

# **Summary of ICO privacy and data anonymisation seminar Wednesday 30 March 2011, London**

## **Background**

Public bodies and other large organisations hold ever increasing amounts of information on individuals in a wide range of information systems and databases. Some information derived from individuals will be published or shared in a form that people believe will not lead to identification. Assessing risks in disclosing information derived from data relating to individuals is difficult in the modern computing era, when new tools and technologies are emerging to “unlock” information, particularly datasets. The issue of assessing whether information in the public domain can lead to identification is becoming increasingly difficult.

The Government’s transparency agenda, demand for greater openness in the public sector and developments in technology are driving the publication of greater volumes of datasets. The ICO welcomes developments to open up public sector data but recognises the need to balance privacy risks and protect individuals’ personal data.

At the ICO seminar on privacy and data anonymisation leading experts presented different approaches to, and perspectives on, this complex subject. The seminar looked at current practice, the risks associated with anonymisation and possible solutions for the future.

Presentations were followed by a facilitated plenary session which allowed speakers and audience members to discuss the issues in greater detail.

The following is a summary of the key points from the seminar. Individual presenters’ slides are also available on the ICO website: [http://www.ico.gov.uk/news/events/events\\_archive.aspx](http://www.ico.gov.uk/news/events/events_archive.aspx)

## **Introduction**

### **Christopher Graham, Information Commissioner**

Christopher Graham welcomed delegates and thanked Sir Mark Walport for hosting the meeting at the Wellcome Trust. He stressed that the seminar would look at current and emerging practical risks of identification rather than being a debate about legal definitions.

Mr Graham said that the meeting was very timely as the government's transparency agenda is driving the publication of large volumes of data. He said that the positive benefits of greater transparency, such as better accountability and better informed research, were not in question, but that we need a debate about how to manage the privacy risks, and risks of re-identification, that arise from the publication of more data.

Mr Graham said that he hoped this seminar would stimulate further debate about privacy and data anonymisation, help to inform the ICO in terms of any guidance or further research which may be needed and feed into the review of the European Data Protection Directive.

### **From data to health**

#### **Sir Mark Walport, Director, The Wellcome Trust**

Sir Mark Walport focused on the challenges around data anonymisation in the medical research field and asserted that good data is inextricably linked to good public health.

Sir Mark outlined some of the benefits of data sharing in a range of contexts, including medical research, and suggested that the public are generally in favour of such sharing.

The ability to link large datasets on, for example, health, housing, the environment and so on will bring many benefits including:

- better public health;
- evidence based policy; and
- personalised medicine.

However:

- this can not be done with truly anonymised data;
- these data linkages will depend on using pseudonymised data;
- proportionality is the key issue regarding the use of personal data;
- it is often impractical to obtain consent, but steps can be taken to ensure confidentiality;
- undertake a risk/benefit analysis - accept that risks exist, but the benefits will often outweigh the risks; and
- the greatest risks to confidentiality are not from researchers.

Sir Mark acknowledged that good governance is crucial in this field and suggested that some sort of safe haven mechanism be established to deal with health data being used for research purposes. Safe havens should be safe and secure mechanisms for health research. Sir Mark stressed that strong penalties should be in place for those who abused the mechanism. He also stressed that the researcher is not normally interested in the identity of the data subject.

A comment from the floor flagged up potential risks of identification arising from the NHS Pseudonymisation Implementation Project<sup>1</sup> that was due to be complete by April 2011.

### **Anonymisation as disclosure avoidance Dr Mark Elliot, University of Manchester**

Dr Mark Elliot set out the conceptual and theoretical background to the concept of anonymisation. He challenged the concept of anonymisation and whether true anonymisation was really possible. He suggested that there is a need to carry out risk assessments when deciding whether or not to disclose data. Risks will always depend on the specific circumstances of the data and the situation. He argued that although risks relate to the uniqueness of an individual the real risk is not necessarily the identification of the individual but the likely consequences of the disclosure. Risk assessments therefore need to look at the intruder's motives and what is likely to happen if the data is disclosed.

Dr Elliot also explained the concept of intuitive demographic knowledge and how this could be used to isolate and identify. He also highlighted how the principles of game theory may be important in understanding risks of identification.

Dr Elliot concluded that:

- complete anonymisation is:
  - best described as the absence of disclosure risk
  - a theoretical state
- pragmatic anonymisation is:
  - low levels of disclosure risk.

Dr Elliot's presentation illustrated that there was an emerging science around statistical disclosure from microdata and aggregated data. A better understanding and application of methods related to data intrusion and evaluation of attribution disclosure are important

---

<sup>1</sup> <http://www.connectingforhealth.nhs.uk/systemsandservices/pseudo>

factors to consider when developing a response to the risks of anonymisation.

A comment from the floor highlighted the risks were not just from certain identification of individuals but also when groups of individuals are identified in groups of 2 or 3.

### **Transparency – opening up Government Nicola Westmore, Cabinet Office**

Nicola Westmore outlined the key aspects of the Government's transparency agenda. The aim of this work is to make to Government more accountable by promoting efficiency and effectiveness. More and more data is being made available to citizens in a number of ways to assist them in making informed decisions.

Ms Westmore explained that the Government is looking at ways of protecting privacy and minimising risks whilst extending transparency and making more data publicly available. A review of the impact of transparency on privacy has just been completed by Dr Kieron O'Hara. The review will:

- support the Government in striking the right balance between transparency and data protection safeguards, and between the interests of wider society and the interests of the individual or corporate body;
- identify the nature of the risk to privacy of the individual posed by transparency of public data, in particular the potential for 'jigsaw' identification; and
- advise the Government on practical approaches to take.

### **Privacy, deanonymisation and transparency Dr Kieron O'Hara, University of Southampton**

Dr Kieron O'Hara gave an overview of the issues he has been asked to consider as part of the review he is carrying out for the Cabinet Office, and discussed whether transparency will pose a threat to privacy.

Public data released as part of the transparency agenda do not include personal data but there will be areas where the state and the citizen interact – such as crime data – which will present privacy issues which will need to be considered carefully. In particular, Dr O'Hara mentioned the problems of potential 'jigsaw identification' whereby data from a number of sources can be combined to enable identification of individuals. He suggested that there will always be

a trade-off between data utility and privacy/anonymity. There is a demand for greater transparency, for example relating to health, education and court data, but we must ensure that transparency of government does not equal transparency of the citizen.

Dr O'Hara suggested the defences of disclosure control, terms and conditions and consent were inconsistent with the aims of the transparency programme and should only be used in particular circumstances.

Similar to other speakers, Dr O'Hara also suggested there were problems with using the concept of anonymisation, that the word suggested success which was illusory. His preferred term is 'disguise'.

Dr O'Hara suggested that transparency depends on retaining the confidence of citizens and in order to achieve this, citizens have to feel that their privacy is protected. Privacy is not just a legal matter and is not limited to data protection, he stressed the importance of perceptions of privacy to the transparency programme. To move forward the debate around transparency and privacy he said that discussions are needed between:

- transparency activists;
- privacy activists;
- technical experts;
- domain experts; and
- information entrepreneurs.

In addition, Dr O'Hara suggested that it is important to describe the processes around decisions regarding the release of data and felt that auditable debate trails would improve the current situation. Finally, he said that privacy should be embedded in any initiative and not bolted on afterwards.

Dr O'Hara also highlighted the lack of empirical research surrounding these issues.

### **Balancing Risk and Utility – Experiences in Official Statistics Dr Marie Cruddas, Office for National Statistics**

Dr Marie Cruddas discussed the issue of balancing risk and utility when releasing statistics. She discussed the problem of not always being able to predict how released data sets will be used and the various ways in which ONS tests data sets to assess the possibility of re-identification. Currently no census microdata is made available publicly.

Dr Cruddas explained the confidentiality protection framework which is used by ONS when determining whether or not to release data. The issues covered by the framework include:

- determining users' requirements;
- understanding the data (e.g. source, sensitivity, age, quality, coverage);
- assessing and managing disclosure risks;
- statistical disclosure control methods; and
- implementation.

Her presentation also set out a range of reasons as to why an intruder would want to discover information about others, such as:

- ID theft
- Commercial gain
- Journalists seeking stories
- Seeking sensitive information – e.g age, salary
- Database enhancement
- People trying to identify themselves

Dr Cruddas acknowledged that risks can not be eliminated but that ONS makes strenuous efforts to reduce them whilst making the data as useful as possible. She suggested that more research is required about what the public think is acceptable in terms of detailed disclosure.

### **Protecting privacy after the failure of anonymisation Professor Paul Ohm, University of Colorado**

Professor Paul Ohm discussed the perception that anonymisation is a concept which is trusted and rewarded by law. However, he went on to present examples from a number of recent studies which challenge assumptions about the robustness of anonymisation. The studies demonstrate how seemingly anonymised data can relatively easily be combined with other information to identify people.

Professor Ohm went on to discuss how policy makers should respond to this situation. He suggested that technology will not provide an adequate solution to this growing problem and instead suggested that regulators should abandon the notion of anonymisation and instead focus on contextual risk assessments which weigh the benefits of access to information against the risk of privacy harm. Such assessments should include a consideration of:

- data handling techniques;
- private versus public release;

- quantity;
- motive; and
- trust.

Professor Ohm suggested that any new privacy regulation principles should move away from promises and requirements to an approach of best efforts and accountability. He also suggested that rather than focusing on personal information we should instead move towards considerations of 'risky data collections' irrespective of whether they contain any personal or identifiable data.

### **Anonymity in Market, Social and Opinion Research** **Barry Ryan, Market Research Society**

Barry Ryan explained the history of market research and how techniques have developed over the years. He gave a picture of some of the features of market and social research today. Firstly, he noted that survey research now includes:

- qualitative groups;
- market research online communities; and
- ethnographic and deliberative techniques.

Mr Ryan discussed some of the current challenges in social research as clients demand more detailed analysis of research data and often hold large amounts of data themselves, such as loyalty card data, from which they would like to gain insight.

Mr Ryan also discussed the implications of the demand for more detailed information in terms of trying to ensure anonymity for the respondent. He said that trust is the key issue and suggested that if respondents can't trust anonymity then perhaps we should move towards an informed consent model and ask respondents to trust the company or trust the safeguards and security measures which the company puts in place rather than trusting that they will remain anonymous.

### **Panel session**

**Chair: David Smith, Deputy Commissioner, ICO**

**Panel Members: Mark Elliot, Marie Cruddas, Nicola Westmore, Kieron O'Hara, Paul Ohm, Barry Ryan**

A number of topics were discussed, including:

- Rights of individuals to redress when things go wrong – different legal systems and approaches. The problem of detection was raised. The benefits of being able to take class

actions in the US were highlighted. The notion of data abuse was also mentioned.

- The use of debate trails in assessing 'recklessness' of data releases.
- Promoting the use of Privacy Impact Assessments (PIAs) in understanding the implications of releasing data.
- Is anonymisation now a myth? If so, should we be focusing on data minimisation and purpose limitation?
- Medical researchers are asking for richer, non anonymised, data – this presents problems around informed consent.
- Trust is the key issue – if people don't trust assurances of anonymity should we focus on encouraging trust in the company holding the data or trust in data security?
- The importance of the US and the rising importance of China in determining how the internet will develop, particularly given the Chinese interest in traceability.
- Individuals have responsibility for looking after their own data – should we raise awareness in an education environment rather than organisations producing materials and putting them on websites?
- The risk of small companies being unaware of the limitations of anonymisation was raised.
- A trade off between privacy and usability had to be considered. The cost of privacy may be a loss of usability in search tools. It was suggested that in order to protect privacy we should accept some reduction in usability.

David Smith thanked the speakers and audience for such a successful and thought provoking day. He suggested that the key message of the day was that whilst complete anonymity may not be possible this should not stop us aiming for 'as anonymous as possible'. Mr Smith stressed the importance of good information governance, including data minimisation, and having privacy protections in place. He acknowledged that personal data will sometimes be released but suggested that the important question is one of privacy risk and stressed the importance of carrying out risk/benefit assessments.

David Smith said that the ICO was pleased to be leading the debate around transparency, privacy and data anonymisation and would be considering plans for further work in this area. Participants were invited to make any further comments or suggestions to the ICO.

## **Overview**

The seminar clearly illustrated that anonymisation is not a risk free option and the false comfort the term may bring. Professor Ohm's



presentation provided some powerful examples of what the risks are and how they may develop.

While challenging some of the assumptions around anonymisation the seminar provided some important illustrations about how the science of anonymisation can often help to provide decision makers with a framework that can significantly reduce the risks to privacy. No anonymisation solution can be 100% safe if data is also to have some value. New solutions such as differential privacy<sup>2</sup> are emerging.

Some of the solutions include disclosure control but this may not be consistent with the aims of open data policy. Sir Mark's approach advocates opening up data to limited audiences rather than advocating open data disclosures to the public, which was Dr O'Hara's focus. Further debate is needed about when disclosure controls should be used or an open data approach followed. Some services that offer researchers controlled access to datasets are offering significant research benefits, such the Administrative Data Liaison and Secure Data Services<sup>3</sup>.

Large scale disclosures of data will continue to throw up difficult questions about risk – when 30,000 lines of data are disclosed should this be stopped if there is a risk that no more than 5 people might just possibly be identified from the dataset?

There was recognition that making data and statistics available for many different purposes had important societal value and data utility would be lost if absolute anonymisation was required. However, the importance of maintaining public trust was emphasized by many speakers.

The possible solutions cover a range of areas – technology and regulation alone not may able to find the solutions. There is need to develop better understanding of public expectations and the basis of public trust as well as the use of social science techniques. There is a need for a better articulation of privacy risk and how it is assessed.

The use of panels of experts/stakeholders to assess risks, already in operation at ONS and proposed by Kieron O'Hara was a valuable practical suggestion that builds on the model envisaged by Privacy Impact Assessments.

---

<sup>2</sup> <http://research.microsoft.com/en-us/projects/databaseprivacy/>

<sup>3</sup> <http://www.adls.ac.uk> and <http://securedata.data-archive.ac.uk/home>

The ideas presented by Professor Ohm about the limitations of current privacy legislation can provoke further debate that can feed into the debate about how new European Data Protection legislation can address this challenge and whether recital 26 in the current Directive is still fit for purpose.

Several speakers mentioned the lack of empirical research in this area.

### **Where next?**

The ICO welcomes feedback from attendees at the seminar and other stakeholders about what the ICO should do next in this area. The ICO would also like to hear from data controllers who are facing difficult challenges related to anonymised data.

Over the coming year the ICO proposes to do the following:

- Develop accessible guidance on anonymisation and the disclosure of statistical information. The ICO will work closely with the Office of National Statistics and other experts on statistical disclosure. The guidance will include a usable framework for assessing risks and making decisions about tolerable risks.
- The ICO will consider issuing the guidance in the form of a Code of Practice under section 51(3) of the Data Protection Act. Recital 26 of European Directive suggests that a Code Practice can be used to provide guidance on identification and anonymisation.
- The ICO will continue to raise the issue of anonymisation as an area the European Commission should address when drafting new European Data Protection legislation
- The ICO will consider whether there is merit in commissioning research to better understand the limitations of anonymisation and how risks can be mitigated. There is also a need to understand the public's expectations about the level of detail of open data disclosures and their view about the risks. Research may also include an overview of technology/privacy science based solutions and the benefits they offer.

- The ICO will continue to provide advice on significant projects that raise issues about the risks of anonymisation, such as crime data disclosure and the disclosure of prescription data.
- Anonymisation may have some application as a solution when managing personal data in cloud computing scenarios. The ICO may feed work on anonymisation into advice and guidance provided on cloud computing. The ICO plans to update on the Personal Information Online Code of Practice in 2011. The ICO acknowledges the outputs of the cloud computing project at Queen Mary University of London<sup>4</sup>, which have raised the issue of anonymisation.

### **Other recent developments**

A recent decision by the High Court in the case *Department of Health v Information Commissioner*<sup>5</sup> (April 2011) has now provided greater legal clarity about how the Data Protection Act should be interpreted when statistical information is requested under the Freedom of Information Act. The judgment found that the focus when disclosing the data was not whether it was personal data in the hands of the data controller but whether a member of the public could identify anyone from the data. This was the approach the ICO argued. The Court also agreed that the finding of the ICO and the Information Tribunal was correct in terms of the risks of identification. The order of disclosure was upheld. The Department of Health has not appealed the judgment and this now offers the ICO greater clarity on some of the legal issues. This approach the Court endorsed can now be incorporated into any guidance the ICO issues.

A recent report published the Ontario Information and Privacy Commissioner<sup>6</sup> offers an important perspective and urges data controllers not to stop using de-identification. The report argues that properly applied anonymisation or de-identification techniques can protect personal information.

---

<sup>4</sup> <http://www.cloudlegal.ccls.qmul.ac.uk/Research/49700.html>

<sup>5</sup> *Department of Health, R (on the application of) v Information Commissioner* [2011] EWHC 1430 (Admin) (20 April 2011)

<sup>6</sup> <http://www.ipc.on.ca/english/Resources/News-Releases/News-Releases-Summary/?id=1085>

## List of attendees

Carole	Abrahams	Office for National Statistics
Carla	Baker	Intellect
Vinod	Bange	Speechly Bircham LLP
Ian	Barker	Ipsos
John	Bates	Department of Health
Karen	Birmingham	Bristol University
Ruth	Boardman	Bird & Bird
Caspar	Bowden	Microsoft
Andy	Boyd	School of Social and Community Medicine
Pete	Bramhall	Hewlett-Packard Laboratories
Dr Tito	Castillo	University College London
Joan	Corbett	Scottish Centre for Social Research
Gina	Coulson	Department for Business, Innovation & Skills
Liam	Curran	HeLEX Centre for Health, Oxford University
Susan	Daly	Symantec Corporation
Anna	Decourcy	University of Greenwich
James	Denman	Department for Communities and Local Government
Keith	Dugmore	Demographic Decisions Ltd
Tony	Dumycz	Munich Re
Barry	Eaton	Learning Records Service
Tim	Edwards	Policy Studies Institute
Jenny	Elkeles	PHSO
Pari	Faramazi	WTG Technologies Ltd
Dawn	Foster	The NHS Information Centre
Maurice	Frankel	Campaign for Freedom of Information
Veronica	Fraser	Department of Health
Dr Kirstin	Goldring	University College London
Peter	Gooch	Deloitte LLP
Andrew	Harris	NIGB
Kuan	Hon	Queen Mary University
Nigel	Hopgood	Intellect security & privacy interest group
Phil	Jones	Consultant
Prof Mark	Josephs	University of Warwick
Kim	Kingham	Scottish Government
Kieron	Mahony	National Statistics Authority
Neil	Mathews	Dell Corporation Limited
Chris	Maude	Tyne and Wear Fire and Rescue Service
Ganka	Mueller	General Register Office for Scotland
Dr Fortune	Ncube	Health Protection Agency
Georgina	Nelson	Which?
Cynthia	O'Donoghue	Reed Smith LLP
Bob	Reid	Which?
Sarah	Ridley	AXA ICAS
Colin	Rogers	IM Services
Patricia	Ruddy	NHS National Services Scotland
Eileen	Ryder	SFW Ltd
Aida	Sanchez	University College London
Gurmit	Sangha	Legal Ombudsman

Prof Steve	Saxby	University of Southampton
Gavin	Sayer	Department for Communities and Local Government
Nicola	Shearman	Office for National Statistics
Ollie	Simpson	Ministry of Justice
Ravinder	Singh	Technology Strategie Board
Victoria	Southern	Astra Zeneca
Adam	Steventon	The Nuffield Trust
Mark	Towers	Ministry of Defence
Lauren	Van Staden	Home Office
Pat	Walshe	GSMA
Ann	Wass	Department for Education
Ken	Watchorn	FCO Services
Colin	Watson	Watson Hall Ltd
Clara	Westbrook	IMS Health Ltd
Dr Edgar	Whitley	London School of Economics
Helen	Wood	Department for Education
<b>Speakers:</b>		
Christopher	Graham	ICO
David	Smith	ICO
Sir Mark	Walport	Wellcome Trust
Nicola	Westmore	Cabinet Office
Dr Kieron	O'Hara	University of Southampton
Dr Mark	Elliot	University of Manchester
Prof Paul	Ohm	University of Colorado
Barry	Ryan	Market Research Society
Marie	Cruddas	Office for National Statistics
<b>ICO staff:</b>		
Richard	Bailey	Solicitor
Jonathan	Bamford	Head of Strategic Liaison
Katie	Johnson	External Relations Officer
Judith	Jones	Senior Policy Officer
Thomas	Oppe	Lead Policy Officer
Simon	Rice	Principal Policy Adviser – Technology
Helena	Szehiridewicz	External Communications Manager
Lyn	Wibberley	Senior Policy Officer
Steve	Wood	Head of Policy Delivery