Imperial College London

Balancing Privacy and Access to Smart Meter Data

An Energy Futures Lab Briefing Paper

Dr Fei Teng
Saurab Chhachhi
Pudong Ge
Jemima Graham
Prof Deniz Gunduz

energy futures lab
An institute of Imperial College London

April 2022

Acknowledgments

Energy Futures Lab is one of seven Global Institutes at Imperial College London. The institute was established to address global energy challenges by identifying and leading new opportunities to serve industry, government and society at large through high quality research, evidence and advocacy for positive change. The institute aims to promote energy innovation and advance systemic solutions for a sustainable energy future by bringing together the science, engineering and policy expertise at Imperial and fostering collaboration with a wide variety of external partners. The Energy Futures Lab Briefing Papers are periodic reports aimed at all stakeholders in the energy sector. They bring together expertise from across Imperial College London to provide clarity on a wide range of energy topics. For more information visit:
http://www.imperial.ac.uk/energy-futures-lab

Energy Futures Lab

Supported by Research England's Strategic Priorities Fund (2021-22)

Table of Contents

List of Acronyms

BEIS	Department for Business, Energy, and Industrial Strategy
CSP	Communications Service Provider
CAD	Consumer Access Devices
DAPF	Data Access and Privacy Framework
DCC	Data Communications Company
DP	Differential Privacy
DNO	Distribution Network Operator
FL	Federate Learning
GDPR	General Data Protection Regulation
GB	Great Britain
HAN	Home Area Network
ICO	Information Commissioner's Office
IHD	In-Home Display
IoT	Internet of Things
MHHS	Market-Wide Half-Hourly Settlement
NILM	Non-Intrusive Load Monitoring
OFGEM	Office of Gas and Electricity Markets
P2P	Peer-to-peer
SECC	Smart Energy Code Company
SMETS	Smart Metering Equipment Technical Specifications
SMIP	Smart Meter Implementation Programme
WAN	Wide Area Network

List of Acronyms

Executive Summary

Digitalising the energy system is expected to be a vital component of achieving the UK's climate change targets. Smart meter data, in particular, is seen a key enabler of the transition to more dynamic, cost-effective, cost-reflective, and decarbonised electricity. However, access to this data faces a challenge due to consumer privacy concerns. This Briefing Paper investigates four key elements of smart meter data privacy: existing data protection regulations; the personal information embedded within smart meter data; consumer privacy concerns; and privacy-preserving techniques that could be incorporated alongside existing mechanisms to minimise or eliminate potential privacy infringements.

Background

Domestic smart metering provides a range of potential benefits for consumers, energy suppliers and network operators as well as opening possibilities for innovative business models and pricing schemes. The Department for Business, Energy and Industrial Strategy (BEIS) projects automated meter reading, improved load forecasting and network management, energy and emissions reductions from real-time informational feedback and demand shifting to bring benefits of up to £19.5bn. Many of these benefits are contingent on or would be enhanced by access to high-resolution consumption data.

Concurrent with the introduction of domestic smart metering, access to consumer data has raised concerns around privacy and data misuse, particularly with the increase of high-profile data breaches and exposures of how seemingly harmless data can be used/misused to influence our daily lives. The introduction of the General Data Protection Regulation (GDPR) is forcing companies and regulators to rethink the way in which they engage with customers and how data is handled. In addition, questions around the value of data and how the benefits of data sharing are being distributed are being debated. Although data sharing can result in benefits to customers and the environment, data could also be leveraged in order to enhance profits through discriminatory pricing.

The Smart Meter Implementation Programme (SMIP), the roll-out of smart meters to all households in the UK, has recognised that smart metering presents a significant privacy risk, instituting several regulatory safeguards. The Data Access and Privacy Framework (DAPF) sets out consumers' rights and choices with regards to how and by whom their smart meter data can be accessed. Although it aims to deliver smart meters to all domestic consumers, it is an opt-in process since customers must request and consent to install a smart meter. Once a smart meter is installed, customers can choose whether they wish to share their consumption data at a monthly, daily, or half-hourly frequency, whether their supplier may use the data for marketing purposes and whether the data can be passed on to third parties. However, there have a been several recent developments which could fundamentally change how and by whom smart meter and associated consumption data can be accessed. Smart appliances and demand response are expected to play a significant role in the energy transition. This will generate complementary data streams with additional, more granular data on energy consumption which are outside the scope of the DAPF. In addition, there are calls for widening access to smart meter data for uses beyond the day-to-day operation of the electricity network as part of a move to digitalise the energy sector. A key component of this is to develop Open Data platforms which will provide access to public-interest actors to inform and shape policy.

Executive Summary

Potential privacy infringements

The types of personal information extracted are highly dependent on the data-sharing options selected by customers. High-resolution data, such as half-hourly data, can be used to infer a wide range of information about a household. Non-intrusive load monitoring algorithms can disaggregate the total consumption data logged by smart meters into individual appliances with high accuracy (up to 95%). As a result, it is possible to understand people's routines and preferences. Even at low resolutions, smart meter data, can also be used to determine socio-economic and demographic information about a household. The potential privacy infringements and the consequences of data sharing are not limited to identifying specific household characteristics. Smart meter data can be linked with other data sources such as social media and other smart devices such as thermostats to provide deeper insights. However, due to the 'black-box' nature of machine learning, even data analysts may not be able to predict what their machine learning algorithms might infer from these combined data sources.

Privacy-preserving techniques

Privacy-preserving data sharing techniques can provide a means to share high-resolution data more widely while protecting customers from potential privacy infringements.

- Data obfuscation techniques such as pseudonymisation and aggregation alter data to remove information that may be considered sensitive. They can provide easily implementable and cost-effective options to reduce the risk of privacy infringements. However, neither offers guarantees of anonymity as they are suspectable to reconstruction and linking attacks.
- Differential privacy is a technique, rooted in cryptography, which introduces noise into datasets to provide mathematical guarantees of anonymity. This overcomes the vulnerabilities in pseudonymisation or aggregation schemes. However, data accuracy is reduced by adding noise, introducing a privacy-utility trade-off.
- Homomorphic encryption is another promising privacy-preserving technique. It allows data analysts to perform arithmetic operations (e.g. addition, subtraction, multiplication, and division) on encrypted data ensuring that underlying data cannot be accessed while maintaining the accuracy of results. It requires neither secure communication channels nor a trusted third party. However, it does require significant computational resources to run the encryption and decryption algorithms.
- User demand shaping is another technique that functions behind the meter by altering actual consumption patterns. Appliances and activities have characteristic consumption profiles. Changing the consumption data seen by the smart meter can hide specific appliances or more generally, limit the amount of information inferred from the consumption data recorded at the smart meter. This is achieved through smart control of flexible assets, such as batteries, to alter actual consumption to hide particular characteristics which may be considered sensitive.
- Distributed learning techniques, such as federated and peer-to-peer learning, ensure raw data (i.e. consumption data) is kept locally (e.g. in the smart meter) thereby enhancing privacy. By reducing the amount of sensitive information that is shared between devices or entities, the potential for privacy infringements is limited. However, even though raw data is not shared directly, federated learning models can indirectly leak information as this framework does not provide any formal privacy guarantees. In most applications, privacy guarantees are provided by incorporating one of the above privacy-preserving mechanisms.

Recommendations

As of September 2021, 47% of domestic meters in the UK are smart meters. However, many barriers remain to the widespread sharing of high-resolution smart meter data and a reliance on permissions controls, making it difficult to safeguard consumer data against misuse, intentional or otherwise. It is essential to clearly communicate the implications, in terms of the personal information being shared, of the different options (half-hourly, daily, monthly) provided within the DAPF and how that data will be used. This is necessary to foster genuinely informed consent. Transparency around how smart meter data will be used, the benefits of data sharing, and how these benefits will be distributed should be communicated to consumers through, for example, Citizen's Advices' proposed Data Dashboard. The benefits of providing more granular data can, for most regulated activities, be quantified and should also be clearly set out. Limiting access to data to only what is necessary and where the benefits have been clearly identified prevents data hoarding and limits potential privacy infringements. Privacy-preserving techniques provide a framework within which such a privacy-utility trade-off can be explored and quantified. These techniques give consumers greater control over their data and force those wishing to access their smart meter data to incentivise them. This will allow consumers to reveal their true preferences, provide them with different levels of privacy protection, and maximise access to high-resolution data. Going beyond permission control, the use of privacy-preserving mechanisms can ensure data is not easily re-identifiable. Even in the case of a data breach, any inferences cannot be linked back to individual consumers. Importantly, this is done while preserving the utility of the data so that the benefits can be realised.

Therefore, policymakers and regulators could consider introducing one or a combination of the privacy-preserving techniques discussed in this paper. This could enable the balancing of privacy and access to smart meter data. Without significant changes to how data is stored, processed, and used, the adoption and resulting benefits of smart meters may not be realised. Differential Privacy, in particular, is a prime candidate to achieve this given that it can be implemented either centrally or in a distributed manner. It provides provable, tuneable and future proof guarantees of privacy protection and has been implemented in other sectors. The US Census Bureau provides a detailed framework which could be built upon to integrate within the UK's existing Smart Meter Implementation Programme.

1. Introduction

Great Britain (GB) is currently upgrading its national electricity and gas metering infrastructure by introducing digital smart meters through the Smart Meter Implementation Programme (SMIP). It is seen as a key component of the digitalisation of the energy sector and facilitates the transition to a more dynamic, cost-effective, cost-reflective, and decarbonised electricity network.

Smart meters allow for logging and accessing high resolution consumption data as well as enabling integration of smart devices and automated load control for the domestic sector. Access to high resolution consumption data provides a range of potential benefits for consumers, energy suppliers and network operators as well as opening possibilities for innovative business models and pricing schemes. These include operational benefits for suppliers through the automation of meter readings and customer switching, consumer benefits in terms of time savings and bill savings from energy usage reductions, and improved system management through greater network visibility and demand flexibility. BEIS has stated that digitalisation of the energy sector is a vital component of the strategy to achieving net zero (BEIS, 2020b, 2021a). At the moment, the programme has been significantly delayed with only 47% of homes and small businesses having installed a smart meter as of September 2021 (BEIS, 2021d).

One major hurdle is the concern raised around privacy and data misuse, as smart meters can collect personal information on individuals at high resolution. Concurrent to the introduction of domestic smart metering, awareness of issues surrounding data privacy and misuse is growing. The introduction of the General Data Protection Regulation (GDPR) is forcing companies and regulators to rethink the way in which they engage with customers and how data is handled (Véliz and Grunewald, 2018a). Infringements of GDPR can result in large fines (€20 million, or 4% of the firm's worldwide annual revenue) and in the three years since the introduction of GDPR over £250 (€303) million have been issued in fines (Data Privacy Manager, 2021). The fines have ranged from lack of valid consent and transparency to inadequate security mechanisms and data misuse. Increasingly high-profile data breaches (BBC News, 2019) and exposés of how, seemingly harmless, data can be used to influence individuals' daily lives are creating a more engaged and cognisant society (Lu, 2020). Additionally, questions around the value of data and how benefits of data sharing are distributed are increasingly being discussed.

The SMIP has taken privacy into consideration right at the outset with the development of a Data Access and Privacy Framework (DAPF) (BEIS, 2018) and multiple privacy impact assessments (Banks and Mcglinchey, 2019). The latest privacy impact assessment was conducted by Ofgem as part of a consultation on data sharing options for market-wide half-hourly settlement which also proposed the use potential for privacy enhancing technologies (OFGEM, 2020). However, a lack of understanding as to the costs and benefits of such mechanisms meant these were not adopted. Given that many of the benefits of smart metering are contingent on widespread adoption and the uses and privacy vulnerabilities of smart meter data are evolving, it is important to consider various techniques to future-proof the privacy preservation of smart meter data. This report provides an overview of the information embedded within smart meter data, the benefits of data sharing and the technical privacy preserving mechanisms which could allow for wider dissemination of smart meter data while also protecting consumer privacy.

This report focuses on electricity consumption data due to the dynamic nature of consumption, market conditions and the wider uses of electricity although similar benefits (Sustainability First and CSE, 2021b) as well as privacy risks exist in relation to high-resolution gas consumption data[1].

Chapter 2 lays out the existing regulatory and technical framework of GB's smart meter roll-out, focusing on dataflows and privacy measures.

Chapter 3 discusses the benefits of smart metering, the potential transformation in the energy sector it enables, and how these depend on data sharing.

Chapter 4 looks at consumer perceptions and privacy concerns with regard to sharing smart meter data.

Chapter 5 describes the types of personal information that can be inferred from smart meter data and explores the potential privacy infringements that could arise.

Chapter 6 explores a range of privacy-preserving techniques that could be introduced to address consumers' concerns while simultaneously allowing for broader access to high-resolution smart meter data.

Chapter 7 presents a summary and offers a range of recommendations on proactive and preventative measures to ensure consumer privacy and increase sharing of high-resolution smart meter data.

[1] The majority (83%) of domestic gas consumption in the UK is for heating and cooking (BEIS, 2020a), and this is expected to be phased-out as part of the UKs decarbonisation plans.

2. The Smart Meter Implementation Programme (SMIP)

The SMIP requires the roll-out of 53 million digital smart gas and electricity meters for the 30 million domestic and smaller non-domestic properties in GB (DECC, 2013). The programme consists of several components (see Figure 1):

- Smart meters: digital electricity and gas meters capable of recording consumption as well as storing tariff and credit information.
- Communications hubs: devices that create a Home Area Network (HAN) to which consumers can connect In-Home Displays (IHD), to see near-real time usage and cost information, and other smart devices. The communications hub connects to the dedicated Wide Area Network (WAN), run by a communications service provider (CSP), through which it is possible to share the data logged by smart meters with other users. Additionally, consumer access devices (CAD) can connect to the HAN providing an alternative means of data access.
- Data Communications Company (DCC): a centralised regulated entity that is responsible for gathering data from smart meters across the country, verifying and processing the data, and providing a gateway for authorised users (e.g. suppliers, distribution network operators (DNO) and others) to access smart meter data.

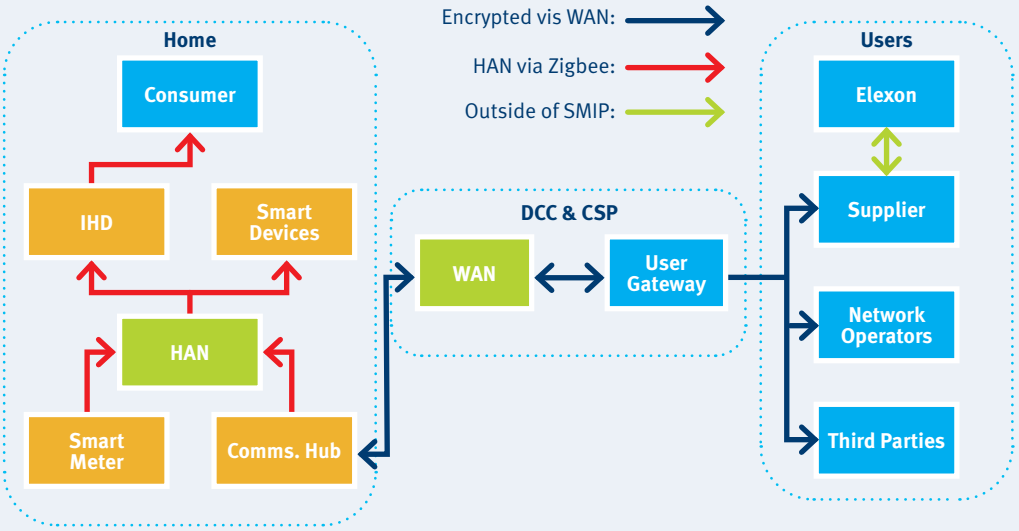


Figure 1 Dataflow for Smart Metering in GB

Several features of the SMIP distinguish it from smart meter rollouts elsewhere (Hledik, Bagci and Chhachhi, 2018):

- Supplier-led: Most national smart meter rollouts have been led by DNOs, as they oversee maintaining network infrastructure and their remit is usually defined by geography. In GB it was decided that the roll-out would be supplier led as they already own existing traditional meters and have existing relationships with consumers.
- Voluntary/Opt-in: Although the SMIP aims to achieve 95% coverage, getting a smart meter is entirely voluntary. This is mainly due to privacy concerns and successful legal challenges to a mandatory roll-out in other countries (Cuijpers and Koops, 2013).
- Mandatory IHDs: One of the major expected benefits of roll-out is energy usage reductions from feedback on energy usage. IHDs play a key role in this as they provide near real-time information on energy usage and costs. The SMIP mandates that energy suppliers must offer consumers a free IHD with their smart meter.
- Data Access and Privacy Framework (DAPF): Through multiple consultations and reviews, privacy and security have been considered throughout the planning and implementation process. This has resulted in a framework to safeguard consumers' privacy interests whilst enabling proportionate access to data (BEIS, 2018).

The process of installing a smart meter under the SMIP can be summarised into the following four steps (Smart Energy GB, 2021):

1. Opt-in: Customers must request a smart meter from their energy supplier. Energy suppliers are therefore required to actively encourage their customers to sign up. In addition, Smart Energy GB, a government entity, provides information around the benefits of smart meters, runs advertising campaigns and research into customer experience.
2. Meter Infrastructure: The smart meter roll-out in the UK is supplier-led. Suppliers are free to choose which smart meters to install as long as they conform to minimum technical specifications and ensure interoperability in the event a customer wishes to switch supplier.
3. Installation: Suppliers book appointments directly with their customers. Upon installation suppliers are required to provide information on data sharing options. In addition, they offer an IHD which allows customers to see their energy consumption and bills in near real-time (Citizen's Advice, 2016).
4. Post-installation: Suppliers are required to remind customers of their current data sharing options on regular basis. Customers can change their options at any time.

2.1 Technical capabilities – SMETS 2

The SMIP allows suppliers to choose which smart meters they want to install and what functionality it should provide. To ensure minimum standards for smart metering infrastructure and interoperability, BEIS set out the Smart Metering Equipment Technical Specifications (SMETS)(BEIS, 2014). It specifies requirements relating to data collection, data transmission and other functionality. The initial specifications, SMETS 1, were set out in 2014 with an updated version, SMETS 2, being published in 2018. Although the core data logging and other functionality are the same, SMETS 2 ensures interoperability across suppliers and enhanced security measures through end-to-end encryption and a dedicated closed communications network.

SMETS 2 certified smart meters can log key parameters including active and reactive energy consumption as well as export (where households have an electricity generation source such as solar panels or storage with grid export capabilities such as electric vehicles or batteries) and voltage readings at high resolution (up to 10 second resolution). The meters are also able to store historical data going back 13 months at a half-hourly resolution. In addition, the smart meter stores tariff, credit and debt information, sent by the supplier, to provide near real-time consumption costs and billing information.

SMETS 2 also incorporates smart control facilities that allow for remote load management and protection as well as remote disconnection of supply. Auxiliary load control switches which can either be connected physically to the meter or wirelessly through the HAN can be programmed to turn on/off based on schedules set by the supplier and stored in switching table or on an ad-hoc basis through a 'Boost' function. Although this functionality is currently limited to on/off events, future versions of SMETS will look to implement Proportional Load Control, which will allow for more fine-grained control actions to take place (BEIS, 2019a).

Table 1 Smart Metering Equipment Technical Specifications Minimum Functionality (SMETS 2)

Data Logging	Data Transmission	Other
Active & Reactive Energy Imports and Exports at 10 second resolution	Two-way communications	Payment Mode (Credit or prepayment)
Time of use pricing rates	Encrypting and decrypting data using the following cryptographic algorithms: Elliptic Curve DSA, Elliptic Curve DH and SHA-256.	Auxiliary load control switches
Meter balance and debt registers	Joining a ZigBee SEP v1.2 Smart Metering Home Area Network	Load Limiting, thresholding and remote disconnection
Historical consumption and cost data at different resolution going back to the previous thirteen months		Maximum meter power consumption of 4 Watts

A key component of the smart meters is their ability to run encryption algorithms to securely verify commands coming from suppliers and other authorised parties and send data over the WAN network these parties via the DCC. Many of the SMETS 2 certified smart meters also include tamper proofing and detection mechanisms which can also be reported back suppliers automatically.

2.2 Data sharing options

Consumers can access the data stored on their smart meter either through the smart meter itself or using an IHD, which must be offered to consumers and provides a more user-friendly interface. The functionality of the IHD varies between suppliers but the SMETS regulations set out minimum standard. The IHD must be able to display near real-time consumption information both in terms of usage (kWh) and cost. It must also facilitate access to billing and debt information and allow those on pre-payment plans to top-up their accounts. The physical infrastructure, smart meters, and communications devices, are owned by suppliers but the consumers have control over how and with whom the consumption data, logged by these devices, is shared. Although SMETS specifications allow for data to be recorded at very high resolutions, the DAPF sets out three data sharing frequencies from which consumers can choose: monthly, daily (default) and half-hourly (see Figure 2 for an illustrative example). These options have been summarised into a Data Guide developed by Citizen's Advice (Citizen's Advice, 2016). Additionally, suppliers require explicit consent to use the data for marketing purposes or to pass it on to third parties. Suppliers are therefore forced to incentivise consumers to provide higher resolution data by explaining how the data will be used and how it may benefit them.

For example, one supplier's information leaflet states that at a monthly resolution suppliers can provide accurate bills, at daily resolution they can also provide useful energy savings and efficiency advice and improve their forecasting (Citizen's Advice, 2016). At half-hourly resolution they state that in addition to the benefits offered at daily resolution, they can provide greater visibility of energy consumption across the day. Of the 47% of GB homes with smart meters, 81% of these are currently in smart mode (BEIS, 2021d) however a 2019 survey by Citizens Advice shows that many consumers are not providing half-hourly data (49%) or simply do not know (37%) what their data sharing options are (Citizen's Advice, 2019).

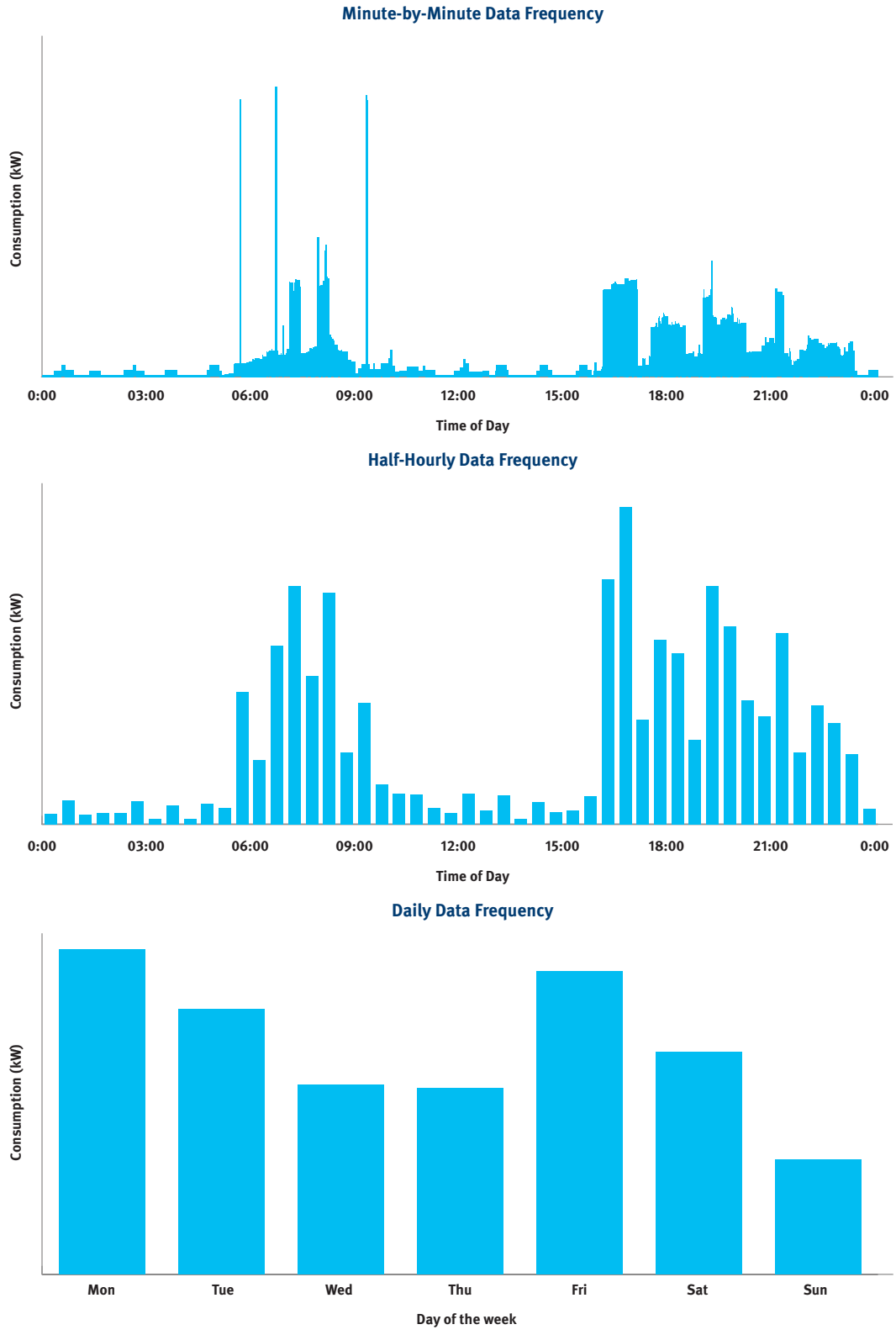


Figure 2 Illustrative Electricity Consumption Data at Different Resolutions[2]

The Data Access and Privacy Framework defines which entities may access smart meter data and how it can be used. Parties and entities are split by those who undertake regulated activities relating to electricity supply and operation and therefore do not require explicit consent to access low resolution, monthly, smart meter data. These are summarised in Table 2. There are currently two exceptions. Firstly, following a consultation by Ofgem on market-wide half-hourly settlement (MHHS), suppliers will now have access to consumers half-hourly consumption data for settlement and forecasting purposes by default with an option to opt-out to daily data sharing (OFGEM, 2021a). This exception has recently been expanded to allow supplier access for business readiness purposes (Ofgem, 2022). Ofgem provides a broad definition of business readiness purposes encompassing forecasting and trading functions as well as the development of new products and services. Additionally, Ofgem states that data used for purposes beyond settlement must be anonymised and aggregated where practicable. Secondly, distribution network operators can access half-hourly data subject to privacy plans approved by Ofgem (Sustainability First and CSE, 2021a). Any other parties such as tariff comparison websites, energy switching services as well as entities outside the electricity sector are required to first register with the DCC and will still require explicit consent from consumers before they can access consumers' data. In addition to the protections laid out in the DAPF, companies are also bound by GDPR regulations as set out in the Data Protection Act 2018, which is currently under review (Department for Digital, 2021), as individual smart meter data is deemed personally identifiable information (OFGEM, 2020). Smart meter data accessed through consumer access devices, such as a suppliers dedicated app would be subject to GDPR but not the DAPF. The DAPF provides consumers with options as to how their data is shared but does not explicitly provide the ability to share anonymous data as is offered in other countries such as the US and Canada (Sustainability First and CSE, 2018).

Table 2 Authorised parties and activities under DAPF (BEIS, 2018)

Authorised Parties	Regulatory Duties
Energy suppliers	Billing
Distribution network operators	Settlement and forecasting (OFGEM, 2021a)
Law enforcement	Investigating suspected theft
Government	Business readiness (Ofgem, 2022)
Authorised third parties registered with DCC	

[2] Synthetic load profiles generated using the CREST Demand Model (McKenna, Thomson and Barton, 2015).

2.3 Beyond the SMIP

The DAPF was developed in 2012 during the early stages of the SMIP and currently remains the main framework for which smart meter data access is governed. Access to smart meter data has been limited mainly to energy suppliers and DNOs. However, there have a been a number of recent developments which could fundamentally change how and by whom smart meter and associated consumption data can be accessed. Smart appliances and demand response are expected to play a significant role in the energy transition. This will generate complementary data streams with additional, more granular data on energy consumption. BEIS has recently set out codes to standardise the operation of such devices and schemes which include specific requirements around data privacy and cybersecurity based on explicit consumer consent, data minimisation, and encryption, however these are not encompassed by the DAPF (BEIS, 2021b, 2021c). As a result, there may be multiple entities (e.g. demand response aggregators, local energy system operators, etc.), other than energy suppliers, who will have access to smart meter data, each with their own data privacy policies. This creates a complex and confusing landscape for consumers to navigate. The concept of a Data Dashboard has been proposed by Citizens Advice, which would provide consumers with a centralised platform to manage access and permissions as well as see how the data is being used (Citizens Advice, 2018). This consent-based dashboard has also been put forward as a key recommendation by the Energy Digitalisation Taskforce (Energy Digitalisation Taskforce, 2022).

In addition, there are calls for widening access to smart meter data for uses beyond the day-to-day operation of the electricity network as part of a move to digitalise the energy sector. A key component of this is to develop Open Data platforms which will provide access to public-interest actors such as government, regulators, local authorities, and other stakeholders to inform and shape policy. How such a platform might incorporate smart meter data is still under discussion but could involve the use of a trusted processor to provide appropriate privacy protections (Sustainability First and CSE, 2021b).

3. Benefits and uses of smart meter data

Smart meters, specifically their data logging and sharing capabilities, are seen as key enablers for a more efficient and cost-effective low carbon electricity network. Access to granular data on consumption and the introduction of time-varying tariff structures will enable a myriad of potential benefits, for customers, energy suppliers and the system overall. As part of the UK's smart meter roll-out BEIS conducts regular cost-benefit analyses. According to the latest edition, completed in 2019, the SMIP is expected to cost £13.5bn with projected benefits of £19.5bn (BEIS, 2019b). This constitutes an average net benefit per household of £250 over the appraisal period (2013 to 2034). Many of these benefits are, however, contingent on high levels of smart meter adoption and data sharing. This chapter summarises some of the key benefits enabled by smart meters with a focus on what level of data sharing is required to achieve them. For details on the innovations enabled by smart metering, as well as the limitations, readers are referred to the Energy Futures Lab briefing paper series on residential demand response, digitalisation of energy and smart electric heating (Carmichael, Gross and Rhodes, 2018; Carmichael *et al.*, 2020; Rhodes, 2020).

3.1 Projected benefits

3.1.1 Automated Meter Readings

A major benefit of smart metering infrastructure is the avoided costs associated with meter readings. Automated meter readings will eliminate the need for customers to take meter readings and avoid estimated bills. Energy suppliers will avoid the costs of sending meter reading operatives to properties in order to obtain a meter reading. Similarly, during the customer switching process, smart meters provide automated meter readings upon change of supplier. It is also expected that accurate, automated billing and streamlined switching process will improve customer experience resulting in fewer inbound customer calls. In addition, debt handling is projected to improve with the possibility for more frequent billing, earlier identification of debt build-up and providing faster follow-up action to help consumers. BEIS estimates the benefits from automated meter readings to be almost £7.4bn consisting of £1.38bn in direct time savings for consumers and £5.12bn in reduced operational costs for suppliers (BEIS, 2019b). Although these benefits are enabled by smart metering, they do not require customers to share their high-resolution consumption data (e.g. half-hourly data) with suppliers or other entities. SMETS meters include tariff, credit and consumption registers allowing customers to see their usage and bills in real-time (BEIS, 2014). As a result, billing and verification can be performed without sharing high-resolution consumption data with suppliers.

3.1.2 Forecasting and Energy Procurement

In an electricity network, demand and supply must be matched at all times to ensure the stability of the electricity grid. Suppliers are responsible for procuring electricity on behalf of their customers. Therefore, they must forecast the expected demand of their customers for each half-hour of the day, known as a settlement period. If there is a mismatch between the amount procured and the amount consumed the imbalance is settled through the Balancing Mechanism. This is a near real-time market where National Grid, the transmission network operator, purchases changes in generation and consumption to correct imbalances.

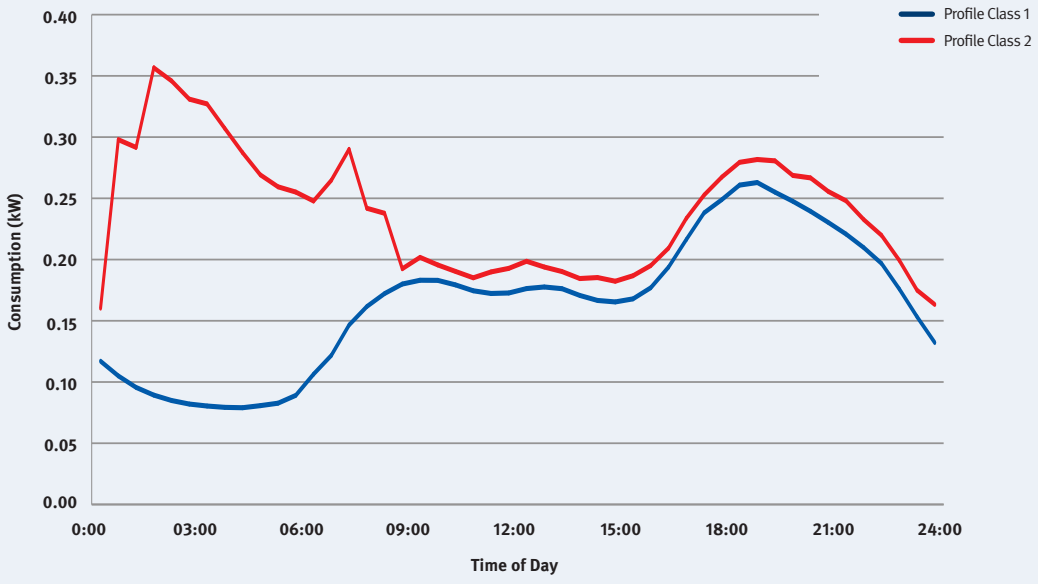
In the absence of smart metering the actual usage of the electricity for each customer in a given settlement period is not known and

customers are assigned to one of eight Profile Classes (for details see Box 1). As a result, costs are not reflective of actual consumption and suppliers are not exposed to the risks half-hourly changes in consumption. Smart metering will enable domestic consumers to be settled on a half-hourly basis, based on actual consumption. Although 47% of homes now have a smart meter installed, very few are being settled on a half-hourly basis (Cornwall Insights, 2019). A move towards mandatory market-wide half-hourly settlement (MHHS) could reduce overall system costs and improve efficiency. It also allows suppliers and domestic consumers to fully harness the benefits of shifting demand from peak to off-peak periods. Ofgem has estimated this to accrue a net benefit to consumers of between £1.56bn to £4.5bn till 2045, with average annual savings per household between £2 and £9 (OFGEM, 2021b).

Market-wide half-hourly settlement will legally oblige suppliers to settle their electricity volumes using actual consumption data instead of the Profile Classes where half-hourly data is available. As smart meters are an opt-in process and consumers will still be allowed to opt-out of half-hourly data sharing when they have a smart meter, such consumers will still

» BOX 1: Electricity Balancing and Settlement Mechanism (Elexon, 2018; OFGEM, 2020)

In an electricity network demand and supply must be matched at all times to ensure the stability of the electricity grid. Suppliers are responsible for procuring electricity on behalf of their customers. They must forecast the expected demand of their customers for each half-hour of the day, known as a settlement period. If there is a mis-match between the amount procured and the amount consumed the imbalance is settled in the through the Balancing Mechanism. This is a near real-time market where National Grid, the transmission network operator, purchases changes in generation and consumption to correct imbalances.



be settled based on estimated load profiles. However, these profiles will be generated on an ongoing basis using the actual half-hourly data available from consumers who do share their data. It is expected that the full transition to MHHS will be completed by October 2025, however Ofgem aims to introduce obligations on the half-hourly settlement on actual consumption sooner than this (OFGEM, 2021a).

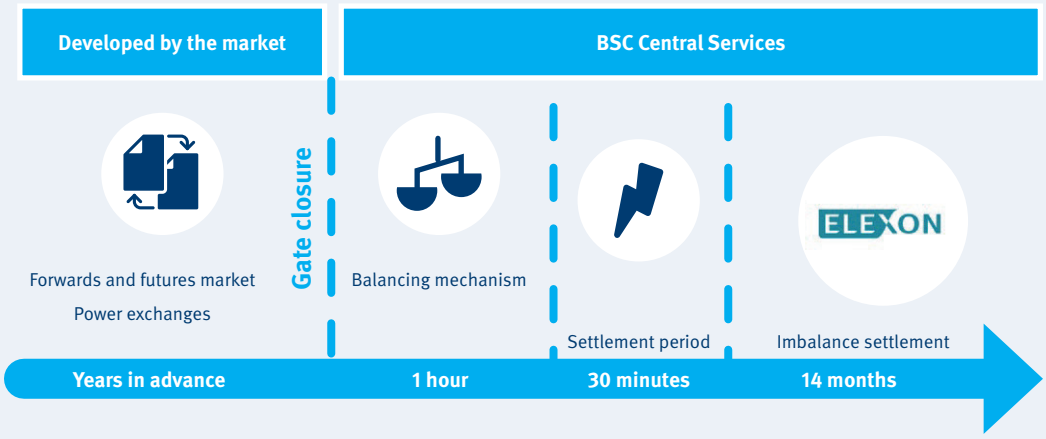
3.1.3 Energy Savings

The smart meter roll-out delivers significant benefits through energy reductions driven by changes in consumers' energy consumption behaviour. Energy usage reductions also reduce the amount of energy that needs to be produced, bringing additional benefits of carbon emissions reductions (estimated at 34.4mln tonnes) and air quality improvements. BEIS has estimated the total benefits from energy and associated carbon reductions due to informational feedback to be £8.273bn (BEIS, 2019b). The average household is expected to have bill savings of £290[3] due to energy reductions and reductions in suppliers' operational costs. Trials have shown that providing real-time information feedback on energy usage can substantially reduce overall

[3] As laid out in (BEIS, 2019b) using 2012 prices based a dual-fuel household which had their smart meter installed in 2020.

In the absence of smart metering the actual usage of the electricity in a given settlement period is not known. Instead, customers are assigned one of 8 Profile Classes. Domestic consumers fall into either Profile Class 1, domestic unrestricted customers with a single rate tariff, or Profile Class 2, domestic Economy 7 customers with a two-rate tariff. The consumption for each settlement period is then estimated based on the average expected consumption for each Profile Class using 'Default Profile Coefficients' generated by monitoring a sample of houses across the country.

Smart metering will enable domestic consumers to be settled on a half-hourly basis if they share their half-hourly consumption data. A move towards market-wide half-hourly settlement could significantly reduce overall system costs and efficiency. It also allows suppliers and domestic consumers to fully harness the benefits of shifting demand from peak to off-peak periods.



energy consumption. A 2019 meta-analysis of trials found that across 130 electricity and gas pilots including around 5.5 million residential customers, feedback trials found an average reduction of 5.4% in electricity consumption and 3.9% gas consumption (Dromaque *et al.*, 2018). There are four main components identified by BEIS that contribute to the realisation of these benefits (BEIS, 2019b):

1. **Direct feedback** – real-time consumption data through In-Home Displays (that are offered to all domestic smart metered households), smartphones, online services, or other platforms. Trials indicate this to be most effective, with average reduction of 7.9% in electricity and 9.6% in gas consumption.

2. **Indirect feedback** – aggregated or non-real-time feedback, e.g. accurate bills and historical or comparative information on bills. Pilots results for this type of informational feedback average reduction of 5% in electricity and 1.8% in gas consumption indicating direct feedback results in higher engagement.

3. **Advice and guidance** – on energy and energy reduction, e.g. advice that installers are required to offer during installations or applications and services that can help interpret data and point towards better choices. Trials providing general tips and advice on ways to reduce energy consumption showed an average reduction of 5.0%. More personalised advice based on disaggregation of consumption exhibited average savings of 7.7%.

4. **Motivational campaigns** – designed to raise energy literacy and motivation to reduce energy consumption. Smart Energy GB, the national communications campaign supporting the roll-out, has an objective to this effect.

None of the components above necessarily require consumption data to be shared with suppliers or third parties as information on consumption can be accessed directly by consumers through their IHD. Although BEIS used conservative estimates of annual reductions of 3% for electricity and 2.2% for gas, recent evidence suggests that energy savings have been lower than expected, and questions remain as to whether these will be sustained in the long term (Carmichael, Gross and Rhodes, 2018; Hledik, Bagci and Chhachhi, 2018). It has been proposed that more personalised and targeted information could help stimulate greater reductions in consumption for which sharing of high-resolution consumption data would be required. This includes detailed bill breakdowns, comparisons with neighbours, personalised advice on energy efficiency tips and appliance level usage information. Such additional functionality would require sharing of high-resolution data with suppliers and/or third parties.

3.1.4 Demand Shifting and Smart Grids

Smart meters are a key enabler of large-scale domestic demand shifting and smart grids. By facilitating intra-day pricing and automated load control they provide the technical infrastructure to unlock potential flexibility in the domestic energy sector. The introduction of more intermittent renewables and the electrification of the heat and transport sectors will present significant challenges for the power system. Demand-side flexibility can reduce system costs by reducing peak demand and consuming renewable energy when it is available, reducing energy procurement costs and carbon emissions. Research has highlighted the high potential value of up to £8bn per year of flexibility (Carbon Trust and Imperial College London, 2016). Within the domestic sector various models have been proposed:

- Time-of-use pricing such as the Octopus Agile tariff, follows day-ahead market prices for each half-hour of the day (Steele, 2019). With prices being linked to half-hourly costs consumers are incentivised to shift consumption to cheaper times of day, lowering costs for consumers with flexibility as well as reducing system costs and emissions by reducing peak consumption and the need for reserve capacity. A 2019 study valued the potential average household savings between £5 per year (assuming current trends) and £90 per year (assuming the electrification of heat and transport and automation)(Hledik *et al.*, 2017).

- Energy-as-a-service models where customers pay a flat monthly fee and sign a performance contract with their energy supplier to, for example, supply of 'warm hours' ensuring a minimum temperature in the home rather than paying for each kWh of fuel consumed (Energy Systems Catapult, 2020). This incentivises energy efficiency rather than consumption while giving energy suppliers control to optimise the heating systems' energy usage.

- Local energy systems or peer-to-peer networks, allow customers to trade electricity amongst each other and manage their consumption while avoiding imports from the rest of the electricity network. Many variations of these schemes have been proposed and are being trialled with centralised and decentralised structures. One of the main components of these schemes is the sharing of high volumes of consumption data in near real-time with peers and operators (Energy Systems Catapult, 2021; Vigurs *et al.*, 2021).

Existing demand response programs and trials have shown mixed results, raising questions around engagement, consumers' responsiveness, and persistence (Carmichael, Gross and Rhodes, 2018). BEIS estimates demand shifting to bring in benefits of £1.363bn based on conservative assumptions on engagement (19% of households). However personalised tariffs, gamification and innovative incentive structures enabled by sharing high resolution consumption data could significantly increase engagement and usage flexibility (Duesterberg and Mirviss, 2021). It is likely that more innovative models such as local energy systems will bring in multiple parties beyond regulated energy suppliers and DNOs which would require wider data sharing with third parties (Maidment *et al.*, 2020; OFGEM, 2021b). Although, on average, the demand shifting schemes discussed above are expected to have a net benefit, individual consumer savings are highly dependent on consumers' flexibility and how and if operational savings are passed on to consumers. For example, a 2016 study on the distributional impacts of different incentive structures found that some can result in increased electricity costs, especially for vulnerable consumers (Hledik *et al.*, 2016).

3.1.5 Network Management

Smart meters offer benefits to the DNOs who manage the infrastructure used for electricity distribution. These benefits come from the increased data that network operators will have available, allowing them to identify faults in the network, restore electricity supply more quickly when outages occur and take better informed investment decisions. Historical smart meter data allows DNOs to identify areas in the existing network which are at risk and might require reinforcement more easily. This will result in investment for network reinforcement being better directed. BEIS has estimated this to be £380mln (BEIS, 2019b). The expected energy savings and demand shifting could reduce overall network losses by reducing the total amount of electricity transported on the distribution network and reduce peak consumption. Access to granular data would allow suppliers to identify patterns of behaviour that may indicate theft allowing them to reduce energy theft more efficiently. This would require access to high-resolution half-hourly data and BEIS estimates this capability to bring benefits of £260mln.

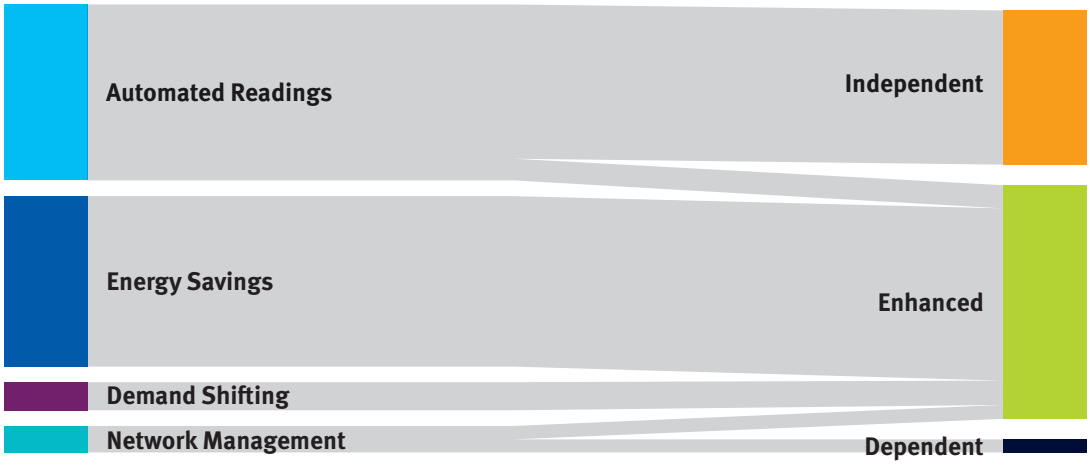


Figure 3 Dependence of Quantified Benefits on Sharing High Resolution data

3.1.6 Dependence on Data Sharing

The projected benefits in the BEIS Cost-Benefit analysis are all enabled by the introduction of smart metering for the domestic sector. However, few are wholly dependent on sharing of high-resolution consumption data. The data dependence can be broken down into benefits which are dependent on, those which may be enhanced with, and those which are independent of access to high resolution data as shown in Figure 3. The direct consumer benefits; the time saving from not taking and submitting meter readings can be achieved by smart meters sending low-resolution data for accurate billing (e.g. monthly which is similar to the data suppliers receive with traditional metering systems) and the expected energy reductions (and associated carbon and air quality benefits) can be realised through informational feedback on the In-Home Display, although personalised advice and recommendations, for which high-resolution data would be required may increase and sustain energy reductions. Similarly, demand shifting can be achieved without consumers having to share their high-resolution consumption data as smart meters are able to store tariff details locally and produce billing information on the In-Home Display. Most of the operational benefits for suppliers are based on the ability to for automated readings and to and avoid site visits but do not require access to high-resolution data. Suppliers' ability to identify theft and DNOs' improvement of network management do require access to half-hourly data to provide meaningful benefits. Although many of the benefits discussed above are not wholly dependent on access to high-resolution, the majority can be enhanced with access which could lead to improved accuracy and ability to personalise recommendations and actions. A summary of the effect of data resolution on the reasonable benefits is shown in Table 3.

3.2 Other Potential Benefits and Uses

Wider access to high resolution smart meter data creates opportunities for expanding energy research and exploring new information and inferences. Currently, most entities with access to smart meter data are directly involved in the operation of the electricity network (suppliers, network operators, switching websites etc.), however, many other entities, such as government, regulators, public interest groups, academic researchers and other stakeholders can benefit from access.

3.2.1 Public-Interest Uses

Smart meter data could play an important part in assessing the impact of policy interventions and reforms in the energy sector. It would, for example:

- Improve BEIS's national energy statistics, which are currently based on annualised consumption estimates, by providing accurate and temporally granular information on consumption patterns (Sustainability First and CSE, 2021b).
- Allow Ofgem to understand distributional impacts of policy choices such as ToU pricing schemes which could adversely affect vulnerable consumers (Hledik *et al.*, 2016; Sustainability First and CSE, 2021b) or lead to discriminatory pricing (Véliz and Grunewald, 2018b).
- Projects such as the Virtual Energy System, which aims to build a real-time replica of the GB energy system led by National Grid, could be enhanced with real-world smart meter data (National Grid ESO, 2021).

3.2.2 Research

The availability of high-resolution smart meter data opens possibilities for new research and insights. For example, a recent publication showed the immediate impacts of the lockdowns instituted in the UK during Covid-19 and their effect of energy consumption patterns, flexibility, and changes to daily routines (Grunewald, 2020). The recently launched Smart Energy Research Lab (SERL) will provide access to over 10,000 customers granular smart meter data and accompanying metadata to authorised researchers with aim to provide new insights (Webborn *et al.*, 2019).

3.2.3 Innovation and Third-Party Access

Smart meter data also has the potential to spur innovation in other sectors beyond energy. BEIS (formerly DECC) is actively encouraging businesses to sign up to the DCC platform (DECC, 2017). Recent work has looked at the potential benefits of using smart meter data to monitor assisted living facilities and dementia patients (Fell *et al.*, 2017; Chalmers *et al.*, 2020; Paxman *et al.*, 2020). Ongoing monitoring can provide relatives, carers and health practitioners with early warnings in the event someone has been incapacitated by a fall. Similarly, energy use patterns can be used to infer living conditions and behaviour changes which may be connected to heath issues. There may also be interest from other sectors and commercial entities such as insurance providers and retailers who could use smart meter data to better understand customers habits and tailor their services.

Table 3 Realisable Benefits at Different Data Resolutions

	≤1 min	Half-hourly	Daily	Monthly
Avoided Meter Readings and Site Visits	✓	✓	✓	✓
HH Load Forecasting	✓	✓		
Energy Savings	✓	✓	✓	✓
Demand Shifting	✓	✓	✓	✓
Network Visibility	✓	✓		
Public Interest	✓	✓	✓	
Research	✓	✓	✓	

4. Consumer Privacy Concerns

The wealth of personal information embedded within smart meter data raise concerns around privacy and data misuse. These have been highlighted within academia, regulators, Citizens Advice, and the privacy impact assessments carried out as part of the SMIP. As a result, several survey and focus groups have been conducted with consumers to understand their awareness, understanding and concerns around smart metering. Although most surveys have found that consumers are generally willing to share their smart meter data, there is significant heterogeneity. A number of key contextual conditions determine this. This chapter summarises the findings of these surveys.

4.1 Data sensitivity – what data?

4.1.1 Type of data

A study conducted in 2020 by Ipsos Mori indicates that 55% of people consider electricity consumption non-sensitive although when younger (respondents under 35) they were more likely to consider it sensitive 49% (SSEN, 2020). Other studies comparing the perceived sensitivity have shown that compared to other types of personal data such as financial details, location data, medical records, social media and contact details smart meter data is considered less sensitive (Knight, 2018; Skatova *et al.*, 2019). However, the majority of consumers are unaware of the personal information that is embedded within smart meter data. A 2019 study by Citizens Advice found that only 30% knew that a smart meter could record the time when a person was in or out of the house (Citizen's Advice, 2019). This was even lower (18%) amongst people from lower socio-economic groups. This clearly reflects inequality in access to information amongst lower socio-economic groups and hence their greater vulnerability to abuse.

When consumers are provided with details of the implications on the type of information being shared when sharing one's smart meter data, they are significantly less willing to share their smart meter data. For example, a 2015 American survey found that when respondents were told that smart meter revealed a lot of personal information demand for smart meters decreased by up to 20% (Horne *et al.*, 2015). Similarly, a multi-stage longitudinal study in Germany investigating different utility subscription models a majority of respondents decided to change (~80%) or cancel (6%) their initial subscription choice once the corresponding privacy implications were described (Jakobi *et al.*, 2019). When looking specifically at the data options within the DAPF, the Citizens Advice study found that only 43% of consumers were comfortable sharing half-hourly or real-time data (see Figure 4), dropping to 28% for those who did not have a smart meter (Citizen's Advice, 2019).

4.1.2 Identifiable or Anonymous

When given the option of anonymisation, a majority of consumers are more willing to share data. An Ofgem commissioned survey from 2018 found that 41% of respondents would be more inclined to share high-resolution smart meter data (Knight, 2018). The 2020 Ipsos Mori study also found that respondents found it important that data is grouped and anonymised so that individual homes cannot be identified (SSEN, 2020).

How comfortable do/would you feel about sharing data from your smart meter with your energy supplier at the following levels?

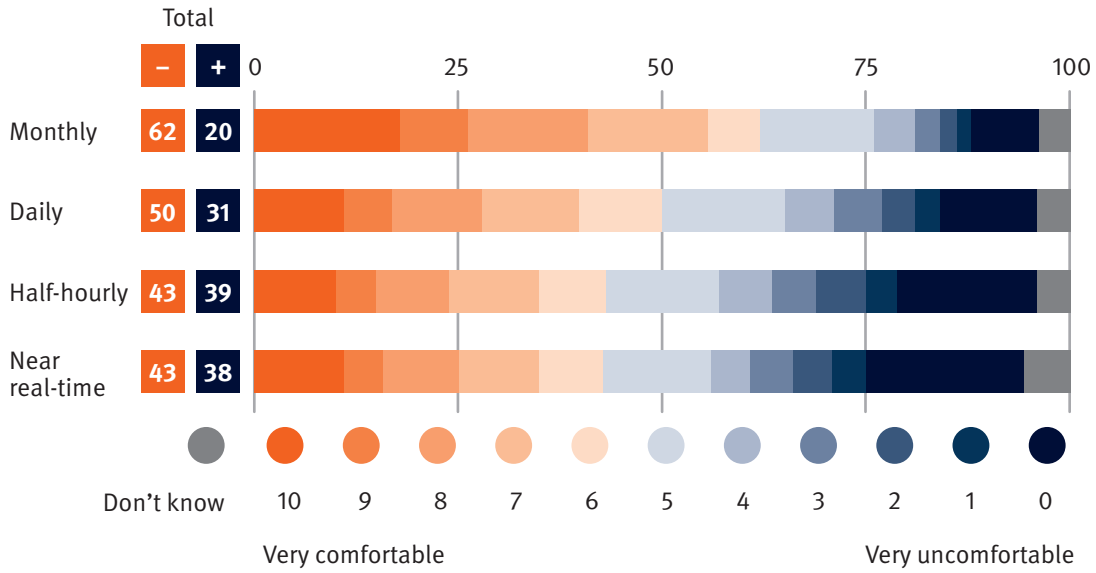


Figure 4 Comfort levels with different data resolutions.(Citizen's Advice, 2019)

4.2 Trust and transparency – who has access and how will it be used?

Willingness to share is also highly dependent on who has access to the data. The Ipsos Mori study showed that 86% said it is important that data is used for improvements and is not sold or shared with other entities (SSEN, 2020). When asked about trust in a variety of organistations to handle their smart meter data, most respondents trusted Ofgem and other central bodies such as the DCC but trust levels for suppliers, their agents and other third parties were significantly less (Knight, 2018). Another important factor is how the data will be used. Most consumers were happy to share their data if it benefits market operations and efficiency. Still, they were most sceptical of consumer facing uses, in particular if the data were to be used for targeted marketing. Trust and control play a significant role in determining consumers' attitudes towards data sharing (Citizen's Advice, 2019; Grünewald and Reisch, 2020; Maidment et al., 2020). Given the wide range of potential uses and misuses of smart meter data, it will be important to provide transparency and clarity on how and by whom smart meter data will be used.

4.3 Willingness to pay/accept for privacy

As smart meter data can provide benefits to consumers, suppliers, and the wider electricity network, how these benefits are distributed is of significance. Overall consumers are either happy to share their consumption data, or willing to share if details on how such data will be used and, importantly, on how it may benefit the system as well as benefit them personally is provided or are reluctant to share data under any circumstances (Dickman and Aslaksen, 2017). A 2015 study found that when offered different electricity service contracts, those requiring smart meter data to be shared would require suppliers to provide a significant discount (Richter and Pollitt, 2018). It is important to note that studies measuring the willingness to pay for privacy tend to exhibit a 'superendowment' effect i.e. the willingness to pay is significantly lower than the willingness to accept (or be paid)(Winegar and Sunstein, 2019). Many consider privacy to be the default and that they should not have to pay for it.

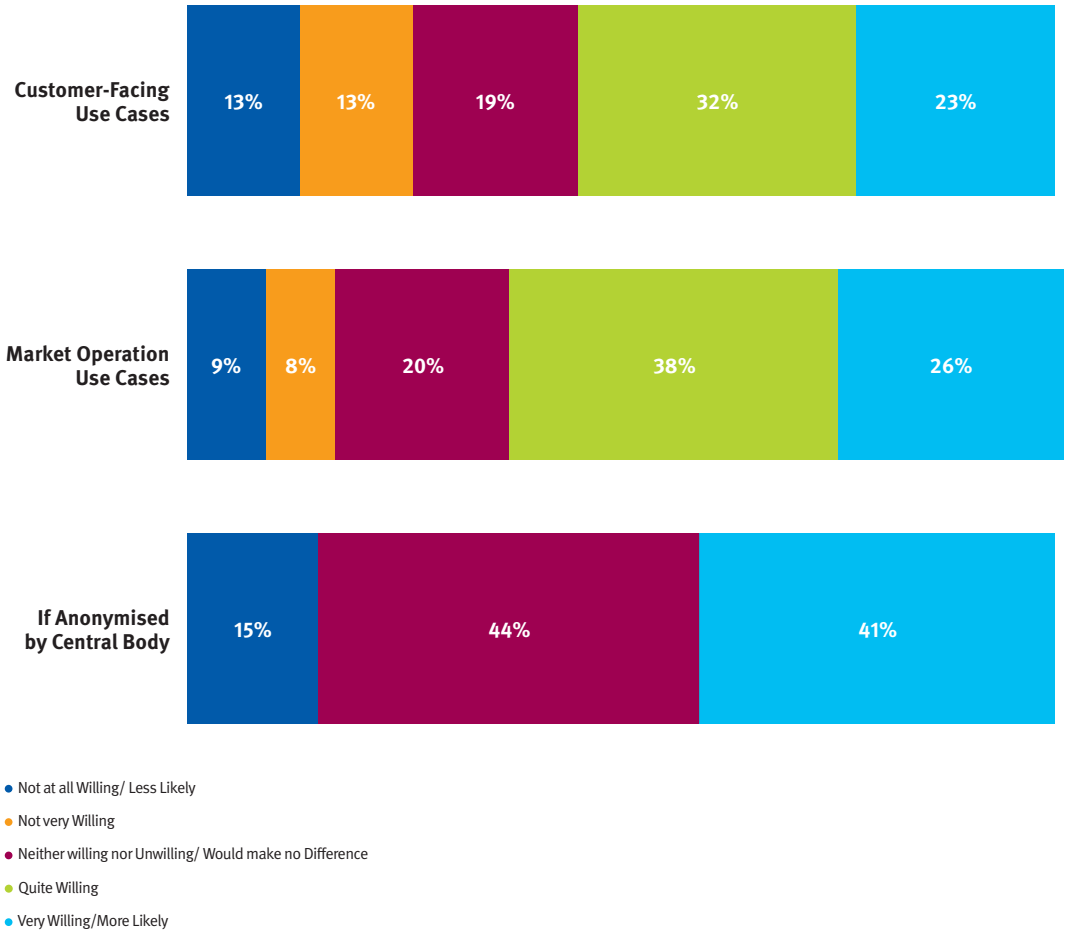


Figure 5 Willingness-to-Share Half-Hourly Smart Meter Data(Knight, 2018)

5. Potential Privacy Infringements and Risk

The types of personal information that can be extracted from smart meter data are highly dependent on the data sharing options selected by customers. SMETS, the minimum technical standards for smart meters in the UK, are capable of recording and displaying consumption data at 0.1Hz (every 10 seconds)(BEIS, 2014). High resolution data, such as half-hourly data, can be used to infer a wide range of information about a household. This chapter presents the potential privacy infringements and risks.

5.1 Load disaggregation

Smart meters record the total electricity consumption of a particular house. This is effectively an aggregated representation of all the different electrical appliances in the home. Specialised techniques known as Non-Intrusive Load Monitoring (NILM) can disaggregate smart meter data to identify and estimate the consumption of different appliances with high accuracy. Appliances have characteristic profiles of energy use which make it possible to classify certain changes in consumption seen at the aggregate level and extract estimates of individual appliance usage. Supervised learning algorithms require a large amount of training data to calibrate. They provide very high accuracy, especially with high resolution data. This includes labelled data for individual appliances and aggregate consumption. Some

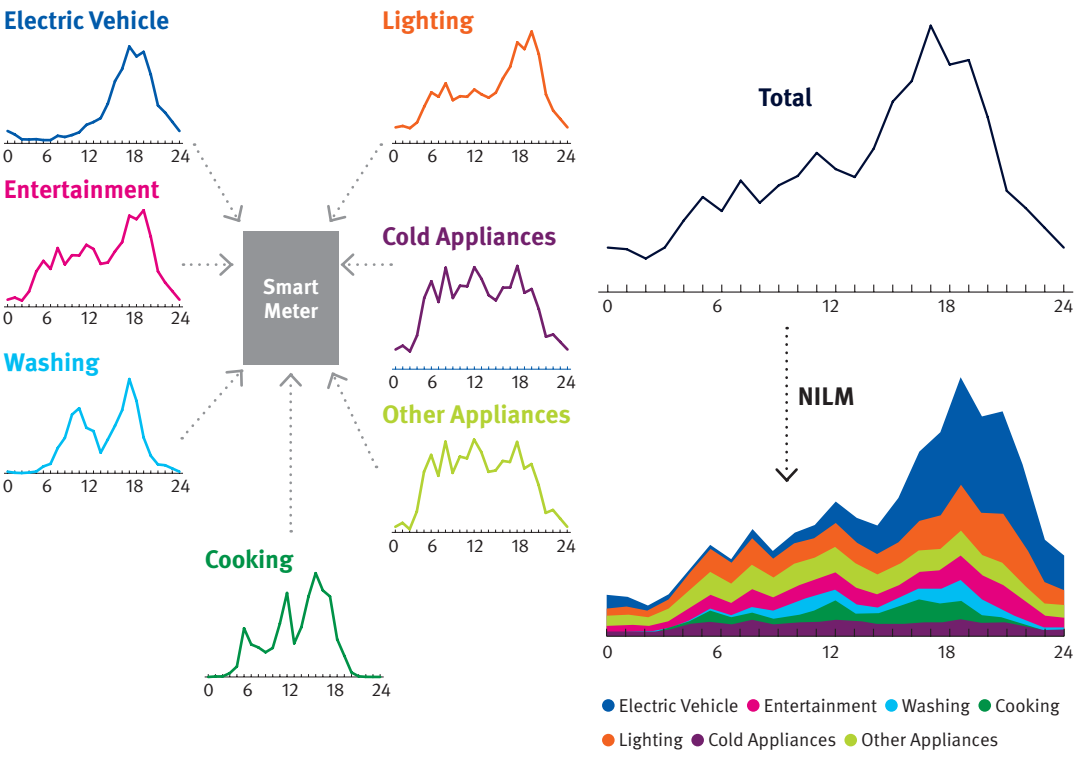


Figure 6 Appliance Signatures and Load Disaggregation. Data from (Munkhammar *et al.*, 2015).

examples of the techniques used are Hidden Markov Models (Dinesh *et al.*, 2017), K-Nearest Neighbour (Cominola *et al.*, 2017), and deep learning (Singhal, Maggu and Majumdar, 2018). Unsupervised learning algorithms can be run without access to such high-resolution labelled data. They perform particularly well for low-resolution datasets. These include adaptive Hidden Markov Models (Kim *et al.*, 2011) and K-means clustering (Abreu, Câmara Pereira and Ferrão, 2012).

Depending on the data resolution it is possible to detect and estimate the power consumption of different appliances with varying degrees of accuracy. Appliances can be broadly categorised in the following:

- Small appliances such as TVs, laptops, lighting, and other consumer electronics devices.
- Cooking appliances such as electric stoves, kettles, and ovens.
- Heating and cooling appliances such as electric space heaters, electric water heaters, refrigerators, and air conditioning units.
- Electric vehicles, batteries, and distributed generation resources such as rooftop solar panels.

The accuracy with which the state-of-the-art NILM algorithms can identify and estimate consumption of these appliances is shown in Figure 7. Heating and cooling typically make up the bulk of a houses' consumption profile and are highly correlated with ambient temperature. As a result, it is possible to infer, for example, whether a house uses an electric or gas heating system even with daily or monthly data. Electric vehicles, distributed generation and electric cooking appliances also have a large impact on a houses' consumption profile. They have characteristic consumption profiles (e.g. solar panels produce electricity during the day reducing the net consumption observed during this period) and can therefore be identified even with hourly data. Smaller electronic appliances and lighting consume less electricity and have more irregular usage patterns making them difficult to identify without granular (sub-minute) data.

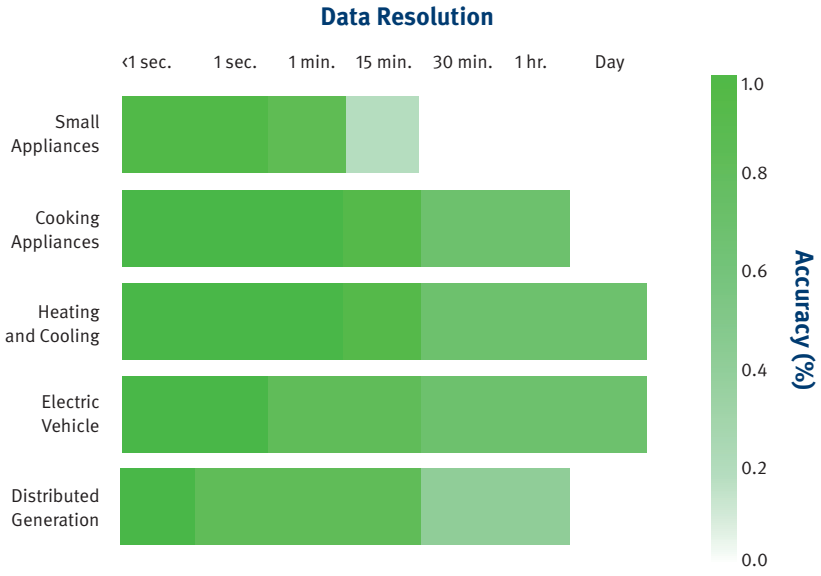


Figure 7 State-of-the-art NILM Accuracy[4]

[4] Maximum reported accuracy for identification and/or disaggregation of appliances summarised from (Parti and Parti, 1980; Abreu, Câmara Pereira and Ferrão, 2012; Zhang *et al.*, 2014; Cetin, Siemann and Sloop, 2016; do Carmo and Christensen, 2016; Cominola *et al.*, 2017; Dinesh *et al.*, 2017; Liu, Luan and Yu, 2017; Zhao, Stankovic and Stankovic, 2018; Hoffmann *et al.*, 2019; Welikala *et al.*, 2019)

5.2 Beyond Energy Use

5.2.1 Demographics and Household Characteristics

Even at low resolutions, smart meter data can also be used to determine socio-economic and demographic information about a household. However, the potential privacy infringements and the consequences of data sharing are not limited to identifying specific household characteristics. Smart meters also store individuals' debt and payment records. In combination, access to such information can lead to unintended or unauthorised practices such as discriminative pricing and unsolicited targeted marketing (Véliz and Grunewald, 2018a). Table 4 summarises the demographic information that can be extracted from smart meter data at different resolutions. As discussed in Chapter 4, most consumers view smart meter data as less sensitive than, for example, financial or medical data. However financial and medical data such as income, employment status, or whether someone is regularly using electrical medical equipment is embedded within smart meter data.

Table 4 Identifiable Demographic Information at Different Resolutions[5]

		< 1 hr.	Daily	Monthly
Socio-Demographics	No. of Residents	■		
	Residents Age	■		
	Marital Status	■		
	Employment Status	■		
	Long-term Illness	■		
	Household Income	■	■	■
	Children and Pets	■	■	■
Dwelling Characteristics	House Type	■	■	■
	No. of Rooms	■	■	
	Size of House	■	■	
	House Location	■	■	
	House Ownership	■	■	■

[5] Considered identifiable if accuracy is greater 50%. Summarised from (Aydinalp, Ismet Ugursal and Fung, 2004; Aydinalp-Koksal and Ugursal, 2008; Beckel, Sadamori and Santini, 2013; Kavousian, Rajagopal and Fischer, 2013; Beckel *et al.*, 2014; Wang *et al.*, 2019)

5.2.2 Activities

Another line of research that is drawing particular attention in the academic community is the linking of smart meter data with time use surveys to understand linkages between activities such as cooking, cleaning etc. and energy usage. A prime example of this is Meter.org, a research project at Oxford University aiming to understand what we use electricity for (Grunewald and Diakonova, 2020a). The researchers record consumption data for a household for a day and ask members of the household to fill out their activities as well as emotions on an app. The project is ongoing but has already built a database of over 10,000 participants. This provides several insights into the drivers of electricity demand and identifying flexibility in energy usage as well as the differences in energy use across gender and household composition (Ramírez-Mendiola, Grünewald and Eyre, 2018; Satre-Meloy, Diakonova and Grunewald, 2018; Grunewald and Diakonova, 2020b). This research has shown strong correlations between smart meter data and peoples' activities and the occupancy of the house. This allows one to develop techniques to infer a much more comprehensive range of information about peoples' day-to-day lives from their smart meter data, beyond which appliances are running (Stankovic *et al*., 2016). For example, high-frequency smart meter data has been used to determine what TV channel an individual is watching (Greveler, Justus and Loehr, 2012). As such, smart meter data has embedded within it significant amounts of personal and sensitive information. As larger high-quality datasets become available the accuracy with which such inferences can be made will increase.

5.3 Linking Datasets

Smart meter data can be linked with other data sources such as social media and other smart devices such as thermostats for further aggregation of multiple data streams. This will enable data analysts to build more detailed profiles of consumers and generate deep insights. Furthermore, the increasing use of Internet-of-Things (IoT) devices, smart energy appliances and the electrification of the heat and transport sector will mean that smart meter data could play a key role in relating these various data sources. Additionally, the 'black-box' nature of machine learning makes it difficult even for data analysts to predict what their machine learning algorithms might infer from these combined data sources. This unpredictability makes it practically impossible to inform consumers about potential future insights and uses of their data (Véliz and Grunewald, 2018a).

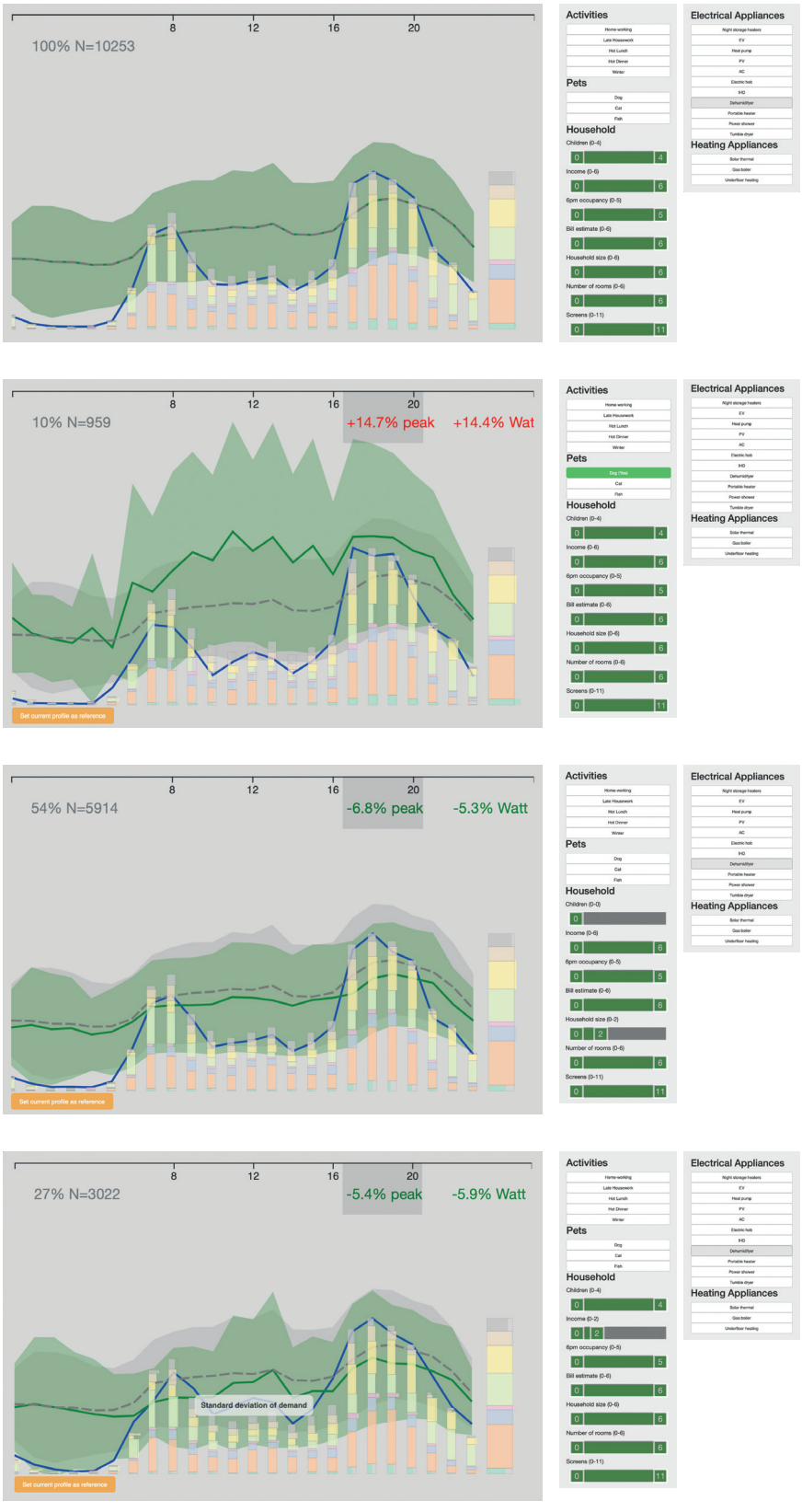


Figure 8 Power Profiler from the METER Project (Grunewald and Diakonova, 2020a)

6. Privacy-Preserving Techniques

The personal information that can already be inferred from smart meter data and the potential for further information to be extracted with advancements in machine learning and availability of big data makes it difficult to convey the potential implications of data sharing. Privacy by Design is a framework and certification developed to overcome such unpredictability (Kingsmill and Cavoukian, 2015). It sets out seven key features of a system to ensure that data breaches and privacy infringements are avoided to the extent possible and that in the event of such infringements the implications are limited (see Box 2). The SMIP does follow many of the principles of the privacy by design framework, although there are differing views on the degree of adherence to these principles (Brown,2014). It has primarily relied on consent and permissions controls.

» BOX 2: Privacy by Design (Kingsmill and Cavoukian, 2015)

Obtaining informed consent in the case of high-resolution smart meter data and specifically for uses outside of the regulated activities outlined in the DAPF is difficult as the extent of potential privacy infringements is not known (Véliz and Grunewald, 2018a; Maidment *et al.*, 2020; Energy Systems Catapult, 2021). The DAPF is supplementary existing data protection legislation (UK Data Protection Act 2018 and GDPR)(OFGEM, 2020). As such, access to, and processing and usage of smart meter data, which is considered personal and identifiable information, is also subject to GDPR. The Information Commissioner's Office (ICO), which regulates data protection in the UK, has raised concerns regarding the current SMIP and specifically around the latest Ofgem consultations on MHHS (ICO, 2016). The ICO also sets out guidelines on how to adhere to existing data protection regulations and are proponents of the Privacy by Design approach (ICO, 2021). The current data access framework does not provide proactive and preventive protection of consumers privacy given the lack of transparency on uses of data, the difficulty in clearly specify the implications of data sharing, and the potential to bypass it through CADs.

Privacy-preserving techniques could overcome these challenges by providing a secure means to share high-resolution smart meter data without compromising on consumer privacy (D'Acquisto *et al.*, 2015). Many of these techniques have been investigated in the context of smart metering and have been considered at various stages of the SMIP by both BEIS and Ofgem but are yet to be implemented (Jawurek, Kerschbaum and Danezis, 2012; OFGEM, 2020). As techniques vary in how and what aspect of privacy they protect, it is important to have consistent metrics which to assess their suitability for smart meter data. These include the privacy guarantees of (Art. 29 WP, 2014):

- Anonymity: Does the technique ensure that no individual can be singled out or identified form the data?
- Link-ability: Does the technique ensure that no individuals' data can be linked across different datasets?
- Inference: Does the technique ensure that information personal information cannot be inferred from the data?
- Data Breaches: Does the technique minimise the potential privacy infringements which may arise in the event of a data breach?

As well as additional desirable properties:

- Individual Level Data: Can the technique still provide access to individual level data, allowing for personalised services to be offered?
- Trusted third-party: Can the technique operate without the need for a trusted third-party such as the DCC?
- Integration: Can the technique be integrated within the SMIP without a significant overhaul of the framework?
- Data Utility: Does the technique preserve the accuracy and utility of the data/ results it outputs?
- Preference Heterogeneity: Can the technique provide different levels of privacy to different users, given the heterogeneity of privacy concerns among consumers?

A wide range of privacy-preserving techniques have been proposed in technical literature and are currently in use in other sectors. These techniques can be categorised into the following:

- Data obfuscation, altering data to provide anonymity or remove information which may be considered sensitive.
- Encryption, limiting access to authorised entities through intra-organisational permissions controls or other enhanced encryption techniques.

- User demand shaping, changing actual consumption to hide certain information which may be considered sensitive.
- Federated learning, performing analysis on data in a distributed manner thereby limiting the sharing of raw data.

The following chapter describes the different privacy-preserving techniques and their suitability for smart meter data applications.

6.1 Data obfuscation

6.1.1 Pseudonymisation

Pseudonymisation involves replacing identifiable features and data with unique identifiers which are consistent across the dataset. Under GDPR regulations, pseudonymised data is still considered personally identifiable information but is removed from real world identity. In order to be considered truly anonymous, data must be stripped of all identifiable features but can lead to data becoming unusable for meaningful analysis. Pseudonymisation provides a middle ground by reducing the possibility of reidentification while maintaining data quality. The ICO code of practice provides guidance on techniques and how to evaluate whether the risk of reidentification is sufficiently remote (ICO, 2012). The UK Anonymisation Network also published the Anonymisation Decision Making Framework which provides practical steps to thinking about the context for each use case to understand the level of risk, manage the risk and effectively inform the anonymisation steps which need to take place (Elliot, Mackey and O'Hara, 2020). For smart meter data this may include, among other things, masking names, addresses and the MPAN (a meter identifier linked to the supply point).

Pseudonymisation has been widely used across various industries and was considered as a privacy enhancing option in Ofgem's latest consultation on data access for energy suppliers (OFGEM, 2020). It could, with modifications, be implemented within the UK's existing data collection framework (Baringa LLP, 2018). It would allow access to individual household consumption data without necessarily revealing any personal information. However smart meter data has a lot of additional personal information embedded within it. These 'quasi-identifiers' can be used to reidentify individuals when linked to other databases such as billing and account information held by suppliers (see Box 3).

Academic literature has shown the risk of re-identification remains high even in country-scale location datasets (Rocher, Hendrickx and de Montjoye, 2019). For smart meter data, it has been shown that given a combination of pseudonymised high-resolution data (e.g. half-hourly consumption data) and low-resolution data (e.g. daily or monthly consumption data extracted from billing), it is possible to de-pseudonymise the data and identify individuals with high accuracy (up to 99%) (Cleemput *et al.*, 2018). Two real-life examples of such linking attacks include:

- US Census Data: A combination of gender, birth dates, and postcodes were used to identify 87% of people in the U.S. (Sweeney, 2000).
- Netflix prize dataset: Even when the dataset on movie ratings of their 500,000 subscribers had names removed and ratings faked, subscriber records and other sensitive information could be identified (Narayanan and Shmatikov, 2006).

6.1.2 Aggregation

Aggregation can be achieved by aggregating consumption data either across multiple periods of time (temporal) or the addition of data pertaining to multiple households (spatial). Aggregation reduces the privacy impact by helping to conceal the pattern of electricity usage inherent in one property's detailed consumption data, either by losing the

detail within a sum totalled over a period of a time, or by masking the detail amongst other properties.

6.1.2.1 Spatial Aggregation

Although no set standards are available on the number of households which must be aggregated to provide privacy protection, a number of schemes have been proposed:

- 15/15 rule - The California Public Utility Commission Decision 14-05-016 defines a limited procedure for anonymizing energy data. It is restricted to monthly average consumption by zip code. If at least 15 meters are aggregated together and no single meter comprises more than 15% of the total energy consumption, then the data are considered "anonymous" (California Public Utilities Commission, 2014).
- DNOs in the UK, as part of their privacy plans, have chosen to aggregate households at the feeder level (Sustainability First and CSE, 2021a). They require, for example, a minimum of 5 households to be aggregated at each low-voltage feeder. If a feeder has less than 5 households it must either be combined with other feeders, or the data cannot be accessed. This was based on an assessment of the coverage a DNO would get across their network (SSEN, 2020).

Spatial aggregation provides a simple mechanism to provide access to high-resolution data as it can be easily integrated into the existing data architecture of the SMIP. For example, one supplier's aggregation framework follows three simple steps; data is requested from the relevant smart meters via the DCC, the individual level data sent back via the DCC is decrypted, validated, and temporarily stored, an aggregation script is run and then the individual level data is deleted (SSEN, 2020). When a sufficient number of households (greater than 20) are aggregated, the diversity of load can significantly reduce the ability of existing load disaggregation algorithms and other inference techniques to accurately extract the personal information embedded within smart meter data (Buescher *et al.*, 2017).

However, there are still a number of vulnerabilities presented by the technique itself as well as the nature of the smart meter roll-out. Given that suppliers and third parties offer their services across different geographical locations aggregation schemes which rely on location, as is being employed by distribution network operators, may result in restrictions on accessing data where there are few customers (OFGEM, 2020). Additionally, aggregated data is vulnerable to reconstruction attacks, where individual level data can be deduced from aggregate information. An assessment of the simple aggregation schemes proposed in the DNO privacy plans showed that it is still possible to determine individuals' information, without complex algorithms, when the level of aggregation is below 10 households (Danezis, 2015). The study did not consider the availability of additional information or complex reconstruction methods thus providing a lower bound on the level of aggregation required. A recent publication has shown that with some additional aggregate information on consumption patterns, such as the average change in consumption between each half-hour, it is possible to reconstruct the entire individual level database with high accuracy (Sheikh *et al.*, 2021).

Under the current framework suppliers and other third parties can incentives customers to provide non-anonymised high-resolution data through the different data options available to customers. If a significant proportion of customers choose to provide such data, then the release of aggregated data would no longer protect those who do not consent to providing such information. As the proportion of those consenting increases, it is possible to deduce the individual consumption data of those who have not consented with greater accuracy. As a result, the privacy protections provided by spatial aggregation depends on the level of aggregation, others' privacy preferences and the type of aggregate data released.

» BOX 3: Linkage Attacks. Examples adapted from (Hoan, 2014).

What is linkage attack?

It is a kind of privacy leakage where the individual information can be locked by obtaining information from another source. In other words, **Sensitive Attribute (SA)** can be matched with **Key Attributed (KA)** through unique **Quasi-Identifier (QID)** information (whether it is anonymised or not).

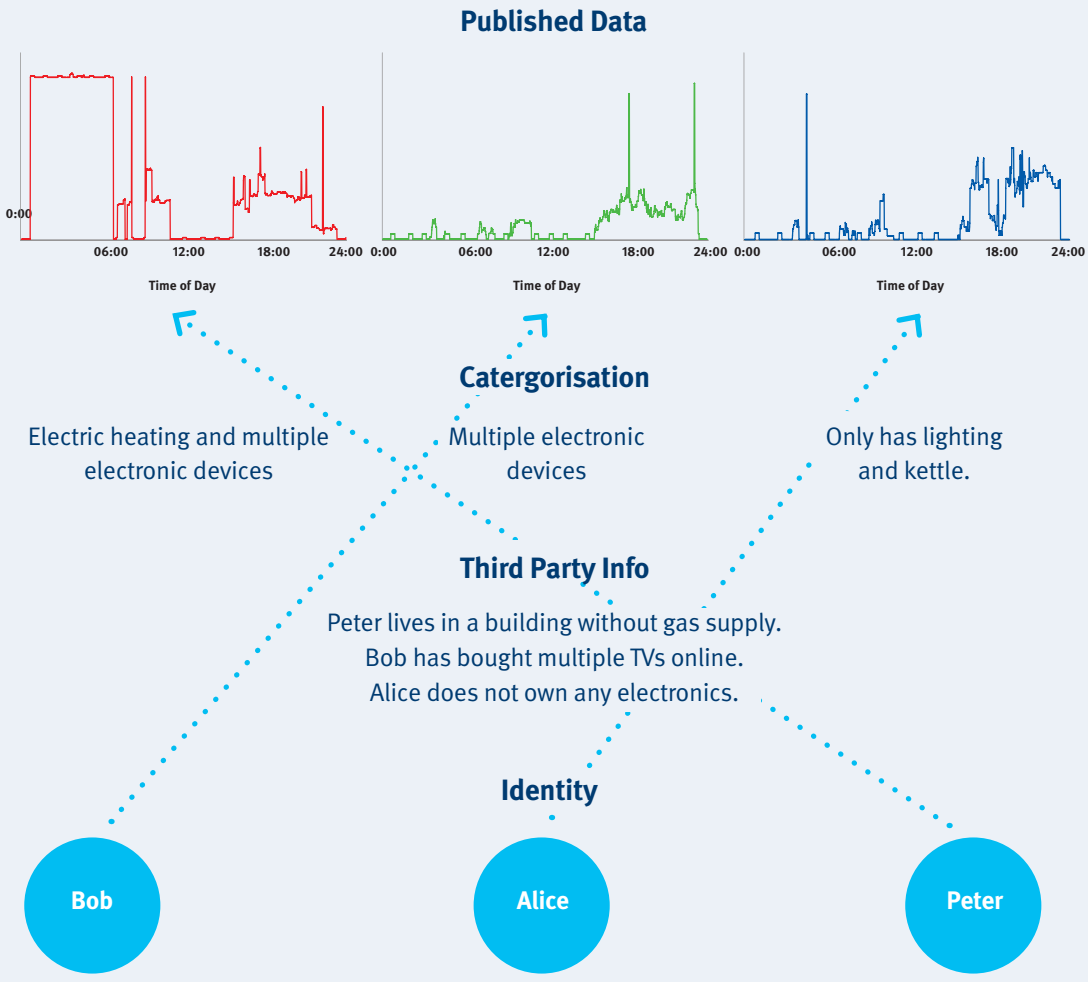
For example, from the figure right, the published data of electricity usage of different users has been anonymised. However, from the third party or other public source, the appliances possibly used can be known, thus the data can be linked with its original identity. Another example of smart meter data can be simply shown in the table below. From the anonymised data, the individual "Carol" can be locked due to its unique initial letters of Postcode.

Original Data			Anoymised Data	
KA	**QID**	**SA**	**QID**	**SA**
Owner Name	*Postcode*	*Data*	*Postcode*	*Data*
Alex	OL2 2XL	5.8	OL2 ***	5.8
Ben	LE3 9EE	4.5	LE3 ***	4.5
Carol	AB41 8UQ	6.7	AB41 ***	6.7
David	OL2 5EE	6.2	OL2 ***	6.2
Emma	LE3 1SR	5.3	LE3 ***	5.3

K-anonymity

A method of protection against linkage attacks. For each record in each version of each piece of data, specify that at least K-1 of the other records have the same semi-identity attribute.

For the example in the upper figure, if “Alice has electric heating”, only Bob’s data can be inferred, and the privacy of Peter and Alice is preserved. Similarly, for data in the left table, except Carol, sensitive information is all privacy preserving because the anonymised Postcode is not unique, and they are all **2-anonymised**.



6.1.2.2 Temporal Aggregation

The resolution or frequency of smart meter data significantly impacts what information may be inferable from smart meter data. Additionally, many of the quantified benefits of smart metering laid out in BEIS's cost-benefit analysis do not require sharing of high-resolution data. As such restricting data access to daily or monthly resolutions can protect from load disaggregation algorithms and limit the information that can be extracted from smart meter data. However, restricting access to higher resolution data precludes many of the envisioned benefits of smart grids, local energy systems and time-of-use tariffs.

6.1.3 Differential Privacy

Differential privacy (DP) is a technique, rooted in cryptography, which ensures the anonymity of individuals within a dataset (e.g. the half-hourly smart meter data of a suppliers' consumers) while allowing aggregated information and statistics about the dataset to be shared (e.g. the average consumption for each half-hour). It works by ensuring that the aggregated information released about the dataset is not altered if an individual is included or excluded from the dataset. It overcomes a number of vulnerabilities posed by using pseudonymisation or aggregation alone as it protects against the worst case; it places no limits on the background knowledge someone analysing the data might have (Dwork and Roth, 2013).

The noise injection needed to achieve this, however, introduces a trade-off between the privacy being offered to individuals and the accuracy of the data being shared (Eibl *et al.*, 2018; Chhachhi and Teng, 2021). The amount of noise to be introduced is inversely proportional to the number of consumers to be aggregated. Simulations suggest that high accuracy can be maintained when the number of customers are of the order of thousands (Eibl and Engel, 2017). However, with application specific knowledge and practical assumptions on the availability of background knowledge, an analyst can significantly reduce the amount of noise that needs to be introduced (Desfontaines and Pejó, 2020). Nevertheless, determining the appropriate level of noise injection to provide privacy remains a topic of research (Hsu *et al.*, 2014).

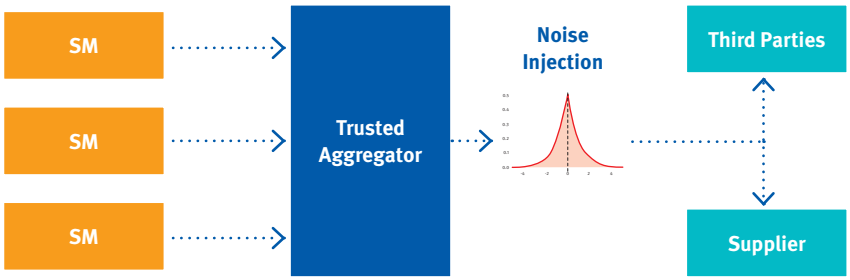
DP has been gaining traction across multiple industries with real-world implementations including:

- Tracking emoji and text usage by Apple (Apple Inc., 2017).
- The Community Mobility Reports published by Google providing insights into the effect of measures to combat Covid-19 (Google, 2021).
- The 2020 US census is being published through a differentially private mechanism (US Census Bureau, 2019a).
- Electricity analytics company Recurve has implemented an open-source DP scheme adapted to smart metering (Recurve, 2021).

DP also has other desirable properties:

- Differentially private algorithms have been developed for complex tasks, clustering, general machine learning.
- As opposed to other data obfuscation and encryption techniques, the level of privacy offered can be controlled through a privacy budget.
- Additionally, variations of DP called heterogenous or personalised DP allows for each individual within the database to be provided a different level of privacy based on their preferences (Jorgensen, Yu and Cormode, 2015).
- It can be implemented with a trusted aggregator (e.g. DCC) performing the aggregation and noise addition, known as global DP, or with an untrusted aggregator where noise is added at the smart meter, known as local DP (Ács and Castelluccia, 2011).

Global Differential Privacy



Local Differential Privacy

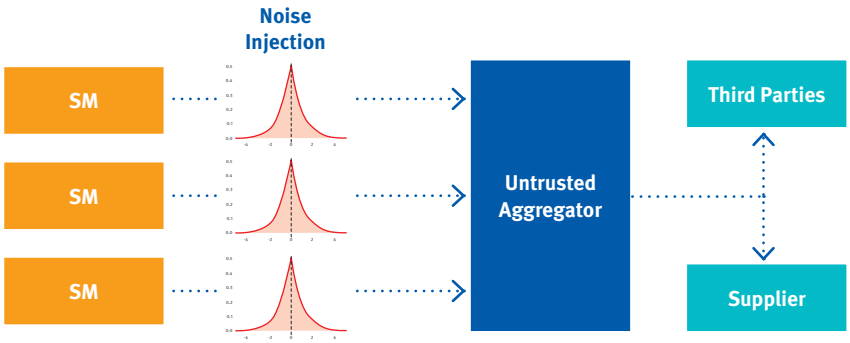


Figure 9 Dataflow Diagram of Differentially Private Mechanisms

6.2 Homomorphic Encryption

The SMETS 2 specifications provide details of the encryption capabilities required by smart meters in the UK (BEIS, 2014). This ensures that only authorised parties have access to consumption data. This is handled centrally through the DCC and the Smart Energy Code (Smart Energy Code Company, 2021). Although encryption provides a level of security and limits access to data it does not in itself preserve privacy. A promising privacy preserving technique is homomorphic encryption (Li, Luo and Liu, 2010). It allows data analysts to perform arithmetic operations (e.g. addition, subtraction, multiplication, and division) on encrypted data without having to first decrypt it ensuring that underlying data cannot be accessed while maintaining the accuracy of results (Asghar *et al.*, 2017).

Furthermore, it requires neither secure communication channels nor a trusted third party (Xue *et al.*, 2020). Such techniques can be categorised into the following (Giaconi, Gunduz and Poor, 2018a):

- Partial such as Paillier or ElGamal which allow only certain operations such as addition and multiplication to be performed (Paillier, 1999).
- Full which allow all operations to be performed. However, existing chemes result in high computational complexity, rendering applications impractical (Gentry, 2009).

Cryptographic techniques such as homomorphic encryption typically suffer from high computational complexity, key distribution issues, overhead, and poor

Homomorphic Encryption

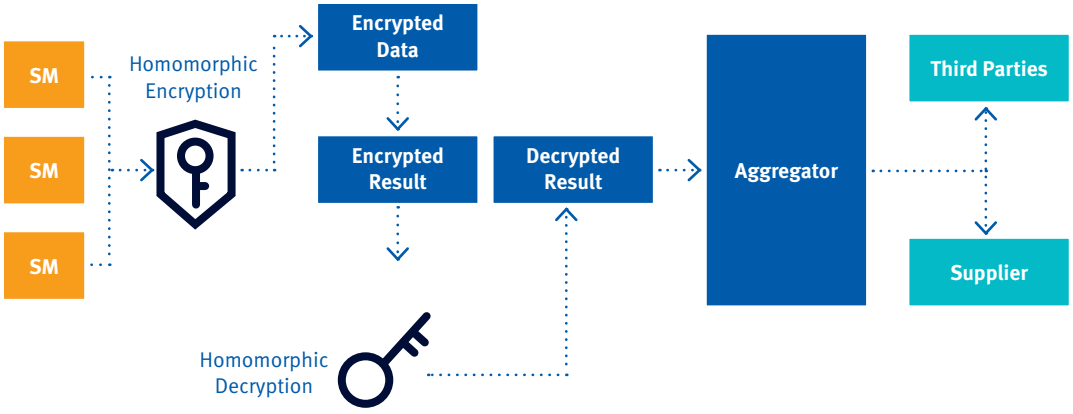


Figure 10 Dataflow Diagram of Homomorphic Encryption

scalability, preventing practical applicability in an SM setting where computational and bandwidth resources are limited. Additionally, cryptographic techniques are vulnerable to statistical attacks and power analysis (Giaconi, Gunduz and Poor, 2018a).

6.3 User Demand Shaping

In contrast to the above techniques, user demand shaping functions behind the meter by altering actual consumption patterns. Appliances and activities have characteristic consumption profiles. Changing the consumption data seen by the smart meter can hide specific appliances or more generally limit the amount of information inferred from the consumption data recorded at the smart meter. This is achieved through smart control of flexible assets such as batteries (Giaconi, Gunduz and Poor, 2018a). There are a number of advantages to such an approach:

- Does not require a trusted third party to verify or perform the privacy preserving actions.
- Protects against physical hacking and unauthorised monitoring.
- Preserving accuracy of data collected and transmitted by smart meter.

User demand shaping as an approach provides customers with direct control over their privacy but by altering the actual consumption of a household the value of the data to make operational insights is greatly diminished. Overall, the main disadvantages of this approach are:

- Requires a battery or generation onsite,
- Reduces potential financial benefits that can be leveraged from battery flexibility (Giaconi, Gunduz and Poor, 2018b),
- Does not provide anonymity,
- By altering the load seen at the grid the accuracy with which inferences about flexibility and usage can no longer be made (Kalogridis *et al.*, 2010).

6.4 Federated Learning

Existing data architectures require data to be collected from smart meters and then processed and analysed centrally. For example, when determining how much energy to buy from the wholesale market on the consumers' behalf, a supplier may want to use historical consumption data from its customers to forecast future consumption to aid their decision-making. They would collect the data,

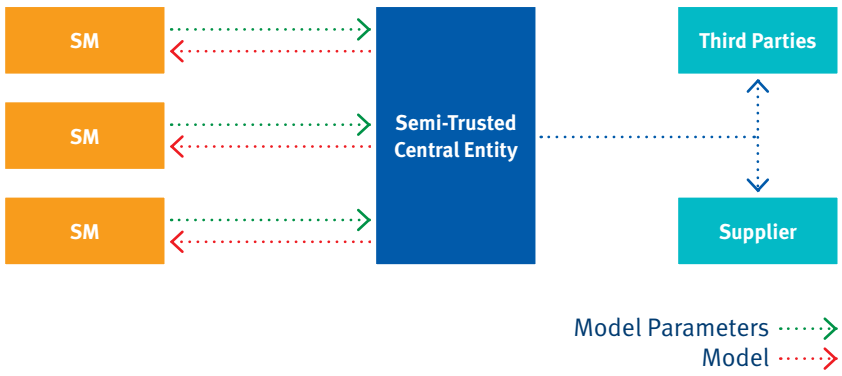
train a centralised model, and then use the trained model to forecast future consumption. While centralised techniques may have worked well in the past, this is becoming computationally challenging with the trend toward Big Data in recent years.

Federated learning (FL) and privacy-preserving forms of peer-to-peer (P2P) distributed learning (sometimes called 'no-peek' (Vepakomma *et al.*, 2018)) are distributed computing techniques that allows such analysis to be performed at many computing nodes without having to first centralise the data.

Not only does this reduce the computational burden at one location, it also reduces the communication requirements as large datasets no longer need to be sent to the cloud, and reduces the potential attack surface as data is no longer kept in one place. Instead, all raw data is kept locally; models are trained locally; and only model updates are shared between participants/with a central entity (Figure 11).

However, even though raw data is not shared directly, FL and P2P distributed learning models can indirectly leak information during their model update stage. This may allow the local

Federated Learning



Peer-to-Peer Learning

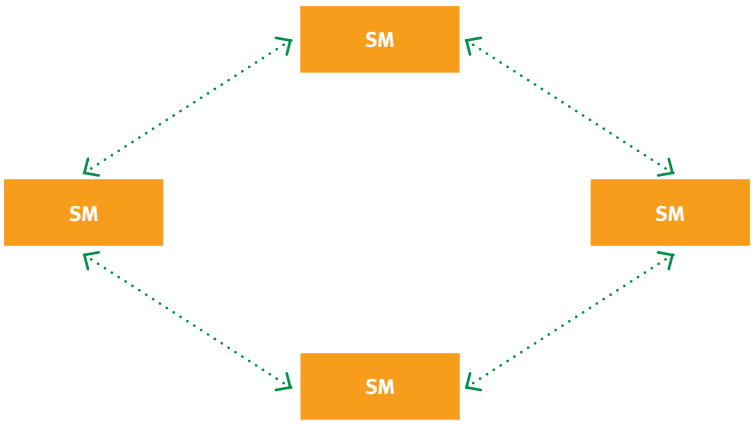


Figure 11 Dataflow for Distributed Learning Techniques

raw datasets to be reconstructed by malicious parties (Yang *et al.*, 2019; Bagdasaryan *et al.*, 2020). Thus, unlike differential privacy (Section 6.1.3), FL and P2P distributed learning do not provide a mathematical guarantee of privacy. In order to reduce the information revealed in model updates, FL and P2P distributed learning are often combined with some of the other privacy-preserving mechanisms discussed in this chapter, for example: homomorphic encryption (Zhang *et al.*, 2020) and/or differential privacy. Additionally, while there may be some accuracy loss when privacy-preserving distributed learning techniques are applied instead of a centralised approach, most studies have found this to be negligible (Yang *et al.*, 2019).

FL is being introduced into a number of fields where privacy and intellectual property rights are of particular concern. Notable examples include:

Keyboard query suggestions for Google devices (Yang *et al.*, 2018);

A platform for multi-institutional collaboration to develop medical imaging diagnostic tools for brain tumours (Sheller *et al.*, 2019);

MELLODDY, an EU-funded project to develop a platform for collaboration between pharmaceutical companies for drug discovery (MELLODDY, 2019).

Furthermore, FL is being actively researched and applied to smart meter data for forecasting and clustering applications (Taik and Cherkaoui, 2020; Jia *et al.*, 2021; Wang et al., 2021). When applied to smart grids, FL is often applied in conjunction with edge computing: a distributed computing paradigm where computation is performed at the 'edges' of the network closer to where the data is collected. In the context of smart metering, such devices include: smart meters, IHDs, and CADs which could connect to the HAN within the home.

6.5 Suitability For Smart Metering

The properties of each technique are summarised in Table 5. Traditional techniques such as pseudonymisation and aggregation do not provide provable guarantees of privacy, although they can be integrated and implemented with minimal change to the existing SMIP architecture. Homomorphic encryption overcomes some of the shortcomings of traditional methods but is computationally intensive and limited in terms of how data can be processed. Federated learning and user demand shaping offer a decentralised approach to privacy preservation. However, neither provide anonymity but instead alter the data being sent. The current smart metering infrastructure does not have the computational resources required for these techniques. The need for interoperability may also hinder suppliers' ability to develop novel machine learning algorithms if such techniques were to be integrated into the SMIP. Differential privacy is considered the gold standard for privacy, and it provides provable guarantees of privacy and anonymity while also allowing the privacy-utility trade-off to be explored introducing flexibility. Although it only allows for access to aggregate data, many of the applications for smart meter data can be performed using aggregated data with marginal improvements provided by access to individual level data. It has already been widely implemented in practice by both private entities such as Apple and Google and public entities such as the US Census Bureau. The next section will explore how the US Census Bureau has implemented differential privacy for the 2020 US census, the associated challenges and lesson for implementing a similar approach for the UK's smart metering programme.

Table 5 Properties of Privacy-Preserving Mechanisms[6]

		Pseudo-nymisation	Aggregation	Homo-morphic Encryption	User Demand Shaping	Differential Privacy	Federated Learning
Privacy Guarantees	Anonymity	✸	✸	✸		✓	
	Invulnerable to Linking					✓	
	Invulnerable to Inference				✓	✓	
	Minimise Impact of Data Breaches		✸	✓	✓	✓	✓
Desirable Properties	Individual Level Data	✓			✓	✸	✸
	No Trusted Third-Party			✓	✓	✓	✓
	Easily Integrated	✓	✓			✓	
	Preserve Data Utility	✓	✸	✸	✸		✸
	Preference Heterogeneity	✓			✓	✓	✓

[6] Orange stars indicate properties that the privacy-preserving technique is purported to have and which, in some cases, may have for practical purposes. However, these properties are not evidenced by theoretical guarantees.

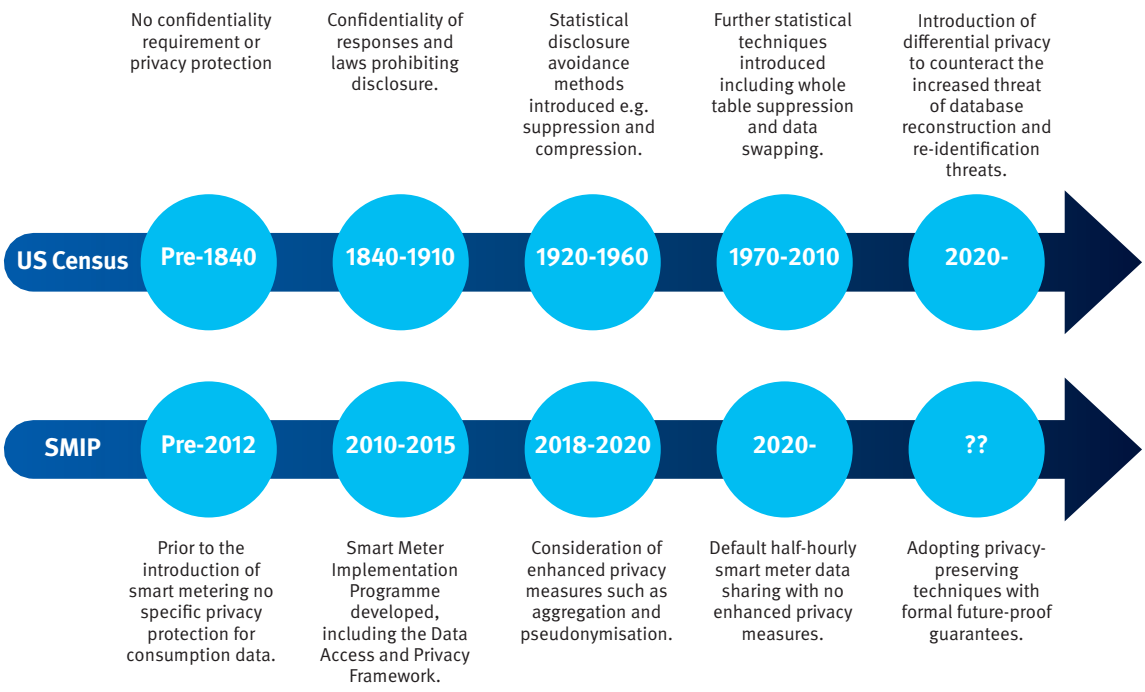


Figure 12 Evolution of Privacy Protection for US Census and SMIP

6.6 Us Census – A Case Study Of Differential Privacy

6.6.1 Background

The US Census collects demographic information on all persons in the US on a decennial basis. This data is used to support a number of essential functions such as budget allocations and apportioning seats in the House of Representatives. Hence, the accuracy of such data is paramount. However, the Census Bureau is also legally obligated to ensure respondents' privacy and data confidentiality, prohibiting the release of any personally identifiable information (Hawes, 2020b).

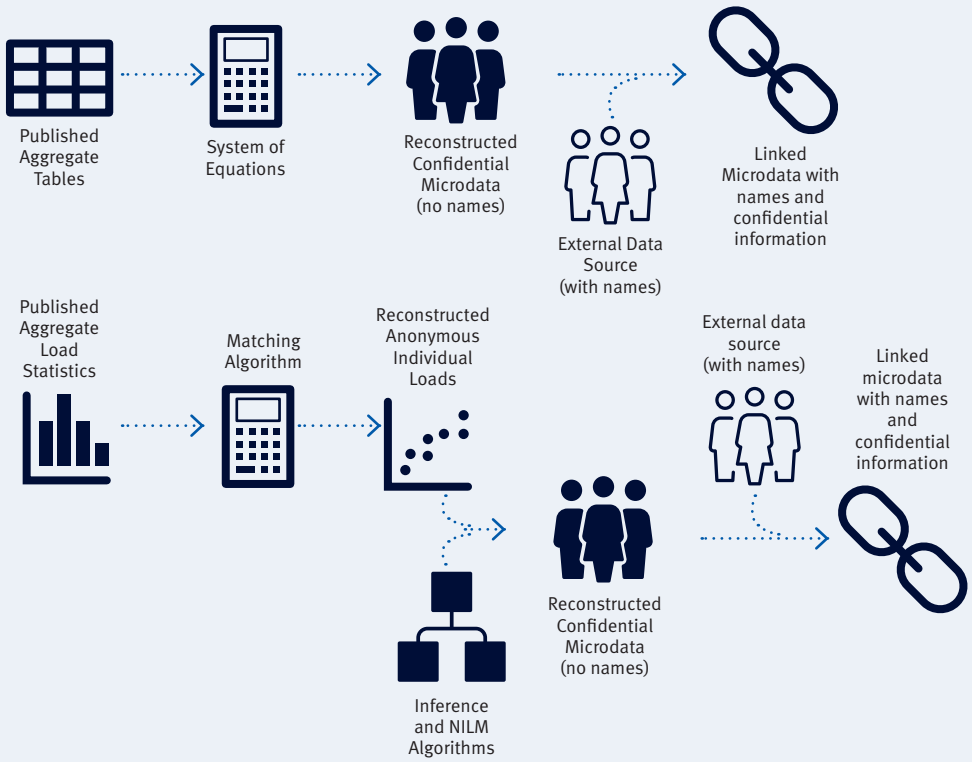
The Census Bureau has constantly been updating its privacy and data protection protocols. Starting with confidentiality for businesses and laws banning census takers from disclosing information in the late 1800s, through the 1900s, a number of aggregation and suppression techniques were introduced to account for the risks of indirect disclosure. More recently, new techniques such as data swapping and top-coding were introduced to reduce the number of tables that would need to be suppressed (US Census Bureau, 2019a). However, advances in computing power and the availability of third-party data sources have meant that existing techniques are no longer able to protect against indirect disclosure (US Census Bureau, 2021a).

In 2016, the Census Bureau conducted an internal experiment to quantify the potential risk of disclosure of personal information under the existing publishing protocols. They used published aggregate data tables from the 2010 census to (Hawes, 2020a):

- Reconstruct the underlying individual records (block ID, sex, age, race, ethnicity) of all 308 million individuals,
- Re-identify the individuals by matching the reconstructed database with commercially available data which include names and addresses.

The results of this exercise were striking. The individual level database could be reconstructed completely with 100% accuracy for 46% of the population and 71% of the population when allowing for a +/- 1 year age range. In addition, the individual variables (block ID, sex, age, race, ethnicity) were unique for more than 50% of the population. Finally, 38% of people's records could be exactly linked to the third-party data sources and their names and addresses identified. Given that aggregate data tables are published publicly, and no authorisation is required to access them, this was deemed a real and immediate threat.

» BOX 4: Database Reconstruction and Re-identification Attacks



6.6.2 Implementation

This experiment showed that traditional statistical methods could not ensure that the Census Bureau was meeting its legal and ethical obligations. As a result, in 2018, they adopted DP for the 2020 Census to protect confidentiality. The Office for National Statistics in the UK is also considering the use of DP for future UK Censuses (Government Statistical Service, 2018). To implement DP the Census Bureau developed a bespoke algorithm, the TopDown algorithm, which accounts for the specific nature of census data and optimises accuracy while ensuring privacy (Garfinkel, Abowd and Powazek, 2018). There were three main challenges:

1. **Determining the privacy/accuracy trade-off:** As DP introduces noise it necessarily reduces accuracy. However, determining the appropriate trade-off and hence privacy budget (ε) remains a question for policymakers. The Census Bureau developed a framework to assess the cost and benefits based on the willingness-to-accept (WTA) privacy loss and the cost of increasing accuracy in terms of foregone privacy (Abowd, 2018). For example, the misallocation of funds due to errors in the population counts can be used to quantify the monetary implications for data accuracy. Similarly, a reference for the WTA for privacy loss can be drawn from existing law on losses incurred due to identity theft or stated preference surveys. The Census Bureau has been actively working with other government departments and demographers who use census data to determine of the optimal privacy budget in an ongoing and iterative process (US Census Bureau, 2021b).

2. **Allocating the privacy budget:** Census data is collected and summarised at four levels of hierarchy – state, county, tract, and block. Each level has fewer and fewer people within it, from millions at the state level to thousands or less at the block level. The amount of noise that DP must add to ensure privacy is inversely proportional to the number of people. As a result, the algorithm was designed to provide block-level aggregates with more of the privacy budget.

3. **Consistency and invariants:** When adding noise to data it is possible to end up with some cases producing nonsensical results. For example, one may end up with -10.5 White females in a block when tabulating counts by ethnicity and sex. In addition, if computed separately, the sum of the population of each county in a state would not match the total population of the state. To overcome this, significant amounts of post-processing was incorporated into the TopDown algorithm. This resulted in additional accuracy loss, which was more than the error introduced by the noise introduced to achieve differential privacy.

4. **Product catalogue and computation:** The census produces a large but finite set of tabulations (e.g. population counts by age and sex or household size counts at each level). Given the requirements described above it is necessary to compute the tables and relevant noise addition all at once for all states. As a result, the algorithm is computationally intensive and requires dedicated computing infrastructure to perform.

Aside from the technical challenges and specific features of the census data, the Census Bureau has also faced issues in convincing and communicating the new approach with data users. There has been significant debate amongst social scientists (Mervis, 2019) and a court case (Brennan Center for Justice, 2021) in Alabama. The accuracy of data still varies across geography (e.g. urban vs. rural) as well as ethnicities (Petti and Flaxman, 2020).

6.6.3 Lessons for Smart Metering

Integrating DP for smart meter data in the UK would require addressing many of the same challenges faced by the US Census Bureau as well as a number of specific issues given the nature of smart meter data and the SMIP structure.

1. **Disclosure Risks and Threats:** Aggregate census data is made publicly available and does not require authorisation for data access. Conversely, access to smart meter data is governed by strict permissions controls allowing only authorised parties registered with the DCC which reduces risks as these are currently heavily regulated entities (e.g. energy suppliers and network operators). The transition to a more dynamic domestic electricity sector will result in more, potentially unregulated, entities such as aggregators and switching websites, and their agents having access to smart meter data. Similarly, the current emphasis on widening access to data for public interest purposes and creating Open Data platforms, could result in comparable risks as publicly available datasets such as the census. Additionally, census data directly includes socio-demographic information which can be used for reidentification whereas smart meter data contains only consumption information. However, significantly more socio-demographic information, can be extracted from smart meter data with relatively high accuracy using standard statistical inference techniques as shown in Chapter 5. As a result, access to individual high resolution smart meter data can be linked to other publicly or commercially available datasets in a similar way to the census data (see Box 4).
2. **Who should implement it?** Several entities have been proposed to act as trusted processors of smart meter data. These include energy suppliers, network operators, Elexon (settlement body) the DCC and the ONS (Sustainability First and CSE, 2019). A trusted aggregator which collects data from individual smart meters, applies differential privacy and then sends this data to authorised parties could follow a similar top-down framework as the Census Bureau. However, the SMIP does not currently provide access to unencrypted smart meter data to these potential trusted processors, so would require a fundamental redesign. Alternatively, it is expected that Elexon will have access to individual smart meter data as part of the market-wide half-hourly settlement, which would allow them to apply appropriate privacy protections. However future energy markets may potentially be highly decentralised with peer-to-peer transactions. In such a scenario a local DP model would be more appropriate with suppliers and network operators implementing their versions of DP. This is similar to the current framework for network operators who submit their privacy plans for approval from Ofgem. However, it is important to note that this would require a technical oversight mechanism to ensure that the proposed DP implementations actually provide meaningful privacy protection. For example, Apple has been criticised for marketing their products as differentially private but the large privacy budget potentially provides little actual protection (Tang *et al.*, 2019). One solution to this would be to require companies to publish their code as the Census Bureau has done (US Census Bureau, 2019b).

3. **Developing a product catalogue:** Unlike the US census data, there is currently no pre-defined catalogue of data products that a user can access. However, there are a set of operational use cases for smart meter data which could provide a basis for developing such a catalogue. For example:

    - Network operators need data at different levels of spatial granularity; the total national demand, demand at the GSP level and the feeder level,
    - Suppliers need data on their total customer base and demand for customers on different types of tariffs,
    - BEIS needs usage data split by different sectors at different temporal resolutions (monthly and yearly) for publications and analyses which rely on DUKES data.

4. **Dynamic nature of smart meter data:** The US census is a static database, data is collected once and the obligation to protect privacy is time-limited to 72 years (after which the individual level records are made public) (US Census Bureau, 2020). It is possible to determine a privacy budget and allocate that budget over time. In contrast smart meter data is constantly generated and the data are correlated in time. As a result, it would be necessary to determine a time limit for which the privacy protection must apply. This could be based on existing data retention storage limitation policies companies have to follow under GDPR (Article 5 1.e). The notion of discounted differential privacy has been proposed to distribute the privacy budget by considering the differences in the sensitivity of historical (e.g. one year old) and real-time data (Farokhi, 2020).

5. **Determining the appropriate trade-offs:** Determining the trade-off between privacy and data utility remains an open question and is highly application and domain specific. In most existing implementations of DP, the trade-off has been chosen based on an assessment by the implementing party. Essentially it has, to date, remained a policy choice. The Census Bureau has focused on the acceptable level of accuracy loss for its' data users (government departments and academia) but not investigated individuals' attitudes towards protecting their privacy. Given the two-way communication offered by smart metering and the dynamic nature of the electricity market, a market-based mechanism to determine the revealed preferences of consumers, who generate the data, and data users (supplier, network operators etc.) could be introduced (Liu *et al*., 2021).

6. Privacy-Preserving Techniques

7. Discussion and Recommendations

The SMIP has considered data privacy and security within its development and implementation. The DAPF, which provides consumers with control as to how and with whom they share their smart meter data, is essential to promote widespread adoption and forces suppliers and other entities to incentivise consumers to share their data. SMETS, which provides technical standards for the smart meters and the centralised data collection and processing framework, ensures security standards for consumers data. These components of the SMIP build a foundation upon which smart meter data can be shared to realise the benefits of smart metering while also providing consumer control over their data and privacy. However, several barriers remain to widespread sharing of high-resolution smart meter data and a reliance on permissions controls make it difficult to safeguard consumer data against misuse, intentional or otherwise. This briefing paper has focused on technical interventions which could address these issues; however, it is essential to complement such techniques with legal and social interventions. A comprehensive review of existing legal regulations and social barriers and recommendations to overcome these can be found in (Maidment *et al*., 2020; Lee and Hess, 2021). This chapter summarises the key issues discussed and provides recommendations on how these can be overcome.

7.1 Fostering Informed Consent

At present, the SMIP and the data sharing regulations are based on informed consent and permissions controls to safeguard consumers privacy. However, consent is currently obtained based on limited information. As shown in Chapter 5, the implications of different data sharing options vary considerably, and high-resolution smart meter data can be used to infer significant amounts of personal and sensitive information. Existing material provided by suppliers, the government, through Smart Energy GB and the data guide developed by Citizens Advice, do not communicate this information. Survey studies have shown that the majority of consumers, especially those in lower socio-economic groups, are unaware of the information embedded within smart meter data. Additionally, when provided with more in-depth information of the implications of data sharing, consumers are less willing to share their smart meter data. In addition, given the quickly evolving nature of machine learning and the increasing availability of rich linked datasets it is difficult to properly define and communicate the potential implications of data sharing. Explaining how machine learning techniques work and especially why they may find specific correlations is in some cases impossible and could lead to unintended consequences such as price discrimination or profiling.

It is important to clearly communicate the implications, in terms of the personal information being shared, of the different options (half-hourly, daily, monthly) provided within the DAPF and how that data will be used. This is essential to foster substantive informed consent. The existing smart meter data guides could be expanded to include such information while acknowledging that, given the evolving nature of smart meter data analytics, it is not possible to detail the personal information which may be inferred in the future. Expanding the data sharing options to include the sharing of anonymous data through use of privacy-preserving techniques, could greatly increase the availability of high-resolution smart meter data.

7.2 Widening Data Access

BEIS sees wider access to electricity data and the development of Open Data platforms as a key component of its net zero strategy. Ofgem's most recent consultation on suppliers' access to high-resolution data has widened the definition of regulated activities beyond billing and settlement to now include forecasting and business readiness. Although Ofgem have stated that for uses beyond settlement, data should be aggregated and anonymised, no guidance on appropriate techniques or mechanisms for ensuring compliance have been discussed.

Wider access would allow the full potential and value of smart meter data to be realised but will also require an increasing number of entities, who may or may not have direct contractual relationships with consumers, to access the data. This includes entities who currently have some access such as suppliers as well as public interest actors (e.g. BEIS, local authorities and Ofgem) who currently do not have access. In addition, smart meter data can have benefits in sectors beyond energy such as healthcare. Providing access to these third parties increases the risks of privacy infringements as well as potential data misuse. As the case study on the US census in Chapter 6.6 details, there is a real and immediate threat of inadvertent disclosure of personal information when even aggregate or pseudonymised data is accessible. With adequate safeguards using privacy-preserving techniques such as differential privacy, the risks of privacy infringements can be limited allowing anonymised data to be shared more widely.

7.3 Transparency Around Benefits And Usage

Many consumers are willing to share their smart meter data if the benefits of doing so, either for the system or for them personally, are clear. As discussed in Chapter 3 some of the benefits of smart metering do not require individually identifiable high-resolution data to achieve results. For example, better forecasting can be achieved with high resolution at the aggregate level with individual level data providing only marginal improvements (Wang *et al*., 2018). Similarly, many of the operational benefits for suppliers such as billing are performed locally on the smart meter and does not necessitate data sharing to function.

Ofgem's recent consultation on MHHS discussed the possibility of including pseudonymisation or aggregation for half-hourly data shared with suppliers for settlement and forecasting purposes. However, this was ruled out as they believed that the costs and complexity of implementation outweighed the potential benefits. In addition, it is expected that suppliers and other parties who access the data will include appropriate safeguards to comply with GDPR regulations (Article 5). Instead, it was decided that access to individual half-hourly data would be provided by default with an option to opt-out to a daily resolution. As a result, one of the core principles of Privacy by Design, that of privacy being the default setting, has not been followed.

Consequently, it is important for regulators, suppliers and third parties to consider how much data and at what level of granularity is required for different uses. Transparency around how smart meter data will be used, the benefits of data sharing and how these benefits will be distributed should be communicated to consumers. The benefits of providing more granular data can, for most regulated activities, be quantified and should also be clearly set out (Maidment *et al*., 2020). Limiting access to data to only what is necessary and where the benefits have been clearly identified prevents data hoarding and the limits of potential privacy infringements (GDPR Article 5 b and c). The introduction of smart appliances, demand response and CADs will increase the complexity of data flows and introduce numerous entities who have access

to smart meter data. A centralised dashboard encompassing all these data flows combined with the possibility to provide anonymised data would provide consumers with clarity on their data sharing options and choices. Citizens Advice's Data Dashboard provides a template for increasing transparency, building trust and potentially increase the sharing of high-resolution smart meter data and the acceptance of smart meters more generally (Citizens Advice, 2018).

7.4 Proactive And Preventative Risk Management

Permissions controls and consent provide a level of privacy protection, but these still leave consumers' data vulnerable to privacy infringements. The evolving nature of machine learning algorithms to generate insights into data and the increasing threats of cyberattacks make it difficult to accurately assess these risks and communicate these to consumers. The centralisation of disaggregated and identifiable high-resolution smart meter data creates a single point of failure. In addition, sharing of such data outside of regulated entities exposes consumers to risks of misuse.

Privacy by Design advocates for preventative and proactive management of such risks. Going beyond permissions control, the use of privacy-preserving mechanisms can ensure data is not easily re-identifiable and that even in the case of a data breach any inferences cannot be linked back to individual consumers. Importantly, this can be achieved with little to no loss of data utility allowing a wide range of benefits, which hinge on access to high resolution data.

Privacy-preserving techniques can help balance the legitimate privacy concerns of consumers and the value and benefits of wider access to high resolution smart meter data. Differential Privacy, in particular, is a prime candidate to achieve this given that it can be implemented either centrally or in a distributed manner, provides provable, tuneable and future proof guarantees of privacy protection and has been implemented in other sectors. The US Census Bureau provides a detailed framework which could be built upon to integrated within the UK's existing Smart Meter Implementation Programme.

7.5 Leveraging Heterogeneity

Privacy concerns regarding sharing of smart meter data vary significantly and depend on the intended use. Some consumers are unwilling to share high-resolution data whereas others are happy to share data at any resolution. Anonymising the data prior to sharing also increases consumers' willingness to share. In addition, many potential uses of smart meter data are not confined to regulated activities such as electricity settlement and billing. A transition towards a more dynamic and active domestic electricity sector will introduce new market actors, such as aggregators and local energy system operators, as well as new uses for smart meter data such as personalised tariffs and product recommendations. The current DAPF options do not allow this heterogeneity in preferences and uses to be leveraged fully.

Differential privacy provides a framework within which the privacy-utility trade-off can be explored and quantified. This provides consumers with greater control over their data and forces those wishing to access consumers' smart meter data to incentivise them. It enables consumers to reveal their true preferences and provides them with different levels of privacy protection while also maximising access to high-resolution data.

7.6 Conclusions

As of December 2020, 42% of domestic meters in the UK are smart meters. However, a number of barriers remain to widespread sharing of high-resolution smart meter data and a reliance on permissions controls make it difficult to safeguard consumer data against misuse intentional or otherwise. It is important to clearly communicate implications, in terms of the personal information being shared, of the different options (half-hourly, daily, monthly) provided within the DAPF and how that data will be used. This is essential to foster truly informed consent. Transparency around how smart meter data will be used, the benefits of data sharing and how these benefits will be distributed should be communicated to consumers.

Privacy-preserving techniques provide a framework within which such a privacy-utility trade-off can be explored and quantified. These techniques give consumers greater control over their data and force those wishing to access their smart meter data to incentivise them. This will allow consumers to reveal their true preferences, provide consumers with different levels of privacy protection while also maximising access to high resolution data. Going beyond permissions control, the use of privacy-preserving mechanisms, can ensure data is not easily re-identifiable and that even in the case of a data breach any inferences cannot be linked back to individual consumers. Importantly, this is done while preserving the utility of the data so that the benefits can be realised.

Policymakers and regulation could therefore consider introducing one or a combination of the privacy-preserving techniques discussed in this paper. Without significant changes to the ways in which data is stored, processed, and used, the adoption and resulting benefits of smart meters may not be realised. Differential Privacy, in particular, is a prime candidate to achieve this given that it can be implemented either centrally or in a distributed manner, provides provable, tuneable and future proof guarantees of privacy protection and has been implemented in other sectors. The US Census Bureau provides a detailed framework which could be built upon to integrated within the UK's existing Smart Meter Implementation Programme.

References

Abowd, J.M. (2018) "Disclosure Avoidance for Block Level Data and Protection of Confidentiality in Public Tabulations."

Abreu, J.M., Câmara Pereira, F. and Ferrão, P. (2012) "Using pattern recognition to identify habitual behavior in residential electricity consumption," *Energy and Buildings*, 49, pp. 479–487. doi:10.1016/j.enbuild.2012.02.044.

Ács, G. and Castelluccia, C. (2011) "I Have a DREAM! (DiffeRentially privatE smArt Metering)," in *Lecture Notes in Computer Science*, pp. 118–132. doi:10.1007/978-3-642-24178-9_9.

Apple Inc. (2017) *Learning with Privacy at Scale*.

Art. 29 WP (2014) *Opinion 05/2014 on Anonymisation Techniques*.

Asghar, M.R. *et al*. (2017) "Smart Meter Data Privacy: A Survey," *IEEE Communications Surveys & Tutorials*, 19(4), pp. 2820–2835. doi:10.1109/COMST.2017.2720195.

Aydinalp, M., Ismet Ugursal, V. and Fung, A.S. (2004) "Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks," *Applied Energy*, 79(2), pp. 159–178. doi:10.1016/j.apenergy.2003.12.006.

Aydinalp-Koksal, M. and Ugursal, V.I. (2008) "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," *Applied Energy*, 85(4), pp. 271–296. doi:10.1016/j.apenergy.2006.09.012.

Bagdasaryan, E. *et al*. (2020) "How To Backdoor Federated Learning," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 2938–2948.

Banks, J. and Mcglinchey, K. (2019) *Access to half-hourly electricity data for settlement purposes: a Data Protection Impact Assessment.*

Baringa LLP (2018) *Access to data arrangements: evaluation.* Available at: https://www.ofgem.gov.uk/sites/default/files/docs/2018/07/baringa_report_for_ofgem_enhanced_privacy_evaluation_for_hhs_published_version_2.0_0.pdf (Accessed: August 25, 2021).

BBC News (2019) "British Airways faces record £183m fine for data breach," 8 July.

Beckel, C. *et al*. (2014) "Revealing household characteristics from smart meter data," *Energy*, 78, pp. 397–410. doi:10.1016/j.energy.2014.10.025.

Beckel, C., Sadamori, L. and Santini, S. (2013) "Automatic socio-economic classification of households using electricity consumption data," in *Proceedings of the the fourth international conference on Future energy systems - e-Energy '13*. New York, New York, USA: ACM Press, p. 75. doi:10.1145/2487166.2487175.

BEIS (2014) *Smart Metering Equipment Technical Specifications 2* (SMETS2) Version 1.58.

BEIS (2018) *Smart Metering Implementation Programme: Review of the Data Access and Privacy Framework.*

BEIS (2019a) *Consultation on Smart Metering System Proportional Load Control Functionality.* London. Available at: https://smartenergycodecompany.co.uk/download/17882/.

BEIS (2019b) *Smart Meter Roll-Out Cost Benefit Analysis 2019.*

BEIS (2020a) *Energy Consumption in the UK (ECUK) 1970 to 2019.* Available at: https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes (Accessed: May 17, 2021).

BEIS (2020b) *Energy White Paper: Powering our Net Zero Future.* Available at: https://www.gov.uk/government/publications/energy-white-paper-powering-our-net-zero-future (Accessed: June 20, 2021).

BEIS (2021a) *Digitalising our energy system for net zero Strategy and Action Plan 2021.*

BEIS (2021b) *PAS 1878:2021 Energy smart appliances –System functionality and architecture – Specification*. Available at: https://shop.bsigroup.com/products/energy-smart-appliances-system-functionality-and-architecture-specification/standard (Accessed: January 13, 2022).

BEIS (2021c) *PAS 1879:2021 Energy smart appliances-Demand side response operation-Code of practice*.

BEIS (2021d) *Smart Meter Statistics in Great Britain: Quarterly Report to end December 2020.*

Brennan Center for Justice (2021) *Court Rejects Alabama Challenge to Census Plans for Redistricting and Privacy.* Available at: https://www.brennancenter.org/our-work/analysis-opinion/court-rejects-alabama-challenge-census-plans-redistricting-and-privacy (Accessed: August 18, 2021).

Brown, I. (2014) "Britain's smart meter programme: A case study in privacy by design," https://doi.org/10.1080/13600869.2013.801580, 28(2), pp. 172–184. doi:10.1080/13600869.2013.801580.

Buescher, N. et al. (2017) "Two Is Not Enough: Privacy Assessment of Aggregation Schemes in Smart Metering," *Proceedings on Privacy Enhancing Technologies*, 2017(4), pp. 198–214. doi:10.1515/popets-2017-0045.

California Public Utilities Commission (2014) *The California Public Utility Commission Decision 14-05-016*.

Carbon Trust and Imperial College London (2016) *An analysis of electricity system flexibility for Great Britain*.

Carmichael, R. et al. (2020) *Smart and Flexible Electric Heat An Energy Futures Lab Briefing Paper.* Available at: http://imperial.ac.uk/energy-futures-lab (Accessed: May 28, 2021).

Carmichael, R., Gross, R. and Rhodes, A. (2018) *Unlocking the potential of residential electricity consumer engagement with Demand Response.* Available at: https://www.imperial.ac.uk/energy-futures-lab/policy/briefing-papers/paper-3/.

do Carmo, C.M.R. and Christensen, T.H. (2016) "Cluster analysis of residential heat load profiles and the role of technical and household characteristics," *Energy and Buildings*, 125, pp. 171–180. doi:10.1016/j.enbuild.2016.04.079.

References

Cetin, K.S., Siemann, M. and Sloop, C. (2016) "Disaggregation and future prediction of monthly residential building energy use data using localized weather data network," in *ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 1–12.

Chalmers, C. *et al.* (2020) "Detecting Activities of Daily Living and Routine Behaviours in Dementia Patients Living Alone Using Smart Meter Load Disaggregation," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1. doi:10.1109/TETC.2020.2993177.

Chhachhi, S. and Teng, F. (2021) "Market Value of Differentially-Private Smart Meter Data," in. Institute of Electrical and Electronics Engineers (IEEE), pp. 1–5. doi:10.1109/isgt49243.2021.9372228.

Citizen's Advice (2016) *Summary report on energy suppliers' communication with consumers regarding smart meter data*.

Citizens Advice (2018) *Smart Metering Data Dashboard*. Available at: https://www.citizensadvice.org.uk/Global/CitizensAdvice/Energy/Smart Metering Data Dashboard .pdf.

Citizen's Advice (2019) *Clear and in control*. Available at: https://www.citizensadvice.org.uk/Global/CitizensAdvice/Energy/Clear and in control - Energy consumers' views on data sharing and smart devices.pdf (Accessed: June 29, 2021).

Cleemput, S. *et al.* (2018) "De-pseudonymization of Smart Metering Data: Analysis and Countermeasures," in 2018 *Global Internet of Things Summit (GIoTS)*. IEEE, pp. 1–6. doi:10.1109/GIOTS.2018.8534430.

Cominola, A. *et al.* (2017) "A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring," *Applied Energy*, 185, pp. 331–344. doi:10.1016/j.apenergy.2016.10.040.

Cornwall Insights (2019) *Market Wide Half-Hourly Settlement: Half-way home or just the first steps on the journey to a smart, flexible, energy system?* Available at: http://www.cornwall-insight.com/market-wide-half-hourly-settlement-half-way-home-or-just-the-first-steps-on-the-journey-to-a-smart-flexible-energy-system/ (Accessed: September 9, 2021).

Cuijpers, C. and Koops, B.-J. (2013) "Smart Metering and Privacy in Europe: Lessons from the Dutch Case," in *European Data Protection: Coming of Age*. Dordrecht: Springer Netherlands, pp. 269–293. doi:10.1007/978-94-007-5170-5_12.

D'Acquisto, G. *et al.* (2015) *Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics*. doi:10.2824/641480.

Danezis, G. (2015) *Smart Meter Aggregation Assessment: Review of the evidence*.

Data Privacy Manager (2021) *20 biggest GDPR fines so far [2019, 2020 & 2021]*. Available at: https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/ (Accessed: August 23, 2021).

DECC (2013) *Smart Metering Implementation Programme: Information Leaflet*.

DECC (2017) *Smart Meters, Smart Data, Smart Growth*.

Department for Digital, C.M.& S. (2021) *Data: A new direction*. Available at: https://www.gov.uk/government/consultations/data-a-new-direction (Accessed: January 18, 2022).

Desfontaines, D. and Pejó, B. (2020) "SoK: Differential privacies," *Proceedings on Privacy Enhancing Technologies*, 2020(2), pp. 288–313. doi:10.2478/popets-2020-0028.

Dickman, A. and Aslaksen, A.P. (2017) *Consumer attitudes to DNO access to half hourly electricity consumption data*.

Dinesh, C. *et al.* (2017) "Non-intrusive load monitoring under residential solar power influx," *Applied Energy*, 205, pp. 1068–1080. doi:10.1016/j.apenergy.2017.08.094.

Dromaque, C. *et al.* (2018) *The Role of Data for Consumer Centric Energy Markets and Solutions*.

Duesterberg, M. and Mirviss, L. (2021) *Reinventing Residential Demand Response Breaking Through Barriers with Gamification and Devices*.

Dwork, C. and Roth, A. (2013) "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), pp. 211–407. doi:10.1561/0400000042.

Eibl, G. *et al.* (2018) "The influence of differential privacy on short term electric load forecasting," *Energy Informatics*, 1(S1), p. 48. doi:10.1186/s42162-018-0025-3.

Eibl, G. and Engel, D. (2017) "Differential privacy for real smart metering data," *Computer Science - Research and Development*, 32(1–2), pp. 173–182. doi:10.1007/s00450-016-0310-y.

Elexon (2018) *Load Profiles and their use in Electricity Settlement*.

Elliot, M., Mackey, E. and O'Hara, K. (2020) *Anonymisation Decision Making Framework: European Practitioners' Guide*. 2nd edn. Manchester: UK Anonymisation Network.

Energy Digitalisation Taskforce (2022) *Delivering a Digitalised Energy System*.

Energy Systems Catapult (2020) *Smart Systems and Heat*. Available at: https://es.catapult.org.uk/case-studies/smart-systems-and-heat/ (Accessed: April 8, 2021).

Energy Systems Catapult (2021) *Enabling Smart Local Energy Systems: The value of digitalisation and data best practice*.

Farokhi, F. (2020) "Temporally discounted differential privacy for evolving datasets on an infinite horizon," in *Proceedings - 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems*, ICCPS 2020. doi:10.1109/ICCPS48487.2020.00008.

Fell, M. *et al.* (2017) *Energising Health: A review of the health and care applications of smart meter data*.

Garfinkel, S.L., Abowd, J.M. and Powazek, S. (2018) "Issues Encountered Deploying Differential Privacy," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. New York, NY, USA: ACM, pp. 133–137. doi:10.1145/3267323.3268949.

Gentry, C. (2009) *A fully homomorphic encryption scheme*. Stanford University.

Giaconi, G., Gunduz, D. and Poor, H.V. (2018a) "Privacy-Aware Smart Metering: Progress and Challenges," *IEEE Signal Processing Magazine*, 35(6), pp. 59–78. doi:10.1109/MSP.2018.2841410.

Giaconi, G., Gunduz, D. and Poor, H.V. (2018b) "Smart Meter Privacy With Renewable Energy and an Energy Storage Device," *IEEE Transactions on Information Forensics and Security*, 13(1), pp. 129–142. doi:10.1109/TIFS.2017.2744601.

References

Google (2021) *COVID-19 Community Mobility Reports*. Available at: https://www.google.com/covid19/mobility/ (Accessed: August 18, 2021).

Government Statistical Service (2018) *Privacy and data confidentiality methods: a Data and Analysis Method Review*. Available at: https://gss.civilservice.gov.uk/policy-store/privacy-and-data-confidentiality-methods-a-national-statisticians-quality-review-nsqr/ (Accessed: September 4, 2021).

Greveler, U., Justus, B. and Loehr, D. (2012) "Multimedia content identification through smart meter power usage profiles," *IN COMPUTERS, PRIVACY AND DATA PROTECTION* (CPDP) [Preprint]. doi:http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.727.4674.

Grunewald, P. (2020) *How has behaviour changed under the COVID-19 lockdown?, Joju Solar*. Available at: https://www.jojusolar.co.uk/opinion/how-has-behaviour-changed-under-covid-19-lockdown/ (Accessed: November 8, 2020).

Grunewald, P. and Diakonova, M. (2020a) *METER: UK Household Electricity and Activity Survey, 2016-2019, UK Data Service, SN:8634*. doi:10.5255/UKDA-SN-8634-1.

Grunewald, P. and Diakonova, M. (2020b) "Societal differences, activities, and performance: Examining the role of gender in electricity demand in the United Kingdom," *Energy Research & Social Science*, 69, p. 101719. doi:10.1016/j.erss.2020.101719.

Grünewald, P. and Reisch, T. (2020) "The trust gap - privacy perceptions of location data for energy services in the UK," *Energy Research and Social Science*, 68, p. 101534. doi:10.1016/j.erss.2020.101534.

Hawes, M.B. (2020a) "Differential Privacy and the 2020 Census: Modernizing Disclosure Avoidance at Scale to Mitigate Growing Privacy Threats," *Zenodo* [Preprint]. doi:10.5281/ZENODO.4122103.

Hawes, M.B. (2020b) "Implementing Differential Privacy: Seven Lessons From the 2020 United States Census," *Harvard Data Science Review*, 2(2). doi:10.1162/99608f92.353c6f99.

Hledik, R. *et al*. (2016) *The Tariff Transition Considerations for Domestic Distribution Tariff Redesign in Great Britain Volume I: Final Report*.

Hledik, R. et al. (2017) *The Value of TOU Tariffs in Great Britain: Insights for Decision-makers Volume I: Final Report*.

Hledik, R., Bagci, P. and Chhachhi, S. (2018) *Two Paths for Advancing Great Britain's Smart Metering Programme. The Brattle Group*.

Hoan, T.Q. (2014) "A Study on Privacy Level in Publishing Data of Smart Tap Network." Niigata: Esaki Labratory, The University of Tokyo.

Hoffmann, V. *et al*. (2019) "Automated detection of electric vehicles in hourly smart meter data," in *CIRED 2019 Conference*. AIM, p. 1531. doi:http://dx.doi.org/10.34890/666.

Horne, C. *et al*. (2015) "Privacy, technology, and norms: The case of Smart Meters," *Social Science Research*, 51, pp. 64–76. doi:10.1016/j.ssresearch.2014.12.003.

Hsu, J. *et al*. (2014) "Differential Privacy: An Economic Method for Choosing Epsilon," in *2014 IEEE 27th Computer Security Foundations Symposium*. IEEE, pp. 398–410. doi:10.1109/CSF.2014.35.

ICO (2012) *Anonymisation: managing data protection risk code of practice*. Available at: https://ico.org.uk/media/1061/anonymisation-code.pdf (Accessed: April 8, 2021).

ICO (2016) *Response to Ofgem's open letter of 17 December 2015: "Half-hourly settlement (HHS): the way forward"*. Available at: https://www.ofgem.gov.uk/sites/default/files/docs/2016/03/information_commissioner_response_-_dec_15_open_letter.pdf (Accessed: August 23, 2021).

ICO (2021) *Guide to the General Data Protection Regulation (GDPR)*. Available at: https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-1.pdf (Accessed: August 20, 2021).

Jakobi, T. *et al*. (2019) "It is about what they could do with the data: a user perspective on privacy in smart metering," *ACM Transactions on Computer-Human Interaction*, 26(1). doi:10.1145/3281444.

Jawurek, M., Kerschbaum, F. and Danezis, G. (2012) "SoK: Privacy Technologies for Smart Grids-A Survey of Options."

Jia, M. *et al*. (2021) "Privacy-Preserving Distributed Clustering for Electrical Load Profiling," *IEEE Transactions on Smart Grid*, 12(2), pp. 1429–1444. doi:10.1109/TSG.2020.3031007.

Jorgensen, Z., Yu, T. and Cormode, G. (2015) "Conservative or liberal? Personalized differential privacy," in *2015 IEEE 31st International Conference on Data Engineering*. IEEE, pp. 1023–1034. doi:10.1109/ICDE.2015.7113353.

Kairouz, P. *et al*. (2021) *Advances and Open Problems in Federated Learning*. Now Foundations and Trends.

Kalogridis, G. *et al*. (2010) "Privacy for Smart Meters: Towards Undetectable Appliance Load Signatures," in *2010 First IEEE International Conference on Smart Grid Communications*. IEEE, pp. 232–237. doi:10.1109/SMARTGRID.2010.5622047.

Kavousian, A., Rajagopal, R. and Fischer, M. (2013) "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior," *Energy*, 55, pp. 184–194. doi:10.1016/j.energy.2013.03.086.

Kim, H. *et al*. (2011) "Unsupervised Disaggregation of Low Frequency Power Measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics, pp. 747–758. doi:10.1137/1.9781611972818.64.

Kingsmill, S. and Cavoukian, A. (2015) *Privacy by Design Setting a new standard for privacy certification*.

Knight, A. (2018) *Consumer views on sharing half-hourly settlement data*.

Lee, D. and Hess, D.J. (2021) "Data privacy and residential smart meters: Comparative analysis and harmonization potential," *Utilities Policy*, 70, p. 101188. doi:10.1016/j.jup.2021.101188.

Li, F., Luo, B. and Liu, P. (2010) "Secure Information Aggregation for Smart Grids Using Homomorphic Encryption," in *2010 First IEEE International Conference on Smart Grid Communications*. IEEE, pp. 327–332. doi:10.1109/SMARTGRID.2010.5622064.

Liu, B., Luan, W. and Yu, Y. (2017) "Dynamic time warping based non-intrusive load transient identification," *Applied Energy*, 195, pp. 634–645. doi:10.1016/j.apenergy.2017.03.010.

References

References

Liu, Jinfei *et al.* (2021) "Dealer," *Proceedings of the VLDB Endowment*, 14(6), pp. 957–969. doi:10.14778/3447689.3447700.

Lu, D. (2020) "The Social Dilemma review: How big tech companies use us for profit," *New Scientist*, 29 September.

Maidment, C. *et al.* (2020) *Privacy and data sharing in smart local energy systems: Insights and recommendations*. EnergyRev.

McKenna, E., Thomson, M. and Barton, J. (2015) "CREST Demand Model." Loughborough University. doi:https://doi.org/10.17028/rd.lboro.2001129.v8.

MELLODDY (2019) *MachinE Learning Ledger Orchestration for Drug DiscoverY*. Available at: https://www.melloddy.eu/objectives.

Mervis, J. (2019) "Can a set of equations keep U.S. census data private?," *Science* [Preprint]. doi:10.1126/science.aaw5470.

Munkhammar, J. *et al.* (2015) "Household electricity use, electric vehicle home-charging and distributed photovoltaic power production in the city of Westminster," *Energy and Buildings*, 86, pp. 439–448. doi:10.1016/J.ENBUILD.2014.10.006.

Narayanan, A. and Shmatikov, V. (2006) "How To Break Anonymity of the Netflix Prize Dataset," *arXiv preprint* [Preprint].

National Grid ESO (2021) *Virtual Energy System*. Available at: https://www.nationalgrideso.com/virtual-energy-system (Accessed: January 13, 2022).

OFGEM (2020) *Electricity retail market-wide half-hourly settlement: consultation*. Available at: https://www.ofgem.gov.uk/system/files/docs/2020/05/mhhs_draft_impact_assessment_consultation.pdf.

OFGEM (2021a) *Electricity Retail Market-wide Half-hourly Settlement: Decision Document*.

OFGEM (2021b) *Market-wide Half-Hourly Settlement: Final Impact Assessment*. Available at: https://www.ofgem.gov.uk/energy-policy-and-regulation/policy-and-regulatory-programmes/electricity-settlement-reform (Accessed: August 21, 2021).

OFGEM (2022) *Statutory consultation on proposals to modify electricity supply licence condition 47: "Smart Metering-Matters Relating to Obtaining and Using Consumption Data."* Available at: www.ofgem.gov.uk. (Accessed: May 9, 2022).

Paillier, P. (1999) "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *Advances in Cryptology — EUROCRYPT '99*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 223–238. doi:10.1007/3-540-48910-X_16.

Parti, M. and Parti, C. (1980) "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," *The Bell Journal of Economics*, 11(1), p. 309. doi:10.2307/3003415.

Paxman, J. *et al.* (2020) *Smart Future of Healthcare*.

Petti, S. and Flaxman, A. (2020) "Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff," *Gates Open Research*, 3, p. 1722. doi:10.12688/gatesopenres.13089.2.

Ramírez-Mendiola, J.L., Grünewald, P. and Eyre, N. (2018) "Linking intra-day variations in residential electricity demand loads to consumers' activities: What's missing?," *Energy and Buildings*, 161, pp. 63–71. doi:10.1016/j.enbuild.2017.12.012.

Recurve (2021) *Real World Use-Cases for Energy Differential Privacy: Using EDP to Track COVID Impacts., Recurve Blog*. Available at: https://www.recurve.com/blog/traditional-approaches-to-protecting-energy-data-dont-work-heres-what-to-do-instead-part-3-of-3 (Accessed: April 8, 2021).

Rhodes, A. (2020) *Digitalisation of Energy An Energy Futures Lab Briefing Paper*.

Richter, L.-L. and Pollitt, M.G. (2018) "Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market," *Energy Economics*, 72, pp. 436–450. doi:10.1016/j.eneco.2018.04.004.

Rocher, L., Hendrickx, J.M. and de Montjoye, Y.-A. (2019) "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature Communications*, 10(1), p. 3069. doi:10.1038/s41467-019-10933-3.

Satre-Meloy, A., Diakonova, M. and Grunewald, P. (2018) "Daily life and demand: New data on behavioral drivers of residential electricity use patterns," *2018 ACEEE Summer Study on Energy Efficiency in Buildings* [Preprint]. Available at: https://aceee.org/files/proceedings/2018/index.html#/paper/event-data/p264.

Sheikh, N. *et al*. (2021) "Trace Recovery: Inferring Fine-grained Trace of Energy Data from Aggregates," in *Proceedings of the 18th International Conference on Security and Cryptography*. SCITEPRESS - Science and Technology Publications, pp. 283–294. doi:10.5220/0010560302830294.

Sheller, M.J. *et al*. (2019) "Multi-institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing, pp. 92–104. doi:10.1007/978-3-030-11723-8_9.

Singhal, V., Maggu, J. and Majumdar, A. (2018) "Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning," *IEEE Transactions on Smart Grid*, 10(3), pp. 2969–2978.

Skatova, A. *et al*. (2019) "Unpacking Privacy: Willingness to pay to protect personal data," *PsyArXiv* [Preprint]. doi:10.31234/osf.io/ahwe4.

Smart Energy Code Company (2021) *The Smart Energy Code: Schedule 9*. Available at: https://smartenergycodecompany.co.uk/the-smart-energy-code-2/ (Accessed: April 8, 2021).

Smart Energy GB (2021) *Smart Meter Installation Process*. Available at: https://www.smartenergygb.org/en/get-a-smart-meter/the-installation-process (Accessed: April 15, 2021).

SSEN (2020) *Smart Meter Data Privacy Plan (Access to household Electricity Consumption Data)*.

Stankovic, L. *et al*. (2016) "Measuring the energy intensity of domestic activities from smart meter data," *Applied Energy*, 183, pp. 1565–1580. doi:10.1016/j.apenergy.2016.09.087.

Steele, P. (2019) *Octopus Energy: Agile pricing explained*. Available at: https://octopus.energy/blog/agile-pricing-explained/ (Accessed: April 8, 2021).

References

Sustainability First and CSE (2018) *Smart Meter Energy Data: Public Interest Advisory Group (PAIG). Stimulus paper 2 - International Experience-Smart Meter Data Access. Maxine Frerk, Sustainability First*. Available at: https://www.smartenergydatapiag.org.uk/ (Accessed: December 8, 2021).

Sustainability First and CSE (2019) *Smart Meter Energy Data: Public Interest Advisory Group (PAIG). Stimulus Paper 7 - Possible routes to the data for a public interest. Maxine Frerk, Sustainability First with support from Judith Ward, Sustainability First & Simon Roberts and Nicky Hodges, CSE*. Available at: https://www.smartenergydatapiag.org.uk/ (Accessed: December 8, 2021).

Sustainability First and CSE (2021a) *Smart Meter Energy Data: Public Interest Advisory Group (PAIG). Annex 1 - Working Paper on DNO Privacy Plans. Maxine Frerk, Sustainbility First*. Available at: https://www.cse.org.uk/downloads/file/PIAG-phase-2-privacy-plans-annex.pdf (Accessed: December 6, 2021).

Sustainability First and CSE (2021b) *Smart Meter Energy Data: Public Interest Advisory Group (PIAG). Final Report - Phase 2. Maxine Frerk, Sustainability First with support from Judith Ward, Sustainability First & Simon Roberts and Nicky Hodges, CSE*. Available at: https://www.sustainabilityfirst.org.uk/images/publications/piag/PIAG-phase-2-final-report.pdf - (Accessed: December 6, 2021).

Sweeney, L. (2000) *Simple Demographics Often Identify People Uniquely*.

Taik, A. and Cherkaoui, S. (2020) "Electrical Load Forecasting Using Edge Computing and Federated Learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. IEEE, pp. 1–6. doi:10.1109/ICC40277.2020.9148937.

Tang, J. *et al.* (2019) "Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12," *arXiv preprint* [Preprint].

US Census Bureau (2019a) *A History of Census Privacy Protections*.

US Census Bureau (2019b) *GitHub 2020 Census Repository*. Available at: https://github.com/uscensusbureau/census2020-das-2010ddp (Accessed: August 18, 2021).

US Census Bureau (2020) *The "72-Year Rule."* Available at: https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html (Accessed: August 18, 2021).

US Census Bureau (2021a) *Comparing Differential Privacy With Older Disclosure Avoidance Methods*.

US Census Bureau (2021b) *Key Parameters Set to Protect Privacy in 2020 Census Results*. Available at: https://www.census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html (Accessed: August 23, 2021).

Véliz, C. and Grunewald, P. (2018a) "Protecting data privacy is key to a smart energy future," *Nature Energy*, 3(9), pp. 702–704. doi:10.1038/s41560-018-0203-3.

Véliz, C. and Grunewald, P. (2018b) "Protecting data privacy is key to a smart energy future," *Nature Energy*, 3(9), pp. 702–704. doi:10.1038/s41560-018-0203-3.

Vepakomma, P. *et al.* (2018) "Split learning for health: Distributed deep learning without sharing raw patient data," *arXiv preprint* [Preprint].

Vigurs, C. *et al.* (2021) "Customer Privacy Concerns as a Barrier to Sharing Data about Energy Use in Smart Local Energy Systems: A Rapid Realist Review," *Energies*, 14(5), p. 1285. doi:10.3390/en14051285.

Wang, Y. *et al.* (2018) "An Ensemble Forecasting Method for the Aggregated Load With Subprofiles," *IEEE Transactions on Smart Grid, 9*(4), pp. 3906–3908. doi:10.1109/TSG.2018.2807985.

Wang, Y. *et al.* (2019) "Deep Learning-Based Socio-Demographic Information Identification From Smart Meter Data," *IEEE Transactions on Smart Grid,* 10(3), pp. 2593–2602. doi:10.1109/TSG.2018.2805723.

Wang, Y. *et al.* (2021) "Electricity Consumer Characteristics Identification: A Federated Learning Approach," *IEEE Transactions on Smart Grid,* 12(4), pp. 3637–3647. doi:10.1109/TSG.2021.3066577.

Webborn, E. *et al.* (2019) "Utilising smart meter data for research and innovation in the UK," *ECEEE SUMMER STUDY PROCEEDINGS* [Preprint].

Welikala, S. *et al.* (2019) "Implementation of a robust real-time non-intrusive load monitoring solution," *Applied Energy*, 238, pp. 1519–1529. doi:10.1016/j.apenergy.2019.01.167.

Winegar, A.G. and Sunstein, C.R. (2019) "How Much Is Data Privacy Worth? A Preliminary Investigation," *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3413277.

Xue, K. *et al.* (2020) "An Efficient and Robust Data Aggregation Scheme Without a Trusted Authority for Smart Grid," *IEEE Internet of Things Journal*, 7(3), pp. 1949–1959. doi:10.1109/JIOT.2019.2961966.

Yang, Q. *et al.* (2019) "Federated Machine Learning," *ACM Transactions on Intelligent Systems and Technology*, 10(2), pp. 1–19. doi:10.1145/3298981.

Yang, T. *et al.* (2018) "Applied Federated Learning: Improving Google Keyboard Query Suggestions," *arXiv preprint* [Preprint].

Zhang, C. *et al.* (2020) "BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning," in *USENIX Annual Technical Conference*. USENIX Association, pp. 493–506.

Zhang, Z. *et al.* (2014) "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," in *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 5419–5425. doi:10.1109/IECON.2014.7049328.

Zhao, B., Stankovic, L. and Stankovic, V. (2018) "Electricity usage profile disaggregation of hourly smart meter data," *4th International Workshop on Non-Intrusive Load Monitoring* [Preprint]. Available at: https://strathprints.strath.ac.uk/63692/.

Energy Futures Lab

Energy Futures Lab is one of seven Global Institutes at Imperial College London. The institute was established to address global energy challenges by identifying and leading new opportunities to serve industry, government and society at large through high quality research, evidence and advocacy for positive change. The institute aims to promote energy innovation and advance systemic solutions for a sustainable energy future by bringing together the science, engineering and policy expertise at Imperial and fostering collaboration with a wide variety of external partners. The Energy Futures Lab Briefing Papers are periodic reports aimed at all stakeholders in the energy sector. They bring together expertise from across Imperial College London to provide clarity on a wide range of energy topics.

For more information visit: **http://imperial.ac.uk/energy-futures-lab**

Suggested citation:
Teng, F., Chhachhi, S., Ge, P., Graham, J., and Gunduz, D. (2022) Balancing Privacy and Access to Smart Meter Data, An Energy Futures Lab Briefing Paper, Imperial College London. Available at: www.imperial.ac.uk/energy-futures-lab/reports/briefing-papers/paper-9/

Contact us

energyfutureslab@imperial.ac.uk
+44 (0)207 594 5865

http://imperial.ac.uk/energy-futures-lab

Twitter: @energyfuturesic
Facebook: energyfutureslab
Instagram: energyfutureslab