SERL data and documentation: Readme

Creation date	2021-04-13
Edition	02
Author	Ellen Webborn
Project	Smart Energy Research Lab (SERL)
Organisation	University College London (UCL)

Table of Contents

Introduction

Smart Energy Research Lab (SERL)

The vision of the Smart Energy Research Lab (SERL) is to 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 by the UK's Engineering and Physical Sciences Research Council (EPSRC) and comprises a consortium of seven UK universities and the Energy Saving Trust.

Our initial work programme focuses on:

- establishing an Observatory panel of ~10,000 smart metered households across Great Britain (GB);
- provisioning of Observatory data to UK researchers via a secure lab environment.

Participant recruitment

Participant recruitment began in August 2019. Approximately 1700 participants were recruited from central and southern England and from Wales as part of a pilot study that tested different recruitment strategies. The second recruitment wave took place in August-September 2020, and the third wave at the start of 2021. SERL aims to recruit around 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.

Table 1: Number of participants in SERL by year and month. Numbers decrease when participants withdraw consent or move house.

Year	*Month*	*Number of participants*	*Change*
2019	8	0	0
2019	9	819	819
2019	10	1,647	828
2019	11	1,706	59
2019	12	1,701	-5
2020	1	1,696	-5
2020	2	1,692	-4
2020	3	1,687	-5
2020	4	1,682	-5
2020	5	1,680	-2
2020	6	1,677	-3
2020	7	1,661	-16
2020	8	1,651	-10

Year	Month	Number of participants	Change
2020	9	3,487	1,836
2020	10	4,671	1,184

SERL data

This document accompanies the SERL Observatory Edition 02 dataset. The data are available in a secure virtual environment to UK-based accredited researchers working on approved projects (see Data Access section for more details).

Smart meter data collection start dates vary by participant; the earliest start date is in August 2018. This data release contains all data available up to and including 2020-10-31 which includes participant households recruited in Waves 1 and 2. Future data releases will include more recent data and data from participants recruited in any future recruitment waves. Participants recruited in Wave 3 will be included in the 3rd edition of the SERL Observatory data which is expected to be released in summer 2021.

The datasets can be linked at the household level using the pseudo-anonymous "PUPRN" (pseudo-UPRN (Unique Property Reference Number)) 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 DCC-enrolled SMETS1). In GB smart meters collect electricity and gas (where available) 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. Note that SMETS1 meters do not provide daily reads, and that the daily sums of half-hourly reads are included in the SERL daily dataset for data imputation and comparison.

As of 2020-10-31, 4,716 participants have daily smart meter reads in the dataset and 4,779 have half-hourly reads in the dataset. 1,149 participants do not have a gas meter that we can access (although some will still have gas central heating we are unable to access the data, perhaps because they do not yet have an installed/commissioned gas smart meter).

Data files

There are four data files relating to smart meter data:

- *serl_smart_meter_daily_edition02.csv*
  - daily electricity and gas readings with some additional derived columns
  - 2,018,336 records
  - 14 fields

- *serl_smart_meter_hh_edition02.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)
  - 96,638,254 records
  - 16 fields
- *serl_smart_meter_rt_summary_edition02.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)
  - 21,933 records
  - 25 fields
- *serl_participant_summary_edition02.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)
  - 4,876 records
  - 40 fields

Documentation

There are two documents accompanying the smart meter data:

- *serl_smart_meter_documentation_edition02.pdf:* smart meter documentation describing the four datasets, the fields, and details about derived variables;
- *serl_smart_meter_data_quality_report_edition02.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 which will contain the most up-to-date versions of all non-disclosive files. The code file is

- *serl_smart_meter_data_prep_edition02.R*

Survey data

When participants sign up they are encouraged to complete a survey about their household and dwelling online or on a paper form. The online and offline versions contain 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).

Survey data exists for 4,753 participants.

Data file

The survey data provided with this release relates to the participants who were recruited before 2020-10-31. In addition to the csv file containing the survey response data we also provide a data dictionary that shows how the response codes relate to the survey response options for each question. The survey-related files are:

- *serl_survey_data_edition02.csv*
  - 4,753 records
  - 152 fields
- *serl_survey_data_dictionary_edition02.csv*

Documentation

The following documents accompany the survey data:

- *serl_survey_documentation_02.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_pilot_recruitment_survey_copy.pdf*: the postal survey sent to households in the first (pilot) recruitment wave (same questions and options as the online survey)
- *serl_main_recruitment_survey_copy.pdf*: the postal survey sent to households in the main recruitment waves (very similar to the survey used in the pilot study, and same questions and options as the online survey).

Code

The R code used to process the raw survey data is available with this data release. All code and documentation is (or will shortly be) available on the SERL GitHub site which will contain the most up-to-date versions of all non-disclosive files, including the survey data cleaning file. The code file is

- *serl_survey_data_prep_02.R*

Energy Performance Certificate (EPC) data

If available, each household's Energy Performance Certificate (EPC) is also provided (the most recent is used). Data linking is done by matching house number and postcode and checking the first line of the address. There are currently 2423 households with an EPC record in the dataset. This reflects the population of the EPC register for England and Wales where a substantial percentage of the domestic housing stock does not have an EPC. We do not currently have access to EPC data for Scotland although we hope to include Scottish EPC data in future editions. Each data release contains the most up-to-date EPC record at time of collection for each household.

For more information about EPC data see the EPC data website.

Data file

The EPC dataset is:

- *serl_epc_data_edition02.csv*
  - 2,423 records
  - 80 fields

Documentation

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

- *serl_epc_documentation_edition02.pdf*

Code

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

Climate data

Climate reanalysis data is also provided for each household. The data from the Copernicus/ECMWF ERA5 hourly reanalysis data. Currently the climate data consists of 24 variables. The data can be linked with SERL participants using the 'grid_cell' 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 latitude and longitude (approximately 28 sq. km). More information about spatial resolution is available here. ERA5 documentation is available here.

Data file

The climate dataset:

- *serl_climate_data_edition02.csv*
  - 7,150,224 records
  - 27 fields

Documentation

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

- *serl_climate_documentation_edition02.pdf*

Code

No processing has been done on the data, therefore there is no code associated with this dataset.

British Summer Time dates

Since the daily data are recorded in local time and the clocks changing affects energy use in UTC time, we have created a file detailing the start and end dates of British Summer Time covering the SERL data collection period and a few years beyond.

This file is called *bst_dates_to_2024.csv* and does not have its own documentation as it's self explanatory. There are 14 rows and 3 columns: *Read_date_effective_local*, *type* (start or end of BST) and *n_hh* (the number of half-hours on the days the clocks change (46 or 50)).

Data access

Access to the SERL Observatory datasets is provided via a secure virtual lab environment and is restricted to accredited researchers working on approved projects. This aligns with the 5 Safes protocols used by the UK Data Service (UKDS). 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) in the UKDS data catalogue to ensure that SERL data is appropriate for your research project before starting the application process.

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)

The Accessing SERL data page on the SERL website has more information for researchers including a set of FAQs.

More information

The following links provide additional information about SERL and data access for researchers. If you require any further information please contact **info@serl.ac.uk**

- SERL website

- SERL GitHub site
- UKDS Observatory Data Record Study Number (SN) 8666
- Utilising smart meter data for research and innovation in the UK (a 2019 ECEEE Conference paper about SERL)