Some thoughts on Government handling of confidential data

I started emailing Henry Porter of *The Observer* about IT issues around the identity and DNA databases in August 2007, because I felt that it was important to have a rational debate about them rather than one based on scientific and technical ignorance and an instinctive hostility to a Government that seems hell-bent on gathering and storing every possible scrap of information about every one of us.

Life got in the way of this endeavour. Then came the catastrophic cock-up involving the loss of discs containing so-called 'Child Benefit data', which was a lot more than that because it included names, addresses and – appallingly – bank account details belonging to millions of individuals.

This got me started all over again, moving me away from what would actually have been a *defence* of a well-designed and competently executed and managed identity database. HM Revenue and Customs had clearly demonstrated that Government departments are so incompetent in the management of data that there should be a moratorium on the development of all major new information systems until a competent Government data-management authority has been put in place. Above all, all planning for an ID database should be put on indefinite hold.

Panic reaction? I don't think so. There is no way that it should have been technically possible – let alone administratively acceptable – for a junior member of staff at HM Revenue and Customs to assemble the lost data at all, let alone download it onto a disc. As for being able to send the disc to another organisation – well, the mind boggles.

Since this first shambles, we have had a succession of similar cock-ups, peaking (in my view) with the loss of a hard-disc containing confidential data about UK citizens in the USA.

First, nobody other than those with very high-level system administrator privileges – in this case the people who designed, created, developed and managed the database systems – should be able to access databases containing any data that can be linked to individual citizens other than by using standard database queries which have been designed for specific approved functions and executed to the necessary high standards of security.

Second, access to every standard query that draws on a Government should be restricted to specified groups of users who have a clearly-demonstrated need to access the information it extracts.

Third, the output of these queries should not be downloadable to any portable media (floppy discs, CDs, DVDs, external hard drives or even paper, apart from the hard copies which would often be the end-product of these queries) under any circumstances whatever. I believe that the individual concerned must simply have been able to save query output to a CD or DVD rewriter drive on his or her own desktop PC, when in any half-secure environment no floppy drives and only read-only CD-ROM drives are fitted to desktop systems.

If data retrieved from a Government system needs to be made available in a department other than that which owns the query that produces it, it should be filtered to ensure that only the data needed by the destination organisation is transmitted – what possible use could the Audit Office have for individual names, addresses *and bank-account* details? Such data should then only leave its host network in encrypted form via a secure link and should only be accessible to properly authorised recipients.

All this is very basic security stuff that should be blindingly obvious to anyone managing confidential information. Whichever official has overall responsibility for data security in HM Revenue and Customs should not just have been sacked for this fiasco but prosecuted under the Data Protection Act. So should those ultimately responsible for the many data-losses which have followed.

If, as it appears, all the data lost was in the possession of HM Revenue and Customs, this tells us that we need a total review of how the Government stores data.

The cardinal rule should be that each item of data about a citizen *should only be stored in one place* (this is actually one of the most fundamental rules of sound database design anyway) and that it should be the responsibility of the department most concerned with it. If this principle had been followed, the child benefit database would not have contained names, addresses or dates-of-birth – let alone bank account details: the only link to individual citizens would be an identifying code such as the National Insurance number and the only data about individuals would presumably be a string of dates and amounts recording Child Benefit payments. If the other data were needed for whatever the National Audit Office was doing, it would have to be extracted from data controlled by other departments, and could therefore not have ended up on the wandering discs even if their misuse were sanctioned.

So how could the data managed by different departments be split up?

Taking the 'lost' data first, obviously the actual Child Benefit data (amounts paid, to whom and when) would be the responsibility of HMRC and each person's record would be identified by his/her National Insurance Number. The name of each individual would be obtained by sending a query to the 'Identity Database' (in quotes because all these data stores might actually be tables within a single database), using the NI number as a key. The address would be found from the Electoral Register data. And the bank details would be the province of HM Paymaster General's department, which is (or should be) responsible for all payments sent out to (and possibly taken from) citizens.

This approach actually offers a positive argument for an identity database. Following the pattern outlined above, the ID system would have the sole purpose of establishing whether or not an individual is who he or she claims to be, so it would contain only a very limited number of data items – the NI number (of course, as I have now established that this should be the key that links all the different bodies of data), the full name (the date- and place-of-birth, along with parents' identities, should belong to the General Register Office) and whatever

actual identifying data is deemed to be needed to establish that a citizen is who he or she claims to be: a digitised photograph (also to be used on passport and driving licence), iris-recognition data, fingerprint data and DNA 'fingerprint', which contains no genetic information that could be used for any of the more sinister purposes envisaged by Henry Porter and many others, since the bits of DNA used are actually redundant in the individual's genome – they don't actually make anything happen in the development of the organism.

In my model, there would be a National Identity Agency responsible for maintaining *and protecting* the identity data.

The police would have a range of standard queries enabling them to check photographs, fingerprints and iris images against names, with the DNA fingerprint as a last resort. Quite how they would go about dealing with forensic use of DNA would probably be controversial, but they could submit a profile for matching – just as they currently do with fingerprints – and then double-check when the suspect was found using the other three kinds of data.

Since the police themselves collect DNA samples and process them to produce DNA profiles, a system needs to be put in place to ensure that once this has been done the original sample (which *does* contain genetic data that could be misused) is immediately destroyed, failure to do so being a serious criminal offence under a revised Data Protection Act.

Of course, there is also a major issue around *building* the identity database. How does the Government propose to ensure that the photographs, iris images, fingerprints and DNA profiles it collects are linked to the correct individuals in the first place? This is a chicken-and-egg problem, since it will be building the first supposedly foolproof means of establishing identity!

Addresses would be stored in a national electoral register database.

Bank account details would, as stated above, be in a database maintained by HM Paymaster General to service only departments authorised by citizens to debit or credit their accounts, and they would only be accessible to systems doing the debits and credits.

I could go on, but I'm only trying to establish the general principles:

1. Each item of personal data would be managed by one Government department, which would be responsible for maintaining and protecting it.
2. Any legitimate process would be implemented by a dedicated query. This would get the citizen's name from the ID system, address from the electoral roll system and so on. Addressed letters and envelopes, tax-return forms, pension statements etc would be printed by dedicated queries and no data from external sources would or could be retained by the department concerned, temporary files on local hard drives being automatically deleted at the end of the query.

There would be other benefits to a structure like this. For example, if a change of address was notified to any department, it would be entered via a dedicated

query *only* in the electoral register system; the new address would instantly become available to all departments' systems and would be checked periodically against electoral registration forms returned by householders. Obviously previous addresses would have to be be retained, with the dates when changes were notified, in case of fraud or error.

At this point I need to state that I have developed these ideas so far despite – or perhaps because of – the fact that I am *not* an IT professional. However, I have been a web developer for BT, a local education authority and the NHS, variously employed and as a freelance, for 12 years (almost as long as there has been a real web). I have developed databases to provide online-editable content for websites, including designing and building from the ground up an entire content management and website administration system providing sites for GP practices.

I would argue that this demonstrates my competence in database design and administration *and* my complete freedom from the preconceived ideas that govern the work of professional database developers.

My concerns about security are informed in part by my experience in the NHS. I have been given administrator access to entire local NHS database servers protected by only the most rudimentary generic logins shared with hundreds of people (admittedly these did not contain any patient data – as far as I was aware). I have been given fairly secure administrator access to a local NHS network - including to the same database servers – from home, via the Internet.

All this has been without a scrap of formal training. I am about to start drawing my State Pension, but I'm sure there are kids not long out of – or maybe still *in* – school who can do, or could rapidly learn to do, everything I can do and a lot more besides.

Finally, the Data Management Authority I proposed at the beginning of this email. It would need serious teeth, so it should be headed by a very senior civil servant ultimately answerable to a minister of Cabinet rank (definitely not in the Treasury, because IT is about a lot more than numbers these days and it is high time we got rid of the automatic assumption that only the money-men should be allowed to control computers). It should be regulated by the Information Commissioner's Office and overseen by a Parliamentary Select Committee. It should recruit a consultative group of real experts, including nominated representatives of major IT companies – experts in hardware, software, database design and data security – who would be well paid for their input. However, this would be as far as the private sector got: I would like to see a highly-paid team of in-house developers and administrators building and maintaining a completely new Government IT structure, beginning – as a matter of great urgency – with all systems handling the personal data which is the property not of Government but of citizens. This would need ultimately to be hosted on a new network of supercomputers, preferably scattered around the country and linked to each other and to client networks by high-capacity landlines.

The entire staff of this Agency should be subject to beefed-up data-protection legislation with the same penalties as those under the Official Secrets Act (if state secrets require such a high level of protection, don't citizens' secrets?).

Then maybe we could trust Government to introduce ID cards and a universal DNA database.

I did email an early version of this document to Henry Porter at The Observer. He did not acknowledge it. Subsequently, I emailed it to Professor Sir Alec Jeffreys, inventor of DNA fingerprinting, who replied within minutes agreeing broadly with what I had written. I forwarded this to Mr Porter, who then responded quite quickly with a brief sentence which indicated clearly that he was not interested in informed, rational debate on the subject: 'I happen to think that all personal information is exactly that - personal. it belongs to the individual. I do not believe in large government databases- certainly not a national DNA database or a NIR or anything that remotely resembles it. This is because government end up abusing data in one way or another whether by losing it, misuing it or selling it.' *I thought this a poor response, though in the light of recent events the final sentence makes a valid point.*

Paul Marsden
19 December 2007