Home Energy Metering Study

Public Data User Guide

Northwest Energy Efficiency Alliance
700 NE Multnomah Street, Suite 1300
Portland, Oregon 97232

Prepared by:
Evergreen Economics
1500 SW 1st Avenue, Suite 1000
Portland, OR 97201

The Northwest Energy Efficiency Alliance (NEEA) is a non-profit organization working to effect market transformation through the acceleration and adoption of energy-efficient products, services and practices. NEEA is an alliance of more than 140 Northwest utilities and energy efficiency organizations working on behalf of more than 13 million energy consumers.

Table of Contents

This guide is provided to the public courtesy of End Use Load Research study participants.

Home Energy Metering Study Overview

The End Use Load Research (EULR) study is a five-year, $12.5 million effort to collect data on selected electric end-uses in both residential and commercial buildings. Respectively, these studies are known as Home Energy Metering Study (HEMS) and Commercial Energy Metering Study (CEMS). As the first end-use load research of its scale in the Northwest since the 1980s, this work will greatly support regional planning and program design. Funding for EULR is being provided outside of NEEA's 2020-2024 Business Plan.

A group of utilities and regional organizations came together to collaborate on these studies of long-term Northwest residential and commercial electricity use. This group is: Avista Utilities, Bonneville Power Administration, Building Technologies Office - U.S. Department of Energy, Clark County PUD, Energy Trust of Oregon, Eugene Water & Electric Board, National Renewable Energy Laboratory, Northwest Power & Conservation Council, PacifiCorp, Portland General Electric, Puget Sound Energy, Seattle City Light, Snohomish PUD and Tacoma Power.

The Northwest Energy Efficiency Alliance (NEEA) is project managing the Home Energy Metering Study with Evergreen Economics.

About this User Guide

This is a data user guide for the residential End Use Load Research (EULR) Home Energy Metering Study (HEMS). It is intended for utility staff, power planners, energy efficiency program managers, policy makers, consultants, trade groups and any other groups who would benefit from an accurate assessment of the electricity use in the Northwest.

The guide includes:

- Instructions for accessing the metering data;
- The database schema showing the relationships between all the tables
- A data dictionary with details on variable types and the unit of measurement; and
- A step-by-step analysis guide to walk new users through the full process of preparing the data for load-shape analysis.

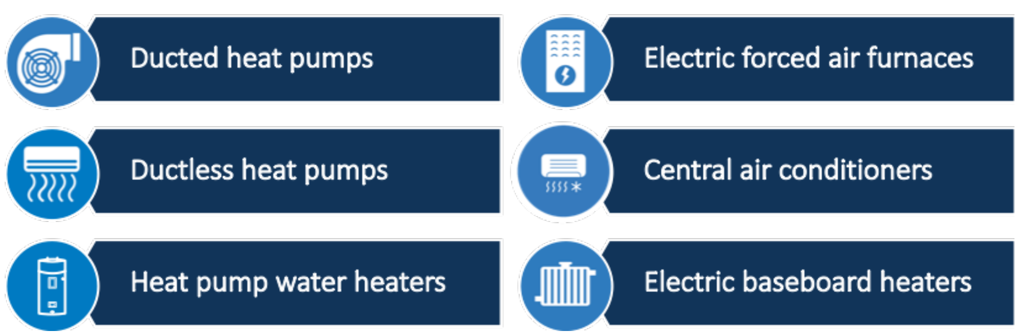
Home Energy Metering Study

The Home Energy Metering Study (HEMS) research, conducted as part of the End Use Load Research (EULR) project, continues a tradition in the Northwest of sponsoring residential end use energy monitoring studies. From 1986 to 1989, the Bonneville Power Administration monitored end use electricity demand at 499 homes across the Northwest as part of the End-Use Load and Consumer Assessment Program (ELCAP).[1] ELCAP collected hourly end use load data, but both buildings and end uses have evolved considerably since 1989. In 2012, NEEA conducted the Residential Building Stock Assessment Metering study (RBSAM)[2] that included end use monitoring at 100 homes across the Northwest. While relatively recent, the RBSAM's smaller sample size means it is difficult to draw statistically significant conclusions for any subregions of the Northwest. NEEA and other agencies in the Northwest have conducted many targeted metering studies focused on specific end uses, as well. These studies are typically smaller in scale and not intended to represent the population of the Northwest.

The main objective of the EULR HEMS study, therefore, is to develop a **current**, **representative**, and **robust** characterization of continuous energy consumption of key heating and cooling measures to support achieving clean energy goals and utility information needs, providing:

- A much more accurate assessment of the contributions of energy efficiency technologies towards reducing peak demand;
- A better understanding of how to integrate renewable energy into the grid, increasing reliability as the deployment of distributed generation and new end use technologies increases over time; and
- Prioritized data by end use for use in a range of utility functions including demand response, load forecasting, and resource planning.

The study includes collecting one-minute interval data by circuit and for the whole house from 400 homes across the region that have one or more of the following targeted equipment types:



At all homes, the study is monitoring multiple circuits and the whole home circuit, as well as indoor temperatures. Exterior temperatures are being monitored for homes where reasonable

[1] https://elcap.nwcouncil.org/

[2] Ecotope Inc. 2014. *Residential Building Stock Assessment: Metering Study*. https://neea.org/resources/2011-rbsa-metering-study

temperature estimates from nearby weather stations were unable to be assigned (i.e., for homes that are far from weather stations).

Sample Design

The objective of the sample design is to achieve confidence/precision levels of 90/10 for each of the six targeted end uses while leveraging available resources to improve the efficiency of the overall study sample. A secondary objective is to achieve robust samples across heating and cooling zones, where possible.

The study plan is to install meters in 400 homes that have at least one of the six target end uses. Our projected sample of 400 homes will result in a sample of approximately 552 target end uses as a result of some homes containing more than one target end use ("overlap").

Table 1: Final EULR Sample Allocation by End Use*

End Use	Number of End Uses
Electric baseboard heater	82
Electric furnace	102
Ducted heat pump	97
Heat pump water heater	88
Ductless heat pump	83
Central AC	100
Total*	**552**

* The total number of homes does not equal the sum of the "Homes with End Use" due to end use overlap. The total number of homes will be 400.

A sample allocation plan for the estimated 552 new EULR sample points was developed to optimize the statistical confidence and precision across all end uses in the analysis database while maintaining a minimum number of new EULR sample points for each end use. The final allocation scheme achieves a balance across the competing priorities of ensuring a sufficient number of sample points per end use across the new EULR sample, factoring in the prior metering data points, and allocating more points to end uses that have greater usage variance and/or uncertainty.

NEEA plans to install monitoring equipment, including temperature sensors and current transducers (CTs) on each circuit, in 400 homes between August 2018 and May 2023. Figure 1 provides an overview of the installation goals from the study plan (actual numbers may vary). Monitoring began in 2018, with equipment installed in 75 homes. By the end of 2019, approximately 225 homes were monitored. Installations were put on hold in early 2020 when the COVID-19 pandemic arrived. Installations will resume as the pandemic allows.

Data Quality Assurance and Quality Control

To ensure that incoming metering data will support rigorous analysis, Evergreen Economics employs a multi-stage quality assurance (QA) and quality control (QC) process on all metering data. While the primary goal of QA/QC is to ensure an accurate metering dataset, this is only possible through regular and thorough inspection of all incoming metering data. Evergreen's approach consists of four broad stages, each adding context to the data and providing new insights:

Stage 1: Data are checked to confirm that metering data are being received from all installed equipment. This simple check is critical, as there is no way to make use of data that were never collected in the first place.

Stage 2: Context is added by checking each individual observation against a set of predetermined expectations. A primary part of this process is to compare end uses to their industry standards for power usage (or generation).

Stage 3: The use of each piece of equipment is compared to how that piece of equipment has been used in the past. This enables us to confirm that changes in usage patterns are not being caused by metering equipment failure.

Stage 4: Individual data streams are contextualized with data from other sources to check the validity of each source. For example, this includes ensuring that indoor air temperature increases correspondingly and as expected with furnace usage.

Combined, these steps produce a robust metering dataset by showing that the real-world data meet expectations in multiple dimensions. These cleaned data provided the backbone to the load shape analysis that follows.

Preparing Data for Analysis

The following data are available for analysis:

- **Sites.** Geographic characteristics of each metered site: state, NW heating and cooling climate zone, time zone, and NOAA weather station assignment.
- **Points.** Technicians record descriptive data for metered circuits during the installation. This includes the circuit label, an equipment description, make/model (if available), and a cleaned end use classification of the circuit. This clean classification will make analysis much easier, as it does not require text parsing of engineer notes.
- **Temperature points.** There are one or more temperature sensors installed in each home. An indoor temperature sensor is always installed, and an outdoor sensor is added if the home is located in an area without a reliable NOAA weather station. These outdoor sensors also record relative air humidity. Temperature probes are added to monitor temperatures near heat pumps. Engineers record the location and type of each temperature sensor during installation.
- **Meter Data.** Each configured circuit at a home collects data for power, apparent power, and power factor. Total harmonic distortion is monitored at the whole home level. Temperature is collected as degrees Fahrenheit. Relative air humidity is recorded for outdoor sensors only.

Data Access

The project funders and NEEA have decided to make 15-minute interval data available to the public. The 15-minute interval data is accessible at https://neea.org/EULRdata for download as a series of CSV files. Statistical software will be needed to compile and prepare these files for analysis.

NEEA will publish metering data in Q2 of each subsequent calendar year. As needed, NEEA will reissue corrections or insertions to historical data on a quarterly basis.

WARNING: As data QC is an ongoing process, users are advised to keep track of the data version number OR download all historical data on a quarterly basis to ensure they are using the most accurate data available.

The sites, points, and temperature points tables are relatively small CSV files. These are regularly updated will act as a consistent guide to all configured circuits and sensors across the metering timeframe. The metering data is stored as a series of zipped directories, one for each quarter of the ongoing metering study as shown in figure 2. In this example, all the metering data from 2018 Q3 through 2020 Q1 is listed as version "v1" (i.e., there have not been any updates to the metering data since it was first published).

Figure 2: EULR HEMs Public Use Data Directory

2018
- 2018_Q3_v1
- 2018_Q4_v1

2019
- 2019_Q1_v1
- 2019_Q2_v1
- 2019_Q3_v1
- 2019_Q4_v1

2020
- 2020_Q1_v1

Data-Points
- points.csv
- sites.csv
- temperature-points.csv

Once unzipped, each quarterly directory contains a series of folders for each metric: power, total harmonic distortion, temperature, and humidity. For a single metric, there are subfolders for each month. Within each of these folders is a series of CSV files containing 15-minute interval measurements for all sites and circuits by day of the month.

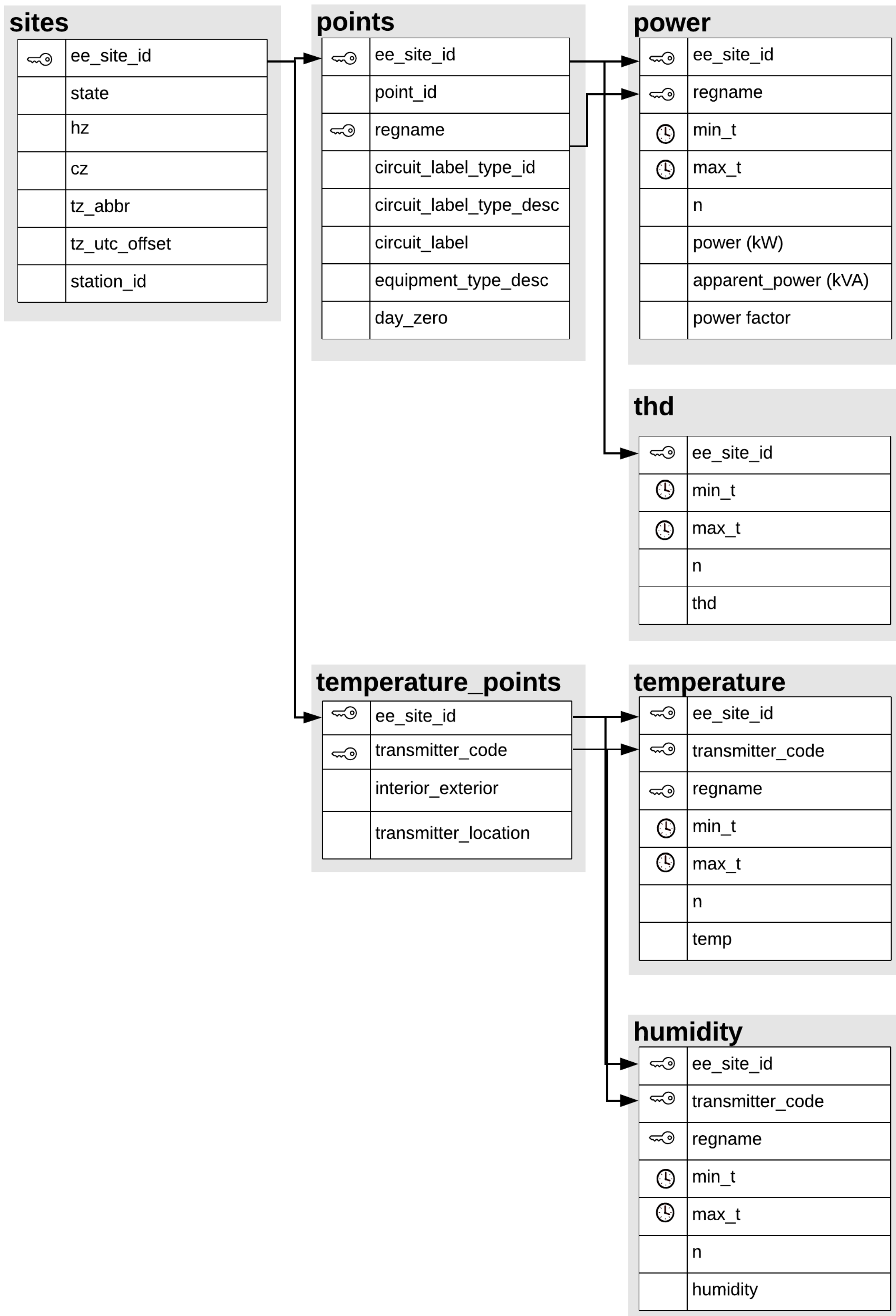
Figure 3: Unzipped Quarterly EULR HEMs Public Use Data File Structure

▼ 2020_Q1_v1	--
▶ Feb	--
▶ Jan	--
▼ Mar	--
▶ humidity	--
▶ power	--
▶ temperature	--
▼ thd	--
thd-fifteen_20200301.csv	949.2 KB
thd-fifteen_20200302.csv	947.7 KB
thd-fifteen_20200303.csv	943.7 KB

Database Schema

As shown in Figure 4, the data has been organized into a series of tables. The points and temperature points tables act as a guide, providing an ID and description of every configured register and transmitter available in the metering data. The unique site ID and circuit register name listed in points table map to the tables of 15-minute interval measurements for power and total harmonic distortion (THD). Similarly, the site ID and transmitter code listed in the temperature points table map to 15-minute interval measurements of temperature and humidity.

Figure 4: HEMS Database Schema



Data Dictionary

Table 2 provides a detailed data dictionary, listing every table and field in the data along with a description of the field.

Warning: all timestamps are listed in Coordinated Universal Time (UTC).

The local time zone of each home (*tz_abbr*) is listed in the *sites* table, as well as the number of hours local time is offset from UTC (*tz_utc_offset*) to allow for a simple conversion between UTC and local time. Please note that UTC does not include Daylight Savings Time.

Table 2: Public Use Data Dictionary

Table	Field Name	Data Type	Description
sites	ee_site_id	integer	unique site ID
	state	character	state of metered site
	hz	integer	heating zone of metered site
	cz	integer	cooling zone of metered site
	tz_abbr	character	time zone of metered site
	tz_utc_offset	integer	time zone offset from UTC
	station_id	character	NOAA weather station ID
points	ee_site_id	integer	unique site ID
	point_id	integer	unique metering point ID
	regname	character	metered circuit ID
	circuit_label_type_id	integer	numeric ID of circuit classification
	circuit_label_type_desc	character	description of circuit classification
	circuit_label	character	actual label of metered circuit
	equipment_type_desc	character	engineering notes about metered circuit
temperature_points	ee_site_id	integer	unique site ID
	transmitter_code	character	temperature sensor ID
	interior_exterior	integer	indicator of temperature sensor location (1 = indoor, 2 = outdoor, 3 = probe)
	transmitter_location	character	engineering notes about temperature sensor location
power	ee_site_id	integer	unique site ID
	regname	character	metered circuit ID
	min_t	timestamp	beginning of 15 min interval (UTC)
	max_t	timestamp	end of 15 min interval (UTC)
	n	integer	count of observations in 15 min interval

	power	numeric	average power (kW) over 15 min interval
	apparent_power	numeric	average apparent power (kVA) over 15 min interval
	pf	numeric	average power factor over 15 min interval
temperature	ee_site_id	integer	unique site ID
	transmitter_code	character	temperature sensor ID
	regname	character	temperature data stream ID (usually 1 per transmitter_code)
	min_t	timestamp	beginning of 15 min interval (UTC)
	max_t	timestamp	end of 15 min interval (UTC)
	n	integer	count of observations in 15 min interval
	temp	numeric	average temperature (°F) over 15 min interval
humidity	ee_site_id	integer	unique site ID
	transmitter_code	character	temperature sensor ID
	regname	character	humidity data stream ID (usually 1 per transmitter_code)
	min_t	timestamp	beginning of 15 min interval (UTC)
	max_t	timestamp	end of 15 min interval (UTC)
	n	integer	count of observations in 15 min interval
	humidity	numeric	average humidity over 15 min interval
thd	ee_site_id	integer	unique site ID
	min_t	timestamp	beginning of 15 min interval (UTC)
	max_t	timestamp	end of 15 min interval (UTC)
	n	integer	count of observations in 15 min interval
	thd	numeric	average total harmonic distortion over 15 min interval
weather	station_id	character	NOAA weather station ID
	min_t	timestamp	beginning of 15 min interval (UTC)
	max_t	timestamp	end of 15 min interval (UTC)
	wind_dir	integer	wind direction; angle measured clockwise between true North and the direction wind is blowing (angular degrees)
	wind_spd	integer	wind speed (meters/second)
	temp	numeric	air temperature (degrees Celsius)
	temp_f	numeric	air temperature (degrees Fahrenheit)
	dew_point	numeric	Dew point temperature (degrees Celsius)
	atm_pres	numeric	Atomic pressure; air pressure relative to mean sea level (hectopascals)

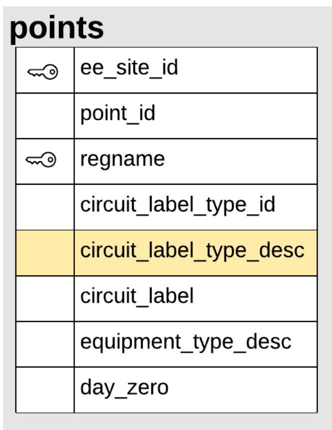
Assembling the Data

1.1 Assembling HEMS Data

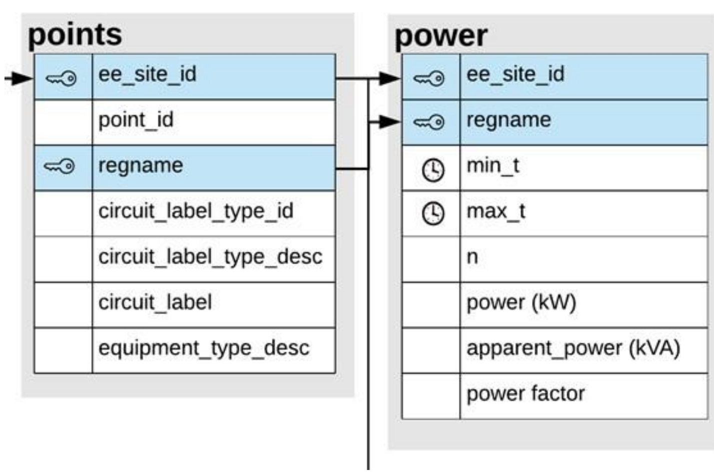
This chapter provides a step-by-step guide to prepare the HEMS data for analysis of load shapes.

After initiating the SQL database connection / downloading and compiling the current metering data (as described in the previous chapter: Preparing Data for Analysis), the basic work-flow is as follows:

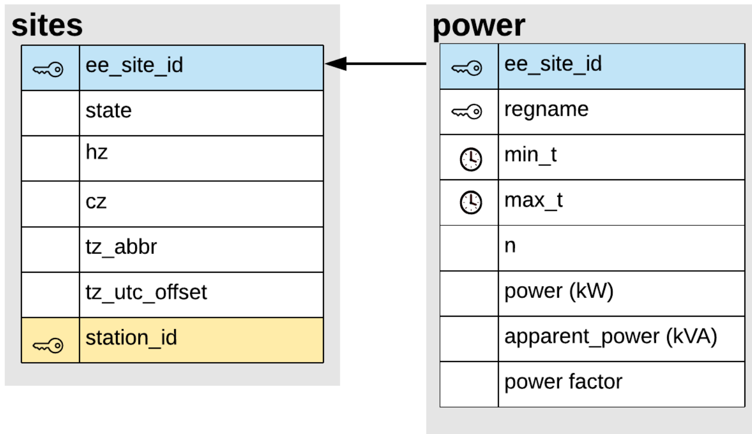
1. Use *points* table to identify all dedicated circuits connected to the end use of interest by filtering on *circuit_label_type_desc* (yellow).

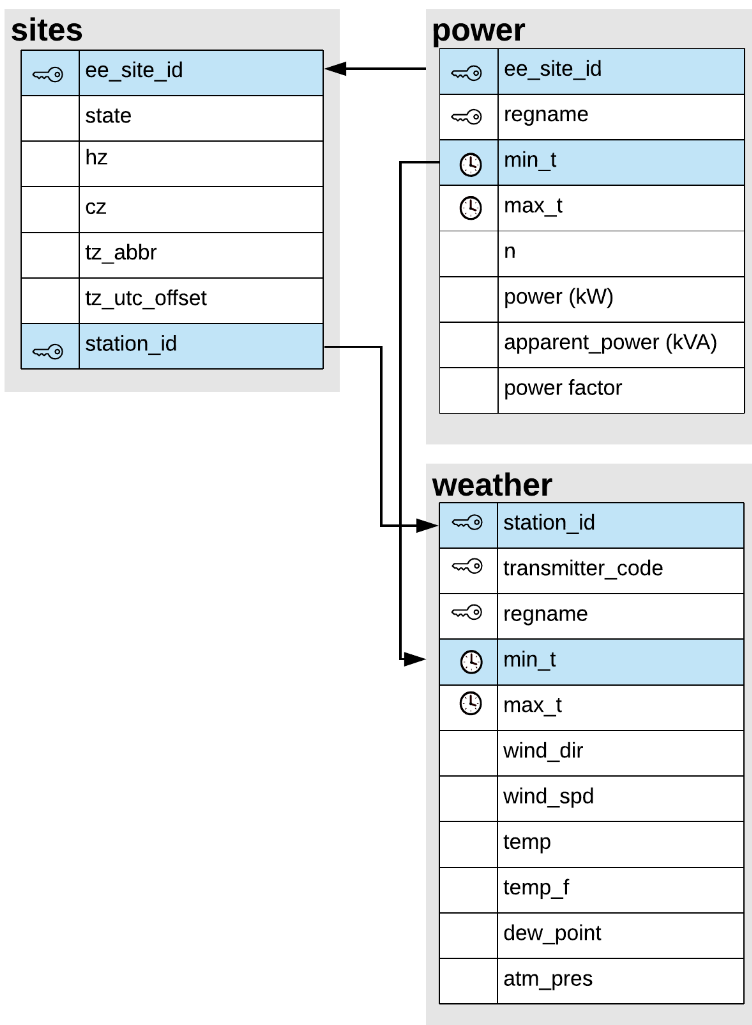


2. Extract metering data for the relevant sites and circuits by joining power on *ee_site_id* and *regname* (blue).

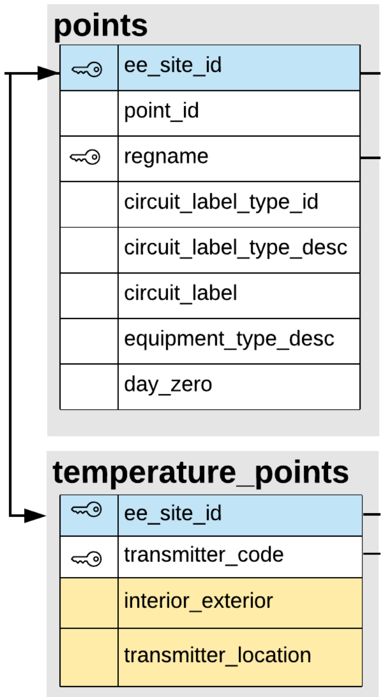


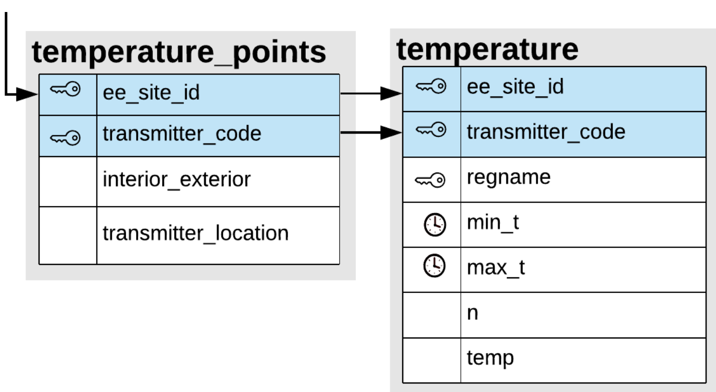
3. Use *sites* table to identify the relevant NOAA weather *station_id* (yellow) by joining your metering table (e.g., *power*) on *ee_site_id* (blue).


sites
ee_site_id
state
hz
cz
tz_abbr
tz_utc_offset
station_id
power
ee_site_id
regname
min_t
max_t
n
power (kW)
apparent_power (kVA)
power factor

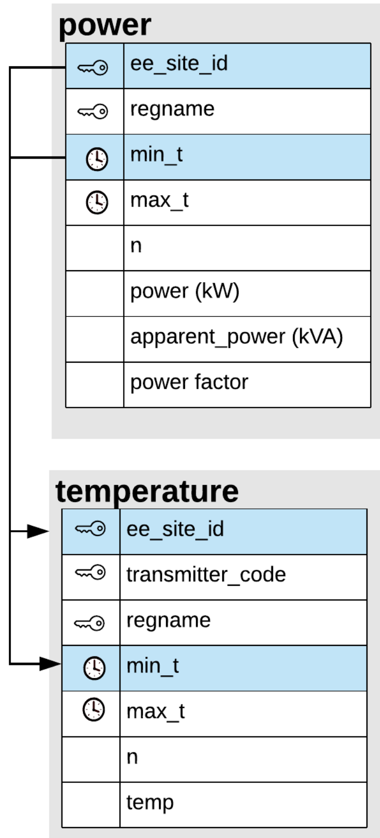

sites
ee_site_id
state
hz
cz
tz_abbr
tz_utc_offset
station_id
power
ee_site_id
regname
min_t
max_t
n
power (kW)
apparent_power (kVA)
power factor
weather
station_id
transmitter_code
regname
min_t
max_t
wind_dir
wind_spd
temp
temp_f
dew_point
atm_pres

5. Use the *temperature_points* table to Identify any relevant indoor or outdoor temperature sensors by filtering on *transmitter_exterior* and *transmitter_location* (yellow) for each *ee_site_id* (blue) with your targeted end use.

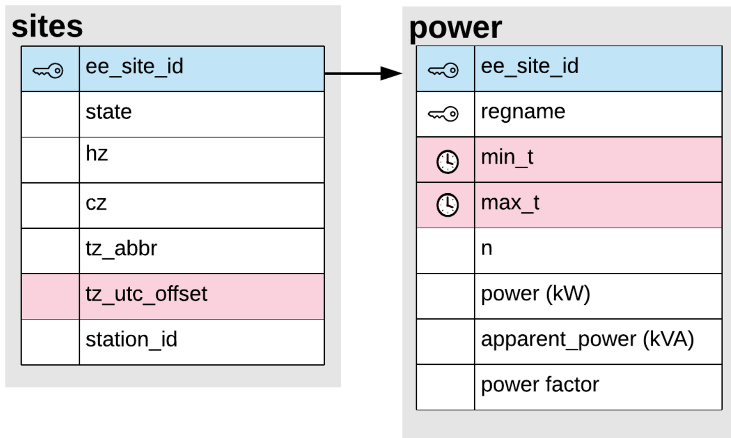


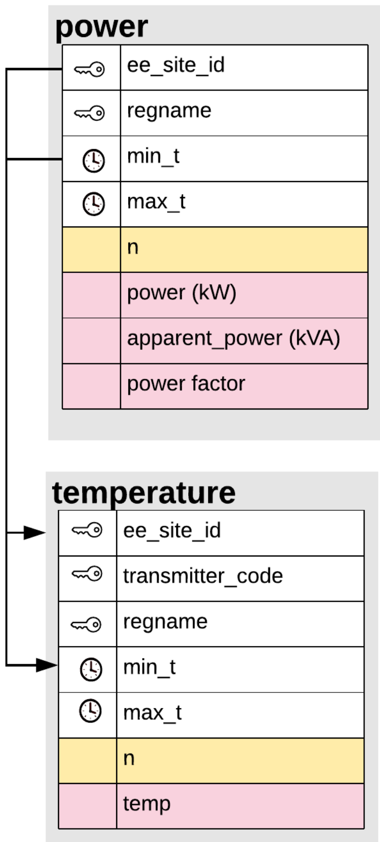


7. Merge *temperature* interval measurements with the metering data (e.g., *power*) by joining on *ee_site_id* and timestamp (e.g., *min_t*).



<u>Warning:</u> UTC does not include Daylight Savings Time. You will need to ensure that the software you are using accounts for this during the time zone adjustment if true local "clock" time is desired.

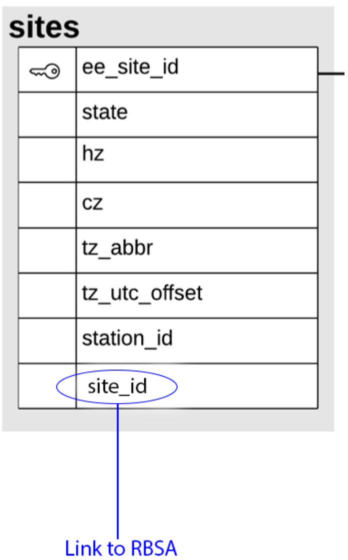

sites
ee_site_id
state
hz
cz
tz_abbr
tz_utc_offset
station_id
power
ee_site_id
regname
min_t
max_t
n
power (kW)
apparent_power (kVA)
power factor


power
ee_site_id
regname
min_t
max_t
n
power (kW)
apparent_power (kVA)
power factor
temperature
ee_site_id
transmitter_code
regname
min_t
max_t
n
temp

1.2 Connecting to RBSA

A unique value of the HEMS data is that it was mostly recruited from participants in the Residential Building stock Assessment (RBSA)[3]. The RBSA is a research study conducted roughly every five years by NEEA and is designed to collect information on the equipment in and characteristics of residential buildings. It contains many variables that may impact energy use including envelope, vintage, appliance, equipment type, lighting, and resident demographics.

The RBSA variables detail specific attributes of the characteristics of the specific equipment on the metered circuits as well as characteristics of the house. By linking to RBSA, we can gain additional insights into the HEMS load shapes. These data can be used for better understanding of the data itself and for modeling.



The sites table is necessary to link the RBSA data to the HEMS data. The sites table must be joined to the power table. From this, the site_id variable is the key to connecting the desired RBSA table.

In RBSA data, site_id is called CK_SiteID in all tables except the SiteDetail table where it is called PK_SiteID. RBSA contains 43 tables, all of which can be connected to HEMS. The RBSA SiteOneLine table will suffice for most analyses.

The RBSA has a large number of variables, many of which are characters instead of numeric, which can increase the overall file size significantly. Additionally, there are many corelated variables, e.g. area, conditioned area, volume, conditioned volume, etc. which can be duplicative. As such, prior to combining with RBSA tables it is advisable to pre-select which variables are of interest to best optimize analyses.

The SiteOneLine table has one line per site, so each HEMS site will link to a single row in this table. The table is designed to include all of the variables necessary for most analyses. In some cases, it isn't possible to include all of the detail collected on a single line so other tables would need to be utilized.

If it is necessary to connect to tables other than SiteOneLine, it is recommended to consult the RBSA Database User Manual [LINK] as the RBSA tables can be more complicated. The Database User Manual can show how to properly use these detailed tables.

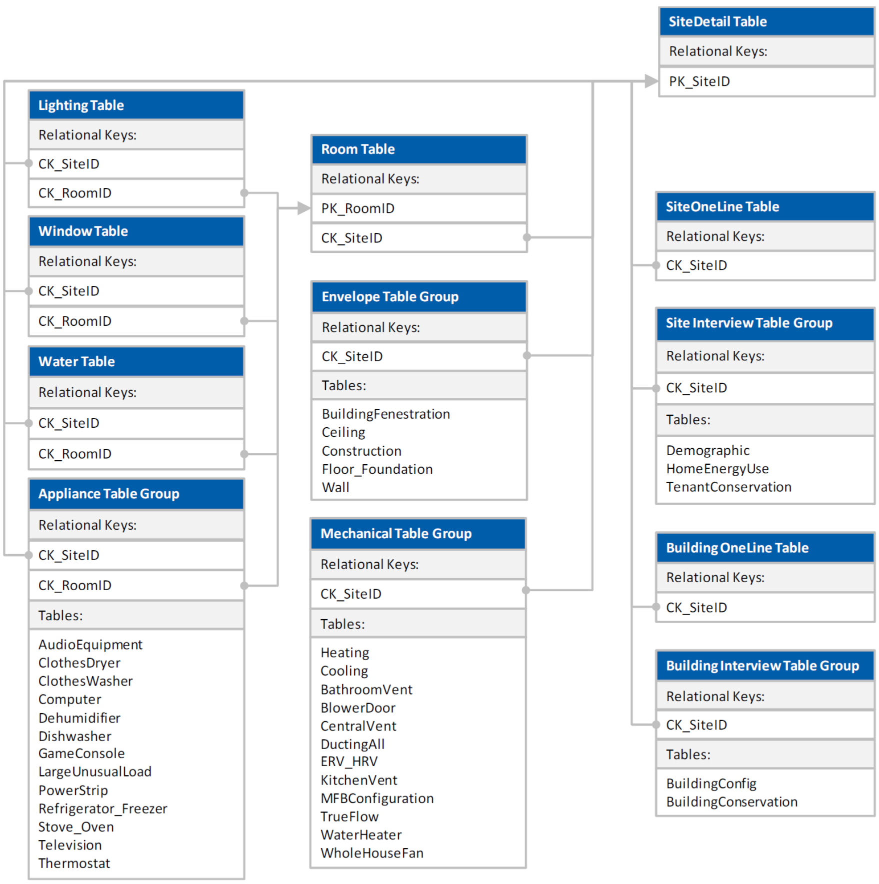
---

[3] Approximately 200 sites (the sites in the program the longest) were recruited from CBSA II (2019). The remainder are recruited from RBSA III (2022). In cases where it is necessary to recruit outside of RBSA, a full RBSA data collection will be conducted when feasible.

The RBSA table structure is shown in the following tables.

Object Definition by Table or Group	
Database Table or Table Group	**Definition of Objects (Line Items) within the Table**
SiteDetail Table	Each line item corresponds to a single entity: either a residence or a building. This table contains high-level information about the entity such as: State, address, strata, and case weight. This table also defines relationships between multifamily buildings and multifamily residences.
SiteOneLine Table	A single residence (e.g., a single-family home, manufactured home, or residential unit within a multifamily building).
BuildingOneLine Table	A single multifamily building.
Site Interview Table Group	A resident interview.
Building Interview Table Group	A property manager interview.
Envelope Table Group	A unique structural component, with each component recorded during on-site visits displayed as an individual line item.
Mechanical Table Group	A unique piece of mechanical equipment. Each piece of equipment recorded during an on-site visit displayed as an individual line item. In some cases, a quantity field is used to indicate the presence of multiple, identical equipment.
Room Table	A single room.
Water Table	A unique water equipment entry (e.g., Faucet, Showerhead, Bathtub).
Window Table	A unique window or door configuration entry. Each entry may represent one or more individual windows with the same characteristics. A quantity field is provided.
Appliance Table Group	A unique appliance entry, with each piece of equipment recorded during an on-site visit displayed as an individual line item. Each entry may represent one or more individual appliances with the same characteristics. A quantity field is provided for large unusual loads.
Lighting Table	An entry that represents a unique light bulb type/base type/shape/wattage entry. Entries with the same CK_Lighting_ID are on the same fixture. Each entry may represent one or more individual light bulbs with the same characteristics. A quantity field is provided.

Figure 5. Entity Relationship Diagram



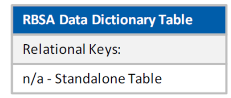


Table 1	Relationship	Table 2
SiteDetails	1:1	Building OneLine
		Site OneLine
		Site Interview Group
		Building Interview Group
	1:M	Envelope Group
		Mechanical Group
		Room
		Lighting
		Window
		Water
		Appliance Group
Room	1:M	Lighting
		Window
		Water
		Appliance Group

Appendix A: Frequently Asked Questions

1. Confidentiality of the public use HEMS data

   The HEMS public use data does not contain any names, addresses, or any information that could be used to identify a specific housing unit.

2. Why use the HEMS data?

   The main value of the EULR HEMS study is that the data provides a **current**, **representative**, and **robust** characterization of continuous energy consumption of key heating and cooling measures to support achieving clean energy goals and utility information needs, such as demand response, load forecasting, and resource planning. Available to the public as 15-minute interval data, it provides higher frequency of observations and greater accuracy than using data from older studies. Acquiring this data involved investing millions of dollars and years of effort from fourteen regional and national organizations.

3. How do I access the public use data?

   The 15-minute interval data is accessible at https://neea.org/EULRdata for download as a series of CSV files / within the HEMs database maintained by NEEA's technical consultant, Marquam. Statistical software is needed to extract and compile these files for analysis.

4. What data is available?

   - **Sites.** Geographic characteristics of each metered site: state, NW heating and cooling climate zone, and time zone. This does not contain any information which could be used to identify a specific housing unit.
   - **Points.** Technicians record descriptive data for metered circuits during the installation. This includes the circuit label, an equipment description, make/model (if available), and a cleaned end use classification of the circuit. This clean classification will make analysis much easier, as it does not require text parsing of engineer notes.
   - **Temperature points.** There are one or more temperature sensors installed in each home. An indoor temperature sensor is always installed, and an outdoor sensor is added if the home is located in an area without a reliable NOAA weather station. These outdoor sensors also record relative air humidity. Temperature probes are added to

monitor temperatures near heat pumps. Engineers record the location and type of each temperature sensor during installation.

- **Meter Data.** Each configured circuit at a home collects data for power, apparent power, and power factor. Total harmonic distortion is monitored at the whole home level. Temperature is collected as degrees Fahrenheit. Relative air humidity is recorded for outdoor sensors only. Additional 15-minute interval measurements will be published in Q2 of each subsequent calendar year.

5. Need help?

   NEEA, which is a non-profit, is making the 15-minute load data available to the public at no charge with the support of the project funders. We are not able to provide technical support for statistical and database software programming, or analytical issues. However, we are glad to try to help with questions of clarification or correcting issues with the data, the website, or the user guide. Please send questions or comments to EULRinfo@neea.org.