

# Overview of the IDEAL Project and IDEAL Household Energy Dataset



## About this document

<b>Purpose</b>	This document provides a very brief introduction to: <ul style="list-style-type: none"><li>the original reason for collecting the data made available in this data release;</li><li>the nature of the sample of households and period covered in the dataset;</li><li>the range of variables and sampling frequencies at which data were collected, and the approaches taken to collecting the data.</li></ul>
<b>Further information</b>	<ul style="list-style-type: none"><li>For technical details of the structure of the data release, including the range of files in the release and the fields present in each, see IDEALdata.pdf.</li><li>A more detailed data descriptor is also available. See: Pullinger, M., Kilgour, J., Goddard, N., Berliner, N., Webb, L., Dzikovska, M., Lovell, H., Mann, J., Sutton, C., Webb, J., Zhong, M. (2021). <i>The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes</i>. Scientific Data.</li></ul>
<b>Creators</b>	Martin Pullinger, Jonathan Kilgour
<b>Organisation</b>	University of Edinburgh
<b>Version date</b>	April 2021

## Project team

The research upon which this data release is based was managed and run by the research team below:

**Principal Investigators and Co-Investigators:** Dr. Nigel Goddard (PI), Prof. D.K. Arvind, Prof. Heather Lovell, Prof. Johanna Moore, Prof. David Shipworth, Dr. Charles Sutton, Prof. Jan Webb.

**Researchers:** Niklas Berliner, Cillian Brewitt, Dr. Myroslava Dzikovska, Edmund Farrow, Elaine Farrow, Jonathan Kilgour, Dr. Janek Mann, Dr. Evan Morgan, Dr. Martin Pullinger, Dr. Lynda Webb and Dr. Mingjun Zhong.

## Acknowledgements

This data release is based on research funded by the UK Engineering and Physical Sciences Research Council projects Intelligent Domestic Energy Advice Loop (grant reference EP/K002732/1) and Data-Driven Methods for a New National Household Energy Survey (grant reference EP/M008223/1). The authors are grateful to the research funding agency.

Thanks also to the participants in the project, whose participation made this dataset and other work in the projects possible, and to Changeworks (<https://www.changeworks.org.uk/>) for identifying and recruiting potential participants, managing participant interactions, and installing and maintaining homes' sensor and feedback systems.



## Overview

### Citing this dataset

If you publish work making use of this dataset, please cite **both** of the following:

#### Data descriptor paper, in the journal Scientific Data:

- Pullinger, Martin; Kilgour, Jonathan; Goddard, Nigel; Berliner, Niklas; Webb, Lynda; Dzikovska, Myroslava; Lovell, Heather; Mann, Janek; Sutton, Charles; Webb, Jan; Zhong, Mingjun. (2021). *The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes*. Scientific Data.

Note: This paper will be published Open Access in mid-2021. For the DOI, when available, see the IDEAL dataset webpage at <https://doi.org/10.7488/ds/2836>.

#### The dataset itself, on the University of Edinburgh's DataShare repository:

- Goddard, Nigel; Kilgour, Jonathan; Pullinger, Martin; Arvind, D.K; Lovell, Heather; Moore, Johanna; Shipworth, David; Sutton, Charles; Webb, Jan; Berliner, Niklas; Brewitt, Cillian; Dzikovska, Myroslava; Farrow, Edmund; Farrow, Elaine; Mann, Janek; Morgan, Evan; Webb, Lynda; Zhong, Mingjun. (2021). *IDEAL Household Energy Dataset [dataset]*. University of Edinburgh. School of Informatics. <https://doi.org/10.7488/ds/2836>.

### About the source projects

The data in the IDEAL Household Energy Dataset were originally collected as part of two related projects funded by the UK Engineering and Physical Sciences Research Council: IDEAL and BIGSMALL. IDEAL aimed to develop, through a co-design process with participants, a wireless digital home energy monitoring, feedback and advice system, and evaluate it using a randomised controlled trial study design, alongside in-depth qualitative interviewing. BIGSMALL aimed to advance Machine Learning methods for Non-Intrusive Load Monitoring, in which appliance usage in a home is inferred from high resolution energy data from the main electricity or gas meter, through the collection of a suitable training and evaluation dataset and research into new NILM methods.



**IDEAL: Intelligent Domestic Energy Advice Loop.** EPSRC grant reference EP/K002732/1. Duration: 31 March 2013 - 31 March 2018.

**BIGSMALL, formally known as Data-Driven Methods for a New National Household Energy Survey.** EPSRC grant reference EP/M008223/1. Duration: 31 December 2014 - 30 June 2018.



### About this dataset release

The IDEAL Household Energy Dataset comprises electricity and gas sensor data along with a diverse range of relevant contextual data from additional sensors and surveys, from 255 homes in Edinburgh and the nearby regions of the Lothians and south Fife in Scotland, UK, collected over a 20-month period ending in June 2018. The start and end times of participation varied by household, with homes having a mean install duration of 286 days over this period.

Sensor monitoring in all homes included 1-second apparent-power electricity data, pulse-level gas data, 12-second temperature, humidity and light data for each room, and 12-second measurements of boiler pipe temperatures to indicate usage of central heating and hot water. In addition, 39 of the homes had an 'enhanced' sensor system installed, which included plug-level monitoring of selected electrical appliances, real-power measurement of mains electricity and key sub-circuits and more detailed temperature monitoring of gas- and heat-using equipment, including radiators and taps (see Table 1 below for a full list of appliances covered). Survey data was collected on occupant demographics, income, values, attitudes, self-reported energy awareness, energy tariffs, and building, room and appliance characteristics. The data is provided in comma-separated format with a custom-built API to facilitate usage, and has been cleaned and documented.

The dataset *includes*:

- household energy data and relevant occupant, home, room, appliance and contextual (including weather) data from participating homes in the IDEAL project, collected during the core study period of the project, from the date of installation of the first home, in August 2016, until the project phase I end date of 30 June 2018.

The dataset *excludes*:

- data specifically related to the study intervention, including click data showing usage of the IDEAL energy feedback and advice app and additional survey data on user perceptions of and responses to the app. These will be released in subsequent data releases.
- The data are also anonymised to reduce the risk of participant identification, in the following ways:
  - Free text fields from surveys are omitted or abridged.
  - All names and contact details are removed.
  - Detailed location data are removed. Homes were recruited from the Edinburgh, Lothians and Fife areas, and a location field indicates location in one of five areas within this range (Edinburgh, East, West or Mid Lothian, or Fife). Additionally, each home's value on an urban rural classification system (the Scottish Government Urban Rural Classification 2016) is provided to indicate the level of urbanisation of the region in which the home is situated.

Figure 1 below gives a visual summary of what is included in this dataset and what will be released in future data releases.

**Figure 1:** Schematic of split of IDEAL and BIGSMALL data releases

<b>Data subject</b>	<b>IDEAL energy feedback study</b> e.g. feedback app usage, user perceptions of app, self-reported behavioural changes	IDEAL Energy Feedback Study Dataset ( <i>release forthcoming</i> )	
	<b>Home and occupant energy use, outcomes and context</b> e.g. energy use; room temperature, humidity and light; building, appliance, and occupant characteristics; weather.	<b>IDEAL Household Energy Dataset</b> ( <i>this data release</i> )	IDEAL Phase II Dataset ( <i>release forthcoming</i> )
		<b>August 2016-June 2018</b> IDEAL phase I (core study period)	<b>July 2018 – June 2019</b> IDEAL phase II (extended study period)
<b>Study period</b>			

The IDEAL Household Energy Dataset is likely to have a wide range of applications. Examples include building performance modelling, Non-Intrusive Load Monitoring, policy and practice related to energy demand management and socio-technical research into understanding energy use.

### Location and usage of data in the data release

Data in the data release are stored in different csv files. To increase ease of use, the data are divided between files according to the type of metric they measure and the subject (e.g. home, person, appliance, sensor) rather than data collection channel.

IDEALdata.md describes the contents and structure of the data release.

The sensor files are individually compressed into zip files to reduce size, and individual files are then also grouped and compressed into one of several separate gzip files that comprise the data release.

For ease of use, the data release includes a Python module for use with the sensor data. This makes use of the file-naming conventions of the sensor data files to provide a useful way to explore and load user-specified subsets of the sensor data and start to work with it. Useful example code that uses the module is also included.

The IDEAL data is provided without any gap-filling for missing data. However, the Python module that loads sensor data and hourly summary data will, by default, fill gaps with NaN so that an unbroken series can be expected. This approach allows users to apply whatever gap-filling approach they consider to be most appropriate.

For those not using the Python module, most operating systems are able to natively decompress both gzip and zip files. For Windows users, there is no native ability to decompress gzip files, but several free programs are available that can do this.

## Sampling and data collection

### Sample selection and participation requirements

Potential participants were identified through a mix of channels, including recent callers to a home energy advice service, on-the-ground recruitment events and online marketing channels. Recruitment and installations were managed by project partner Changeworks (<https://www.changeworks.org.uk/>).

Several requirements needed to be met for potential participants to be eligible to join the study. These were primarily to increase the probability of collecting high quality data for the duration of the project, and to reduce certain sources of variability in home energy use that would overly complicate the development and delivery of energy advice (notably too much variation in home heating technology, and the presence of micro-generation and certain high power devices). Full details of the eligibility criteria are provided in Box 1 below.

Participants were given an information booklet prior to agreeing to participate to facilitate informed consent. An initial information leaflet, the information booklet and the consent form used in the study are all provided accompanying this data release, and can be found in the Supplementary\_material directory.

The data descriptor paper describes the recruitment strategy and the resultant sample characteristics in more detail.

**Box 1.** Criteria for participating in the IDEAL project

Located in Edinburgh, Lothians or south Fife areas of Scotland, and living in a non-moveable home not shared with other households.
Willing to participate until June 2018, and unlikely to have any change of occupants during that time; and home usually occupied by same occupants at least nine months of the year. (Initially this was proxied by disallowing private rented accommodation and paying guests, but this restriction was removed in February 2017 to increase recruitment rates).
Home heated with gas central heating in majority of rooms, with a gas combi boiler.
Sensor-readable model of gas meter (we could read a range of pulse-enabled gas meters), and non-smart electricity meter (to help avoid confounding effects of another In-Home Display system); willing to keep these until June 2018 unless needing changed for safety reasons.
Home has broadband and occupant is willing to leave router on at all times, and let project use the connection, until June 2018. (Potential participants were advised of a chance of increased broadband monthly costs in the rare cases they still had capped data allowances).
No electricity or gas prepayment (to reduce risk of system downtime arising from gaps in energy supply).
None of the following (and no plans to acquire before June 2018): microgeneration; hot water heating other than via combi boiler; storage heaters, air conditioning or heat pumps, Mechanical Ventilation with Heat Recovery, solid fuel fires unless used infrequently; Agas; electric vehicle charging.
No plans for major changes to property before June 2018 (e.g. removing walls, building extensions, fitting double glazing).
Adequate access to combi boiler pipes, electricity and gas meters and (for enhanced installations) most of the appliances to be monitored and willing to allow the required electricity rewiring.
Good signal propagation for the sensor system: For standard installs, this was managed by evaluating based on home's WiFi propagation and placing a recruitment constraint that outdoor gas meters had to be close to the property and sheltered from rain; for enhanced installs, this was evaluated based on the propagation rates from IDEAL standard sensors over the first few weeks after the standard installation visit.
Willing for anonymised data to be lodged in a data archive after project end.
Moderate or higher level of self-reported digital literacy for at least one occupant in the home (to increase chances that the IDEAL app could be used by the occupants).
No children under 7 or pets (enhanced homes); all potential participants made aware of need to keep system components out of reach of any children under 3.

**Box 2.** List of categories of variable for which data were collected in the IDEAL project.

1. Outcomes: energy use/power (overall and by selected appliances); indoor temperature, humidity and light levels; expenditure; self-reported comfort and satisfaction
2. Energy using activities
3. Consideration of energy use, including use of the IDEAL app
4. Values and attitudes
5. Sociodemographic characteristics
6. Household dynamics
7. Energy information and awareness
8. Materials and resources: personal (e.g. rooms, outdoor space, appliances, vehicles); local natural and built environment; public services
9. Weather conditions

## Data collection

### Choice of metrics for which data were collected

The choice of metrics to collect measurements on was informed by the research aims described above and by existing theory on relevant factors to consider to understand and explain changes in energy using activities in participating homes. Box 2 above indicates the top-level categories of variable for which data were collected in the project. The numbering presented in Box 2 is used in the data release to indicate which category survey variables belong to, e.g. within the naming of the majority of survey *uniquequestionids* (see below).

### Data collection channels

Data in this data release were collected through one of three channels.

### Sensors

Used for frequently changing variables. Sensors collected data on:

- **Whole home electricity use.** 1-second apparent power in all homes; plus 5-second real power in enhanced homes.
- **Whole home gas use.** Per pulse (data point whenever a fixed volume of gas was used; either 1dm<sup>3</sup> or 1ft<sup>3</sup>, depending on the meter model).
- **Room temperature, humidity and light levels** for every room in the home. 12-second frequency.
- **Selected appliances and electricity subcircuits** (see Table 1 below). Varying frequency (see below).

Usage of electrical appliances was measured by monitoring electricity use by the appliance. As such, time of use and appliance energy use were measured and recorded. Plugged-in appliances have data points whenever their power use changed by a fixed small amount. Appliances hard-wired into subcircuits of the home's electricity wiring had their subcircuits measured and recorded at 5-second intervals.

Usage of gas or hot water-using appliances, where monitored, was measured by monitoring temperature on or near to the appliance. As such, only time of use was measured and recorded, indirectly; appliance energy use was not measured. These data were collected at 12-second intervals.

Note:

- For sensors other than real power and for appliances and electricity subcircuits, 1 hour summary data are also available for convenience of analysis, in the auxiliary data download. These are simple totals or means (as appropriate) of the data available for a given hour for a given sensor. As such, for periods with high levels of missing data, these may not be the most precise estimates of totals/means that could be calculated.
- There is no electricity data for one home in the data release (*homeid* 223). It was found after installation of the sensor system in this home that it was not possible to obtain electricity data due to the location of the home's electricity meter in relation to its internet router.

**Table 1.** List of appliances monitored by sensors in IDEAL participating homes. **Appliances in bold** were monitored in all homes; the remainder were monitored only in enhanced system homes, where present. Sensor monitoring of appliances was dependent in each case on occupant permission and technical requirements being met.

Activity	Appliances monitored directly or by proxy measures
<b>Laundry</b>	Washing machines, tumble dryers, and combined washing machine-tumble dryers
<b>Personal washing</b>	Inlet pipes for hot water taps for baths, showers and bathroom sink(s), or outlet pipes or underside of the unit
<b>Space heating and cooling</b>	<b>Boiler pipes (central heating and hot water inflow and outflow)</b> <b>Room temperature and humidity in all rooms</b> Radiator pipes (inflow and outflow) <b>Solid fuel fires/stoves (e.g. wood or charcoal burning) in working condition and reported by participant to be used at least occasionally</b> Fixed gas fires Electric heaters Dehumidifiers
<b>Hot food and drink preparation</b>	Cookers, ovens and hobs Microwaves Kettles Kitchen sink(s)
<b>Washing up</b>	Dishwashers Inlet pipes for hot water taps for kitchen sink(s), or outlet pipes or underside of the unit
<b>Other cleaning</b>	Vacuum cleaners
<b>Lighting the home</b>	<b>Light levels in all rooms</b>
<b>Having things running in the background</b>	Fridges, freezers, and fridge-freezers Aquariums
<b>Leisure-related</b>	Hot tubs Wine coolers

A mix of sensor types was used. Real power was monitored using Open Energy Monitor (OEM) technology. Plug-level appliance monitoring was done using Z-Wave Individual Appliance Monitors. All other sensor data was collected using sensors developed by the project team. The sensorbox schematic for the project-developed sensors is included in the Supplementary material download. For a full description of the individual sensors and full sensor system used in the IDEAL project, see the data descriptor paper accompanying this data release.

#### *Surveys with participants*

This data release includes data from six surveys delivered to participants in the project, delivered in three main waves. Surveys were delivered either to only the 'primary participant' of each home, or made more widely available to other occupants too. The primary participant was the person who was present during the initial installation visit and agreed to be the primary point of contact for the project. Primary participant surveys focused on gathering data on the building characteristics, appliances and sociodemographic information about occupants, whilst the more widely distributed surveys gathered information on respondents' attitudes, values, motivations and perceptions of their home and wider environment. They also collected data on aspects of the IDEAL app intervention such as the participants' perceptions of the systems' usability and usefulness.

The current data release:

- Includes responses to the majority of multiple choice and matrix question in the surveys listed in Table 2 below.
- Excludes responses to multiple choice and matrix questions where they relate to aspects of the IDEAL app intervention, notably questions from the survey *all\_web\_end* covering users' perceptions of different aspects of the IDEAL app, the feedback and advice provided through the app, and behaviour changes they have made or attempted over the course of their participation in the study. These will be included in the forthcoming IDEAL Energy Feedback Study Dataset.
- Includes responses, where available, from every person in homes that are included in the final data release.

Table 2 below summarises how and to whom each survey was administered. Where surveys were delivered to all occupants or appusers (i.e. those occupants who had an account in the IDEAL app, which as a minimum included all occupants aged 15 or over on the date of installation), primary participants and any other participants who had provided email addresses received email notifications that a new survey was available and requesting them to complete it. Email recipients were also asked to pass on details of the survey to other participants in the home who had not received the email. For web surveys, this comprised a link to the online survey. For surveys delivered within the IDEAL app, this prompted participants to log in to the app using their personal account. For in-app surveys, recipients would also be notified on the home page at login if they logged into the app.

Each survey question in the data release has a *uniquequestionid* to identify it, which takes the format *[questionid]\_[surveywave]* where:

- *questionid* comprises a name intended to be indicative of the question that was asked plus, in most cases, a code signifying which kind of construct the question is intended to tap (see Box 2 above for a list). Where the same question is asked across multiple survey waves, this component of the *uniquequestionid* remains unchanged.
- *surveywave* describes which of the three main waves of survey data collection the question is from, taking one of the following three values:
  - *initial*: Two surveys during the installation visit, or soon afterwards – one face to face, the other within the IDEAL app.
  - *mid*: Repeat questions from the initial surveys, and meter readings from within the IDEAL app.
  - *end*: Repeat of the initial in-app survey, plus questions on usefulness of and responses to the app.

Note that as the delivery format changed between waves for some questions (e.g. from face-to-face interviewing to web survey; from feedback app to web survey), some changes in question and response wordings were necessary between waves.



**Table 2.** Survey data collection overview

<i>surveyname*</i>	Criteria for selecting homes	Criteria for which occupants within eligible homes were selected	Available for completion from when	Available for completion until when	Delivery channel	Notifications and reminders (reminders sent only to participants/households which had not completed the survey)	Key survey contents
<b>primary_facetoface_initial</b>	All homes	Primary participant only	Varying - during standard installation visit	N/A	Face-to-face computer assisted interview	N/A	Occupant sociodemographics, working patterns; household income; building characteristics; movable energy using appliances.
<b>all_inapp_initial</b>	All homes	All appusers	First login (usually during installation visit for primary participants)	Prior to October 2017: until completion. From October 2017: Six weeks after first login	IDEAL app	Primary participant asked to complete this during installation visit by interviewer and requested to ask other appusers to complete too. Popup prompts within app until completed/survey closed.	Attitudes towards different energy-using activities; values
<b>primary_web_mid</b>	All homes in study 3+ months on survey launch date	Primary participant only	21 September 2017	5 October 2017	Web survey interface. Personalised link emailed to participants	21 September 2017: Initial email notification to primary participant. 28 September & 3 October 2017: Reminder emails.	Repeat questions from primary_facetoface_initial
<b>all_inapp_mid</b>	All homes in study 6+ weeks on survey launch date	All appusers who completed or partially completed the initial feedback app survey ( <i>all_inapp_initial</i> above) 3+ months ago OR had been in study 6+ weeks and had not started that survey.	10 October 2017	2 November 2017	IDEAL app	10 October 2017: Popup prompt within app on first login and visible in app's message box. 13 October 2017: Initial email notification to all eligible participants with email addresses on record, including request to request completions from other appusers in the home. 20 October 2017: Reminder email. 24 October - 2 November 2017: daily in-app popup notification on login.	Repeat questions from all_inapp_initial
<b>all_inapp_meters_mid</b>	All homes installed up to 6 December 2017	All appusers created up to 6 December 2017	7 December 2017	14 December 2017	IDEAL app	In-app display in message panel. 7 December 2017: Initial email notification to all eligible participants with email addresses on record. 12 and 14 December 2017: reminder emails.	Electricity and gas meter readings
<b>all_web_end</b>	All homes	All occupants	23 May 2018; 5 June 2018 for homes unallocated to control, treatment or enhanced cohorts at start of May 2018.	30 June 2018	Web survey interface. Personalised link emailed to participants + generic link with request to pass it to other occupants in the home.	23 May or 5 June 2018: Initial email notification to all eligible participants with email addresses on record, including request to request completions from all others in the home. 1 June, 12 June, 22 June, 29 June: Email reminders. From 4 June 2018: Follow-up reminder phone calls.	Repeat questions from all_inapp_initial ( <i>Not included in this data release: perceptions of IDEAL app and participation in study; self-reported attempts to change energy using activities</i> )

\* *surveyname* is a survey identifier used to uniquely identify each of the surveys throughout the data release and documentation. It is composed in the format [*recipients*][\_*delivery\_channel*][\_*surveywave*], where:

- *recipients* has one of two values:
  - *primary*: only delivered to primary participants of eligible homes
  - *all*: delivered or made available to (nearly) all participants of eligible homes
- *delivery\_channel* has one of three values:
  - *facetoface*: face to face interview with an interviewer (undertaken during the installation visit).
  - *web*: delivered via a web interface, using the Qualtrics survey service.
  - *inapp*: delivered via the IDEAL app (and hence also available through a web browser if the participant logged into their IDEAL app online rather than via the project-provided tablet).
- *surveywave*: the same wave indicator, *initial*, *mid* or *end*, as described above for *uniquequestionid*.

### Survey questionnaires

The full survey questionnaires, including non-question text (e.g. guidance and information presented to the respondents), routing and questions whose responses are not included in this data release, are included in this data release in the Surveys directory, as described below:

- Web surveys (including the initial face-to-face interview) are provided as PDFs and as QSF format, one each per survey. (QSF is a proprietary survey file format used by Qualtrics, the survey service used in the project. It can be imported into a suitable Qualtrics account to reproduce the fully functioning survey).
- In-app surveys are provided as xml files, one each per survey, as used in the IDEAL app itself. Also in the Surveys directory are three images that provide illustrative examples of how the in-app surveys appeared within the IDEAL app, as a popup screen within the main display:
  - *survey\_all\_inapp\_initial\_information\_screen.png* shows an example information screen, which does not require a response;
  - *survey\_all\_inapp\_initial\_multiple\_choice.png* shows a single multiple-choice question, requiring a single response;
  - *survey\_all\_inapp\_initial\_matrix.png* shows a matrix question, requiring multiple responses – one per subitem.
- The prompt cards, used as a visual aid by the interviewer to show participants the response options for certain questions, for the initial primary participant survey during the installation visit (*primary\_facetoface\_initial*) are provided in PDF and Word formats.

For the subset of questions from these surveys for which responses are included in this data release, two csvs (*survey\_question\_wordings.csv* and *survey\_response\_wordings.csv*) provide a machine-readable version of the question wordings and response options.

**survey\_question\_wordings.csv** is laid out in a wide format with one row per *questionid*, with columns as described in Table 3 below.

Uses:

- Provides a complete list of the exact question wordings for each survey question for which participant responses are available in the data release.
- Provides an indication of which question were asked across multiple waves of survey, to enable evaluation of change over time.
- Provides a machine-readable way to identify which csv of the data release any given response can be found in.
- Provides an 'attribution' indicating the source of the original survey question.

**Table 3:** Fields present in survey\_question\_wordings.csv

Field	Description
<b>questionid</b>	The first part of the <i>uniquequestionid</i> , without the wave identifier
<b>uniquequestionid_initial</b> <b>uniquequestionid_mid</b> <b>uniquequestionid_end</b>	The <i>uniquequestionid</i> in the initial, middle and end waves of surveys respectively (blank if not asked in that wave). Note that meter readings, stored in the <i>meterreadings.csv</i> , do not have <i>uniquequestionids</i> ; instead these fields indicate the values in the table's <i>provenancedetail</i> and <i>energytype</i> columns where the responses are recorded.
<b>surveyid_initial</b> <b>surveyid_mid</b> <b>surveyid_end</b>	The <i>surveyid</i> in the initial, middle and end waves of surveys respectively in which the question was asked (blank if not asked in that wave).
<b>superquestion_wording_initial</b> <b>superquestion_wording_mid</b> <b>superquestion_wording_end</b>	The wording of the question for matrix questions in the initial, middle and end waves of surveys respectively (blank if not asked in that wave or if the question is not a matrix-format question).
<b>question_wording_initial</b> <b>question_wording_mid</b> <b>question_wording_end</b>	Wording of the multiple-choice question or of the matrix question sub-item in the initial, middle and end waves of surveys respectively (blank if not asked in that wave).
<b>response_location_initial</b> <b>response_location_mid</b> <b>response_location_end</b>	The table in the data release in which the responses to this <i>questionid</i> are located for the initial, middle and end waves of surveys respectively (blank if not asked in that wave). Responses are stored in <i>person.csv</i> , <i>home.csv</i> , <i>other_appliance.csv</i> , <i>meterreadings.csv</i> or <i>survey_responses.csv</i> depending on the nature of the construct being measured by the question. Note that <i>survey_responses_numeric.csv</i> is also available – the content of this is identical to that in <i>survey_responses.csv</i> , except that Likert-type responses have been converted to numeric values, which may ease some analyses (e.g. "7: Strongly agree" is represented as "7").
<b>attribution</b>	Indicates which survey(s) and survey question(s) the <i>questionid</i> is from. Questions may appear as they did in the original source or be altered for the purposes of this project. Full references corresponding to these attributions can be found in the References below.

**survey\_response\_wordings.csv** is laid out in a long format with one row per *uniquequestionid* and *response\_wording* combination, with columns as described in Table 4 below. For each *uniquequestionid*, the set of valid responses is laid out one row per response option, in the same order in which they appeared in the original survey. (NB. responses for *uniquequestionids* that had numeric, Boolean or free text response options are not included in *survey\_response\_wordings.csv*).

Uses:

- Provides a complete machine-readable list of the exact response wordings for each survey question for which participant responses are available in the data release.
- Can be used to facilitate analyses of variables, where the original order of response options is needed (e.g. for Likert scales) and/or where some response options were never selected by participants so do not appear in the table of participant responses (e.g. extreme ends of some scales).

Notes:

- Income band values: *income\_band* and *income\_band\_mid* values are abridged in the *survey\_response\_wordings.csv*, *home.csv*, *survey\_responses.csv* and *survey\_responses\_numeric.csv* tables: ends of the scales are shortened to improve anonymisation of outliers; only annual income bands are presented, for increased readability. To see how these response options were presented to participants, see the original survey files (*survey\_primary\_facetoface\_initial.pdf* and

survey\_primary\_web\_mid.pdf) and the prompt card shown to participants during the initial primary participant survey (card hhincome in prompt\_cards\_survey\_primary\_facetoface\_initial.pdf). See also table\_of\_income\_bands.csv, described below.

**Table 4:** Fields present in survey\_response\_wordings.csv

Field	Description
<b>uniquequestionid</b>	The <i>uniquequestionid</i> that the response wordings are associated with.
<b>response_wording</b>	The text value of the response option e.g "7: Strongly agree".

**table\_of\_income\_bands.csv** provides the income band response options in a machine readable format, described in Table 5 below.

Uses:

- Facilitates analyses using income responses (home.csv: *income\_band*; survey\_responses.csv and survey\_responses\_numeric.csv: *income\_band\_mid*).
- Shows the full representation of the response wordings shown to participants.

**Table 5:** Fields present in table\_of\_income\_bands.csv

Field	Description
<b>data_release_value</b>	Responses as presented in the data release. These present only the annual income figures, for ease of analysis. High and low income bands are also collapsed to improve anonymisation: all bands of less than £10,800 are collapsed into one category; all bands of £90,000 or more are collapsed into one category.
<b>response_options_initial</b>	Approximate representation of the full response option shown to participants during the initial primary participant survey, <i>primary_facetoface_initial</i> . (The actual prompt card used, showing the full response options as seen by participants, is available in the supporting documents – see prompt_cards_survey_primary_facetoface_initial.pdf).
<b>response_options_mid</b>	The full response option shown to participants during the middle primary participant survey, <i>primary_web_mid</i>

### Installation apps

The technicians who installed the IDEAL sensor and feedback system in participating homes made use of dedicated iOS and web apps developed by the project, to:

- collect and record further data about the home, rooms, and appliances, and the location and function of installed sensors;
- allocate home ids, room ids, appliance ids and sensors ids and to programme sensors and other system equipment as required.

In enhanced homes, further sensors were set up using dedicated web apps modified from open source apps for the purpose.

### Weather data

Weather data were sourced from Weather Underground (<https://www.wunderground.com/>).

## References

### Sources of survey questions and response options

(as referred to in the *attribution* column of *survey\_question\_wordings.csv* to identify sources of questions used in the project)

- **ARCC-Water Survey, 2011.** Browne, A., Anderson, B. and Medd, W. (2011) *Lancaster SPRG/ARCC water survey*. University of Lancaster, UK. <https://eprints.soton.ac.uk/439637/>
- **Axsen et al, 2012.** Axsen, J., TyreeHageman, J. & Lentz, A. (2012). *Lifestyle Practices and Pro-Environmental Technology*. *Ecological Economics*, 82, 64–74. <https://doi.org/10.1016/j.ecolecon.2012.07.013>
- **English Housing Survey, 2013-2014.** Ministry of Housing, Communities & Local Government (2015). *English Housing Survey 2013 to 2014: Questionnaire and Physical Survey Form*. <https://www.gov.uk/government/publications/english-housing-survey-2013-to-2014-questionnaire-and-physical-survey-form>
- **Living Costs and Food Survey, 2011. Volume B: Household Questionnaire.** Office for National Statistics (2012). *Living Costs and Food Survey, User Guide Volume A – Introduction. January – December 2011. Great Britain & Northern Ireland*. [http://doc.ukdataservice.ac.uk/doc/7272/mrdoc/pdf/7272volume\\_a\\_introduction\\_2011.pdf](http://doc.ukdataservice.ac.uk/doc/7272/mrdoc/pdf/7272volume_a_introduction_2011.pdf)
- **Schulenberg and Melton, 2008.** Schulenberg, S. E., & Melton, A. M. A. (2008). *The Computer Aversion, Attitudes, and Familiarity Index (CAAFI): A Validity Study*. *Comput. Hum. Behav.*, 24(6), 2620–2638. <https://doi.org/10.1016/j.chb.2008.03.002>
- **Third European Quality of Life Survey, 2011-2012.** European Foundation for the Improvement of Living and Working Conditions. (2012). *Third European Quality of Life Survey: Questionnaire*. <https://www.eurofound.europa.eu/publications/other/2012/quality-of-life/third-european-quality-of-life-survey-questionnaire>
- **Understanding Society, Wave 6, 2014-15.** University of Essex, Institute for Social and Economic Research. (2019). *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 12th Edition*. UK Data Service. SN: 6614, <http://doi.org/10.5255/UKDA-SN-6614-13>
- **World Values Survey, Wave 6, 2010-2014.** Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E. & Puranen, B. et al. (eds.). (2014). *World Values Survey: Round Six - Country-Pooled Datafile*. Version: [www.worldvaluessurvey.org/WVSDocumentationWV6.jsp](http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp). Madrid: JD Systems Institute.
- **Wynford Survey, 2012.** Webb, J., Hawkey, D., McCrone, D. & Tingey, M. (2010-2014). *Bringing Warmth to Wyndford: Household Survey of New District Heating. Heat and the City Project*. Sociology, School of Social and Political Science, University of Edinburgh. Funded by the Research Councils' UK Energy and Communities Programme. Grant RES-628-25-0052. <https://heatandthecity.org.uk/project/heat-and-the-city/>

### Scottish Government Urban Rural Classification 2016

As described in the documentation, this is provided in the *home.csv* table.

- Scottish Government Geographic Information Science & Analysis Team, Rural and Environment Science and Analytical Services Division. (2018). *Scottish Government Urban Rural Classification 2016*. <https://www2.gov.scot/Resource/0054/00542959.pdf>. Accessed 26 February 2020.
- Scottish Government (2019). *Scottish Government Urban Rural Classification 2016 - Postcode Lookup* (<https://www2.gov.scot/Resource/0054/00544931.csv>, via <https://www2.gov.scot/Publications/2018/03/6040/downloads>). Accessed 26 February 2020.