Utilising smart meter data for research and innovation in the UK

Ellen Webborn, Simon Elam,
Eoghan McKenna & Tadj Oreszczyn
UCL Energy Institute, The Bartlett
University College London (UCL)
Central House, 14 Upper Woburn Place
London WC1H 0NN
United Kingdom
e.webborn@ucl.ac.uk
s.elam@ucl.ac.uk
e.mckenna@ucl.ac.uk
t.oreszczyn@ucl.ac.uk

Keywords

smart meters, energy demand, domestic energy, research, web portal, big data

Abstract

Great Britain aims to install 53 million smart electricity and gas meters in around 27 million domestic properties by 2020. Smart meters provide the potential for high-resolution electricity and gas consumption data that has never previously existed on a national scale. To leverage this national investment, UK Research and Innovation has funded a sizeable project to develop a Smart Energy Research Lab (SERL) to provide access to smart meter energy data for UK researchers. A primary objective of SERL is to develop a secure research portal for researchers to access energy data, linked to relevant contextual data (e.g. socio-demographics, building characteristics and weather data), thereby facilitating high-quality scientific research. This paper focuses on data availability and will discuss the benefits, challenges and methods for developing a national data resource that will support a wide range of research across the energy sector. The paper provides practical guidance to researchers who want to utilise SERL data directly, as well as insights for researchers, policy-makers or other organizations who wish to utilise smart energy data more broadly. Issues discussed include the complexities of data governance and quality associated with smart energy data, innovative approaches to research design (enabling both Observatory and Laboratory functions) and practical solutions to sector-wide issues such as smart meter consumer authentication.

Introduction

This past decade has seen the implementation of large-scale smart meter rollouts in many countries around the world. For example, by 2020, Great Britain (GB) aims for every home and small business to have been offered smart meters by their energy supplier (Ofgem, 2018), amounting to a target of 53 million meter installations across electricity and gas (DECC, 2015). GB meters record electricity/gas consumption at half hourly resolution and can store 13 months of historic data which can be accessed through a central system (the DCC Gateway) – a significant advance on traditional metering systems that require manual reads that typically occur only once or twice a year.

It has been argued many times over the past few decades that the impact of energy policies needs to be rigorously evaluated using a strong evidence base, with "relatively little success" (Skea, 2012), (Allcott & Mullainathan, 2010), (Frondel & Schmidt, 2005), (Train, 1994) and (Hartman, 1988). Smart metering has the potential to provide access to data from millions of homes with a time resolution hitherto inaccessible without the installation of additional monitoring equipment. This kind of data are valuable to researchers, businesses and policy makers seeking to understand, manage and potentially shift energy use in domestic buildings (Summerfield & Lowe, 2012). Historically, common barriers to high-quality research using this type of data have included a lack of raw and contextual data, data quality issues, and low sample size (Hamilton, et al., 2015). For example, the Low Carbon London project (UK Power Networks, 2014) provides energy consumption readings for 5,567 London households, but contextual data are reduced to a single 'CACI Acorn Group' meaning that details such as location (postcode), building, socio-demographics and consumer attitudes data are inaccessible to researchers.

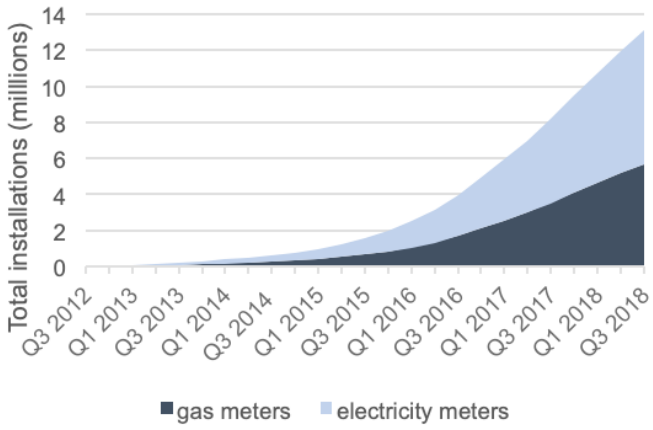

Figure 1. Total number of installations by large energy suppliers of domestic gas and electricity smart meters in GB. Data from (Department for Business, Energy & Industrial Strategy, 2018).

In this paper we introduce a UK government funded initiative, the Smart Energy Research Lab (SERL), which will provide accredited researchers in the UK with access to long-term, high-resolution, high-quality energy data linked to contextual factors where home owners have consented to data access. The paper starts by describing the smart meter rollout in GB and how data can be accessed independently of energy suppliers. Next, we describe the project vision and research design. The third section discusses the recruitment and sampling methods including our upcoming pilot studies intended to optimise the recruitment design for maximum participation levels. We then describe our data governance framework, our work on the issue of consumer authentication, and the challenges facing SERL. Lastly, we discuss how researchers can access SERL data, and summarise the paper.

The smart meter rollout in Great Britain

BACKGROUND

In 2012 the UK government introduced new legislation that began the smart meter rollout in England, Wales and Scotland. Figure 1 shows the GB progress, up to October 2018 (the latest figures) of domestic smart meter installations in GB up to October 2018 (13.2 million meters). An additional ~20,000 smart and 'advanced' meters have been installed in the non-domestic sector (Department for Business, Energy & Industrial Strategy, 2018).

The UK has been unique in its smart meter rollout approach in that the rollout is led by energy suppliers rather than network operators. Suppliers are required to offer the smart metering equipment (smart gas and electricity meters, a communications hub and an in-home display) to all consumers but adoption is voluntary (Hledik, et al., 2018). The goal is for every home in the country to have an electricity and gas (if applicable) smart meter by the end of 2020. However, as acceptance of a smart meter is not mandatory for households (Hierzinger, et al., 2012), consumers have to be convinced of the benefits of smart metering. Once installed, access to data stored on smart meters is strictly controlled, and informed consent from the consumer is required to access consumption data. There is no central database of smart meter data. The challenge for researchers is that consent is required from every participant in a study collecting smart meter data and the highly secure systems needed to access smart meter data via the DCC Gateway are expensive to develop and implement.

The DCC Gateway

The Data Communications Company (DCC) is "leading the design, build, test and integration of the data and communications infrastructure to connect smart meters with … authorised users" (Data Communications Company, 2018). This central infrastructure is known as the DCC Gateway, which acts as a secure messaging service between smart meters (where data is stored) and authorised system users. SERL will collect data via a DCC "Other User" role and has to meet a strict set of criteria set out by UK legislation, the Smart Energy Code (SEC), in addition to the General Data Protection Regulations (GDPR).

Implications for researchers

It is difficult to undertake empirical energy research if you do not have access to high quality energy data. This ranges from research into consumer energy use including their flexibility, energy efficient technology evaluation and the potential for demand side response. Historically there have been two routes to such research; to use an existing anonymised dataset of estimated energy use or install monitoring equipment in individual buildings to collect consumption data. For example, every 2 years since 2004, the National Energy Efficiency Data-Framework (NEED) has been using electricity and gas data collected as part of energy billing in England and Wales and linking this to administrative building and occupant socio-demographic data which has been collected at a different time. Each sample consists of approximately 4 million properties (around 16 % of the population) (Department for Business, Energy & Industrial Strategy, 2018). However, because traditional energy meters only record a running total of electricity or gas data, the only data stored for future use are the readings provided to energy suppliers on an ad-hoc basis.

Utility companies are only required to do at least one physical meter reading every two years. In the absence of regular energy readings, energy suppliers, government, and datasets such as NEED rely on estimated annual consumption (EAC – for electricity) and annual quantity (AQ – for gas). Annualised estimates fit energy readings (if and when they exist) to an energy consumption curve for the 'end user category' of each consumer, correcting for between-year seasonality effects. This allows for data to be compared between years but removes a researcher's ability to study the effects of weather on energy consumption. Another issue is that these records are fully anonymised and may undergo statistical disclosure control, which means there is very little contextual data available at the individual household level for general research.

With smart metering, energy use is recorded every half hour, which for consumers means more accurate billing, and for researchers means more accurate data sources for studies, and far greater scope for valuable research. Installing additional monitoring equipment will still be useful for studying individual household appliances, occupant behaviours, comfort levels and preferences, and for acquiring consumption data on the or-

der of minutes, seconds, or sub-seconds. However, the benefits need to be weighed against the costs of installing equipment, the intrusion into people's homes, and the reduced (and potentially biased) sample sizes likely to result from these issues.

In addition to electricity or gas readings, the latest generation of UK smart meters' (SMETS2) data will also include information about the current energy tariff, separately metered large appliances such as storage heaters, devices connected to the smart meter such as the in-home display, and exports readings from any micro-generation such as solar panels. Energy readings from the previous 13 months can be retrieved once consent has been granted[1]. Similarly, micro-generation export readings can be accessed for the previous 3 months.

Despite the wealth of data created by the GB smart meter rollout, there are many barriers facing researchers wanting access to raw data, and in particular data that can be linked to other contextual data sources, such as survey information completed by householders, external data sources relating to building characteristics, and national surveys. The Smart Energy Research Lab intends to overcome these difficulties for researchers in the UK by providing a large longitudinal data set that can be linked with contextual data sources.

Smart Energy Research Lab (SERL)

INTRODUCTION

The Smart Energy Research Lab (SERL)[2] is a £6m (€6.67 m) 5-year project funded from September 2017 by the Engineering and Physical Sciences Research Council (EPSRC) in the UK. Led by University College London (UCL), the partners include seven universities[3] and the Energy Saving Trust; a social enterprise with a focus to help people save energy. The initial goal is to recruit a representative sample of around 8,000–10,000 participants for the core 'Observatory' panel by the end of 2019.

VISION

The SERL vision is to deliver a high quality multi-disciplinary research programme, facilitated by a smart energy research portal. The long-term provision of high quality, high-resolution energy data via SERL will support the development of a reliable evidence base for intervention, observational, and longitudinal studies across the socio-technical spectrum.

The goals of the lab are to provide:

- A consistent, trusted, and sustainable channel for researchers to access large-scale, high-resolution energy data, thereby providing a reliable empirical dataset for research;
- An effective mechanism for collecting energy data alongside other variables from national surveys or individual research projects;
- A confidential, ongoing repository of smart meter data enhanced with contextual dwelling, household and neighbourhood attributes for use in secondary data analysis.

The ambition of the research programme is to undertake research that will:

- Support government policy;
- Kick-start the development of new products, services and energy markets;
- Help provide solutions to the energy trilemma (security, affordability and environmental sustainability);
- Facilitate better research by developing best practice guidelines and methods to improve data security and enable innovative uses of smart meter data.

Specific examples of the types of research that could be facilitated by SERL include:

- Evaluation of energy efficiency programmes
- Development of Smart Energy Performance Certificates (EPC's)
- Development of tailored energy advice services for consumers
- Understanding habitual energy consumption
- Determining the energy responsiveness of dwellings to external temperature
- Disaggregation of end-use energy consumption using smart meter data
- Impact of socio-demographic factors on energy demand profiles
- Impact of smart home technologies such as HEMS (home energy management systems) or smart heating controls
- Distributional impact of switching suppliers/tariffs on bills and energy consumption
- Identification, targeting and mitigation of fuel poverty.

SMART METER DATA

SERL will be able to access smart meter data via the DCC 'Other User' (OU) role. The DCC User Interface Specification V2.0[4] states that an OU can make requests to read, for example: half-hourly active and reactive imports, half-hourly exports, daily consumption, and energy tariff information. An OU can also request to join consumer access devices (CADs) to the smart meter. CADs can monitor energy use down to sub-second resolution, monitor individual appliance use, and link to other sensors in the home e.g. temperature sensors. Over time, we expect the number of CADs linked to smart meters to increase in GB homes. Eventually SERL will look to utilise the additional data streams provided by CADs to enable a wider variety of research than can be achieved using smart meter data alone.

1.If a tenant moved in more recently than 13 months ago then the smart meter data can be accessed back to the move in date rather than the full 13 months.

2. Formerly known as the Smart Meter Research Portal (SMRP).

3. The seven university partners are UCL, the University of Essex (UK Data Archive), University of Edinburgh, Cardiff University, Loughborough University, Leeds Beckett University and the University of Southampton.

4. https://smartenergycodecompany.co.uk/the-smart-energy-code-2/

RESEARCH DESIGN

SERL will collect half-hourly gas and electricity data via informed consent (see Recruitment section below for more detail) from homes in GB, request completion of a survey about the occupants and dwelling, and link these data with other available sources of contextual data. Examples may include the dwelling's Energy Performance Certificate (EPC), local weather data and national surveys where possible. The data will be available to accredited researchers in the UK in a secure environment and more widely available after thorough processing for anonymisation. Researchers will also be able to use SERL as a data pipe for collecting and managing data generated from participants recruited by their own projects.

The core SERL data set of (initially) 8,000–10,000 participants will form an 'Observatory' panel that can be used for observational and longitudinal studies, as well as hypothesis testing. This panel may be suitable to act as an external control group in some scenarios. Over time the smart meter rollout will become more representative of GB, and continued recruitment waves to SERL should reflect this.

Figure 2 shows a graphical illustration of the main structures and processes for SERL. The secure environment is represented by a solid blue line with access via three functions (Data Dispensary, Library and Workshop) in the output layer. Data are drawn in from the DCC Gateway (smart meter data) and specific contextual data sources for SERL-curated (Observatory) and project-curated (Laboratory) datasets. All data drawn into SERL will meet rigorous data access and governance requirements. To make maximum use of the data from experimental studies without necessitating onerous standards for data collection and curation, data from Laboratory studies can be taken back into the Observatory in a 'periphery' area for other researchers to use where consent allows. The arrows indicate principle movements of data though, for clarity, not all data flows are shown.

Data Management Layer

SERL-curated data are data that have been collected by the SERL project or data collected from other sources (e.g. Energy Performance Certificates) that utilise methods for data collection, management, sampling and recruitment that align with core SERL methods and its overall research design framework. **Project-curated data** are data that derive from approved projects that utilise methods that are different to SERL-curated data but still meet basic standards e.g. an intervention study collecting smart meter data via SERL and linking it to project-specific data. These data may still be of some value to future SERL projects even they are not fully compliant with SERL data sampling and recruitment protocols.

One of the key advantages of UK smart meter data for intervention studies is the 13 months of historic data that is stored on the smart meter. This means that interventions can start immediately without having to wait a year for pre-intervention data to be collected. This can almost halve the time of intervention studies. The Observatory will primarily consist of SERL-curated data, while the Laboratory will primarily consist of project-curated data. All data will be appropriately tagged to ensure that data provenance and alignment to SERL protocols

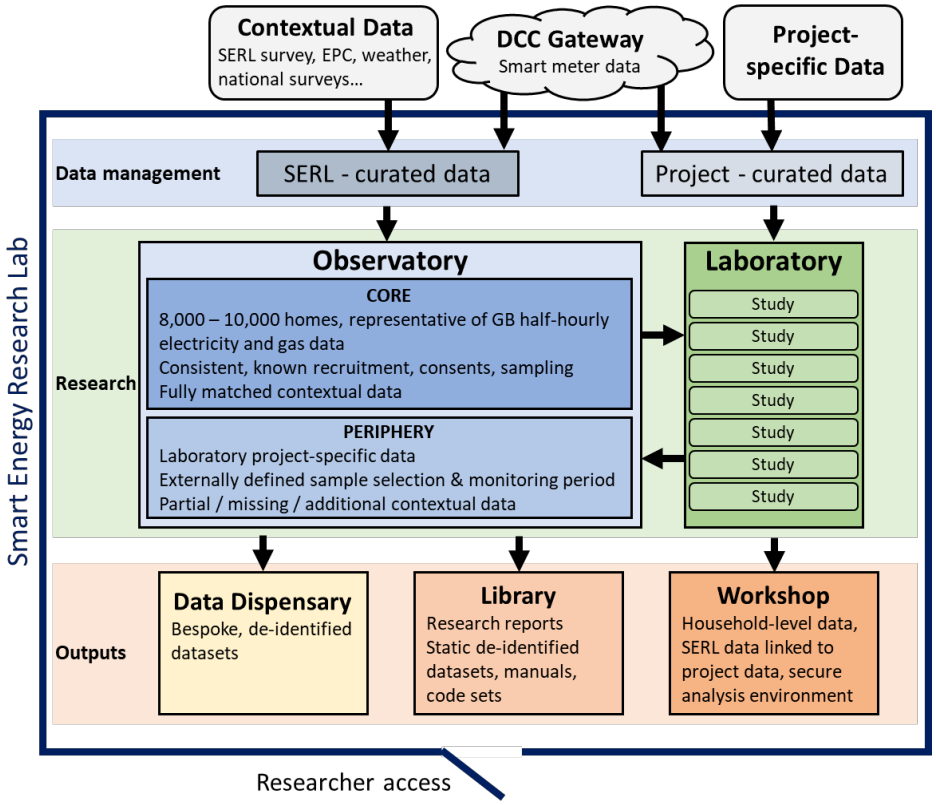

Figure 2. The structure and processes of the Smart Energy Research Lab. The main data flows are indicated by arrows.

is clear and documented. The protocols regarding data curation will be detailed in SERL's Data Access and Governance Framework.

Research Layer

Observatory

The Observatory function itself is comprised of two components: SERL-curated data fed into what we call the '**core**' Observatory: this is the part of the Observatory that is built to look at energy use according to certain SERL-defined standards of data collection and processing. Project-curated data are data used for experimental studies in the Laboratory. These data can then be used to augment the core data sets by providing data collected under different circumstances. However, to manage the range of approaches to sampling and consent, different and new types of contextual data, and to provide researchers with a clear understanding of what data they might get from SERL, any project-curated data from the Laboratory will go into a special area within the Observatory that we call the '**periphery**'. The key features of the core and periphery of the Observatory are outlined below.

The Observatory core is essentially a large database of smart meter data and linked contextual data – the source of smart meter data will be from SMETS-compliant meters that can be accessed via the DCC Gateway. For each participant, half-hourly gas and electricity data will be stored alongside contemporaneous contextual data from other sources at household or local area level, outdoor temperature data, and EPC data. The initial analysis for a suitable sample size is around 8–10,000 homes. The database will form the foundation for high quality observational studies and trend analysis for energy use in the UK and will be a high-quality source for matched control data in Laboratory studies.

The Observatory periphery shares some of the characteristics of the core but is distinguished by being comprised of less-validated and less-complete data. Data stored here will represent a patchwork of data from Laboratory studies. Such data will have limited historical trajectory, bespoke sampling, and bespoke associated contextual data. The data will be available for anyone to download via the appropriate output channels. A potential role for these data is to cross-validate the core data: by analysing the bespoke contextual data it becomes possible to see if there are fundamental biases in the core dataset that arise from the approach to sampling and whether the kinds of contextual data are really the most useful to explore and explain energy use patterns and implications. The periphery will take inputs from the Laboratory and will include all the SMETS-compliant data within the remits of the consents gained for undertaking the research. This may mean for instance, only a year's worth of data, only gas data and so on. Any bespoke contextual data, such as that from a researcher's own survey of inhabitants or technical measurements will also be included.

Laboratory

The Laboratory is different to the Observatory in that it is not a standing database with a consistent sample. The primary purpose of the Laboratory is to facilitate intervention studies. Intervention studies will use participants recruited for each project so as not to "contaminate" the core Observatory panel. It will comprise a range of parallel datasets that exist temporarily while extracted from the DCC according to the needs of the researchers, based on consents of householders who own the data. These samples will not be homes that are currently in the Observatory core (but could be homes that have been in the periphery) and will likely be recruited specifically for each project using SERL's Laboratory. This means that sample sizes will vary according to the needs of individual research projects.

The key input to the Laboratory will be smart meter data via the DCC Gateway for participants recruited by the relevant researcher (providing consent has been obtained). Alongside this, researchers can also request data from the Observatory core (or periphery where this suits their needs) to provide control or other comparison data for their project.

Output layer

Researchers will be able to access data from the Laboratory and Observatory via the Workshop – a secure analysis environment which at the UK Data Archive or online (depending on the level of data sensitivity). The online portal will also have a Library function where reports, de-identified data sets, manuals for using SERL and sets of code can be accessed. The Data Dispensary will be an online tool where bespoke, de-identified datasets can be requested from SERL and released following project assessment and according to strict criteria. The following sections describe these different functions in more detail

Data Dispensary

The Data Dispensary is a simple mechanism for researchers to download specific, customised subsets of Observatory data. Smart meter and matched contextual data will be available to select for download, at which point the datasets will be anonymised and further data matching will no longer be possible. All datasets will go through robust statistical disclosure control and other de-identification processes before download.

Workshop

The Workshop is designed to be a secure online analytical area where accredited researchers are able to access home-level data in order to match data from SERL to data they have collected in allied surveys. This enables researchers to conduct analysis which they wouldn't otherwise be able to do via the dispensary while not undermining the integrity of the data security.

Library

Researchers are also able to access the secure Library space to download a range of materials related to SERL, including anonymised static datasets, manuals, code and research reports. Static datasets together with derived variables relating to particular research reports will be held to enable a complete record of how data from SERL has been used. This will enable scrutiny of prior analysis, repeat analysis using similar data (together with the stored code for the analysis). SERL will be fully documented to enable external researchers to understand what the data within it comprises, its source, how it is validated (or not) and curated. There will also be documents related to the interface SERL develops to enable online self-service. For those researchers using statistical analysis packages like R or SPSS there will be a place to deposit code relating to particular derived variables or undertaking different classes or types of analysis. This also leaves

open the potential for there to be a kind of open source repository (e.g. similar to Github) for code development in relation to SERL data analysis. Reports, research articles, technical reports relating to the functioning of SERL and analyses developed by the SERL team based on data from the Observatory core will be indexed for easy discovery in a research library. Ideally code and datasets used to construct the reports will be linked to relevant items in the Datasets and Code sections.

Recruitment and sampling

We aim to recruit from a random sample of homes who have a smart meter installed that is accessible via the DCC Gateway i.e. a SMETS2 meter or a SMETS1 meter once they have been migrated into the DCC Gateway. We will contact such addresses by letter and invite households to participate in SERL. The invitation letters will have to be addressed 'to the occupier' since we will only have access to addresses, and not the names of the occupants. This will reduce the response rate to the invitation letters as many people in the UK throw away this type of mail without opening it.

PILOT STUDIES

To maximise our response rate and minimise response bias we are conducting a series of pilot studies to test variants that might maximise recruitment rates. This will form the initial wave of recruitment to the SERL Observatory, and inform us about expected response rates for the main phase of recruitment. The pilot studies will be conducted during Spring-Summer 2019.

Testing response rates and channels

We would prefer participants to sign up to SERL by completing the consent form and survey online. However, we do not wish to exclude people who are unable/unwilling to go online, and we do not wish to unnecessarily bias our sample. We will test sending each participant an initial invitation letter and up to 3 reminders containing instructions for how to respond online. We will also send a paper consent form and survey (with prepaid return envelopes) in 1 or 2 of these mailings. This will inform us about how many reminders to send to non-responders, and when best to include the paper documents.

Communication content and wording

The content of reminder letters may need to vary in order to appeal to a range of people with different motivations across all demographics and thus increase the chances of participation. Different letters will be written for the reminders and invitations with different styles of emphasis, such as a 'community/future benefits' approach or a focus on the loss of wider public benefit by not participating.

Incentives

SERL is an academic research project that aims to facilitate innovation for the public good. This theme will be heavily emphasized in communications to potential participants to encourage participation and thus we expect "public good" or altruism to be the primary motivation for recruitment. However, we will also test the use of small incentives to understand the impact on response rates, such as conditional incentives with monetary value e.g. (£5 voucher) or unconditional, low-value, non-financial incentives such as a pen or fridge magnet.

Survey

All participants will be asked to complete a paper or online survey when they consent to participate in SERL. The aims of the survey are to:

- collect key contextual data variables that help explain observed variability in energy demand;
- identify sub-groups and assess sample representativeness.

The paper survey will be fairly short (a few pages) and contain the same questions as the short online survey. We will ask for basic details about the household (such as occupants' ages, occupations, relationships), and the dwelling (such as the number of rooms, main sources of heating and presence of micro-generation). We will also test offering some online respondents a longer survey.

Consumer attitudes

We are currently considering a second, separate pilot project that would not involve (or recruit) potential participants for SERL. This project would seek to understand public views on sharing smart meter data, such as:

- the willingness of the general population to share smart meter data for academic research, and in particular for the SERL project;
- how data granularity affects willingness to share;
- whether offering a tailored energy advice service would increase willingness to participate in SERL.

These questions would be asked of a nationally representative sample of the GB population.

Data governance

SERL will process personal data, which is regulated by the General Data Protection Regulation (GDPR) and UK Data Protection Act 2018 (DPA 2018). Some of the personal data that SERL will process is smart meter data, which is additionally regulated by the Smart Energy Code (SEC). SERL is a UK University research project and requires ethics approval by an independent University Ethics Board.

A SERL Data Governance Framework has been developed, led by the UK Data Archive team. The framework consists of robust protocols that govern both inbound and outbound data governance:

- Inbound data governance – issues relating to the ingestion of data into the portal – primarily data privacy issues and compliance with GDPR and the SEC;
- Outbound data governance – issues relating to the provision of data out to researchers e.g. "5 Safes" protocols (Desai, et al., 2016)
- Endogenous data management – SERL's internal data management processes that are designed to ensure that we securely manage data in compliance with inbound and outbound governance protocols.

Table 1. Assessment of possible authentication processes for SERL. In some scenarios, a consumer should be offered several authentication options. E.g. if they can provide an In Home Display (IHD) ID this process could be used but, if not, a meter reading could be chosen, or a notification letter sent.

	Authentication method	Channel	Criteria				
			Risk of Harm	Credibility	Consumer Burden	Skew	Impact on DCC User
Solution	No additional Authentication	Mail, face-to-face	Very Low	High	Very Low	Very Low	Very Low
	Existing Authentication	Assess original source	Very Low	Assess original source	Assess original source	Assess original source	Assess original source
	IHD Device ID	Phone, digital	Very Low	High	Low–Moderate	Low–Moderate	Low–Moderate
	Meter reading	Phone, digital	Very Low	Moderate–High	Moderate	Low–Moderate	Low–Moderate
	Letter	All	Very Low	High	Low	Low	Moderate
	CIN	Phone, digital	Very Low	Moderate	High	Moderate–High	High
	Energy Bill	Phone, digital	Very Low	High	High	Moderate–High	High

The Data Governance Framework will be used by SERL's Data Access and Governance Board to approve research projects using the portal and ensure they comply with the framework.

Consumer authentication

THE ISSUE

Before SERL can collect smart meter data from participants who have provided consent there is an obligation on DCC Other Users to undertake an additional authentication check, as specified in SEC I1.5[5]:

> **I1.5** Each User shall put in place and maintain arrangements designed in accordance with Good Industry Practice to ensure that each person from whom it has obtained consent pursuant to Section I1.2 to I1.4 is the Energy Consumer.

Existing examples of Good Industry Practice arrangements (e.g. use of the Customer Identification Number (CIN) or checking energy bills) place a burden on the consumer that is likely to lead to low completion rates of the consumer authentication process and present a significant barrier to the successful delivery of SERL, and potentially other services offered by DCC Other Users. Examples of why this may affect completion rates include lost energy bills or difficulty in accessing the meter which in some cases may be locked away and only accessible by the landlord.

THE SOLUTION

We have proposed that Good Industry Practice arrangements for consumer authentication should include a range of "approved" processes that are contextual and pragmatic thus benefiting consumers and DCC Other Users alike. Factors for evaluating authentication processes include:

- The **channel** by which consent has been obtained.
- The **risk of harm** from data being obtained by an incorrectly-authenticated party.
- The **credibility** delivered by the authentication process i.e. the certainty that the consumer, meter and address are linked to consent obtained.
- The **burden** on the consumer and their ability (and preference) to respond to different authentication processes.
- The risk of an authentication process disadvantaging or under-representing some socio-demographic groups more than others, such as the vulnerable, elderly, fuel-poor etc who may be less able/willing to complete some authentication processes than other groups. This would **skew** the sample of respondents.
- The **impact on the DCC Other User**'s operations and service offered:
  - Impact of low completion rates;
  - Cost of implementing the authentication process;
  - Scalability of the authentication process.

The benefits of using a more stringent authentication process must be weighed against any negative impacts on the consumer and DCC other user. For example, a high-risk scenario might require authentication using a scanned energy bill or a site visit, whereas a low-risk scenario may only require a letter notifying the household that consent has been received.

We believe a suite of "approved" authentication processes that take into account relevant criteria would provide a pragmatic, flexible solution that would benefit both the consumer and DCC Other Users alike. The authentication processes that might be applicable to SERL are summarised in Table 1 against the criteria described in the previous section.

The SERL team have discussed this issue extensively with the UK government Department for Business, Energy & Industrial Strategy (BEIS), the Smart Energy Code Administrator and

5. The most recent version of the Smart Energy Code (SEC) is available here: https://smartenergycodecompany.co.uk/the-smart-energy-code-2/.

Secretariat (SECAS), the DCC independent data privacy auditor (Deloitte), and many SEC Parties and responded to the formal consultation on SECI1.5 obligations. This resulted in changes incorporated into v2.0 of the Privacy Controls Framework (Smart Energy Code, 2018) which gained SEC Panel approval in August 2018. Deloitte have provided verbal assurance that SERL's approach to consumer authentication will comply with SECI1.5 obligations.

SERL proposal for consumer authentication

The key accepted change for consumer authentication in the Privacy Controls Framework (discussed above) is that no additional authentication checks will be necessary where the initial consent process is inherently self-authenticating, i.e. the consent can only have come from an occupant of the smart meter's registered address. Examples would include where consent has been obtained by a letter posted to a specific address or a face-to-face survey at the relevant address. Previously only the CIN and energy or council tax bills were accepted for consumer authentication, which would have placed a huge burden on participants in this and other projects.

Most SERL participants will be recruited by a letter sent to a specific address. The letter will contain a unique code allowing the consumer to opt-in to the SERL research project by providing consent via a web portal or by alternative means, such as by mail. We have successfully argued that for most SERL scenarios no further consumer authentication is required, due to:

- No burden on consumers and thus 100 % completion rates;
- No additional cost of implementation (above the initial consent process);
- No risk of disadvantaging/under-representing certain groups;
- High credibility authentication – as the initial process is inherently self-authenticating.

Where SERL obtains consent via alternative recruitment methods (e.g. by phone or email) that do not have an inherent link to the meter's registered address, SERL will ensure that a context-relevant authentication process takes place.

Challenges for SERL

The Smart Energy Research Lab is still in development, and as the project is the first of its kind in the UK (with no obvious peers internationally), there are a number of challenges to overcome to reach our goals.

- There have been several delays to the rollout of SMETS2 meters, and as yet no firm timelines for migrating SMETS1 meters (the vast majority of the 12+ million installed meters to-date) into the DCC Gateway (enabling SERL access).
- The smart meter rollout (in particular the SMETS2 rollout) is not yet nationally-representative, and certain areas and types of dwelling are more likely to have smart meters installed than others. SERL participants will therefore initially be representative of the smart meter rollout rather than GB households.
- DCC Adaptor services for Other Users (OU) (such as SERL) are either immature or still in development, and SERL will likely encounter technical or operational issues by merit of being one of the first DCC Other Users actively collecting data via the DCC Gateway.
- Response bias is a challenge for SERL because participants must be prepared to complete a consent form and survey online or by post. Certain population demographics are more likely than others to open envelopes sent to their address, engage with the project and agree to data collection and linking. By capturing householder and dwelling characteristics on the survey we hope to identify sources of bias in our sample. The large sample size will also increase the statistical power of the data we collect.

The pilot studies described above will provide us with valuable information about consumer response rates and bias, as well as highlighting any practical/technical challenges using the DCC Gateway to access data.

Accessing SERL data for research

Researchers at UK universities will be able to access SERL data so long as the pre-requisites for data access are met. These pre-requisites are important for safeguarding the data privacy of the individuals whose personal data we hold. As a result, we will not share personal data with third parties other than to accredited researchers for approved scientific research. We will use best practice measures to safeguard personal data as developed by the UK Data Archive. For example, identifiable personal data will be pseudonymised and accessed within a secure lab environment. International researchers will be able to access SERL data indirectly through collaboration with academic UK researchers and providing all other safeguards are met.

PRE-REQUISITES FOR ACCESS TO DATA

The following are essential pre-requisites to access SERL data:

- Research projects will need to have prior ethics approval from, for example, their university's research ethics committee.
- Researchers will need to have undertaken a Safe User of Research Data Environments (SURE) training course. This will cover using the secure lab at the UK data archive, statistical disclosure control for making statistical outputs safe, and data security and personal responsibility including legal background, breaches and penalties (UK Data Service, n.d.).
- Registration requiring an institutional address and agreement to an End User License. Users will be required to log on to the online data portal with two-factor authentication.
- Application process submitted to our Data Governance Board for approval. Each application for data will be treated as a separate project that must be in the public interest.

ACCESS ROUTES

There are two pathways for researchers to physically access SERL data:

- Directly download (anonymised) files from the online portal to a local computer;

- Access in a "secure lab" (sometimes referred to as a virtual desktop) which prevents data from being downloaded. The secure lab environment will include a range of analytical tools, software and High Performance Computing facilities to enable researchers to conduct analyses of high resolution data at scale.

The decision as to whether data in a project can be downloaded or must stay within the secure lab environment, will be determined by the Data Governance Board on a project-by-project basis, according to the level of disaggregation (by time, geography, etc.) and types of linked data requested. Disclosure control will be applied according to the access route and identifiability of the requested data.

Discussion

Reducing the world's carbon emissions to combat climate change requires energy consumption to reduce and become more flexible. Research and innovation are needed to develop demand-side technologies, energy policies, and better understand consumer attitudes and behaviours. Smart metering has the potential to underpin research in areas including (but not limited to) dwelling-level consumption patterns, building improvements for energy efficiency, dynamic and locational energy pricing, demand shifting and reduction, and peer-to-peer energy trading.

In many countries researchers face barriers to accessing energy consumption data of sufficient quality, time and spatial resolution, longevity, sample size and detail (inclusion of relevant contextual factors). For example, studies that monitor energy consumption in homes typically suffer from small sample sizes due to the cost of installing monitoring equipment, and sample bias from self-selection into trials such as energy advice services or time-of-use tariffs. Comparing datasets from different sources can be extremely difficult or impossible when the raw data are unavailable, and key factors are hidden behind non-standard classifications of consumer 'type', for example.

The Smart Energy Research Lab is unique in providing both 'Observatory' and 'Laboratory' functions for researchers. The Observatory gives access to half-hourly smart meter data with contextual variables such as location, weather, and dwelling and resident information. The Laboratory allows researchers to recruit their own participants, collect their own data, have it linked to existing SERL data resources, and compared against the Observatory control group. In allowing the data to be accessible to accredited UK researchers, SERL takes on the considerable costs and practical challenges of becoming a DCC 'other user', acquiring a DCC adaptor service, developing a data governance framework, and storing and managing the data on behalf of all UK research institutions.

SERL is still being developed and will begin data collection in the summer of 2019. The secure lab environment is expected to be ready for researchers by the end of 2019. SERL will continue recruiting participants and undertaking research for at least a further 3 years and has been set up so that it has a long-term future with minimal operational costs, to provide the UK with an independent observatory of energy demand in GB housing stock over time.

While there are many individual research projects utilising smart meter data across the globe, we are not aware of any other projects that are developing a large-scale, shared resource focussed on academic research in a similar manner to SERL. Please contact the authors if you wish to share knowledge in this area.

Acronyms

CAD	consumer access device
CIN	consumer identification number
DCC	Data Communications Company
EPC	Energy Performance Certificate
GB	Great Britain
GDPR	General Data Protection Regulations
HEMS	Home energy management system
IHD	In-home display
OU	Other User (of the DCC Gateway)
SEC	Smart Energy Code
SERL	Smart Energy Research Centre
SMETS	Smart metering equipment technical specification

References

Allcott, H. & Mullainathan, S., 2010. Behaviour and energy policy. Science, Volume 327, pp. 1204–1205.

Data Communications Company, 2018. Products and Services. [Online] Available at: https://www.smartdcc.co.uk/products-services/ [Accessed 2 January 2019].

Department for Business, Energy & Industrial Strategy, 2018. Domestic NEED: Methodology Note, s.l.: s.n.

Department for Business, Energy & Industrial Strategy, 2018. Statistical release and data: Smart Meters, Great Britain, quarter 3 2018. [Online] Available at: https://www.gov.uk/government/statistics/statistical-release-and-data-smart-meters-great-britain-quarter-3-2018 [Accessed 18 December 2018].

Department of Energy and Climate Change (DECC), 2015. Smart Meters, Smart Data, Smart Growth.

Desai, T., Ritchie, F. & Welpton, R., 2016. Five Saves: designing data access for research. s.l.:s.n.

Frondel, M. & Schmidt, C. M., 2005. Evaluating environmental programs: the perspective of modern evaluation research. Ecological Economics, 55 (4), p. 5150526.

Hamilton, I. et al., 2015. Co-benefits of energy and buildings data: The case for supporting data access to achieve a sustainable built environment. International Conference on Sustainable Design, Engineering and Construction.

Hartman, R. S., 1988. Self-selection bias in the evaluation of voluntary energy conservation programs. Review of Economics and Statistics, 70 (3), pp. 448–458.

Hierzinger, R. et al., 2012. European Smart Metering Landscape Report 2012 – update May 2013, Vienna: Österreichische Energieagentur – Austrian Energy Agency (AEA).

Hledik, R., Bagci, P. & Chhachhi, S., 2018. Two Paths for Advancing Great Britain's Smart Metering Programme: A Discussion Paper. s.l.: The Brattle Group.

Skea, J., 2012. Research and evidence needs for decarbonisation in the built environment: a UK case study. Building Research & Information, 40 (4), pp. 432–445.

Smart Energy Code, 2018. Privacy Controls Framework. [Online] Available at: https://smartenergycodecompany.co.uk/privacy-controls-framework/ [Accessed 3 January 2019].

Summerfield, A. J. & Lowe, R., 2012. Challenges and future directions for energy and buildings research. Building Research & Information, 40 (4), pp. 391–400.

Train, K. E., 1994. Estimation of net savings from energy-conservation programs. Energy, 19(4), pp. 423–441.

UK Data Service, n.d. Access to the Secure Lab. [Online] Available at: https://www.ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab/train [Accessed 3 January 2019].

UK Power Networks, 2014. SmartMeter Energy Consumption Data in London Households. [Online] Available at: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households [Accessed 3 January 2019].

Acknowledgements

This work has been funded by the Engineering and Physical Sciences Research Council through grant EP/P032761/1. There are over 30 individuals across 8 organisations in the SERL Consortium (University College London, the University of Essex (UK Data Archive), University of Edinburgh, Cardiff University, Loughborough University, Leeds Beckett University, the University of Southampton and the Energy Saving Trust) who have contributed to the development of SERL and thus the content of this paper.