Smart Meter Energy Data:
Public Interest Advisory Group

A policy dialogue and work programme led by
Centre for Sustainable Energy & Sustainability First

Stimulus paper 8

Capability requirements of 'public interest' data user organisations

Author : Nicky Hodges. CSE

with support from

Simon Roberts and Josh Thumim. CSE

& Judith Ward and Maxine Frerk. Sustainability First

Status of this Draft

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

An earlier version was discussed with PIAG members at their workshop on 4 April 2019.

Last revise – 01 May 2019

PIAG paper 8

Capability requirements of 'public interest' data user organisations

Introduction

The aim of this paper is to explore what organisational capabilities would be required of a 'public interest' user to make use of smart meter energy data for a 'public interest' purpose.

This includes the resource and capability – tangible and intangible – that an organisation would need in order to manage their data requirements responsibly, where anonymised smart-meter data was a key input. The paper therefore looks at requirements common to public bodies handling data-sets which include some component of personal data: technology capability, culture, motivation, specialised skills and expertise. The paper is informed by our earlier PIAG papers[1], and in particular draws on earlier stakeholder interviews for paper 4 about potential 'public interest uses' of smart meter data.

Public interest use cases identified in paper 5 are:

- Improved national energy statistics.
- Better data to support regional and local-level energy-system planning.
- Better evidence to support policy-makers in modelling the impacts of new policies.
- Access to higher resolution data to enable better development of services.

Paper 5 also drew a clear distinction between the **input data** required for each public interest use-case and the resulting **output datasets**, which would become available for use by policy-makers, public interest organisations and others once suitably aggregated and anonymised by a trusted processor.[2]

In this paper, we only consider potential access to, and use of, output datasets by public interest users.

This in turn means that the capability requirements discussed are relatively contained. The data-handling requirements for any 'trusted processor' role are of course different and self-evidently far more demanding.

[1] General landscape for smart meter data access: privacy (paper 1); ethics (paper 3). Identification of key 'public interest' use cases (paper 5). (Possible routes to the data (paper 7).

[2] This distinction between Input data and Output data is explored in more detail on page 3 of Stimulus Paper 5

Snapshot: What stakeholders told us on data-handling capabilities

Interviews to inform paper 4 drew out some stakeholder reflections on the organisational capability requirements for public interest users to realise use cases. Stakeholders identified a number of organisational strengths, weaknesses, opportunities and threats to their ability to use smart meter data for public interest purposes, as well as revealing stakeholder's own uncertainties about the capabilities required. These are briefly summarised below.

Elected city and regional bodies: This refers to Mayoral Offices, city councils and combined authorities, which already manage large data-sets with potentially sensitive or personal data. They also have the weight to influence the future energy transition and an appreciation of the potential value afforded by using smart meter data alongside other data streams to inform significant infrastructure decisions: from planning of heat networks and siting of renewable energy to energy storage assets and EV charging provision. These elected bodies also have an interest in how smart meter data could improve the accuracy of their models for carbon reduction and the targeting of energy efficiency services locally.

However, elected city and regional bodies saw some barriers to their engagement with energy sector incumbents. They described the energy and the built environment sectors as lagging behind other sectors such as health, education, and transport, in their engagement with a wider 'smart cities' agenda. They also cited little in-house recognition of a commercial case for participation in the energy sector and for investing in the necessary data-management capabilities.

Other public bodies: Stakeholders highlighted the lack of a data-driven culture and some public bodies did not currently make best use of existing energy data to which they have access. Such bodies would need support to develop a more data-led culture – including technical capabilities – before they could derive value from access to smart meter energy data.

Other practical barriers identified by stakeholders included resource constraints, access to specialist skills, and technology. Outsourcing was considered one route to addressing some constraints. Confidence in local authority in-house data skills varied. That said, public sector stakeholders generally felt confident in handling personal or sensitive data.

New entrants: Compared to elected city and regional bodies, stakeholders judged new entrant innovators to be disadvantaged by their lack of access to smart meter data (in contrast to energy sector incumbents), though examples were noted of new entrants developing smart energy services using precisely the sorts of capabilities which are needed to handle and create value from such data.

This 'snapshot', based on a limited number of stakeholder interviews about potential uses for smart meter data[3], suggests a mixed picture in terms of resource, will, culture, technical capability and

[3] The interviews were principally focused on potential public interest use-cases for smart meter data. The interviews were not seeking to assess the capabilities of a full range of potential users but interviewees did raise some of the capabilities and barriers that would influence likely use-cases.

financial capacity among potential 'public interest' users of smart meter data; though leading city authorities would appear to be relatively well-positioned.

Data management capabilities

Against this short background, the remainder of this paper focuses on what capabilities would be needed for users to make good use of smart meter **output** data should this become available for public interest purposes. In so doing it:

- Briefly identifies the range of organisations that may want to use smart meter data for public interest purposes.
- Draws on smart city, big data and open data literature understanding of the capability requirements to generate public value outputs from data.
- Discusses the different capability – and associated capacity, technology and connectivity requirements – for organisations to use the data for public interest purposes.
- Provides a set of 'example uses' of smart meter data to illustrate the associated skills, knowledge and technology requirements.
- Notes a number of initiatives to develop relevant user capabilities within the smart city and open data spheres.

Overview of potential public interest users

Organisations that may want to use smart meter data for public interest purposes include: government departments[4], statutory and non-statutory advisory bodies[5], regional bodies, local authorities, community energy groups, intermediary energy organisations, universities and other research organisations, housing associations, and new entrants wishing to develop new services.

Organisational capabilities

The wider organisational capabilities for good decision-making using smart meter output data produced by a trusted processor are likely to be the same as those for smart city or open data uses – see Annex 1 for a summary of relevant academic literature. These can be summarised as: an enabling organisational culture, governance capabilities, legal compliance, collaboration, personnel with appropriate skills, decision-making quality, and appropriate technical infrastructure.

Legal compliance capabilities to satisfy privacy protection, security and data ownership regulations will be important considerations, particularly for more advanced uses of smart meter data. Users will need to develop a compliance strategy which is proportionate to risks associated with the granularity of data and proposed techniques for data analysis.

[4] This is understood to include government departments with responsibilities for energy and climate change, for housing and planning policy, health.
[5] Examples include Committee on Climate Change, Committee on Fuel Poverty, Energy Saving Trust

For public bodies facing financial constraints and the pressure of multiple competing demands, the strategic fit and resource allocation will have considerable importance. How an organisation wants to use smart meter data will influence what scale of resource is required.

The next section of this paper specifically focuses on the knowledge, skills and technology requirements for a number of example smart meter data use cases.

Knowledge, skills and technology requirements for using smart meter data

A typical PIAG public interest use-case which draws on smart-meter output data would require general organisational capabilities for good decision making and relevant energy domain understanding to interpret the data. Generic problem solving and analysis skills would be required to support use cases where domestic energy consumption statistics are used in combination with other forms of data.

More advanced expertise in data management, problem solving and research methods would be required for use cases involving large representative or anonymised aggregated output data as an input to more complex analysis or modelling. In such cases, the following skills, including use of types of software tools are likely to be relevant:

- Statistics.
- Energy system domain technical knowledge.
- Energy system modelling.
- Computer programming.
- Data handling.
- Use of data management and analysis tools (e.g. SQL, HiveQL, R, Python, Java).
- Use of spatial data analysis tools (e.g. Quantum GIS, Spatialite, PostGIS).
- Research and evaluation skills.
- Cooperation competencies and other ‘soft skills’.

Requirements may include specialised technologies to integrate and analyse very large volumes of data. Examples here include Hadoop, and other “NoSQL” tools. For very large scale uses, investment in large scale hardware may be needed (Kim, Trimi and Chung, 2014). However for cases which are smaller scale, the data analysis and storage capacity required may be much more limited and not require special tools.

Under the proposed PIAG model to firmly separate **input data** from **output data,** the **output datasets containing smart meter data joined to** other household-level or neighbourhood level data would be anonymised before release. Even so, appropriate security technology and procedures would be important to guard against residual risks of re-identification.

It is important to note that smart meter data is not inherently special in terms of the sorts of data processing skills and technological requirements to which it gives rise. It is also not necessarily ‘big’ in the context of the conventional discourse of ‘big data’. It comprises essentially:

- A unique meter-level identifier (e.g. MPAN for electricity, MPRN for gas).
- A time series of consumption values (typically half-hourly but potentially also peak demand).

Stored in a comma-separated text file, this would look something like:

MPAN, date time, kwh
123456789012, 2015-01-01 08:30, 1.234
123456789012, 2015-01-01 09:00, 0.123
… etc

But many of the data outputs released by the trusted processor would no longer look like this, having been transformed into summary statistics, as a single figure, a simple table or a set of related simple tables.

For example:

- National average domestic gas consumption – a single figure for GB.
- Weekly national peak domestic electricity demand over a year – 3 columns x 52 rows output table.
- Domestic gas and electricity consumption by local authority / by year / by season / trends over time: several tables which could be stored in separate worksheets within a single spreadsheet file.

The users of such statistics would simply need to understand them, and follow general good practice in interpreting and presenting statistics for their intended purpose, be that, for example, in a briefing to a Minister, proposals to use battery storage to help manage local peak demand, or a local authority evaluation of how far investment in domestic energy efficiency programmes has contributed towards emissions reductions.

Scale and complexity for more advanced uses

The specific capability and technology requirements for more advanced uses will vary depending on the context, but the following two dimensions are relevant:

1. Scale of data inputs required to address question.
2. Complexity of methods required to address question.

There are expected to be around 50 million smart household and SME energy meters in the UK. This itself does not qualify as 'big' data – however scale can become an issue when the consumption time series are taken into account.

For example, a year's worth of half-hourly consumption for all 50 million meters would comprise:

50×10^6 meters x 17,520 half-hours/year = 876 billion records

As illustrated above, each record would at a minimum contain the meter identifier, the timestamp, and the consumption in the half-hour period associated with the timestamp. A single record would require approximately 40 bytes to store in text form (each ASCII character requires a single byte). A single household's gas and electricity meter readings would therefore require approximately 1.4MB (the capacity of a 3.5" floppy disk!), while the whole country would require 35 TB.

This latter figure is large enough that specialised tools would be required to store and query the data. The data would get bigger again with other data (e.g. household and neighbourhood attribute data) linked to the smart meter data prior to anonymisation by the trusted processor.

However, in PIAG's proposed approach, the trusted processor would undertake this work, generating much smaller size, anonymised **outputs,** made available to public interest users. Only in certain use cases would public interest users need to have specialist data management, modelling and technology requirements.

Example 1: Build a model for Commission on Fuel Poverty (CFP) to enable distributional analysis of the impact of time-of-use tariffs on different households.

This would enable the CFP to understand the potential fuel poverty impacts of the growth in time-of-use tariffs, so that it can act to influence associated government policy and regulation.

Data output from trusted processor: One year's worth of smart electricity meter data for a nationally representative sample of 10,000 households, linked to household and neighbourhood attribute data, as an anonymised dataset.

Skills / technology requirements: Problem solving skills, energy domain expertise, database and SQL skills, and a programming language such as Python or R.

Model design & output: Applies TOU tariffs to the sample dataset, and reports on which types of households are made better and worse off relative to a base case.

Example 2: Design a city-wide, high resolution, spatial analysis of demand profiles for heat and power for city planning and investment body.

This would inform strategic planning and investment decisions to support clean growth.

Data output from trusted processor: Half-hourly postcode-level demand profile for a city, covering an entire year (or an average over multiple years), with no associated household data. For Greater London, approximately 200GB.

Skills / technology requirements: Problem solving skills, energy system technical knowledge, modelling skills, reasonably high performance relational database server, programming language such as Java or Python.

Database / analysis output: The identification of which different heat-power system couplings are most appropriate given system constraints in different parts of the city. Examples of such couplings include:

- Heat networks with CHP.

- Individual building heat pumps.
- Centralised heat pumps serving a heat network.
- Electrical thermal storage serving a heat network.

The set of example uses above show that there is a wide range in the skills and technology requirements for public interest use cases. In many cases, ‘public interest’ users of smart meter data would rely on statistical outputs or ‘ready-made’ analytic outputs by others (national or specialist) bodies. For such users, the skills and technology requirements are likely to be less important than more generic organisational capabilities. It is only in certain cases that more advanced capabilities will be required.

Building capabilities

A range of initiatives have sought to build the capabilities of public bodies and other organisations to use open / big data for public benefit, including:

- Business and Local Government Data Research Centre (BLGDRC) offers capacity building services and training in analytic techniques.
- Catapult Future Cities has produced a toolkit (2018) which sets out a high level process for organisations to share city level data, covering: building a business case, updating legal, privacy and ethical approvals, updating technology, confirming governance arrangements, and supporting a sharing culture.
- The Open Data Institute has, amongst other things, produced a data roles framework to help people identify what combination of skills they need to play their part in engaging with data.
- Various corporate entities, such as IBM, have developed “smart city platforms” and engaged with regional / local communities to develop opportunities to use aggregated data for public (and presumably private) benefit.

Conclusion

The smart city, open data and big data agendas have generated recent new thinking and insights around the capability requirements of organisations wanting to use data for public good, much of which is transferable to the use of smart meter data for a public interest.

This includes recognition that capability strengthening may need to extend across different levels of an organisation, to encompass leadership and decision-making, organisational culture, governance and legal compliance systems, as well as technical knowledge, skills and technology.

Whilst there will be some domain expertise requirements, the technology and data skills and capabilities are likely to be generic – there is nothing specific to smart meter data that is likely to create obstacles to success. That said, the scale and complexity of the intended use of smart meter data (which as we have seen can vary significantly by use case) will shape the specific technology and analytical requirements.

A range of initiatives exist to support 'public interest' organisations to understand and address their capability requirements. Local authorities active in smart city or other open data initiatives are likely to have already engaged with many of these challenges.

For some organisations, in-house resourcing constraints may require them to outsource the work to realise use cases. However, the publication of official statistics and analysis outputs by specialist bodies would enable organisations with limited specialist skills or technology to make use of smart meter data for public interest purposes.

Annex 1 : Smart city, big data and open-data literature on the capabilities required to generate value

The literature on big data, open data and on smart cities has included consideration of how to achieve economic, social, cultural, political or ecological value of benefit to citizens and communities (Benington 2011, in Jetzek, Avital and Bjørn-Andersen, 2013), and outlined the activities typically involved and the requisite capabilities for doing so.

Jetzek et al (2013) provide helpful definitions of capability and technical connectivity, which can be applied to thinking about public interest uses of smart meter data:

- Capability: the ability of an organisation to use data to achieve value.
- Technical connectivity: the availability of technologies that allow public interest users to store, access, combine and analyse the data. This includes the "*availability of the infrastructure and use of semantic technologies, as well as data management; analytics and discovery software, plus the use of multiple platforms to enable general access to content.*"

Klievink et al (2017) present a flow diagram of the main activities in the process of big data analysis[6], as well as a more detailed list of the activities involved:

FIGURE 1: MAIN STEPS IN BIG DATA ANALYSIS PROCESS

TABLE 1: LIST OF ACTIVITIES INVOLVED IN BIG DATA ANALYSIS

	Big data activities
Collection	Collect, annotate Acquire, record Generate Choose, select Sense
Combination	Extract, clean, prepare, process Combine Organize Store Integrate, represent
Analytics	Analyse, model Visualise Interpret
Use	Initiate Deploy Make decisions Apply, produce insight Evaluate

6 These steps apply equally to data of any size, so the word "big" is somewhat redundant, as well as being subject to changing definition over time, as available computing power increases, and data storage costs fall.

Arguably, the identification of the purpose of (smart meter) data analysis and corresponding data requirements and analysis design requirements should be appended to the top of this list, with problem-solving skills being a key generic skill requirement.

Open / big data literature recognises that there is a broad range of organisational capabilities that are required to make use of big data to generate public value (Kleivink et al 2017; Jannsen et al 2017; Pappas et al 2018) and that data skills and technology resources alone are not sufficient.

Kleivink et al (2017) set out a summary table of organisational capabilities required for big data use, as reproduced in Table 2.

TABLE 2: ORGANISATIONAL CAPABILITIES FOR BIG DATA USE (KLIEVINK ET AL 2017)

Capability	Explanation
IT governance	Capability to design and develop IT strategy, decision-making and responsibility structures, supporting the organization, including integration of new IT systems
IT resources	Capability to design, develop and maintain suitable IT infrastructure and expertise to facilitate current and new IT systems
Internal attitude	Capability to develop internal commitment and vision for new processes and systems, especially openness towards data-driven decision-making
External attitude	Capability to develop external commitment and support for new processes and systems with important stakeholders
Legal compliance	Capability to design and develop a compliance strategy including process design, monitoring and redesign of processes, especially regarding privacy protection, security and data ownership regulations
Data governance	Capability to design and develop a data strategy including collection, acquisition, quality control and data partnerships
Data science expertise	Capability to bundle/acquire, develop and retain data science knowledge in the organization, especially bundling knowledge on IT, business, statistics, computer science and mathematics

Having an enabling organisational culture is widely identified as important, including alignment between the organisation's strategic priorities and the proposed data purpose, having a culture of organisational learning and a 'data-driven culture' (Pappas et al).

With the appropriate organisational alignment and governance, organisations can more confidently focus on the capabilities and technology requirements. Janssen et al (2017) focus particularly on the factors influencing the quality of decision making using big data. Again, this highlights how the mechanics of data analysis needs to be supported by wider organisational capabilities.

TABLE 3: FACTORS INFLUENCING QUALITY OF DECISION MAKING (JANNSEN ET AL 2017)

Factors	Description
Contractual governance	Agreements and contracts with big data providers to increase data quality, ensure mutual understanding, create clear responsibilities and procedures and improve communication.
Relational governance	To build trust and ensure the sharing of relevant knowledge necessary to interpret big data
Big data analysis capabilities	This includes finding and selecting the right tools for analysis of big data and how to visualise big data. Domain knowledge, big data and analytic skills.
Knowledge exchange	Transfer of knowledge about how data is collected and processed to aid the finding of patterns and relationships.
Collaboration	Collaboration between data producers, analysts and users can create a big data chain and enable the creation of valuable applications
Process integration and standardization	This results in lower efforts and costs to use big data and big data analysis.
Routinizing and standardization	This can speed up big data analysis and decision making
Flexible infrastructure	This determines the ability and the amount of effort needed to handle and process data. Systems integration can improve the handling of big data, reducing lead times for results.
Staff	Specialists able to handle big data, with analysis skills and able to communicate with users are scarce. Partnerships with external businesses may be needed to address in-house skills gaps.
Decision maker quality	More experienced decision-makers can make better and faster decisions. They need to be able to interpret the analytic outcomes.

Kleivink et al conclude by identifying areas where public bodies may fruitfully focus their efforts to improve their ability to generate value from big data:

- Identify how specific data applications align with an organisation's strategy, structure and main activities.
- Collaborate with relevant organisations to enable more accurate and informative insights.
- Focus on developing data science expertise, data governance and IT governance capabilities.

Pappas defines the skills requirements of personnel involved in data analytics and decision making using big data to include: computer programming, research methods, data handling, visualization tools, soft skills, domain knowledge, and strong cooperation competencies.

Both Pappas et al and Kleivink et al encourage organisations to start small to allow organisational learning, iteratively develop their analytic capabilities, build trust and minimise risks.

The findings from this smart city, open data and big data literature are likely to be transferable to smart meter data, though it will only be organisations wanting to undertake more advanced modelling, testing of algorithms or distributional analysis use cases that are likely to require specialised technology and data analysis capabilities.

References

- Janssen, M., van der Voort, H. and Wahyudi, A. (2017) Factors influencing big data decision-making quality.
- Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2013) *Generating Value from Open Government Data*.
- Kim, G.-H., Trimi, S. and Chung, J.-H. (2014) ‘Big-data applications in the government sector’, *Communications of the ACM*. doi: 10.1145/2500873.
- Klievink, B., Romijn, B-J., Cunningham, S. and de Bruijn, H. (2017) Big data in the public sector: uncertainties and readiness.
- Pappas, I. O. *et al.* (2018) ‘Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies’, *Information Systems and e-Business Management*. doi: 10.1007/s10257-018-0377-z.