ICO consultation on the draft anonymisation code of practice.

Open Rights Group response. August 23 2012

For more information contact Peter Bradwell, peter@openrightsgroup.org

The consultation took the form of a series of questions about a draft code. Below are our answers.

More detail on the consultation can be found at the ICO website[1].

1. **Do we adequately explain how the Data Protection Act relates to the issue of anonymisation?**

We certainly welcome the draft code and the work the ICO has put into addressing this issue.

At a general level, the code is clear that sometimes data that is being considered for publication publicly may contain personal information and thus would engage data protection principles. The code explains the Commissioner's view that anonymisation can be a route to publishing such data without breaching the Data Protection Act.

We also believe however that the draft code is somewhat lacking in detailed specifics and has a tendency to deal in general terms. We do not believe it would provide data controllers with the requisite level of clarity and certainty about how to anonymise and make judgements about the publication of data sets. First, it is not completely clear what role the guidance will have, or how binding it is, and therefore to what standards an organisation will be judged. For example, on page 4 the code states that:

> *'….the Information Commissioner...will take the good practice advice in this code into account...it will certainly stand an organisation in good stead if it can demonstrate that its approach to producing and disclosing anonymised data has been done with due technical and organisational rigour'.*

We suggest this may leave data controllers feeling uncertain about the standards to which they will be held.

Second, there are also some cases where the code gives general guidance on an issue, rather than more detailed advice that may be more helpful. For example, on page 27 the code states that:

[1] http://www.ico.gov.uk/about_us/consultations/our_consultations.aspx

> “always removing numbers relating to five or ten individuals or fewer may be a reasonable rule of thumb for minimising the risk of identification in a proactive disclosure scenario, but in the context of a specific FOIA request a differentWe certainly welcome the draft code and the work the ICO has put into addressing this issue.

Whilst this is useful to highlight the nature of the issue, and serves as an example, we would suggest that the code would be more useful and effective if it specified why such decisions would be more or less reasonable with reference to the Data Protection principles.

Third, there is a danger that as written the code assumes a certain level of knowledge of the DPA. We feel that the code could more clearly connect specific case studies of data anonymisation with data protection principles, and explicitly explain why data protection principles are relevant. We develop this point further in our answers to question 10 and 11. The code could, for example, go through the data protection principles more clearly, one by one, explaining how anonymisation relates to each and referencing possible scenarios and the issues that are engaged.

Consent

We also believe that the issue of consent could be treated in more depth. Different types of information will have been shared with varying levels of consent. Again, it would be more helpful were the code to discuss specific examples of varying types of data, the consent mechanisms involved and the implications for anonymisation. This is particularly problematic as the purposes for which data will be used cannot be controlled once published.

We address issues about the clarity of the guidance further in answer to questions 10 and 11.

2. **Does the code explain adequately what anonymisation is, its technical aspects and how it’s used in practice?**

We believe that the code does not fully acknowledge that anonymisation is in practice never going to be 100% effective, and tends to give the impression that anonymity is a binary issue. It would be helpful to highlight this more clearly, and stress that the disclosure of information is a question of judgement and disclosure control, which is a task requiring a significant level of expertise.

Whilst we appreciate the “motivated intruder test” provides a useful method of analysing potential risk, it may be difficult for non-experts to use that model and to consider it as a sufficient test of whether a data set is safe to be published. Considering the behaviour of a motivated trespasser requires a difficult mental exercise involving second guessing the motives, knowledge and technical expertise of the intruder, and assumptions about the usefulness of the data.

We would recommend network or panel of experts capable of dealing with referrals of hard cases would be useful. Given the impossibility of predicting the vulnerabilities involved in applying anonymisation mechanisms to specific data, scrutiny should be possible through engaging the technical community via disclosure and open review similar to those found in other computer security areas.

The code does not give sufficient attention explanation or guidance on the problem of a 'jigsaw' of data sets, and how combinations of data can re-identify people.

Ultimately a suite of further resources will be necessary to ensure the requisite expertise is available to those making anonymisation decisions. Whilst some of this may not fit within a code, we suggest the ICO identifies what additional resources are needed and helps to make sure they are in place for when the code is launched.

3. **Are there any key anonymisation techniques which the code does not cover?**

Anonymisation is a developing field and there are likely many techniques not covered by the code. Any guidance on techniques and best practice will require frequent revision.

We suggest that it will not be possible to include a definitive list of techniques in a static code. We suggest instead references to comprehensive lists of techniques and methods that reflects the most up to date practice, with links to sources and further expertise. This should be accompanied by work to develop these resources.

4. **Does the code strike the right balance between the protection of individuals' privacy and the benefits of making information publicly available?**

The code is not the ideal mechanism to determine the balance between protection of privacy and the benefits of greater information sharing. It is best thought of as a tool to help controllers take decisions about publication through frameworks such as the flow diagram.

The trade offs between privacy and innovation or economic growth – or whatever benefits are envisaged from the publication of data are better settled through the strategies for open data, formulated in a democratic way. This should set the aims and desired outcomes of the open data agenda. These are broader issues about democratic oversight and control over the use of varyingly personal data. We are concerned that the government has not fully tackled such privacy concerns – something we note further in our response to question 12.

5. **Does the code cover the use of anonymisation techniques in all the key sectors?**

See answer to question 10.

6. **Do we satisfactorily explain the issue of anonymisation and spatial**

information?

See answer to question 10.

7. **Does the code adequately explain the difference between publication and limited forms of disclosure?**

8. **Is the section 33 research exemption clearly explained**

9. **Is the flow diagram useful?**

The diagram is useful but could be far more so. We would advise it is more obviously linked to the text of the code. We suggest that it is better annotated, or that the guidance more clearly follows the flow diagram in terms of structure. As we note further in our response to questions 1 and 10, this would be an area where more detail on specific examples of how certain types of data engage data protection principles would be beneficial.

10. **Do you think further diagrams, examples or case studies should be used?**

We feel that the code should involve a much wider, detailed typology of data. That should include associated case studies and example scenarios. This could include consideration of the possible harms and risks from disclosure and how exactly data protection principles apply. At the moment the code lacks detail. Spatial data is singled out, for example, but a similar analysis for a broader typology of data would be more helpful.

11. **Is the code easy to understand?**

One of the most important questions to ask regarding this code is a consideration of who it is written for. Does it hit the right tone and language for likely users? Does it go into sufficient detail, in the sufficiently clear terms, to enable users to make informed and privacy aware decisions?

Our broad concern is that the open data and open government agenda will lead to many more institutions being asked, encouraged or required to publish data. This will put many more people in the position of having to make judgment about whether to release data and how effectively it may be 'anonymised'. There is a greater likelihood that more datasets including variously personal data will be considered potentially publishable. Many of these will not be specialists in the field of anonymisation.

Given that anonymisation is effectively a question of judgement and risk (considering actual anonymisation to be an ideal rather than a practical possibility) it is important to think about who is making those judgements and on what criteria. The code needs to primarily address the needs of those who will be dealing with the issue of anonymisation.

Anonymisation and the process of disclosure to control to assess and manage disclosure risks is a very expert, developing technical field. Expecting non-experts to manage this is a problem. Many of the institutions that will be expected to handle requests for data, such as hospitals, GPs and schools, are likely not prepared to deal with routine generation and processing of anonymised data.

We recommend a user-focused process for the next iteration of the code, that develops the text with groups most likely to use it. Pro-actively working alongside such groups would help to ensure maximum usefulness and effectiveness and ensure that the code is fit for purpose. This goes beyond a consultation, requiring work to seek out a sample of people from relevant organisations. It should help to identify gaps in controllers' knowledge, their technical skills and the support they are likely to need. We suggest that as part of this the ICO work with the Cabinet Office to develop a form of skills audit of those who will be dealing with open data.

It is likely that to cover knowledge and skills gaps, data will be sent in an identifiable form to “anonymisation havens” provided by specialist units or commercial providers. Attention also needs to be paid therefore to those organisations offering anonymisation services, specifically looking at how they are scrutinised and regulated. With that in mind, it is important to note concerns about a general lack of high-level skills in anonymisation in UK will not allow for proper scrutiny of such services and of new techniques.

12. **Is there anything else the code should cover or are there any other ways in which the code could be improved?**

We are pleased the ICO are moving to address the issue of anonymisation. However, it is worth remembering that this will likely act as a spur for more publication of data. And this code sits in a wider regulatory and policy context.

The Government have embraced the open data and open government agenda, which is to be applauded. However, there is a concern that without the right strategic and regulatory framework, this will lead to privacy being sacrificed for the perceived benefits of wider publication of information. We are extremely concerned that a failure from government to place privacy concerns at the centre of a stronger and more coherent open data strategy may lead to significant and unforeseen privacy harms.

Whilst the code is a step towards addressing this, there are broader issues such as expertise, redress, and strategy that the ICO should bear in mind when considering the effectiveness of the code.

There may be a limit to the extent which the ICO feels it can be the sole regulator for the anonymisation of data. We would recommend an expert body that can help manage decisions about the release of data, and analyse the risks of publication and how appropriate anonymisation techniques are.

We also suggest a further study of the possible privacy risks from the publication of a variety of data types, which should inform the typologies suggested in response to earlier questions.

It would be helpful for the ICO to consider what broader set of resources are necessary to ensure that privacy is a strong consideration for those dealing with the possible release of sets of data.

The strategy behind open data and open government

The government has been somewhat unclear about what they see as the principle aims and benefits of the push for open government, citing innovation, economic growth, transparency and accountability. There has been a tendency to conflate three issues:

- open data (uncontrolled free online access with an open license for reuse and technically accessible)
- data sharing (synergies and efficiencies of increased semi/controlled access)
- big data processing (generating unique new insights by processing, combining and mining huge datasets)

In his report on privacy and the open data agenda, commissioned by the Cabinet Office, Professor Kieron O'Hara states: “Under the current transparency regime, in which public data is specifically understood not to include personal data, most data releases will not raise privacy concerns. However, some will, especially as we move toward a more demand-driven scheme.”

There is a risk that a commitment to a broad set of goals, from innovation to transparency, will lead to a confusion between personal and non-personal data sets and how the publication of either type satisfy the aims of the data agenda.

Some of these are issues that lie beyond ICO's remit, and which can not be dealt with in this kind of code. However, it is important to consider this broader context which the code will sit within, and for the ICO analyse whether there are sufficient resources available, from expertise through to clarity of political strategy, to enable data controllers to make the most privacy appropriate decisions possible.