Smart Meter Energy Data: Public Interest Advisory Group

A policy dialogue and work programme

led by

Centre for Sustainable Energy & Sustainability First

Stimulus paper 5

Public interest use-cases: data attributes, data requirements, and associated privacy and access implications

Author: Simon Roberts, CSE

Status of this Document

This paper was prepared as an input to the work programme of the Public Interest Advisory Group on access to smart meter energy data.

Last revision 17 October 2018

1. Introduction

Analysis by the PIAG team and a programme of engagement with PIAG members and other stakeholders has produced a set of six high-level 'archetype' use-cases for smart meter energy data which can be considered broadly 'in the public interest' (See PIAG Stimulus Paper 4 at 'Downloads' at www.smartenergydatapiag.org.uk).

To meet PIAG's purposes, there is a need to understand in more detail how each use-case could be served and the potential implications of doing so for access to smart meter data and consumer privacy concerns. This requires analysis of:

a. the attributes of the smart meter data required to serve the use-case, including spatial and temporal resolution and frequency of data capture;

b. the attributes of any other data (such as building data, socio-demographic data, other energy data etc) that would be required alongside the smart meter data to meet each use-case;

c. the data access, analysis and processing requirements to prepare the dataset required to serve each use-case;

d. the attributes of the dataset that would need to be made available to users to serve each public interest use-case, and;

e. the extent to which the necessary dataset to deliver each use-case might risk compromising consumer privacy.

This paper examines these issues, covering the six 'archetype' use-cases which capture the essential qualities of the range of the public interest use-cases identified for PIAG (Xref Nicky's paper).

2. Key considerations

2.1 Absence of a database of smart meter data which could serve public interest uses

In considering how smart meter data could serve 'public interest' uses, it is vital first to understand that, there are no current plans for a database which might hold all of the domestic smart meter data for Great Britain. This was a choice made in the design of the smart meter roll-out in response to concerns about potential risks to privacy and data security.

Rather than being stored as part of a 'complete' national database or data system, the data from each domestic smart meter will be stored only on the individual household meter (for 13 months) and by each customer's electricity and gas supplier. It is not clear how long suppliers plan to store this data or at what resolution (e.g. half-hourly) the data will be held. Plans and systems are likely to vary between suppliers.

Correspondingly, there is no requirement on any organisation either to capture or to maintain domestic smart meter data within a single database or data system which could then, in principle, be utilised to serve a defined public interest use.

The implication of this situation for our analysis of how to use smart meter data to serve the public interest use-cases is simple:

To create any dataset to enable the realisation of any of the six public interest use-cases we have identified would require the individual level smart meter records (1) to be captured and (2) to be processed.

In the absence of a mass programme of securing individual-level household consent for 'public purpose' access, this suggests a need for a 'trusted processor' (or equivalent capability) to capture, curate, analyse and process the individual smart meter records in a secure environment (i.e. one that carries virtually no risk of individual privacy being breached by release to other parties). The Office for National Statistics undertakes this role in relation to a range of sensitive household socio-demographic and income data.

A future PIAG paper will examine potential options for data-capture and how the question of a trusted processor might be addressed, including characteristics required of the trusted processor function and options for how this might be realised in practice, taking into account the Data Access and Privacy Framework (DAPF), GDPR and other privacy considerations.

However, for the sake of the analysis in this paper, we have assumed that this function is in place. As indicated above, it is difficult to see how, without this further step, most of the public interest use-cases we have identified for smart meter data could be served.[1]

2.2 Distinguishing between data *input* and data *output* requirements

In order to understand the privacy implications of potential approaches to realising the use-cases, it is important to distinguish between (a) the data required for processing to create the dataset which can then be used to serve a use-case and (b) the nature of the data then available in that dataset.

This is because, as will be demonstrated in this paper, the privacy issues potentially associated with the data processing exercise (what we term 'Inputs') can be very different from the privacy issues potentially associated with the dataset subsequently available for a public interest purpose (what we term the 'Output').

As explained in Section 0 above, the *Inputs* for every use-case would always require the processing of individual household smart meter data (if only to aggregate it to the scale required for the use-case).

However, as explained below, **for each of the use-case archetypes identified, the *Output* dataset can serve the use-case *without* containing any data which has identifiers for individual households** (and therefore offers no potential for a breach of their privacy).

Later PIAG papers and discussions will explore the privacy and consent issues associated with this sort of secure data processing to serve a public interest purpose and where the associated dataset cannot itself compromise privacy.

1 This is because the public interest use-cases generally require data at a scale which would make securing individual household consent for access to smart meter data unfeasible. While there are initiatives such as the UCL Smart Meter Research Portal which will be building up a longitudinal database of household smart energy data having secured individual consents to access smart meter data, there will be challenges in confirming the representativeness of the participating households and significant limitations in terms of use-cases focused on local energy system planning.

3. Public interest use-cases: the six archetypes

The six archetype public interest use-cases are introduced briefly below; their requirements of smart meter and other data are explored in Section 5 below.

Use Case 1: National and sub-national domestic sector energy statistics

Smart meter energy data could enhance significantly the detail available in national and sub-national statistics about domestic energy consumption. It could also reduce delays in the production of the statistics, providing policy makers and others with the opportunity to observe and respond to trends more promptly. For example, using just smart meter data could provide a wider range of typical daily profiles for each season (for consumption and peak demand), peak demand for electricity and for heat, levels of export from domestic generation etc.

Use Case 2: Local level energy system planning

There is a public interest in enabling local actors (such as local authorities, community interest groups etc) to establish a detailed picture of their local energy system and the fine-grain patterns of energy supply and demand within it. This will enable better infrastructure planning (for example, to install heat networks or EV charging points) and the design and targeting of more effective interventions to reduce carbon emissions, improve system efficiency and secure local economic benefit (such as balancing local electricity supply and demand more effectively, aggregating neighbourhood demand response, or monitoring the impact of local low carbon housing refurbishment initiatives).

Use Case 3: Data for analysis and modelling to support policy making, research and insights

Smart meter data could offer a significant opportunity to improve the quality of the data analysis, modelling and research which supports policy-making by providing much finer grain energy data alongside other household attributes of policy interest. These improvements could be made in relation to, for example: policy-makers understanding of energy use in different types of household and dwelling type; the potential distributional impacts of policy and market changes (some of which may only occur at particular times of the year or for households with particular energy use profiles); the monitoring evaluation of policy impacts; etc.

Use Case 4: Improved intervention design and testing

Many public interest stakeholders are interested in how smart meter data could help them to improve the design and targeting of their planned interventions (for example to tackle fuel poverty or encourage take up of demand flexibility services locally).

Use Case 5: Local electricity system 'live' monitoring to trigger reactions/interventions in real time

Several public interest stakeholders were keen to have access to 'live' (i.e. virtually real time) data on the local electricity system in order to drive interventions such as local triggers for demand shifting or reduction or participation in local (or national) balancing and associated markets.

Use Case 6: Service innovation and development and testing of early stage design/algorithms etc

A number of public interest stakeholders are interested in the development and/or provision of services which used smart meter data to help individuals to gain some sort of personal benefit (from thermal safeguarding which warns people – or their carers - when their home is unhealthily cold and/or damp, to peer-to-peer trading services). While we have concluded that provision of smart meter data to enable the delivery of these services should be a matter for individual consent, there is a public interest in such services being developed so that they become available for individuals to choose whether to participate.

It should be noted (as outlined in PIAG Stimulus Paper 4 – see 'Downloads' at www.smartenergydatapiag.org.uk) that stakeholders also identified potential public interest value in the research and policy insights which could be gained from including individual household smart meter energy data as an 'administrative dataset' alongside other individual household datasets. The smart meter data would then be available for the type of 'big data' analysis typically undertaken by ONS as a 'trusted processor' on behalf of government and which is becoming an increasingly common feature of policy design processes following the introduction of the Digital Economy Act. However, this public interest use for smart meter data has not been included here as an 'archetype' because the use-case ('better policy evidence') is too generic to enable a clear description of either the required date inputs or the required data outputs. Nevertheless, it should be borne in mind within PIAG's wider consideration of the potential value of smart meter energy data being accessed and applied for public interest purposes.

4. Understanding different data attributes

To provide a consistent and intelligible framework for considering the data attributes of smart meter data, we have identified three key dimensions on which they can vary:

Temporal resolution: ranging from a single annual figure (kWh or peak kW demand) to below half-hourly

Spatial resolution: from national level down to individual property level

Data capture frequency: ranging from yearly down to live feed of real-time data

It should be borne in mind that the data captured by smart meters includes both gas and electricity data and includes potentially useful data such as peak-use and maximum load (kW) as well as consumption data (kWh).

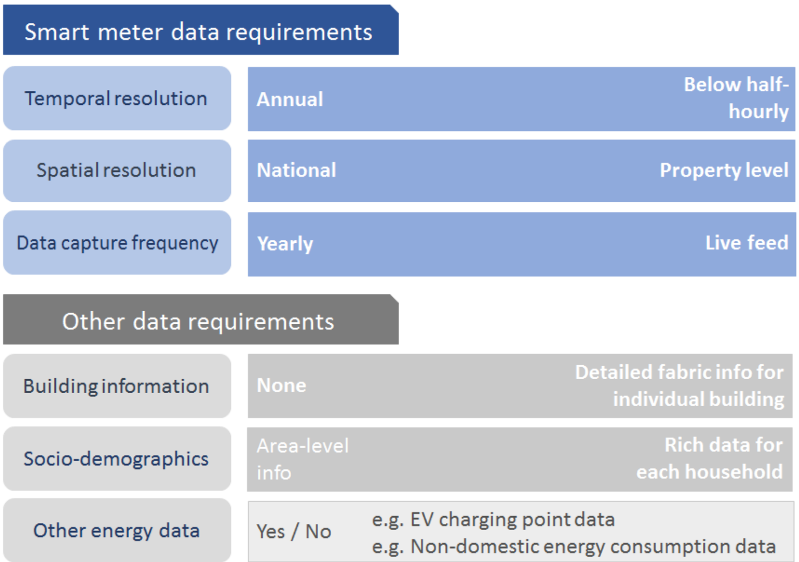
In addition, the other types of data potentially required to serve at least some of the use-cases have similar dimensions which vary by granularity of resolution:

Building information: from no use of building-related information to use of detailed data for individual buildings (such as size, built form, energy performance, use, tenure etc)

Socio-demographics: from use of area-level (e.g. LSOA) information to use of rich data for each household

In addition, to realise some use-cases will require the domestic smart meter data to be combined with what we've categorised as **'Other energy data'**, such as non-domestic energy consumption data, local energy generation data, EV charging point data.

For each use-case these can be represented graphically (see below). By indicating where on each dimension the required data sits (for example, half-hourly data at property level captured annually), the data attributes associated with each use-case can be illustrated and simply compared. As discussed above, these need to be considered separately for the data **Inputs** for processing and analysis and the resulting **Output** dataset which is made available to serve the use-case. This distinction is important because there are significant differences between the two, with very different implications for the potential risk of compromising the data privacy of an individual household.



5. Archetype use-cases and their data requirements

Using the framework of data attributes outlined above, we can review each use-case archetype in terms of its required data processing Inputs and the attributes of the required Output dataset which enables the use-case to be realised.

Use Case 1: National and sub-national domestic sector energy statistics

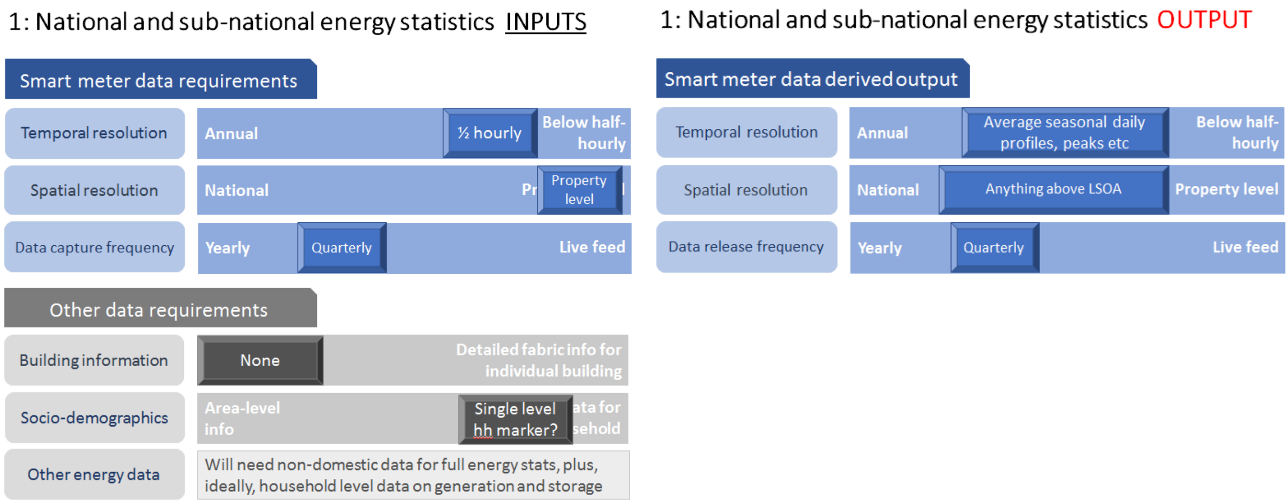
The Digest of UK Energy Statistics – DUKES – currently uses monthly, quarterly and annual survey data from energy generators and producers, energy suppliers and major energy users to create a 'top-down' data set of energy supplied to different end-use sectors (including domestic). The survey does not directly collect any meter data (though the data from energy suppliers and users is likely to be based on aggregated meter data they have collected). The sub-national domestic sector energy statistics (which are used in Energy Consumption in the UK – ECUK – and the National Energy Efficiency Database) are compiled from a single annual consumption record for each household as collected from (largely) non-smart meters by Xoserve (in case of gas) and from a number of meter data aggregators (in the case of electricity). These household records are aggregated to national level annually and also published at various other levels of spatial resolution for annual data (from regions and local authority areas down to Lower Layer Super Output Area LSOA).[2]

2 Electricity consumption data has also been published by BEIS (as combined domestic and non-domestic consumption) at full postcode level in an experimental dataset for 2015 (see here, published in March 2017).

Both of these approaches (DUKES and the sub-national data) provide the basis for calculating average annual domestic consumption and, by comparing datasets over time, observing trends in consumption. However, because the data is either directly or indirectly sourced from surveys or 'dumb' meters and only contains a single annual consumption figure, it can provide none of the detail of typical usage patterns across the year and associated peak demands etc. In addition, the household-level dataset includes a degree of inaccuracy because, in its compilation, any estimated meter reads in the annual dataset are ignored (reverting instead to the most recent 'real' meter read for that household in a previous dataset, and therefore likely to be at least a year out of date, albeit weather corrected).

Given the scale and speed of change in the energy system, there are potentially public interest benefits from improving these important statistical inputs to public policy and regulatory practice. Using domestic smart meter data as the input to these statistics could, in time, greatly reduce delays in their availability and significantly improve the detail available. For example, using just smart meter data, it could provide a wider range of typical daily profiles for each season (for consumption and peak demand), peak demand for electricity and for heat, levels of export from domestic generation etc. [3]

The attributes of the data Inputs and the dataset Output are shown below, using the graphical representation developed for this analysis.



The important distinction to be made from a privacy perspective between 'Inputs' data attributes and those of the 'Output' dataset is well illustrated by this use-case. The 'Inputs' data required for this use-case is property level and half-hourly smart meter data (and therefore arguably the most sensitive from a household privacy perspective). However, the Output dataset for publication (to meet this use-case) would contain data aggregated both spatially and temporally to such an extent that it would become impossible to identify any individual record. The implications for privacy thus relate only to the use of individual records in the initial data capture and 'trusted processing' to create outputs to serve this public interest use-case, not in the published output itself.

Consumption for postcodes with as few as 6 meters are reported (those with fewer having been amalgamated into neighbouring codes).

3 However, as domestic generation (eg PV) and in-home 'behind the meter' storage become more prevalent, it will also be important to capture data on the contribution these technologies are making; if they are ignored, meter readings will not reflect 'final consumption' by the household but rather 'final demand on the energy system' made by the household. Both represent important information but they are different

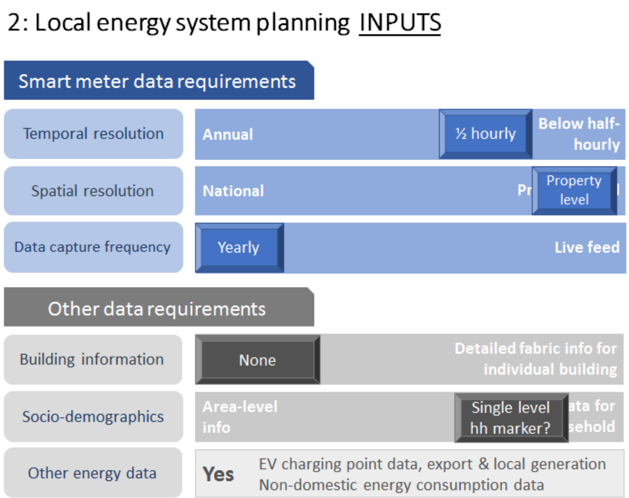
Use Case 2: Local-level energy system planning

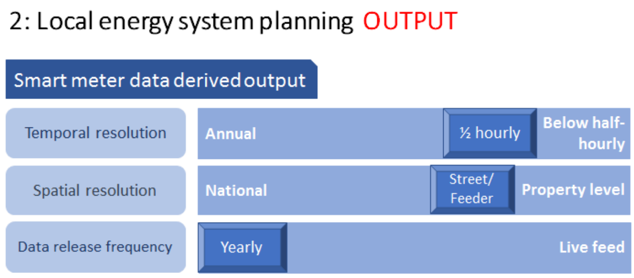
There is a public interest in enabling local actors (such as local authorities, community interest groups etc) to establish a detailed picture of their local energy system and the fine-grain patterns of energy supply and demand within it. This will enable better infrastructure planning (for example, to install heat networks or EV charging points) and the design and targeting of more effective interventions to reduce carbon emissions, improve system efficiency and secure local economic benefit (such as balancing local electricity supply and demand more effectively, aggregating neighbourhood demand response, or monitoring the impact of local low carbon housing refurbishment initiatives).

At present, this data is not available at sufficient temporal resolution (currently a single annual consumption figure) to enable effective local system planning. Similarly, LSOA level data is arguably too large (at c. 500 homes – or a population of 1,000) to enable effective identification of opportunities for interventions or new infrastructure. Smart meter data could help to change this.

By way of Inputs, the smart meter data attributes required are very similar to the national and subnational data use-case. However, in order to provide a full picture of the local energy system which local planning would require, the Inputs also need to extend beyond domestic data to include other energy data such as non-domestic energy consumption data, generation/supply data, EV charging etc and information about local electricity and gas networks (both their topology and their current capacity and constraints).

To serve use-case 2, the Output dataset needs to be at fine grain both temporally and spatially. Half hourly data is needed to provide the full picture required for local system planning (for smart meter data and the other energy data). Aggregating to street (or DNO low voltage substation feeder) level (with care taken with those areas with very low population density) should provide sufficient detail without risking compromising individual privacy. For the purposes of planning (as opposed to operation) of the local energy system, a yearly dataset from the previous year will be adequate.





In this approach, the Output dataset would be purely energy data, but it could be combined at local level with other socio-demographic and building datasets (e.g. Index of Multiple Deprivation) to improve understanding and aid targeting of interventions (for example at more disadvantaged neighbourhoods).

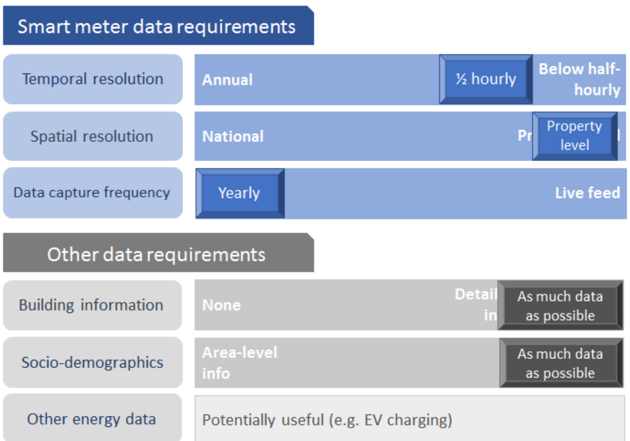
Use Case 3: Data for analysis and modelling to support policy making, research and insight

Smart meter data could offer a significant opportunity to improve the quality of the data analysis, modelling and research which supports policy-making by providing much finer grain energy data alongside other household attributes of policy interest. These improvements could be made in relation to, for example: policy-makers understanding of energy use in different types of household and dwelling type; the potential distributional impacts of policy and market changes (some of which may only occur at particular times of the year or for households with particular energy use profiles); the monitoring evaluation of policy impacts; etc.

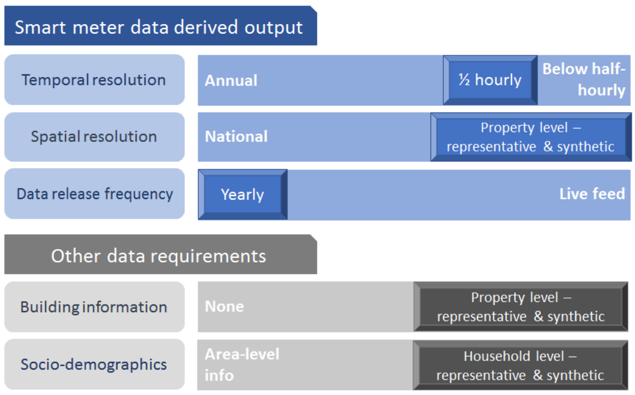
Serving this public interest would require a dataset that is representative of the population by socio-demographic characteristics, building type (and energy performance levels), and energy usage and half-hourly profile. Developing such a dataset will require (as the Inputs) the full smart meter dataset (at property level and half-hourly resolution) for a year (i.e. historic data) to be combined with as much socio-demographic and building data as possible available at individual property/household level.

This combining of datasets, matched at address level, would need to be done at individual property level to create a rich dataset of characteristics for each household for further analysis. Given the current limited understanding of the how energy use profiles vary by socio-demographic characteristics and building type, it will only be by analysing these individual records that it becomes possible to be confident in the creation of an adequately representative sample dataset. As mentioned 'representativeness' would need to be achieved to an adequate degree for socio-demographic, building and energy use profile features, the nature of which can only be established by looking at the full 'raw' dataset. The analysis of the full dataset will also provide guidance for what 'representative' looks like for other sampled datasets (such as being developed for the SMRP).

3: Data for analysis and modelling INPUTS



3: Data for analysis and modelling OUTPUT



To serve use-case 3, the Output dataset will need to retain this level of granularity so that socio-demographic and building 'markers' are associated with individual smart meter records. However, to avoid the risk of individuals being identifiable from the resulting dataset, the individual records would need to be anonymised (for example by losing their property-specific geographical marker) to create a 'synthetic' (i.e. not reflecting 'real' cases) but representative dataset. Other publicly available official datasets (such as the Food and Expenditure Survey and the English Housing Survey) have these characteristics. Further exploration is needed with key PIAG members (such as ONS) to be sure that such anonymization techniques can be effective without compromising the value of the data for analytical purposes.

Use Case 4: Improved intervention design and targeting

As discussed elsewhere (see PIAG Stimulus Paper 4 at 'Downloads' at www.smartenergydatapiag.org.uk), many public interest stakeholders are interested in how smart meter data could help them to improve the design and targeting of their planned interventions (for example to tackle fuel poverty or encourage take up of demand flexibility services locally).

We have concluded that the use of an individual household's smart meter data to support improved delivery of services to that household **should require the household's consent**. However, the design and targeting of these services could be improved through the use of datasets produced for other use-cases outlined above, specifically the local dataset Output from use-case 5.2 and the nationally representative dataset from use-case 5.3. Between them, they would give prospective service providers extensive insight into the patterns of energy using behaviour of their target socio-demographic and also, potentially, in which neighbourhoods they will tend to be found locally.

Use Case 5: Local electricity system 'live' monitoring

Several public interest stakeholders were keen to have access to 'live' (i.e. virtually real time) data on the local electricity system in order to drive interventions such as local triggers for demand shifting or reduction or participation in local (or national) balancing and associated markets.

However, domestic smart meters will not be collecting or transmitting this 'live' data routinely so, irrespective of the potential public interest that would be served, this use-case cannot be realised by use of smart meter data. Instead, the local electricity distribution network operator may potentially provide this sort of data as it develops its role as Distribution System Operator. This data would probably be provided at LV substation or feeder level (see, for example the Network Innovation Competition funded project, Open LV). While the DNOs are also in the process of establishing the basis on which they will have access to smart meter data (see PIAG note on energy networks and smart meter data at 'Downloads' at www.smartenergydatapiag.org.uk), this would not captured be in real-time at the level of individual records (aside from a meter's 'last gasp' notification signal in the event of a power cut); it will therefore tend to be used by DNOs/DSOs for planning purposes and, over time, trend analysis, rather than operational management.

Use Case 6: Service development innovation test-bed

A number of public interest stakeholders are interested in the development and/or provision of services which used smart meter data to help individuals to gain some sort of personal benefit (from thermal safeguarding which warns people – or their carers - when their home is unhealthily cold and/or damp, to peer-to-peer trading services – see PIAG Stimulus Paper 4 at 'Downloads' at www.smartenergydatapiag.org.uk).

As with use-case 5.4 above, **we have concluded that there is no public interest argument to impose such services on householders without their consent** or, therefore, to enable service providers to access a household's smart meter data without their explicit consent.

However, we believe there is a public interest in many of these types of service coming to market and to them being well-designed and targeted when they do. This may only be possible if those developing the services have access during their development phase to decent, fine grain smart meter data (of the sort they would have from a consenting household) which is representative of their target market. This would enable them to test algorithms and service design and examine the potential for unintended negative impacts for some potential customers. It would avoid the need to secure consent from households (and its associated costs and risk) for access to their data for a service which does not yet exist or for one which is not yet tested.

The representative dataset that is created as the Output for use-case 5.3 above would provide such an 'innovation test bed'. Making this dataset available (possibly with some sort of public-interest test for those seeking access) would improve the quality of products and services coming to market and reduce the cost of innovation by removing the need for innovators to secure consent to access data at the early and most uncertain stages of service development.

6. Concluding remarks

The analysis here shows that smart meter data could: (a) add significant value to existing public interest use-cases (e.g. national/sub-national energy stats, representative dataset for analysis and modelling) and; (b) enable new use-cases to be realised which are currently unfeasible or costly (e.g. local energy system planning, intervention targeting, and test-bed for development of new services).

By understanding the data attributes required to meet each archetype public interest use-case , our analysis also demonstrates that each use-case can be realised with ***Output*** datasets which would not risk any compromise of individual privacy.

However, as already emphasised, all new Output datasets created to meet the use-cases we have identified each require Inputs which would involve the data capture and processing of individual smart meter records (and, for some, data-matching with other sensitive data at individual record level). This suggests that realising these public interest use-cases would need a 'trusted processor' to create these Output datasets.

There is no such 'trusted processor' in the current arrangements for smart meter data access and privacy. The PIAG will therefore consider in a later paper what this role/function might involve, how it might be realised, and the basic characteristics required of any such processor (e.g. secure systems etc). It will also consider associated data access and privacy concerns and how these could be addressed.

The PIAG will also consider in due course (a) whether the Output datasets are made available publicly or accessible only by those who pass a clear 'public interest' test (and whether such rules would differ by use case) and (b) the capabilities (from IT systems to data analysis skills) which public interest actors will need to be able to access to make effective use of the Output datasets for each use-case.