
Customers should be informed by the service provider regarding its privacy and data security policies and practices.	Links to security and privacy terms are included on the Data Hive customer authorization form and the website page.
Customers should have complete access to their own data.	While Data Hive's focus is authorization of vendors to have access, customers can authorize themselves as users of the data, which ensures access.
Customers should have the ability to control third party access to their data.	All Data Hive interactions are driven by customer consent (opt in). When a vendor requests data from a customer the vendor must identify what they will use it for; the customer decides if they want to share. A vendor can select all uses. A customer can also authorize data sharing directly from a vendor webpage.
Prior customer consent for data disclosure to a third party should be required.	SVCE validates customer consent via a customer authorization form.
Customer data should be secured against unauthorized access.	Front-facing Internet security via https and oauth2.0 protect against unauthorized access. Back-end security measures are also part of the Data Hive system.
Service providers should adopt a standardized data exchange protocol for the authorized transfer of customer energy usage data to a third party.	Data Hive is built on UtilityAPI's certified Green Button Connect My Data.

Third parties seeking access to customer data should be required to register with the service provider and execute a contractual agreement to binding terms of use.	SVCE requires vendors to register prior to accessing customer data and using Data Hive. The vendors are not allowed to use the system until SVCE reviews their registration submission and gives them access.
Third party access to customer data should be enabled without prior consent if the shared data set cannot be used to reveal the identity of an individual customer.	N/A
The methodology employed to prevent customer identification should conform to accepted standards of data aggregation or anonymization.	N/A
Enforcement	Because there is registration, vendors are vetted by SVCE and can be removed from the Data Hive. Also, customers have the ability to revoke their authorization to access to the data.

General Observations

The SVCE Data Hive demonstrates the feasibility and scope of effective sharing of individual customer metering data sharing. Useful insights from this example of best practice include:

- A well-designed, efficient and secure system for primary data sharing can leverage resources and minimize burdens on customers, utilities, aggregators and prospective service providers by streamlining and standardizing processes. It can also level the playing field for market participants.
- Developing a flexible user interface that provides the option of user-specified selections enhances the applicability of the source data.

Appendix: Published Sources - Data Access Guidance

State and Local Energy Efficiency Action Network (2012) *A Regulator's Privacy Guide to Third-Party Data Access for Energy Efficiency*. Prepared by M. Dworkin, K. Johnson, D. Kreis, C. Rosser, J. Voegele, Vermont Law School; S. Weissman, UC Berkeley; M. Billingsley, C. Goldman, Lawrence Berkeley National Laboratory.

Data Privacy and Smart Grid: A Voluntary Code of Conduct (VCC) (January, 2015) U.S. Department of Energy.

*Energy Data Unlocking Innovation With Smart Policy (*December, 2017) Michael Murray, Laura Kier And Bob King, P.E. Mission:data Coalition, available at: http://www.missiondata.io/s/Energy-data-unlocking-innovation-with-smart-policy.pdf.

Alexandra Klass and Elizabeth Wilson, *Remaking Energy: The Critical Role of Energy Consumption Data*, 104 CAL. L. REV. 1095 (2016), available at https://scholarship.law.umn.edu/faculty_articles/557.

*Knowledge Is Power: How Improved Energy Data Access Can Bolster Clean Energy Technologies & Save Money (*January, 2015) Ethan N. Elkind.