

# Participant information:

## Technical Documentation

Creation date	2021-08-23
Edition	03
Author	Ellen Webborn, Jessica Few
Project	Smart Energy Research Lab (SERL)
Organisation	University College London (UCL)

### Introduction

This document describes the participant summary file available to researchers with secure access to the SERL Observatory datasets. This file gives basic information for each participant as well as a summary of the smart meter data quality.

The filename of this dataset is: *serl\_participant\_summary\_edition{number}.csv*, where *{number}* is the number of the data release, e.g. "02" (note that the first data release files are labelled with the release date rather than edition number).

The file has one row per participant and there are 13321 rows in the latest data table. There are 38 columns, such as the region and LSOA of the participant, how many survey questions they answered, information about the start and end dates of each schedule and the number and percentage of reads that were valid for each schedule. Details of all columns in this data table are provided below. Note that we no longer provide column detailing the number of invalid read times for each participant.

The read-related column names take the form "text\_W\_X\_Y\_Z" for electricity readings and "text\_W\_X\_Z" for gas readings. Rather than explaining every single column, here is key to the variable component of the name:

- W = device type (either electricity meter ("ESME") or Gas Proxy Function i.e. gas meter ("GPF"))
- X = whether the read is half-hourly ("HH") or daily ("D")
- Y = active ("Act") or reactive ("React") power (electricity reads only)
- Z = import ("Im") or export ("Ex")

For example "NumValid\_ESME\_D\_Act\_IM" is the number of valid daily electricity active import readings. The text in the first part of the name is described in the table below.

Table 1:

<i>Field (or field name format)</i>	<i>Description</i>	<i>Units</i>	<i>Class</i>	<i>Example value</i>
PUPRN	Pseudonymised participant identifier	NA	character	1VUXXXF1
Region	Region in GB	NA	character	East Midlands
LSOA	Lower Super Output Area in GB	NA	character	E01015916
grid_cell	Grid cell for linking to climate data	NA	character	38_31
IMD_quintile	Index of Multiple Deprivation quintile (1 is most deprived, 5 is least deprived)	NA	integer	2
EPC_exists	TRUE if a record exists in the EPC dataset for the participant	NA	logical	TRUE
EPC_rating	EPC rating (original column name 'current_energy_rating') - a letter between A and G inclusive	NA	character	C
pilot_test_cell	Test cell (1 to 12) assigned to participants for the pilot study (testing recruitment approaches)	NA	integer	12
N_survey_ans	Number of questions answered in the survey (30 relevant to all, a further 9 may be relevant depending on other answers)	NA	integer	35

<i>Field (or field name format)</i>	<i>Description</i>	<i>Units</i>	<i>Class</i>	<i>Example value</i>
Perc_survey_ans	% of relevant questions answered in the survey (may be greater than 100% if question relevance was not possible to determine or if question skipping was not done correctly)	NA	numeric	98.1203
Start_W_X_Y_Z	First valid read date for data type W_X_Y_Z	%Y-%m-%d	Date	2019-11-01
End_W_X_Y_Z	Last valid read date for data type W_X_Y_Z	%Y-%m-%d	Date	2020-02-29
NumValid_W_X_Y_Z	Number of valid reads (error code 1) for data type W_X_Y_Z	NA	integer	98
PercValid_W_X_Y_Z	Percentage of possible reads (using theoretical start and end dates rather than actual valid read start and end dates) that are valid (error code 1) for data type W_X_Y_Z	NA	numeric	95.2