Anonymisation: managing data protection risk code of practice

Consultation summary

This document summarises the responses to the consultation on the Anonymisation: managing data protection risk code of practice. It was held between 31 May and 23 August 2012. There were 74 responses to the consultation from both organisations and some individuals.

The code has been substantially rewritten following the consultation exercise; indeed some parts will be fairly unrecognisable from the consultation draft. The code should now be much clearer and better explained. In particular, we have explained the relationship between the Data Protection Act (DPA) and data anonymisation more clearly. Whilst we have stuck to our initial line – that anonymisation is possible and desirable – we have made it clear that we are approaching the subject using the tests in the DPA.
We extend our thanks to all those who took the time to respond.

Following receipt of the responses, we felt it was necessary to make the following main changes:

- Strengthening considerably the transparency aspects of anonymisation and how organisations should explain what they are doing to the public;
- Ensuring that the data protection liabilities of those who perform 're-identification' are clearly emphasised;
- Ensuring that the terminology and 'tests' are consistent throughout;
- Enhancing the section on 'governance', in particular emphasising the difference between publication and limited disclosure;
- Including a range of case-studies and examples – many sent to us by consultation respondents – that make the 'legal issues' much easier to understand in real-world contexts; and
- Including several references to the ukanon.net anonymisation network.

We have also included a new section in Annexe 1 which explains the process of turning personal data into various forms of anonymised data

and looks at matching data through unique patters and assessing re-identification risk. We hope that this provides the detailed worked-through example which many respondents called for.

On the whole the majority of the respondents were welcoming of the code, stating that it was well written, useful and a very good document which was easy to understand. The code was described as "a useful and highly pertinent document" which is "timely and very valuable" and "not too legal or corporate – a breath of fresh air". It is clear that freedom of information and the open data agenda has stimulated the need for guidance in this area.

A number of parties welcomed the introduction of guidance setting out how to assess the risks of identification and how information can be successfully anonymised. It was generally felt that the code would help to promote a better understanding of the approach to be taken when anonymising personal data that it sets out a clear and useful framework to help with what – many respondents acknowledged - can often be very "difficult judgments". The clarity provided by the code that data which that has been anonymised effectively falls outside the scope of the DPA was also welcomed, as was our line on the status of pseudonymised data.

Most respondents found the code to be concise, relevant and well defined, although it was recognised that developing a single code for all sectors was a challenge.

Many responses highlighted specific points which were helpful and which have been incorporated into the final version of the code. We made a very large number of specific changes to the text and we hope it is now more consistent with the work of others in the area. However, it is clear that there is different terminology in use – 'de-anonymisation' Vs 're-identification' for example - and complete consistency with others' work was not always possible. We do try to make it clearer, though, what we mean by the various terms we use.

One respondent, however, felt that the code attempted "to sell anonymisation as a 'cure- all for privacy problems, ignoring the relevant science and engineering'." This was very much a minority view but we did try to accommodate it by explaining more clearly that we are using the test of identification/identifiability in the DPA - we are not approaching the subject in terms of the absolute impossibility of re-identification at any point in the future. This is why the 'motivated intruder' test adopted in the code was generally considered useful.

Several respondents made similar comments and suggestions:

Comment:

The terminology in the code is unclear and clearer definitions would be useful. For example, it was stated that the "lack of a definition of research

is unhelpful." A summary section or overview of the main points would be useful, as would bullet points summarising the key points by section. Overall the structure of the code works well but certain sections of the code (such as the examples) should be re-arranged and more closely associated with others.

ICO response:
We have introduced 'key points' as an introduction to each section. We have also redrafted those parts of the code, where it was felt that the terminology was unclear and inconsistent, to ensure consistency throughout the document. We also included a definition of 'research' and restructured the way the examples and case-studies are presented.

Comment:

Others commented that the code was less of a code of Practice "more a set of principles for a particular purpose", and that it read more like guidance than a code of Practice.

ICO response:
We do not think there is an accepted definition of what a code of practice is or how one should approach a particular topic. The distinction between a code of practice and 'ordinary' guidance can be unclear. However, we are satisfied that the status of the code, and how it is meant to be used, is explained clearly.

Comment:

We received comments that considered the code was "lacking in detailed specifics and has a tendency to deal in general terms". Many noted that the code would benefit from further more detailed case studies across a variety of different sectors.

ICO response:
We accept that the initial draft could be somewhat theoretical and abstract, so we hope that the many examples and case studies we have now included in the annexes to the code will help to address that – as will the tightened-up language.

Many of the examples we have used in the updated version have been provided by respondents. The first annex consists of a detailed description of how a set of personal data can be converted into various forms of anonymised data and used in various ways. It also illustrates the difference between publication and limited disclosure and explores re-identification risk. The second annex consists of case-studies showing how various anonymisation techniques can be used in practice. Finally, the third annex consists of a set of practical examples of some anonymisation techniques drawn up by

experts at the University of Southampton. We are particularly appreciative of staff at the University of Southampton and to those respondents who provided us with examples – the majority of which we included in the code.

Comment:

Some respondents felt that specific information for specific sectors, particularly research, would be advantageous. Several responses noted that the code was focused predominantly on the public sector and research organisations and stated that it would be helpful if the code had a greater emphasis on how anonymisation can benefit private sector organisations.

ICO response:

The scope of the code is necessarily broad. As the DPA applies to all sectors, it would not be feasible for us to write detailed sector specific guidance. We do now say, though, that organisations may wish to use our code to produce their own sector-specific products. We accepted the point though about bias towards the public sector and did include some private sector examples, from retailing and other areas that were provided to us

Comment:

We also received comments about the technical level and content of the code. Some respondents considered that the code was not aimed at the average practitioner and that a certain level of technical knowledge would be required to understand it. However, it was also felt that pages 7 to 9 in particular were "accessible to the various specialist audiences who use anonymisation but who are not data protection and privacy practitioners."

ICO response:

We recognise and acknowledge that not all parts of the code will be relevant to all organisations. We were also trying to write this code for different organisations and people, with very different levels of expertise – ranging from those that may have to carry out relatively simple anonymisation occasionally to ones that make large-scale data releases routinely. We have made every effort, particularly following the consultation responses, to ensure that the code is not too complicated to read or understand. Some of the more 'technical stuff' will be addressed through the ukanon.net anonymisation network that is being set up.

Comment:

Many suggestions were received about other key anonymisation techniques which could be covered in the code for example the use of 'keys', sampling, cell suppression and techniques relating to qualitative data.

ICO response:

We have made it clear though that the code is not meant to be a security engineering manual nor to describe every anonymisation technique. We have tried instead to give a flavour of the techniques available and of how they might be used. As some of the responses themselves acknowledged, it would not be possible to have "a definitive list of techniques in a static code." Again, ukanon.net will help to address this. We accept that we did not deal with qualitative data adequately and have included a new section and an example relating to this.