SERL data and documentation:  
Readme

Table of Contents

|  |  |
| --- | --- |
| Creation date | 2020-09-02 |
| Version | 2020-08 |
| Author | Ellen Webborn |
| Project | Smart Energy Research Lab (SERL) |
| Organisation | University College London (UCL) |

# Introduction

## Smart Energy Research Lab (SERL)

The [Smart Energy Research Lab](https://www.serl.ac.uk) (SERL) will provide a secure, consistent and trusted channel for researchers to access high-resolution energy data, which will facilitate innovative energy research for years to come. SERL is [funded](https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/P032761/1) by the UK’s Engineering and Physical Sciences Research Council (EPSRC) and comprises a [consortium](https://serl.ac.uk/about-who-we-are/) of seven UK universities and the Energy Saving Trust.

Our initial work programme focuses on:

* establishing a Smart Energy Observatory of ~10,000 smart metered households across Great Britain (GB);
* developing a mechanism for other researchers to access smart meter data for other household samples from whom they have obtained informed consent.

## Participant recruitment

Participant recruitment began in August 2019. Approximately 1700 participants were recruited from central and southern England adn from Wales as part of a pilot study that tested different recruitment strategies. The second recruitment wave takes place in August-September 2020, and the third wave at the end of 2020. SERL aims to recruit 8-10,000 households to be regionally representative across England, Scotland and Wales. Recruitment is also designed to be representative of each Index of Multiple Deprivation (IMD) quintile; an area-based relative measure of deprivation.

# 2020-08 SERL Observatory dataset

This document accompanies the 2020-08 release of the SERL Observatory dataset. This is the first Observatory dataset release and is available in a secure virtual environment to UK-based accredited researchers working on approved projects (see below for more details on data access). Smart meter data collection start dates vary by participant; the earliest start date is August 2018. This data release contains all data available up until 31st July 2020. Future data releases will include more recent data and data from participants recruited in the second and third recruitment waves.

The datasets can be linked at the household level using the “PUPRN” pseudo-anonymous identifier included in each dataset. The climate data can be linked through the “grid\_cell” provided with the participant summary (smart meter) dataset.

## Smart meter data

In order to participate a household must have a DCC-enrolled electricity smart meter (SMETS2 or upgraded SMETS1). In GB smart meters collect electricity and (mains) gas data at 30-minute resolution. If a household has embedded generation such as solar PV, the electricity readings show the net demand for each half-hour (always non-negative) and separate readings show net export (always non-negative). Reactive power data is also available. In addition to the half-hourly data, daily reads are also provided, which can be used for data quality checking, or enhanced by the half-hourly reads if daily reads are missing.

Currently 1612 participants have daily smart meter reads in the datset and 1688 have half-hourly reads in the dataset. n\_no\_gas participants do not have a gas meter (that we are able to access).

### Data files

There are four data files relating to smart meter data:

* *SERL\_smart\_meter\_daily\_v2020\_08.csv:* daily electricity and gas readings with some additional derived columns;
* *SERL\_smart\_meter\_hh\_v2020\_08.csv:* half-hourly electricity and gas readings with some additional derived columns (note that this dataset also includes reactive readings and export readings where available);
* *SERL\_smart\_meter\_rt\_summary\_v2020\_08.csv:* data quality summary for each read type for each participant (such as number of errors found by type) and basic read statistics (such as mean and maximum);
* *SERL\_participant\_summary\_v2020\_08.csv:* data quality summary for each participant (less detail than the read-type summary data, but also including basic participant information such as region and number of questions answered on the survey).

### Documentation

There are documents accompanying the smart meter data:

* *SERL\_smart\_meter\_documentation\_v2020\_08.pdf:* smart meter documentation describing the four datasets, the fields and details about derived variables;
* *SERL\_smart\_meter\_data\_quality\_report\_v2020\_08.pdf:* smart meter data quality report summarising various data statistics, data availability, and the prevalence of missing and potentially erroneous reads.

### Code

The R code used to process the raw half-hourly and daily datasets (creating the four datasets outlined above) is available with this data release. All code and documentation is (or will shortly be) available on the [SERL GitHub site](https://github.com/smartEnergyResearchLab) which will contain the most up-to-date versions of all non-disclosive files. The code file is

* *SERL\_smart\_meter\_data\_prep\_v2020\_08.R*

## Survey data

When they sign up participants are encouraged to complete a survey about their household and dwelling. It is not compulsory but most participants complete the survey. The survey is completed online or on a postal form. Each asks the same questions and response options, and the data are provided together, with a variable indicating the response mode. The online survey has the advantage of automatic question routing and basic answer checking (e.g. flagging up the wrong input type).

### Data file

The survey data provided with this release relates to the participants who were recruited during the first recruitment wave in 2019. The file is

* *SERL\_survey\_data\_v2020\_08.csv*

### Documentation

Two documents accompany the survey data:

* *SERL\_survey\_documentation\_v01.pdf:* documentation for the survey data, summarising the responses to each question and comparing the results with national surveys where possible, commenting on evidence of bias;
* *SERL\_survey\_questions\_v01.pdf:* postal survey sent with some of the mailings/reminders (same questions and options as the online survey).

### Code

The R code used to process the raw survey data will be made available on the [SERL GitHub](https://github.com/smartEnergyResearchLab) observatoryData repository.

## Energy Performance Certificate (EPC) data

If available, each household’s Energy Performance Certificate (EPC) is also provided (the most recent is used). There are currently 902 households with an EPC.

### Data file

The EPC dataset is:

* *SERL\_epc\_data\_v2020\_08.csv*

### Documentation

The EPC data documentation details all fields included in the dataset and gives information about the values included.

* *SERL\_epc\_documentation\_v2020\_08.pdf*

### Code

After address identifiers were removed and replaced by our PUPRN pseudo-identifiers no further processing was performed, therefore there is no code available for this dataset.

## Climate data

Climate reanalysis data is also provided for each household. The data from the Copernicus/ECMWF [ERA5 hourly reanalysis data](https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview). More variables will be provided in future releases, but at the moment only temperature at 2 metres is part of the data. It can be linked with SERL participants using the ‘gridCell’ variable.

This is reanalysis data based on recorded data from many weather stations across GB, at a horizontal resolution of 0.25 x 0.25 degrees latitute and longitude (approximately 28 sq. km). More information about spacial resolution is available [here](https://confluence.ecmwf.int/display/CKB/ERA5%3A+What+is+the+spatial+reference). ERA5 documentation is available [here](https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation).

### Data file

The climate dataset: \* *SERL\_climate\_data\_v2020\_08.csv*

### Documentation

Documentation describing the dataset and the fields available is given in the file:

* *SERL\_climate\_documentation\_v2020\_08.pdf*

### Code

No processing has been done on the data so this datset does not have associated code.

# Data access

Access to the SERL Observatory datasets is provided via a secure lab environment and is restricted to accredited researchers working on approved projects. This aligns with the [5 Safes](https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-safes) protocols used by the UK Data Service (UKDS) for all access to any controlled data via a secure lab. Currently, SERL only provides controlled datasets via a secure lab environment.

The process to obtain access to controlled data is appropriately rigorous so it is strongly recommended that researchers read the detailed documentation associated with the SERL catalogue record ([SN 8666](https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8666)) in the UKDS data catalogue to ensure that SERL data is appropriate for your research project before any other steps.

Summary of steps and rough timelines:

1. Accredited Researcher status (safe Researcher training and exam): *1 month*
2. University Ethics approval: this varies by institution but allow at least *3-4 weeks*
3. Project application (UKDS triage and SERL Data Governance Board review): *4-6 weeks*

Therefore, if starting from scratch, it is recommended that researchers allow at least 3 months for completion of the above processes before access to the data is required.

Full details are available here: [Accessing SERL Observatory data – information for researchers (PDF)](https://serl.ac.uk/wp-content/uploads/2020/08/Accessing-SERL-Observatory-data-info-for-researchers-v04.pdf)

The [Accessing SERL data](https://serl.ac.uk/researchers/) page on the SERL website has more information for researchers including a set of FAQs.

# Participation over time