The Foundation for Information Policy Research

Written evidence to the Information Commissioner on

The Draft Anonymisation Code of Practice

The Foundation for Information Policy Research (FIPR) is an independent body that studies the interaction between information technology and society. Its goal is to identify technical developments with significant social impact, commission and undertake research into public policy alternatives, and promote public understanding and dialogue between technologists and policy-makers in the UK and Europe.

The draft anonymisation code of practice starts from the wrong place. It is described thus:

> *"The code is intended to demonstrate that the effective anonymisation of personal data is possible, desirable and can help society to ensure the availability of rich data resources whilst protecting individuals' privacy."*

This is well known not to be the case. The recent authoritative Royal Society report on "Science as an Open Enterprise" (of which the Director of the Wellcome trust, and Government Chief Scientific Adviser designate, Sir Mark Walport, is an author)[1] gives an overview of the scientific and research policy issues. The Information Commissioner should have read the section on privacy. For example at p 53 the report states

> *"It had been assumed in the past that the privcacy of data subjects could be protected by processes of anonymisation such as the removal of names and precise addresses of data subjects. However, a substantial body of work in computer science has now demonstrated that the security of personal records in databases cannot be guaranteed through anonymisation procedures where identities are actively sought."*

It is disgraceful that the draft Code ignores the relevant science as summarised in that report – from the pioneering work of Dorothy Denning and others over thirty years ago; through the many well-publicised incidents of anonymity failure, including the Netflix incident and Latanya Sweeney's work on re-identifying medical records; to the more recent framework for analysing anonymisation properly developed by Cynthia Dwork and her colleagues. The writers of the draft Code disregard not just this science, but the policy lessons that follow from it (set out in Paul Ohm's widely-cited paper[2]).

---

[1] "Science as an Open Enterprise", 21 June 2012, at http://royalsociety.org/policy/projects/science-public-enterprise/report/

[2] "Broken Promises of Privacy – Responding to the Surprising Failure of Anonymization", Paul Ohm, UCLA Law Review v 57 p 1701 (2010), at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

The draft Code is also introduced as

> *"The code of practice will provide guidance on how to assess the risks of identification and how information can be successfully anonymised."*

Again, it fails completely. If the ICO wishes to train people contemplating the use of anonymity as a privacy mechanism to do a proper security engineering job, then that training had better include (a) an introduction to security engineering methodology (b) an overview of the relevant science plus pointers to where the curious can learn more (c) a clear warning that anonymisation is hard, and that advice had better be sought from experts in information security and inference control. These are simply lacking.

Security engineering

Security engineering methodology involves first setting out what the system is supposed to do (its concept of operations); second, the bad outcomes to be avoided (the threat model); third, the strategy to be used to prevent these bad outcomes (the security policy); fourth how this policy is to be implemented (the technical mechanisms) and fifth how the data subjects and other stakeholders can reasonably rely on all this (the assurance)[3].

As an example, consider what is probably the most important anonymisation process in the UK: that run by the MHIA for the CPRD, the gateway through which our medical records will be "anonymised" and made available for research from September 2012. The Government's privacy 'tsar', Tim Kelsey, promised in public that its anonymity mechanisms would be public. Yet a Freedom of Information Act request to the MHIA was declined; on review the agency provided a copy of a paper they had submitted for publication yet had failed to disclose in response to the initial request. Fuller information has still been refused – and the case papers show that the MHIA were so clueless about security engineering that they needed the simplest concepts such as 'security policy' and 'threat model' spelled out to them[4]. Although we will refer this to the ICO and the Tribunal in due course, the ICO might care to take a more active interest in this particular system. If CPRD fails then the credibility of the Government's strategy on anonymisation (which the ICO seems so anxious to promote) will be at stake.

Basic concepts

To educate bodies such as the MHIA which propose to make highly sensitive data available without consent on the basis of technical mechanisms which they feel unable to defend in public, the ICO had better educate its audience about the basics of anonymity and privacy. First, the *anonymity set* is the set of all individuals with whom a data subject might be confused; thus if instead of being named I am merely described as "a Cambridge Professor" the anonymity set consists of the 200-odd professors at

---

[3] Multilateral Security, 2008, at http://www.cl.cam.ac.uk/~rja14/Papers/SEv2-c09.pdf

[4] Case documents online at http://www.whatdotheyknow.com/request/privacy_mechanisms_in_cprd

Cambridge. Similarly, let the *privacy set* be the set of people to whom a data subject requires that a given sensitive datum not be disclosed. For most data subjects and most sensitive data, the privacy set will consist of friends, family, colleagues and enemies – perhaps a hundred individuals (though for celebrities and in some particular contexts the privacy set may be essentially everyone). Privacy fails if the anonymity set is reduced to one from the viewpoint of anyone in the privacy set.

Yet at page 20 the Code downplays the risk of re-identification by family members or work colleagues who have partial knowledge of a data subject. On p 22 it says "It is generally reasonable to assume that those constrained by professional or legal obligations and close family and friends, and in some circumstances, colleagues, are likely to be motivated to defend the identity of an individual to whom the anonymised information related." This is just wrong. We help occasionally with a health privacy NGO where a typical hard case is a girl ostracised by her family after a termination of pregancy became known to them via an uncle who worked at a health authority. Yet the ICO continues on p 22 that re-identification from recorded information may be a data protection violation, while re-identification from personal knowedge is not a data protection violation even if it is a privacy violation. This is an astounding position for a privacy regulator to take. It implies that if a famiy were identifiable from a published anonymous medical record because they included "a 55-year-old woman with a 27-year-old daughter both of whom have psoriasis" then this would only be a privacy violation but not a data protection violation so long as this remained tacit verbal knowledge among family and friends – but as soon as any of them typed their knowledge into a computer it would be a data protection violation too. This is evidently ridiculous.

How did the ICO get into such a mess? Successive Information Commissioners have preferred to argue from the letter of the Data Protection Act and ignore Recital 26 of the Data Protection Directive which requires Member States to assess deanonymisation risks "by any other person" as well as by the controller. (This recital is misrepresented on page 7 of the Code.) Under EU law, the UK is bound to properly implement the Directive; the least that the ICO should do is apply the UK Act as much as possible in accordance with the Directive. But the ICO has consistently refused to do so.. In addition, there is the Human Rights Act, which implements the ECHR, including Article 8. The ICO is actually required to have regard for the case law of the Strasbourg Court. So it is not satisfactory for its Code to say, as it does at p 30, "It is advisable to seek specialist advice if you believe a disclosure has Article 8 implications". Although the HRA may only apply to public-sector bodies and firms carrying out functions of a public nature, the public sector amounts for over 40% of GDP; and most of the sensitive data to which the Code refers will involve at least one public-sector player – such as the MHIA. (The proposed new Data Protection Regulation will remove this loophole anyway.)

Expertise

The code has many other defects. For example, much of its discussion is in the context of the Freedom of Information Act. Yet it agonises at length about what other information might be available to a recipient. But information released under FOI must be assumed to

be public, so the re-identification risk must be analysed with respect to all other information that is public, or that is reasonably likely to become public in the subject's lifetime, or that might reasonably be available privately to the subject's privacy set.

Another disturbing example of the ICO's lack of expertise is on page 16 which suggests that GPs' surgeries and supermarkets could use a shared encryption key to create a common pseudonym from people's names and addresses so that patients' diabetic status could be correlated with supermarket purchases by a research company. The implication is that GPs are making specially sentitive information available without consent on the assumption that it is anonymous. Yet in the proposed design the cryptographic keys are also available to local supermarkets who can re-identify diabetic patients who are their employees, or who have handed over their names and addresses to get a loyalty card. The idea of providing a database of all diabetic patients to a research firm when this database can be trivially re-identified by any of the supermarkets that are the research firm's customers is breathtaking. If the Information Commissioner were on the ball I would expect him to take enforcement action against any GP who participated in such a scheme; for him to suggest it in an official publication is simply astonishing.

A further serious defect is the Code's failure to be sufficiently explicit about the transparency of the anonymisation process. Given that anonymisation is usually much less effective than claimed, data subjects need to understand the process so that they can decide not to supply sensitive personal information if they wish. Indeed the Code at p 24–5 sets out conditions under which personal data can be anonymised and disclosed without consent including that "the data controller's privacy policy – or some other form of notification – explains the anonymisation process and its consequences for individuals". This skates over the critical question of whether the data controller should provide a full technical explanation that can be assessed by critics, or merely ill-informed hand-waving assurances of the kind we see in the Code – and which are offered publicly by the MHIA in respect of CPRD. Given the recommendation (pp 35–6) that anonymisation schemes should be subjected to third-party penetration testing (which is sound and which, on its own, we welcome), we would be concerned that an organisation which had performed such testing might be permitted by the ICO to wave its test certificate rather than explaining in detail to affected data subjects how the 'anonymisation' was carried out. Security by obscurity has been known since the 19th century to be unreliable; and in the case of the CPRD, any data perturbations are likely to become public anyway as medical researchers using the data will have to publish how it was processed.

In short, inference control and security engineering are both hard. The ICO does not have adequate expertise, and must not mislead readers into thinking that they can learn these subjects by reading its Code. They cannot, and must be directed to seek advice.

Legal effects

We are seriously concerned at the potential effects should this Code, despite its technical inadequacy, be approved by the Secretary of State under Section 52 of the Data Protection Act. In particular, the code promises to mitigate the legal effects of negligence

by a data controller. *"In the event of the Information Commissioner investigating an issue arising from the anonymization of personal data, he will take the good practical advice in this Code into account."* The ICO has a duty under the Act to promote good practice; this code promotes bad practice by giving poor technical advice and advising firms that if they follow it they will be less likely to be subject to enforcement. Thereby the Commissioner is failing to discharge his duty and we call on him to withdraw this Code.

If he does not, we will write to the Secretary of State pointing out that he is duty bound to withhold approval under section 52B(2) of the Act on the grounds that approval would place the UK in breach of its Community and other international human-rights obligations as described above.

We will now discuss the specific questions raised in the consultation.

1. The Code does not adequately explain how the Data Protection Act relates to anonymisation; as described above, this is the wrong question to ask anyway.
2. The Code's explanation of anonymisation is totally inadequate. It fails to communicate even the basics of security engineering let alone the subtleties of the underlying computer science. It may convince the untutored reader that they can make private data fit for release by choosing from among a few technical tweaks described in an appendex; it fails to emphasise that this is a complex engineering problem for which high-quality professional advice is necessary.
3. Significant numbers of anonymisation techniques are not covered; see any textbook for more (e.g. [5]).
4. The Code is not balanced, or even honest. It tries to market anonymisation as a cure-all for privacy problems, ignoring the relevant science and engineering.
5. The Code does not come close to covering the use of anonymisation across different sectors of business and government. That would take a large textbook.
6. No.
7. The explanation of limited disclosure is woefully incomplete. It might give the reader the impression that it's OK to let any medical researcher interested in psychiatry to have the records of all Britain's millions of people with diagnoses of anxiety or depression sitting on her laptop, minus only their names, so long as she'd signed an NDA and got ethics committee approval (i.e. disclosure is limited). Yet when such a laptop is left on a train and its contents published online, the consequences for many individuals could be catastrophic.

8–12:

The Code is so far from what's needed that the Commissioner should abandon it. The ICO has never had adequate technical expertise and is simply not able to undertake exercises of this kind.

Ross Anderson FRS FREng
Chair, Foundation for Information Policy Research

[5] Multilateral Security, 2008, at http://www.cl.cam.ac.uk/~rja14/Papers/SEv2-c09.pdf