SCRIPT-ed

Volume3, Issue1, March 2006

Data Protection: Too Personal to protect?

Dr Mark J Taylor[*]

Abstract

This article analyses the application of data protection legislation to sensitive personal data, particularly to genetic data, and asks whether the present framework may be adequate to respond to the very sensitive issues involved .

DOI: 10.2966/scrip.030106.71

[*] Lecturer in Law, University of Sheffield.

1. Introduction

Article 1 of the Data Protection Directive 95/46/EC identifies as one of its objectives the protection of 'the fundamental rights and freedoms of natural persons, and in particular their right of privacy with respect to the processing of personal data.'[1] As the Data Protection Act 1998 aims to implement this Directive we can expect this to be one of the aims of the Act too: to protect the fundamental rights and freedoms of *identifiable* natural persons, in particular their right of privacy, with respect to the processing of data that *relates to* them.[2]

There is, however, reason to doubt the ability of the current law of data protection to consistently achieve this objective. In particular, an identifiable individual might enjoy no protection from the 1998 Act if data which can be related to them, and is capable of affecting their enjoyment of, *inter alia,* their right of privacy, is either,

Not *primarily* related to them but rather to somebody else *by the data controller,* i.e. the data controller considers the data to be 'about' somebody else, or,

The individual that the data relates to is not likely to be personally identifiable *by the data controller* but are identifiable by a third party (with access to the data)[3]

The current law on data protection encourages the view that specific data can only be the personal data of *a* particular individual (whom I shall describe as a 'primary' data subject to contrast with 'secondary' etc.) *and* that this individual must be identifiable by the data controller themselves. This leaves open the possibility that an individual's fundamental rights and freedoms can remain unprotected. It is possible for the relevant relationship between data and an identifiable individual to be judged from a perspective other than that of the data controller, and, if it is, then there are currently circumstances in which individuals could be linked to data, capable of affecting their fundamental rights and freedoms, without enjoying any rights in relation to the processing of that data.

2. An Example

The limitations of the law can be illustrated through the facts of a case reported first in the New Scientist magazine last year. The magazine told how a 15 year old boy in America identified a previously anonymous sperm donor as his biological father.[4] The boy had used the services of an internet-based company, called FamilyTreeDNA.com, which offers through its website a 'genealogy driven DNA testing service'.[5] As well

---

[1] The fundamental rights to be protected are stated within the Directive to be those 'in the constitution and laws of the Member States and in the European Convention for the Protection of Human Rights and Fundamental Freedoms' (recital 1).

[2] Italicised words draw upon the definition of personal data contained within the Directive: ''personal data' shall mean any information relating to an identified or identifiable natural person' (see below)

[3] This might extend to circumstances in which the data controller can anticipate, but is unlikely to effect, such identification.

[4] http://www.newscientist.com/channel/sex/mg18825244.200 (last accessed 30th November)

[5] http://www.familytreedna.com/.

as analysing subscribers' DNA for particular characteristics, the service puts people who share certain characteristics into contact with each other.

The company put the boy in contact with two men who had Y chromosomes closely matching his own. [6] While they were strangers to each other the genetic similarities between them suggested that there was "a 50 percent chance that all three had the same father, grandfather or great-grandfather". [7] The boy used the fact that the men had similar surnames to guess the surname of his biological father. He then used information about the date and place of birth of his biological father, which had been provided to his mother at the time she received the donated sperm, to track him down (using the services of another online company called omnitrace.com).

The case helps illustrate circumstances in which data held upon a database, while capable of linking to an identifiable individual and seriously affecting their privacy, would probably not be considered to be *their* personal data. In this case, the data that was held by FamilyTreeDNA.com was not linked to the boy's father by anybody other than the boy himself. Indeed, a European company, in the position of FamilyTreeDNA.com, would almost certainly have not considered the information *about the two men* to be the *personal data* of anybody other than the two men themselves. Even the boy, who *was* linked to the two men's data by the company due to their shared genetic characteristics, would not likely have been considered a *subject* of the two men's data and accordingly would have enjoyed no rights in relation to it under the law of data protection.

At first sight this might be considered quite proper. Clearly the data, as processed by the company, was not *'about'* the father. It might seem inappropriate for him to have the same rights in relation to the data as the person whom the data *was* actually 'about'. The point is however that *who* data is 'about', who it 'relates to', can be described as a matter of both context and of law. The fact that data is 'about' X in one context does nothing to prevent if from 'relating to' Y in another. There is nothing necessarily unforeseeable about this and there is nothing to prevent the law from recognising it and granting both X and Y rights in relation to it in appropriate circumstances. Indeed, failure to recognise the fluid nature of interpretive contexts, and the possibility that data may be used in more than one way, would leave the law incapable of recognising that more than a 'primary' data subject could have a legitimate interest in the processing of data that might be 'related to' them.

It is, in fact, entirely consistent with the *definition* of personal data given by the Data Protection Directive (and the 1998 Act) to recognise that data *could be* the personal data of more than one person simultaneously.

---

[6] http://www.newscientist.com/channel/sex/mg18825244.200 (last accessed 30th November)

[7] Alison Motluk "Anonymous sperm donor traced on internet" NewScientist.com news service (03 November 2005) accessible at http://www.newscientist.com/article.ns?id=mg18825244.200 (last accessed 05 November 05)

3. Somebody else's data could be my data too: the conceptual possibility of multiple data subjects:

The English law on data protection is largely determined by the Data Protection Act 1998. The 1998 Act was brought in to implement the European Data Protection Directive[8] (95/46/EC). 'Personal data' is defined by Article 2(a) of the Directive thus,

> *'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*

Clearly whether data can be the personal data of more than one individual turns on an interpretation of the term 'relate to'. [9] If data is capable of relating in relevant fashion to more than one 'identified or identifiable' individual then the data must be considered to be the personal data of each of them.

Unfortunately, the Directive does not explain how the term 'relate to' ought to be understood. There is a strong argument however that, given the aims and terms of the Directive, the *most restrictive* interpretation of 'personal data' would hold that data capable of *both* 'affecting an individual's enjoyment of their fundamental rights and freedoms' *and also* 'contributing to their identification'[10] would 'relate to' an individual in relevant fashion.[11] While the relative importance of the capacity to 'affect', or the capacity to 'identify', might be disputed,[12] any data capable of doing both has a strong claim to being regarded as 'personal data'.

In this context then, the case reported by the New Scientist demonstrates two things:

1. Genealogical data, capable of affecting enjoyment of an individual's fundamental rights and freedoms, can be related to an identifiable individual

---

[8] Full title is Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[9] The 1998 Act defines personal data as data "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" [emphasis added]. For a discussion of how this relates to the Directive's definition see further below.

[10] Any suggestion that the data must render an individual uniquely identifiable is denied by the fact that 'indirect' identification is sufficient: in order to be sufficiently identifiable it is not necessary for an individual be identifiable by the data alone, identification can follow locating the data in a particular context. From this we can reason that more than one individual could be indirectly identified from a particular piece of data if the context were to change.

[11] See Booth, Jenkins, Moxon, Semmens, Spencer, Taylor and Townend (2004) 'What are Personal Data?' (A study conducted for the Information Commissioner, available at http://www.informationcommissioner.gov.uk/cms/DocumentUploads/What%20are%20personal%20data%20research.pdf), esp. pp.78-79. The English Court of Appeal has suggested that 'In short, [personal data] is information that affects [the claimant's] privacy, whether in his personal or family life, business of professional capacity' see the comments of Auld LJ in *Durant* v *Financial Services Authority* [2003] EWCA Civ 1746, at para 28. For more on *Durant* see below.

[12] *Ibid.*

2. The individual who might be identified as the individual to whom the data relates can change as the interpretive context changes.

The significant point is that data can be capable of affecting an individual's fundamental rights and freedoms *and* possessing 'identificatory potential'[13] *in relation to more than one person simultaneously*.

Despite the fact that the definition of 'personal data' is consistent with the possibility of multiple data subjects there are nevertheless good reasons to suspect that, if the case had happened in Europe, the data held by FamilyTreeDNA.com would not have been considered to be the personal data of the Father. While it might clearly be said to have 'related to' him in way that was capable of affecting his privacy, and it clearly possessed identificatory potential (for his son used it to identify him), he would nevertheless not likely have been considered to be a 'subject' of the data.

Before exploring why the Father might not be considered a data subject *because* the data was considered to be *about* somebody else, there is another reason why he might not be considered a data subject in the circumstances of this case. If data is to satisfy the requirement of 'identifiability' imposed by the law, identification has to be more than simply a bare possibility. It is improbable that the father's identification would have been considered *sufficiently likely* for the data to have been regarded as his personal data.

4. 'Means reasonably likely to be used'

The Directive establishing the framework for European Data Protection law suggests that when establishing whether an individual is 'identifiable',

> *account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person (Recital 26)*

It seems that given that what the boy did "had never been done before"[14] any data controller in the position of FamilyTreeDNA.com would probably have been justified in not treating the data they held as the personal data of the father. The combined use of the databases and services by the boy would probably not have been considered 'means reasonably likely to be used'.

The case does raise interesting questions about what means are 'reasonably likely to be used' to draw connections between individuals and data in an era of growing online resources that are manipulable by increasingly sophisticated tools. There is also a question now of course, as this has happened, of whether it would still be reasonable

---

[13] The data has the potential to be placed within a context that enables it to be linked with a particular individual. It is important to note that within the context of information systems "Identification is the association of data with a particular individual" Clarke R. (1994) 'Human Identification in Information Systems: Management Challenges and Public Policy Issues' *Information Technology and People*, (Vol. 7, No.4, April, 1994, pp 6-37 at 8).

[14] Wendy Kramer, whose online registry for children trying to find anonymous donors had been used by the boy in the New Scientist report, quoted by Stein (2005) 'Found on the Web, With DNA: a Boy's Father' *The Washingon Post.com* (http://www.washingtonpost.com/wp-dyn/content/article/2005/11/12/AR2005111200958_pf.html) (last accessed 21st February 2006)

to regard the kind of searching done by the boy to be means sufficiently *unlikely* to be used again? It has been reported that,

> *Since word of the case emerged, several other offspring registered on […] http://www.donorsiblingregistry.com/, have clicked the link to the Family Tree DNA site (http://www.familytreedna.org/) in hopes of locating their biological father* [15]

Might data controllers in the position of FamilyTreeDNA *now* be considered to hold the personal data of those who *could* be identified by cross-referencing data (primarily) 'about' somebody else? I would suggest not. Not because identification would *still* be considered sufficiently *unlikely* (the question of relevant likelihood is one that doesn't actually need to be pursued here). They would probably not be considered to hold the father's personal data because, in cases such as the one described, even *if* the requirement of sufficient likelihood had been surpassed there is *still* reason to believe that the father would not have been considered to have been a data subject in his own right. Indeed, *even if the controller had themselves identified the father and linked him to the data* they held about the other two men (as they did with the boy) there is still reason to believe he would not be considered to hold any rights in relation to the processing of the data. While it is entirely *consistent* with the Directive to consider there to be multiple data subjects in relation to a single piece of data, the law serves to effectively undermine their recognition in practice.

5. The reality of multiple data subjects: marginalised secondary data subjects

The Data Protection Directive recognises that a data subject is entitled to certain rights in relation to their personal data. If the law is to encourage the recognition of multiple data subjects in relation to a single piece of data it must make clear how the responsibilities of a data controller are to be appropriately managed in the case of a *prima facie* conflict between their rights. Inadequate guidance on this issue will encourage a data controller to privilege the interests of the 'primary' data subject; this being the individual who has the more obvious claim. The interests of any secondary data subject, rather than being appropriately qualified, would be pushed out of adequate consideration.

To give just one example of why such conflict might occur and a 'primary' data subject's interests might be unduly privileged to the detriment of a 'secondary' data subject; Section IV of the Directive places a data controller under a legal duty to provide data subjects with certain information. Article 10 lists what must be made available to a data subject if the data is obtained from them. Article 11 lists what must be made available to a data subject if the data has not been obtained from them.

> *The information referred to within Article 11 includes (a) the identity of the controller and of his representative, if any; (b) the purposes of the processing; (c) any further information such as the categories of data concerned, the recipients or categories of recipients, the existence of the right of access to and the right to rectify the data concerning him.*

---

[15] *Ibid.*

The possibility of multiple data subjects raises in turn the possibility of both Article 10 and Article 11 placing responsibilities upon a data controller in relation to the same piece of data but in relation to *different* data subjects. It is easy to imagine circumstances, such as with genealogical data, where the provision of information under Article 11 to a 'secondary' data subject would constitute a *prima facie* breach of a 'primary' data subject's right of privacy.

Article 11 provides however that this information *must* be provided to the data subject unless it is 'impossible', 'would involve disproportionate effort', or, 'if recording or disclosure is expressly laid down by law'. No mention is made within Article 11 of what should happen if, in the circumstances, providing the information would constitute a violation of a 'primary' data subject's fundamental rights and freedoms.

Article 13 of the Directive, titled 'Exemptions and Restrictions', *does* provide for Member States to restrict the scope of the Directive and of the scope of the obligations provided for in Article 11, *inter alia,* in cases where the operation of implementing legislation would otherwise conflict with the fundamental rights and freedoms of others. [16]

It is therefore *open* to Member States to appropriately manage any conflict through their domestic laws, but, if they fail to do so, the Directive serves to encourage the marginalisation of 'secondary' data subjects. This is because, in circumstances where Member States have not taken the opportunity given by Article 13 to appropriately qualify the rights that a secondary data subject would otherwise have, *denying* the status of 'data subject' to anyone other than the 'primary' data subject might be the *only* way to protect their fundamental right to privacy.

The effect of this however is to deny a 'secondary' data subject *any* rights in relation to the data processed. Even in circumstances where recognition of them as additional data subjects would not conflict with the interests of a primary data subject (or indeed in any circumstance where *their* interests would be accorded priority in the case of such conflict), their interests would be denied.

---

[16] The Article states that, 'Member States may adopt legislative measures to restrict the scope of the obligations and rights provided for in Articles 6 (1), 10, 11 (1), 12 and 21 when such a restriction constitutes a necessary measures to safeguard, […](g) the protection of the data subject or of the rights and freedoms of others.

It is possible to imagine other potential conflicts[17] but, hopefully, this one example illustrates how a failure to clearly specify how the interests of multiple data subjects should be managed encourages a restrictive interpretation of the phrase 'relate to'. The possibility of multiple persons having rights in relation to the data has to be avoided, absent action by the Member States, for the sake of the rights of a primary data subject. We can in fact see a move away from recognising 'secondary data subjects' within recent English case-law.

6. Focussing on Primary Data Subjects

In 2003 the Court of Appeal in England heard the case *Durant* v *Financial Services Authority.* [18] Through this case the court suggested that two notions might be of assistance when establishing whether a piece of data crosses the threshold of 'personal data'. It is the second notion that is of relevance here[19],

> *The second [notion] is one of focus. The information should have the putative data subject as its focus rather than some other person with whom he may have been involved or some transaction or event in which he may have figured or have had an interest […]* [20]

The suggestion that a data subject should be the 'focus' of the data seems clearly to undermine the possibility of recognising multiple data subjects. It denies identifiable individuals rights in relation to the processing of data that can be linked to them in ways that could affect their fundamental rights and freedoms, especially privacy. Not only does this prevent the legitimate interests of putative secondary data subjects from being taken into account but it also appears to unduly privilege the perspective of the data controller. As can be seen from the case described, the fluid nature of interpretive

---

[17] For example, Article 7 states that data may only be processed legitimately if one of a number of conditions is satisfied. The first of these is that 'the data subject has unambiguously given his consent' (Article 7(a)). The Directive does not specify how this requirement shall be understood in the case of multiple data subjects. We must consider whether it is necessary for each to give consent or whether it is sufficient for just one of them to do so. Part f of Article 7 does provide that processing may also be legitimate, in the absence of consent, if it "is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)" (emphasis added). This would seem to support the idea that consent need not be obtained from all data subjects, if processing would be consistent with the 'legitimate interests' of a data subject (being an individual to whom the data are disclosed), for as long as 'the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)' are appropriately protected. Protection would seem to require consent if it might be obtained, but, if it were impracticable, or, obtaining consent would present a disproportionate risk to the fundamental rights and freedoms of another, then part f would seem to justify not seeking it. The fact that this is uncertain however leaves open the possibility that different member states would adopt disparate practices and the harmony that the Directive seeks would not be achieved.

[18] [2003] EWCA Civ 1746; [2004] F.S.R. 28; Times, January 2, 2004; 2003 WL 22826914

[19] The first [notion] is whether the information is 'biographical in a significant sense', that is, 'going beyond the recording of the putative data subject's involvement in a matter or an event that has no personal connotations, a life event in respect of which his privacy could not be said to be compromised', Auld LJ, *ibid* at para 28. It is certainly arguable that an event might be considered a 'life event in respect of which' more than one individuals' privacy could be simultaneously compromised.

[20] Auld LJ. *Op.Cit. n*. 17 at *para* 28

contexts renders such a static and one dimensional attitude particularly ill-equipped to deal with modern information technologies.

This is not the only way in which English Law seems to unduly personalise data from the perspective of the data controller. In a way that seems contrary to the requirements of the Data Protection Directive it also limits the notion of an 'identifiable individual' to one who might be identified *by the data controller* his or her self. This extends to cases where identification by a third party might be entirely foreseeable.

7. Identification is foreseeable by a third party but unlikely by the data controller

The 1998 Act states,

> *'personal data' means data 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,*
>
> *and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual* [21]

This definition appears to vary in at least one significant respect from that given in the Directive. The Directive's definition (see above) does *not* explicitly state that the data subject must be identifiable *by the data controller*. In fact, as noted above, recital 26 to the Directive states 'account should be taken of all the means likely reasonably to be used either by the controller *or by any other person* to identify the said person' [emphasis added].

The 1998 Act would seem to indicate that the data in the case described could not be the Father's personal data because he had not been identified, and was not likely to be identifiable, *by the data controller*. The Directive on the other hand would seem to suggest that so long as identification were sufficiently likely (i.e. through such means as were reasonably likely to be used) by *anyone* then the data ought to have been considered the Father's personal data (if other requirements were satisfied).

8. What does this tell us about how the system needs to be reformed?

It was a principle of the Directive that data protection legislation should assist Member States of the EU to "protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data." [22] The facts of this case help to illustrate some circumstances in which

---

[21] s1(1) Data Protection Act 1998.

[22] Article 1(1), *ibid*.

there are reasons to doubt the ability of the Directive, and the 1998 Act implementing it, to adequately provide this protection.

If information that I submit to a database is capable of assisting in the identification of my relatives, and telling you something significant about them, then the information should be regarded as *their* data as well as mine. This assumes that the information would have to be held in a context that would enable their identification as well as mine by 'means likely reasonably to be used either by the controller or by any other person', but, this might be quite possible and foreseeable. An understanding of what means *are 'likely reasonably to be used'* must be kept under review as information technology continues to progress: it is however likely that the class of 'secondary data subjects' will not shrink as technology allows more sophisticated searching and cross referencing.

Given the fluid nature of interpretive contexts it may well be that the law has made a mistake in trying to personalise data protection. It has personalised data protection in a way that undermines the possibility of individuals other than those whom the data is most obviously about (arguably from the perspective of only the data controller) having their interests taken into account when a data controller is considering their duties under data protection regulation. English Law has *also* personlised it by making (likely) identification *by the data controller* a *sine qua non* of personal data.

It might have been preferable for the law to have sought to ensure that individuals' fundamental rights and freedoms are protected by *requiring* reflection upon the possible *impact* that the acquisition and use of *any* information might have upon *any* individual's fundamental rights and freedoms. Such reflection might be prompted by a realisation that any acquisition and use of information that *did* unjustifiably infringe an individual's fundamental rights and freedoms in a foreseeable way would be actionable.

Acknowledging such a general responsibility within law need not place any undue burden upon persons. In the vast majority of cases a moment's reflection might provide absolutely no reason to suspect that the acquisition or use of particular information would in any way engage another's fundamental rights and freedoms. A putative data controller would then have a good defence if it subsequently transpired that, in a way that it was not reasonable to have expected them to anticipate, their actions had in fact done so. To place a general responsibility upon persons to act in a way which is consistent with certain interests held by other people is not unusual within law and need not be particularly onerous in fact.

Even though the law did not take this approach it is *still* possible for the law to incrementally develop in this general direction. The possibility of multiple data subjects is entirely conceptually consistent with the definition of personal data provided within the Data Protection Directive, and, if Member States were to take action to appropriately manage any potential conflict of interests so as to ensure fundamental rights and freedoms were protected, the category of data subjects could be greatly, and appropriately, expanded.

The English law on data protection is in particular need of reform. This is not only because it fails to take adequate account of the interests of other than those whom are the 'focus' (as in *Durant)* of the data held by a data controller (even when data is capable of affecting another identifiable individual's fundamental rights and freedoms in entirely foreseeable ways). The English law of data protection also inappropriately

releases a data controller from *any* responsibilities in relation to putative data subjects who might foreseeably be identifiable by a third party but not by the data controller his or her self given the information that is *likely* to come into their possession. This is irrespective of the harm that might be expected to follow.