Smart Meter Energy Data: Public Interest Advisory Group

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

PIAG Workshop 1 – 19 March 2018

Stimulus paper 1

Background to ICO Guidance on Anonymisation - & annex on data access privacy legal framework

Author: Maxine Frerk

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.

Following the workshop on 19 March a short annex has been added on the Digital Economy Act 2017

Last revise – 10 August 2018

1. Introduction

The basic privacy framework governing access to smart meter data comprises the Data Protection Act (DPA), being superseded in May by the General Data Protection regulation (GDPR) and the smart meter specific Data Access and Privacy Framework (DAPF). These were discussed briefly at the PIAG kick-off meeting and a recap is provided in the Annex to this paper.

The Information Commissioner (ICO) is responsible for oversight of the DPA / GDPR and produces guidance on a range of issues. Of particular relevance from PIAG's perspective is the guidance produced on anonymisation[1] back in 2012. This paper draws out the salient points but the full guidance includes more detail in particular on technical options.

The guidance is based on the premise that underpins PIAG – that open data can be of real public benefit but that care is needed to maintain privacy protections. The guidance sets out the range of factors that the ICO has considered in cases where these tensions arise and hence provides valuable insights for thinking about access to smart meter data

"*If we assess the risks properly and deploy it in the right circumstances, anonymisation can allow us to make information derived from personal data available in a form that is rich and usable, whilst protecting individual data subjects*".

2. Overview

The EU Data Protection Directive (from which the DPA stems) makes clear that the principles of data protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.

The test set out in the DPA is that this means that there is not a "reasonable likelihood" of the individual being identified from the anonymised data. The ICO recognises that it is impossible to be 100% sure that re-identification cannot happen and courts have taken a balanced approach. The guidance captures the sorts of factors to take into account in weighing the risks.

The GDPR on the face of it introduces a higher hurdle around the use of anonymised data but it still refers to considering "all the means likely to be used" and that time and cost can be taken into account. Whether this represents a fundamental difference in practice will be dependent on the court's interpretation of the legislation. However the ICO has said that the current guidance on anonymisation is consistent with GDPR.

Given the test is essentially a risk-based assessment the guidance sets out the sorts of factors that should be considered in any Privacy Impact Assessment. These are summarised below.

[1] https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

3. Key considerations

The motivated intruder test

In thinking about the risks of data being re-identified the ICO advocates the use of a "motivated intruder test". The idea is that in judging the risks of re-identification you should look at it through the lens of someone who has a motivation to link an individual and their data (which could be for a variety of reasons) and uses a range of public information to do so. However they are not assumed to have particular technical expertise such as computer hacking skills.

The 'motivated intruder' test sets the bar higher than considering whether a 'relatively inexpert' member of the public can achieve re-identification, but lower than considering whether someone with access to a great deal of specialist expertise, analytical power or prior knowledge could do so.

In developing the Data Access and Privacy Framework for smart metering it was acknowledged that there could well be individuals who were motivated to undermine the rollout by demonstrating that smart meter data was insecure. The concept of a motivated intruder thus has clear relevance to smart metering.

The ICO also advocates "penetration testing" – attempting to re-identify personal information using the sorts of information that would be available to a motivated intruder.

Forms of anonymisation

The term 'anonymised data 'is used to refer to data that does not itself identify any individual and that is unlikely to allow any individual to be identified through its combination with other data.

There are in essence two approaches that can be taken to anonymisation:

- Pseudonymisation (or de-identification): where individual records are retained but the link to the individual is removed. In some cases an alternative identifier may be used that is only known to the originator of the data but that may help with data linking in future[2];
- Aggregation: where individual records are combined to present statistical information.

In terms of re-identification, pseudonymisation carries the greatest risks, including where personal knowledge (eg that a neighbour was on holiday for a certain period) would allow the individual's data record to be identified. The ICO acknowledges that for some purposes individual level data is needed but where this is to enable the provision of services for example, consent may be the best route.

The risks with aggregated data arise if aggregation is done on a number of different bases which mean that by comparing sources "jigsaw re-identification" can occur.

[2] Pseudonymisation is defined in the GDPR but was not mentioned in the previous directive. The GDPR encourages the use of pseudonymisation as a general security measure. By holding the de-identified data separately from the "additional information," the GDPR permits data handlers to use personal data more liberally without fear of infringing the rights of data subjects.

Clearly the risk of combining information to produce personal data increases as data linkage techniques and computing power develop, and as more potentially 'match-able' information becomes publicly available. A judgment therefore has to be made as to the level of risk involved.

Publication versus limited disclosure

The ICO draws a clear distinction between the publication of data and providing data to a closed community, for example for research purposes, where the use of the data can be controlled eg by contract.

Clearly the risks are higher with publication of data.

The ICO concludes that in general data for publication should be aggregated. However pseudonymised data may be made available within a closed community where more protections can be put in place governing how the data is to be used:

"In general, the more detailed, linkable and individual-level the anonymised data is, the stronger the argument for ensuring only limited access to it. ….The more aggregated and non-linkable the anonymised data is, the more possible it is to publish it".

Where data is being made available on a limited access basis then safeguards can be put in place covering for example the uses of the data, training of staff, encryption and physical security, restrictions on disclosure or copying, prohibition of re-identification, and penalties for breaches

For aggregated data, governance arrangements are still needed to oversee the different levels of eg geographic aggregation to minimise the risk of jigsaw re-identification.

Other considerations around risk

As part of a risk-based approach the ICO also encourages thought to be given to the level of detriment that could result in the event that data is re-identified. Clearly the more sensitive the data the more risk averse one should be.

Some specific factors that the ICO suggests should be considered are:

- That historic data is typically less sensitive than real time data (especially old historic data);
- That samples from an individual's records will typically be less sensitive than the full record;
- Banding or "blurring" of data can reduce its sensitivity and the risk of re-identification.

If aggregating across geography then the level of postcode is relevant. For example full postcode equates to approx 15 households (although some postcodes only relate to a single property). The postcode minus the last digit equates to approx 120/200 households

Any organisation involved in the anonymisation and disclosure of data, should have an effective and comprehensive governance structure in place that will address the practical issues surrounding the production and disclosure of anonymised data. The ICO encourages organisations to be transparent about what they do in terms of anonymisation.

Practicalities around consent

The guidance is explicit that consent is not needed to legitimise the anonymisation of data (although it could be viewed as a form of processing). In part this is an argument about practicality but primarily it reflects the fact that anonymisation could not be said to cause unwarranted damage or distress (the only basis on which processing can be prevented under the DPA, although that changes with GDPR). That said the guidance does talk about the need to consider the basis on which the data was originally collected and that if very restrictive commitments were given as to future usage then these should be respected.

Moreover, if there is any prospect of re-identification then the individual's consent should be sort.

The DPA provides an exemption for the processing of data for research or statistical purposes.

Trusted third parties

The guidance discusses the role of trusted third parties. A trusted third party is an organisation which can be used to convert personal data into an anonymised form. This is particularly useful in the context of research, as it allows researchers to use anonymised data in situations where using raw personal data is not necessary or appropriate. Trusted third parties can be used to link datasets from separate organisations, and then create anonymised records for researchers. Examples of how this is done in the health sector are included in the PIAG stimulus paper on data ethics.

Further guidance

The ICO guidance document contains a range of case studies and further details on different anonymisation techniques. This includes examples such as cryptographic hash techniques to support pseudonymisation and data perturbation (eg adding noise into the data to reduce the sensitivity). We may wish to explore these technical options later in the PIAG programme.

An Anonymisation Network led by a number of academics has been established to explore issues in greater depth. This has its own website and has produced a handbook[3]. It may be worth engaging with this group as PIAG thinking develops.

[3] http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf

Annex 1: Smart Metering Privacy Legislative Framework

Data Protection Act[4]

The Data Protection Act 1998 (DPA) is concerned with 'personal data'. It says that 'personal data' means: data which relate to a living individual who can be identified— (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller. The DPA sets out a number of principles:

1. Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless –
   (a) at least one of the conditions in Schedule 2 is met, and
   (b) in the case of sensitive personal data, at least one of the conditions in Schedule 3 is also met.
2. Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.
3. Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed. This links with the concept of "data minimisation". An organisation should identify the minimum amount of personal data they need to properly fulfil their purpose. They should hold that much information, but no more.
4. Personal data shall be accurate and, where necessary, kept up to date.
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes. The GDPR provides specific exemptions to this around statistical data as noted below.
6. Personal data shall be processed in accordance with the rights of data subjects under this Act.
7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.
8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

The conditions for processing are set out in Schedules 2 and 3 to the Data Protection Act. Processing, in this context means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data. Unless a relevant exemption applies, at least one of the following conditions must be met whenever you process personal data:

- The individual whom the personal data is about has consented to the processing.
- The processing is necessary in relation to a contract which the individual has entered into or because the individual has asked for something to be done so they can enter into a contract.
- The processing is necessary because of a legal obligation that applies to you (except an obligation imposed by a contract).
- The processing is necessary to protect the individual's "vital interests". This condition only applies in cases of life or death;
- The processing is necessary for administering justice, or for exercising statutory, governmental, or other public functions
- The processing is in accordance with the "legitimate interests" condition.

[4] https://ico.org.uk/for-organisations/guide-to-data-protection/

General Data Protection Regulation (GDPR)[5]

The General Data Protection Regulation comes into force on 25 May 2018. As a Regulation it has direct effect but the Data Protection Bill 2017-19, which has just concluded its second reading in the Commons, will formally implement the GDPR by clarifying how certain derogations apply in a GB context, establishing the ICO as the relevant national body and repealing the DPA. The Bill also covers other data related issues around immigration and the press. The House of Commons library briefing[6] provides a good overview including on the impacts of Brexit.

The definition of personal data and the lawful basis for processing information have not changed materially under the GDPR.

The data protection principles are revised but are broadly similar to the principles in the DPA: fairness, lawfulness and transparency; purpose limitation; data minimisation; data quality; security, integrity and confidentiality. A new accountability principle makes controllers responsible for demonstrating compliance with the data protection principles.

There is a new (or arguably re-framed) condition for processing around performance of a public task where processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. The public task must be set out in law.

Personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes shall not be considered incompatible with the original processing purposes

The key changes of most relevance to smart metering are probably around consent and the right to data portability. There are also now explicit references to pseudonymisation.

Consent

The GDPR is clearer that an indication of consent must be unambiguous and involve a clear affirmative action (an opt-in). It specifically bans pre-ticked opt-in boxes. It also requires individual ('granular') consent options for distinct processing operations. Consent should be separate from other terms and conditions and should not generally be a precondition of signing up to a service.

The GDPR has references to both 'consent' and 'explicit consent'. The difference between the two is not clear given that both forms of consent have to be freely given, specific, informed and an unambiguous indication of the individual's wishes. Explicit consent however may require a positive opt-in or declaratory statement.

[5] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/

[6] http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-8214#fullreport

Although the higher thresholds for consent have been flagged as a concern in the smart metering context, the ICO guidance makes clear that:

- Consent is one lawful basis for processing, but there are alternatives (as set out above). Consent is not inherently better or more important than these alternatives.

- Consent is appropriate if you can offer people real choice and control over how you use their data, and want to build their trust and engagement. But if you cannot offer a genuine choice, consent is not appropriate.

- Public authorities, employers and other organisations in a position of power over individuals should avoid relying on consent unless they are confident they can demonstrate it is freely given.

The ICO observes that confusion can arise as to whether individuals have to be given a choice and have to agree to their data being used in a particular way. In a strict data protection sense, the law generally provides alternatives to individual consent for data usage. In the ICO's view policy makers need to be much clearer as to whether they are giving people a choice, or whether they are going to go ahead without consent – or even in the face of objection – because it is in the public interest to do so.

From a smart metering perspective, the importance of consent and giving consumers a choice over the use of their data is driven as much by the need to encourage take-up of smart meters as it is by strict privacy and DPA concerns.

Data Portability

The right to data portability allows individuals to obtain and reuse their personal data for their own purposes across different services.

Data controllers must provide the personal data in a structured, commonly used and machine readable form. Open formats include CSV files. Machine readable means that the information is structured so that software can extract specific elements of the data. The information must be provided free of charge.

If the individual requests it, the data controller may be required to transmit the data directly to another organisation if this is technically feasible.

The Midata arrangements in GB effectively meet these requirements.

Pseudonymisation

The GDPR includes explicit reference to pseudonymisation. As noted above it is seen primarily as a way of minimising security risks where companies are holding data over a long period for example.

The GDR instructs controllers to implement appropriate safeguards to prevent the "unauthorised reversal of pseudonymisation." To mitigate the risk, controllers should have in place appropriate

technical (e.g. encryption, hashing or tokenization) and organisational (e.g. agreements, policies, privacy by design) measures separating pseudonymous data from an identification key.

The GDPR also talks more generally about the reidentification risk and the need to consider whether a method of reidentification is "reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly." Such an analysis is necessarily contextual and "account should be taken of all the objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."

Smart Metering Data Access and Privacy Framework (DAPF)

The DAPF was put in place by government to provide clarity around the arrangements for smart meter data and to build confidence in the smart meter rollout. Government has made clear that consumers will not be forced to have a smart meter and for that decision to be meaningful consumers need to be clear how their data will be used. Changing the arrangements subsequently raises potential concerns (which the ICO has flagged) about the basis on which consumers who already have smart meters agreed to the installation.

The basic principle (which is reflected in the Article 29 working party opinion on smart metering) is that consumers should have a choice about the use of their data except where that data is required for regulatory purposes. Guidance on Privacy Impact Assessments for smart metering issued by the A29 Working Party[7] made clear that some of the issues with smart meters were about wider human rights not just privacy.

The DAPF is focussed on consumption data and is geared towards domestic consumers although microbusinesses are also within scope.

In developing the DAPF DECC carried out a privacy impact assessment[8] which explored the nature of the risks involved and which provides a helpful starting point for what an updated smart metering PIA might cover.

The DAPF sets out the arrangements for supplier access to data in that:

- Suppliers are entitled to access monthly data for billing purposes;
- Suppliers are entitled to daily data for any purpose other than marketing but consumers have a right to opt out;
- Suppliers can access half-hourly data (or other data for marketing) only if the customer opts in.

[7] http://www.garanteprivacy.it/documents/guest/normativa%20internazionale/Articolo%2029/WP209%20Opinion%2007%202013%20Opinion%20smart%20grids.pdf

[8] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/43044/7226-sm-privacy-ia.pdf

Ofgem are currently considering the options for enabling half-hourly settlement which it was always acknowledged would need a change to these rules.

Under the DAPF Distribution Network companies can access half-hourly data for regulated purposes if they can satisfy Ofgem that the data will be suitably anonymised / aggregated.

Third parties can access the data with customer consent. However they also need to become SEC (Smart Energy Code) parties and meet stringent security requirements. In particular, third parties will be required through the Code to:

- Take steps to verify that the request for third party services has come from the individual in question;
- Obtain explicit (opt-in) consent from consumers before requesting data from the DCC; and
- Provide reminders to consumers about the data that is being collected.

To ensure compliance with these requirements, third parties will be subject to audit arrangements.

The Privacy Impact Assessment makes clear that the DAPF will be subject to review:

"As the roll-out of smart meters continues consumers will become more familiar with how smart metering data will be processed and used, therefore their attitudes and perceptions may change. In addition, the data access and privacy framework will evolve to take account of learning and best practice"

BEIS are undertaking a review of the DAPF to ensure it remains fit for purpose.

As part of the Privacy Impact Assessment, DECC discussed the use of data for its own monitoring and evaluation purposes. It also published a separate document on its approach to monitoring[9]. This focussed on monitoring progress around the rollout but also discussed the benefits of eg quarterly consumption data to help in evaluating energy efficiency savings. It left open the options of how this would be achieved but noted that one option was to use the Statistics for Trade Act 1947 which is the basis on which it currently collects annual data for the National Energy Efficiency Database (NEED).

The processing of personal data by appropriate authorities is permitted under Schedule 2 (5) of the Data Protection Act, if it is deemed to be necessary for administering justice, or for exercising statutory, governmental, or other public functions exercised by any Government department if it is in the public interest. It is important to note that data that is produced by conventional meters is currently legitimately accessed by public bodies.

[9] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/43136/7206-gov-resp-cons-sm-monitor-evaluation.pdf

Annex 2: Digital Economy Act 2017

ONS Access to Information from Public and Private Sectors

The Digital Economy Act 2017 amends the Statistics and Registration Service Act 2007 to provide the UK Statistics Authority and its executive office, the Office for National Statistics (ONS), with greater and easier access to a range of data sources held within the public and private sectors.

The ONS can only seek access to data for the purposes of fulfilling one or more of its statutory functions, including to produce official statistics and undertake statistical research that meets identifiable user needs for the public good.

The Act requires the ONS to set out principles that it will adhere to in exercising its new powers. It has done this and the principles cover confidentiality of data, transparency, ethics and the law, public interest, proportionality and collaboration.

Disclosure of information for research purposes

The Digital Economy Act 2017 also facilitates the linking and sharing of datasets held by public authorities for research purposes. To ensure data are processed and made available in a safe and secure way the legislation sets out six conditions under which this can take place.

In particular it creates a gateway to enable public authorities to make data available to researchers for research that is in the public interest using a trusted third party model. Under this model, a data holding public authority discloses identifiable data to an accredited third party processor (or the public authority itself acting in this capacity), who is then responsible for processing the data (that is, linking, de-identifying, storing, making data securely available or related procedures) before the de-identified data are made available to an accredited researcher.

Researchers must be accredited (but do not have to come from academic institutions) and projects must be accredited as being in the public interest.

The ONS is required to produce a code of practice which it has done and which includes examples of research in the public interest as being to:

- Provide or improve evidence bases that support the formulation, development or evaluation of public policy or public service delivery;

- Guide critical decision-making with anticipated impacts on the UK economy, society or quality of life of people in the UK;

- Significantly extend existing understandings or social or economic trends or events, either by improving knowledge or challenging accepted analyses; or

- Replicate, validate or critically analyse existing research (including official statistics) in a way that leads to improvements in the quality, coverage or presentation of existing research.