

# Information for Academic Researchers

1. Prerequisites for making a request.
2. Writing your data description.

## Prerequisites For Making A Request

The CPUC decision that established EDRP imposes two requirements on the researcher and three additional requirements on the specific research for which the data is requested. All five of these requirements must be met in order for any data request to be valid under the EDRP program.

The two requirements on the researcher are:

1. The researcher is affiliated with a non-profit college or university accredited by a national or regional accrediting agency and the accrediting agency is formally recognized by the U.S. Secretary of Education.
2. The researcher is a faculty member or is sponsored by a faculty member and the researcher and the sponsoring faculty members are responsible for carrying out the terms of the data release and a non-disclosure agreement.

The three requirements on the specific research project are:

1. The researcher must demonstrate that the proposed research will provide information that advances the understanding of California energy use and conservation.
2. The project must be certified to be in compliance with the federal government's "Common Rule" for the protection of human subjects by an Institutional Review Board.
  - a. Note that if your data is anonymized consistent with CPUC guidelines, then PG&E will waive this requirement.
3. To ensure data privacy, the academic institution must:
  - a. Sign a Non-Disclosure Agreement (NDA)
  - b. Successfully complete PG&E's Third Party Security Review (TSR), which evaluates the security of the supporting storage and computing infrastructure.
    - i. The complexity and detail of the TSR varies according to the sensitivity of the data. If data is anonymized consistent with CPUC guidelines,

then the TSR is much shorter and simpler than a TSR for data with Personal Information (PI).

## Writing Your Data Description

1. How much data can I request?
2. What data elements are commonly requested?

## How Much Data Can I Request

PG&E has a general obligation to protect our customers' privacy as much as possible; because of this we ask that all requests consider both the minimum number of customers and the minimum amount of data per customer necessary to fulfill the research objectives.

In general, PG&E limits data requests to data for no more than 5% of our overall customer base.

- This might be all the customers in some defined geographic area within our service territory (county, city, census block group, or zip code) or customers defined by some combination of geographic area and sector (e.g., all the residential customers in Sonoma County).
- It might also be, for instance, a random sample of 5% of the customers in our entire service territory. This is roughly 250k residential customers and/or 35k non-residential customers

There are special considerations if a researcher is requesting electric interval data; this is because interval data file sizes can become very large, very quickly. This can make the handling of the files much more difficult for both PG&E and the research team. For this reason, the data restrictions on interval data are tighter:

- If the interval data is "raw" i.e., meters are reported at whatever interval we have available no more than 150,000 customer-years' worth of interval data (e.g., 75k customers for two years or 30k for five years).
- If all the interval data is provided in one-hour intervals (even if 15 minute data is available), then no more than 300,000 customer-years' worth of interval data (e.g., 100k customers for 3 years or 30k for ten years).

## What data elements are commonly requested?

The most commonly used fields are:

<b>DATA CATEGORY</b>	<b>DATA ELEMENT</b>	<b>DESCRIPTION</b>
CUSTOMER	ACCOUNT NUMBER	Identifier of a customer account in the billing system (may include several addresses)
	COMMUNITY CHOICE AGGREGATOR	Indicator for CCA participation
	CUSTOMER CLASSIFICATION	Customer sector as determined by rate schedule code
	DIRECT ACCESS	Indicator for Direct Access participation
	NAICS CODE	Standard industry code associated with a business
PREMISE	SERVICE ADDRESS	Physical street address
	CITY	City
	ZIP CODE	ZIP code
	ADDRESS ID	Identifier for a customer address
	CEC CLIMATE ZONE	CEC Climate Zone
	CENSUS BLOCK GROUP	Census Block Group
	ELECTRIC SERVICE LOCATION ID	Identifier for the location of the electric meter tied to an address
	GAS SERVICE LOCATION ID	Identifier for the location of the gas meter tied to an address
	JURISDICTION/TERRITORY	Specify Town, City, County, Incorporated, Unincorporated
	LONGITUDE/LATITUDE	Longitude and latitude of the address
	MASTER METER	Indicates a master meter
	MULTIFAMILY	Indicates a multifamily building
	SOLAR CUSTOMER INDICATOR	Indicator for customer billed on net energy metering rate
	SOLAR INTERCONNECTION DATE	Photovoltaic system interconnection date
METER DATA [6]	COMMODITY	Specify gas and/or electric commodity
	DAYLIGHT SAVING TIME INDICATOR	Daylight Saving Time Indicator
	INTERVAL LENGTH	Length of interval read
	INTERVAL READ DATE	Meter reading date
	INTERVAL READ TIME	Meter reading timestamp
	INTERVAL USAGE	Usage recorded during interval
	METER ID	Identifier for the meter associated with a gas or electric service point
	METER TYPE	Meter equipment type
	MONTHLY BILLED USAGE	Billed usage (gas or electric)
	OFF PEAK USAGE	Off peak usage
	ON PEAK USAGE	On peak usage
BILLING	BASELINE TERRITORY	Baseline territory
	BILL AMOUNT	Amount billed for gas or electric commodity
	BILL END DATE	Billing period end date
	BILL ESTIMATED INDICATOR	Indicates billed usage was estimated
	BILL START DATE	Billing period start date

	BILLED RATE SCHEDULE CODE	Rate schedule in effect at time of billing
	CARE INDICATOR	Indicates billed on CARE rate
	ELECTRIC SERVICE ID	Billing identifier for the electric service
	FERA INDICATOR	Indicates billed on FERA rate
	GAS SERVICE ID	Billing identifier for the gas service
	HEATING TYPE/ALL ELECTRIC	Electric heating source indicator
	MONTHLY BILLED USAGE	Usage billed for the month indicated
	TOTAL BILL AMOUNT	Total amount billed for the month indicated
<b>ENERGY EFFICIENCY (EE) PROGRAM DATA [8]</b>		
	EE MEASURE CODE	Unique identifier in database
	EE MEASURE DESCRIPTION	Description of measure/technology installed
	EE PROGRAM	Statewide Program (ie, Residential)
	EE SUBPROGRAM	Statewide Subprogram (ie, Home Upgrade)
	EE MEASURE INSTALLATION DATE	Date EE application was paid (not the actual installation date of the measure or measures)

Notes:

- PG&E will often provide a UUID, or Unique Unidentifiable ID, in lieu of a customer's actual ID (e.g., an Account ID). These are functionally interchangeable with actual identifiers for research purposes, but protect customer privacy.
- PG&E does not necessarily retain data on customers who have not had service with PG&E for five years or more. PG&E will provide usage and usage related data for customers as far back as 2010, but can only provide data for our entire customer base for the last five years.