



Caledonian Research Foundation
Prize Lectures 2007

Can Information be Personal?

Baroness Onora O'Neill
President, The British Academy

Can Information be Personal?¹

Onora O'Neill

The title of this talk is deliberately provocative, and my claim to expertise is slender. I am not a lawyer but a philosopher, I have no special insight into UK data protection legislation, which is meant to protect personal information, and thereby aspects of personal privacy.² My practical experience of the legislation has led to no more than average frustration – that is to say, to considerable frustration but little clarity. However, I have chosen this topic because I think it is too important to be left solely to those with legal expertise. I shall argue that current legislation saddles us with a cumbersome, dysfunctional and sometimes incoherent approach to informational privacy, sketch an explanation of why the legislation fails to achieve its aims, and suggest some reasons why alternative approaches might do better.

Data protection legislation, in the UK as elsewhere, is based on the idea that we can protect informational privacy by establishing special systems for handling data or information that are personal, which we do not need to handle data and information that are not personal. I shall argue that this distinction cannot be well drawn, and that by making it the basis of legislation we have saddled ourselves with a deeply frustrating and in many respects impossible task. In effect, the legislation seeks to base informational privacy on an inept, if not impossible, classification of informational content. In arguing against this approach to data protection, I do not challenge the thought that privacy, including informational privacy, is important and should be secured. Rather I shall argue that it might be better secured not by regulating speech content, but by regulating speech acts.

Informational Privacy and Data Protection

Data protection legislation aims to secure informational privacy by regulating the 'processing' of specific types of information that are seen as intrinsically *personal*, or in some cases as both *personal* and *sensitive*. It does so by imposing obligations on those who hold the relevant type of information ('data controllers') and assigning rights to those to whom the relevant type of information pertains ('data subjects').³ In effect, the UK *Data Protection Act 1998* construes informational privacy as a matter of individuals having rights to control their 'personal' information by prohibiting its 'processing' for purposes to which they do not consent, unless there are special reasons for setting aside demands for prior consent (such reasons might include audit, or criminal investigations). The notion of 'processing' used in data protection legislation is something of a term of art, since it covers activity such as acquiring, organizing, altering, retrieving, consulting or using data. Indeed, the accompanying *Legal Guidance* to the Act states that 'The definition [of 'processing'] in the Act is a compendious definition and it is difficult to envisage any action involving data which does not amount to processing within this definition'.⁴

Given that almost any use of data is to count as 'processing', it is important to have a clear way of picking out *which* information is to count as *personal*, or as *personal and sensitive*, and so is to be regulated. Yet the Act is not helpful in explaining what personal data are. It states that they are data

1 This text formed the basis of a lecture given on 28th May 2007 at The Royal Museum, Edinburgh. I am grateful to the Caledonian Research Foundation for sponsoring this and other lectures given that week, and to the Royal Society of Edinburgh for organising them. The work draws on one strand of the argument of Neil C. Manson and Onora O'Neill, *Rethinking Informed Consent in Bioethics*, Cambridge University Press, 2007. The work that underlies this book and a range of related papers was supported by the Wellcome Trust.

2 For the full text of the Data Protection Act 1998 see <http://www.hms.gov.uk/acts/acts1998/19980029.htm> ; see also Data Protection Act 1998: Legal Guidance at: www.ico.gov.uk/documentUploads/data%20Protection%20Act%201998%20Legal%20Guidance.pdf

3 Schedules 2, 3 and 4 of the Data Protection Act 98 confer on data subjects rights of access to data held that pertains to them; rights apply to a Court to order a data controller to rectify, block, erase or destroy personal details if they are inaccurate or contain expressions of opinion based on inaccurate data; rights to ask data controllers not to process their data if doing so will lead to damage or distress; rights to prevent use of their data for direct marketing purposes (here the SNP recently faced some problems); and rights to compensation for damage or distress caused by breaches of the duties of data controllers.

4 See Data Protection Act 1998: Legal Guidance., p. 15
[Data%20Protection%20Act%201998%20Legal%20Guidance.pdf](http://www.ico.gov.uk/documentUploads/data%20Protection%20Act%201998%20Legal%20Guidance.pdf)

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.⁵

What are 'data that relate to a living individual'? And when do data and other information make an individual identifiable? Neither idea is particularly clear.

2. "Relating to a Living Individual who can be Identified"

What is meant by the phrase 'data relating to a living individual who can be identified'? At least one point is clear here: data protection does not extend to the dead, to whom it affords no protection. But everything else is far from clear.

Clearly, not all information that is true of a living individual counts as personal. A lot that is true of each of us is general information that is equally true of all or of many others. Each person has red blood cells, and each was born at some time in the past. Personal information does not cover this sort of general information that is true of all individuals. On the other hand, personal information cannot be confined to information that is *uniquely* true of a person to whom it pertains. For example, I was the first person born in the Townland of Aughafatten for over a century – but that fact isn't personal, and could be ascertained from the public record in the (Northern) Irish *Register of Births, Deaths and Marriages*. Clearly a lot of the information that we think ought to be private is not unique to the individuals to whom it pertains: everyone with cancer might well reasonably hold that this fact about their health is personal and ought to be treated as private. So informational privacy cannot plausibly be construed as privacy for those matters in which we are unique.

So the crucial element in the legislative definition of personal information must lie in the idea that certain information makes an individual *identifiable*. Yet it is not clear how we are to determine when an individual "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 'other information' held by, or likely to be held by, somebody holding (supposedly) personal information will vary, and the inferences that can be drawn and the identifications that will be possible will vary correspondingly. Even the most commonplace piece of information about an individual may be the crucial bit of evidence that makes him identifiable to those who hold other information which allows them to make an identifying inference. If the police are holding blond and brown-haired suspects, reliable evidence that the perpetrator had brown hair may make him identifiable and enable investigators to infer whom to release and whom to detain.

The emphasis on data that render individuals 'identifiable' by data controllers in the *Data Protection Act 1998* has led to great controversy, particularly in the area of medical research. On one view, if medical data are anonymised – or 'pseudonymised' i.e. reversibly anonymised – they will not count as identifiable, and so will not count as personal. The individuals to whom the data pertain cannot be identified by those without access to the code or key, since they will be unable to reverse the anonymisation. However, those with the key will be able to identify the individuals to whom the information pertains, and this fact leads others to conclude that reversible anonymisation is too little to satisfy data protection requirements.

This unclarity about what makes data *identifiable* and so *personal*, or *personal and sensitive*, can play havoc with medical research, where data are often used for *impersonal* ends. Epidemiologists, and those who do secondary data analyses, typically have no specific interest in particular individuals, or in finding out to whom the data that they use applies. However, the data must be indexed to identifiers in order to link different bits of information about the same individual. Without linked data, many lines of inquiry in epidemiology and many secondary data analyses would be impossible. Reversible anonymisation, which does not destroy the possibility of linking information, was traditionally thought to provide adequate informational privacy. However, the *Data Protection Act 1998* applies to all uses of

⁵ Data Protection Act 1998, Part I, Section 1. This formulation is closely based on that of the European Directive 95/46/EC which states that "personal data" shall mean any information relating to an identified or identifiable natural person ('data subject')" (Chapter 1, Article 2 (a)).

data in which the data subjects are 'identifiable' even by indirect means,⁶ and some interpreters of the legislation conclude that research may be done without consent *only* if 'personal' data are subjected to a stronger form of anonymisation, which removes links that *could be* used to identify data subjects. In that case the data must be *de-linked* or *irreversibly anonymised*, nobody can reverse the anonymisation and the data subject will be unidentifiable not only by the researcher but by others. Irreversible anonymisation or de-linking of data makes many sorts of research on human subjects in biomedicine and the social sciences impossible.

Moreover the burden of these requirements not only affects complex research, for which elaborate processes of consent might perhaps be devised and implemented. Exactly parallel problems can arise when a doctor wishes to revisit information about the treatment of a past patient in order to inform treatment of a current patient. Such valuable and routine activity will apparently breach data protection requirements, unless consent is obtained from all the past patients. Of course, sometimes consent can be obtained, but it is unclear why data about medical treatment used for impersonal purposes should be regarded as personal data. Each of us is always treated on the basis of information obtained by treating earlier patients, so why should any of us be thought to have a right to refuse to allow reversibly anonymised information about his or her own case to be used to inform treatment of future patients? My suspicion is that compliance is – fortunately – poor!

Interlude: Some Stories

The problems that I have sketched may seem arcane. In fact they are everyday affairs, as can be illustrated by few short stories that highlight some problems generated by data protection approaches to informational privacy.

1. Suppose that you are waiting at your GP's surgery and the receptionist calls out your home address in front of other patients. In doing so she discloses personal information, which the surgery legitimately hold for purposes connected with your medical care, but which they should not have communicated to others or used for other purposes without your prior consent. Yet on the other side of town there is an electoral register containing your name and address, which is there for the public to consult without needing your consent. So is your home address personal information? Or is it not? Note that in this case the information that was disclosed was not medical information, but simply personal information that happened to be held in a medical setting. **Conclusion: Data cannot be reliably classified as personal or non personal.**

2. The second story recounts an episode that happened at the GP's surgery where I am registered, which installed an electronic indicator board on which a notice flashed up the words 'Mr Smith – Wart Clinic'. The notice did not read 'Mr Smith has a Wart'. Arguably it disclosed no personal information pertaining to Mr Smith. Arguably it could have been interpreted as a request to Mr Smith to go to the Wart Clinic – possibly to take a message or to deliver supplies. Yet anyone reading it would probably infer on the basis of a routine understanding of the way things are done that the said Mr Smith had a wart, and was about to have it looked at, or perhaps removed. There was a certain amount of fuss about this notice, despite the fact that it did not strictly speaking state anything about Mr Smith's warts, or disclose personal information – but it did make Mr Smith identifiable as a patient with a wart. Since that fuss the indicator board carries more guarded notices such as: 'Mr Smith – Room 3'. What people come to know from a given piece of information depends on what else they can infer using available information and assumptions – which vary for different persons. **Conclusion: we cannot define personal information as information that makes someone identifiable.**

3. Thirdly, consider that time-honoured medical practice, the taking of a family history. Imagine that you are talking to your GP, who asks you whether any member of your family has heart trouble. Without any hesitation you start telling her some medical details about your relatives – without seeking their prior consent; and your GP then writes some notes containing (unverified and possibly inaccurate) medical information about living persons who have not consented to its disclosure or to its being recorded, and who do not even know that their health problems are now known to your GP, listed in

⁶ For definitions of identifiable data and of reasonably identifiable data see Department of Health, Confidentiality: NHS Code of Practice, 2003, p 9; <http://www.dh.gov.uk/assetRoot/04/06/92/54/04069254.pdf>; the Code has been endorsed by the Data Protection Commissioner.

your medical records, and may become known to others. Medical practice routinely uses information that is personal to A in treating B, thereby violating the requirement to use information only for the purposes for which it was originally provided. **Conclusion: it is normal to use information for purposes for which it was not originally obtained, and hard to prevent this.**

4. Finally consider the case of a doctor who treats a series of patients with an unusual disorder and wants to write a case note. Strictly speaking, the information contained in each patient's file was provided only for the use in that patient's treatment so should not be used for further purposes without prior consent. Of course, the identity of the patients referred to in a case note would be duly (reversibly) anonymised in the published case-note, and the patients would probably be pleased that information about the treatment of their disease is being made more available. Yet, strictly speaking, such publication will meet data protection standards only if each patient provides explicit consent to publication. **Conclusion: Clinicians and researchers routinely reuse and publish information that was collected in order to treat a given patient, thereby violating the requirement to use information only for the purposes for which it was originally collected unless there is specific consent to a further use.**

Each of these everyday stories is about a use of information that is taken to be personal for purposes other than those for which it was originally collected and held. If we are to decide how to handle these and many other cases, we need to be sure that we can distinguish personal from non-personal information, and that demands for consent to any further uses of information are workable. Yet this is surprisingly hard. What could be done?

Rethinking informational privacy

It is important to bear in mind the *reasons* why Data Protection legislation was enacted. Both the European Directive and *the UK Data Protection Act 1998* responded to changes in the way that we acquire, store and use knowledge. The acquisition, processing and linking of information have become amazingly fast and cheap, with significant practical implications for informational privacy. It was no doubt tempting to try to deal with this problem by restricting and regulating the use of specific types of 'personal' information, thereby securing a better standard of informational responsibility than prudence and economic competitiveness alone were likely to secure. So there were good reasons for enacting new protection for informational privacy.

However, looked at with hindsight, this objective might have been secured by other approaches. In particular, it might have been more coherent to focus on regulating what is done – above all on regulating communicative transactions – rather than regulating *all* types of activity ('processing') that use supposed, yet ill-defined types of informational content. Speech acts are more readily regulated than speech content.

An alternative approach could focus on the communicative actions and transactions by which information is obtained and communicated, and the norms and obligations relevant to such communicative transactions. Obligations of *confidentiality* provide a good example of communicative obligations that bear on what is done with information, rather than on the 'processing' of all information of some putative type. We generally think that communicative transactions must meet a range of standards. For example, they must be intelligible and relevant to their intended audiences, accurate and honest; the commitments entered into by means of communicative transactions must be observed. Ethical, professional and legal requirements for confidentiality provide good examples of obligations to use information *of any sort* only as agreed. Confidentiality may provide a more coherent and robust basis for securing informational privacy than can approaches that rely upon putative privacy rights over ill-specified types of informational content and their correlative obligations.

Obligations of confidentiality are generally said to hold where there is a well-defined – *not* necessarily *legally* defined – relation between two or more parties. They may hold between friends and relations; between business partners; between doctors and patients; between lawyers and clients; between bankers and account holders; between employers and employees; and so on. In a confidential relation, the confider discloses, or permits the confidant to acquire, information not taken to be a matter of public knowledge. The information acquired may or may not count as personal information, in various meanings of that obscure term. In return, the confidant assumes obligations not to use that knowledge to harm the confider, and not to communicate that knowledge to third parties without the consent of the confider.

The basic aim of the law of confidentiality was classically set out by Lord Denning, in discussing whether legal action could be brought for breach of confidence in cases where no explicit contract exists between confider and confidant. He took the view that the legal notion of confidentiality 'depends upon the broad principle of equity that he who receives information in confidence shall not take unfair advantage of it. He must not make use of it to the prejudice of him who gave it without obtaining consent'.⁷ Although the law of confidentiality traditionally governed confidential disclosure in professional and commercial contexts, recent court judgments extend its scope beyond formally constituted or legally recognized relationships to relationships of other sorts.⁸

Confidential relations and confidential communication – whether protected by law or not, can be of value to confider and confidant, as well as to third parties. For example, patients might be reluctant to seek medical treatment, and clients reluctant to seek legal advice, without confidential relations between client and professional. Businesses would be at risk unless employees could be told facts in confidence, which it would be damaging for competitors to know. Confidentiality and the laws that define it arose in well-defined relationships, but are now being interpreted more broadly by the courts as applicable in contexts where there is no formal relationship, but there is a reasonable presumption that certain information is not for wider consumption and will be not be *made available to others* or *used for other purposes* without the agreement of the confider. Informational privacy may be better protected by requirements of confidentiality than by ill-defined requirements for data protection.⁹

This protection reflects the fact that relations of confidentiality impose substantial obligations. A confidant may not *tell* others certain things that he has come to know, and in particular may not use that information to the disadvantage of the confider, unless the confider consents. If this obligation is met, confiders will have reason to trust confidants not to publicise or communicate information that was imparted in confidence without their consent. Like other trust relationships, confidential relationships may be supported by systems of accountability which add legal and regulatory force to obligations of confidentiality – provided that these systems of accountability are well designed for this purpose (and some are not).

The basic principles that underlie confidentiality are quite different from those that underlie data protection. Confidentiality focuses on regulating *types of action* – specifically – *types of speech act* – rather than on all 'processing' of *types of information*. An approach to informational privacy based on extending the law of confidentiality does not require anyone to determine which information is or is not *personal*, or which is *personal and sensitive*, and does not require anyone to determine just what it takes to make an individual identifiable to one or another party. Rather than defining and protecting intrinsically 'personal' content, confidentiality is a way of protecting content of *many types* that the parties to a communicative transaction seek to protect, have agreed to protect, or are required to protect. It can be invoked for specific aspects of professional, commercial or other relationships, and can be waived by seeking consent from the confider. Confidentiality also standardly receives second-order professional and legal backing.

7. Some Conclusions

Data protection legislation creates substantial difficulties for medical and social research, and especially for research that re-uses legitimately acquired, lawfully held data (so-called 'secondary use'). Such re-use is held to breach informational privacy rights, unless informed consent is given to the relevant further uses of the information. Given the breadth of the conception of 'processing' in the *Data Protection*

7 Lord Denning, *Seager v Copydex Ltd (No. 1)* [1967] RPC 349

8 Gavin Phillipson, 'Transforming Breach of Confidence? Towards a Common Law Right of Privacy under the Human Rights Act'. *Modern Law Review*, 66, 5 (2003) 726-758.

9 See note 8, and also Mr. Justice Scott, who notes that "the law of confidentiality can, in conjunction with the law of trespass and the law of nuisance, go a long way to remedy the alleged absence of a right to privacy under English law" in his introduction to *Confidentiality and the Law*, p. xxiii; Gavin Phillipson and Helen Fenwick, who argue that the legal 'doctrine of confidence is able to offer far more protection [of privacy] than is generally recognised' and explore how the legal notion of confidence can do work in protecting the Article 8 'right to respect for private life' in their 'Breach of confidence as a Privacy Remedy in the Human Rights Act Era'. *Modern Law Review*, 63,5 (2000), 660-693 (p. 662).

Act 1998, clinicians and researchers are apparently required to seek specific consent *even where the purpose of an investigation is not to find out anything about, or to do anything to, the individuals to whom the data refer*. It is hardly consoling that the legislation permits such investigation when all source subjects consent to them, and (in the case of medical research) exceptionally if permission can be obtained from the *Patient Information Advisory Group*. Obtaining further consent from all source subjects is often impractical, while selective re-consenting is likely to damage research findings by skewing their statistical basis. Moreover, even where it is in principle possible to re-contact and to seek renewed consent for some range of further work, doing so may not be feasible given the gaps between the informational complexity of the consent required and the real capacities and limitations – of human individuals to understand complex information.

I conclude that there are good reasons to rethink informational privacy and its enforcement. We could best do so by focussing on the epistemic and ethical requirements on communicative transactions, rather than upon supposedly distinctive types of information content. If we focus on communicative action and transactions, and so on speech acts rather than speech content, we can make use of a robust framework for thinking about epistemic and ethical norms for informational, and in particular for communicative action. Norms of epistemic responsibility, ethical norms, and second-order legal and institutional requirements that reinforce norms of both sorts, constitute obligations; they define and clarify rights. Action in accordance with such norms can protect the informational privacy that data protection legislation is meant to protect, without invoking the flawed and ill-defined assumption that some types of information have intrinsic ethical significance while others do not.