SHARE AND SHARE ALIKE

The UK government's plans to share data more freely are bound up in a knot of political, social, legal and technical issues

by Christine Evans-Pughe

TWO CHILDREN died in Soham, Cambridgeshire, in 2002, in part because national and local police intelligence systems were unable to share information effectively, according to the 2004 Bichard Inquiry into the murders. Ian Huntley, their killer, had been accused of sexual offences against children, but the allegations were held in a different computer system to that used to check his suitability for working in schools. In 2003, the Victoria Climbié Inquiry drew attention to similar issues surrounding her death, pointing the finger at the complex data-sharing arrangements that allowed her case to slip through the net.

Human, not computer, error was at the heart of these appalling cases, but the inquiries' findings helped strengthen the government's case that data held on individuals by its departments and other public bodies should be more widely shared. A 2002 report by the Cabinet Office, entitled 'Privacy and Data-Sharing – the Way Forward for Public Services', had already explained the need for new powers to share data, to make e-government possible. The report also mapped out how these powers could easily be implemented using 'data-sharing gateway' clauses within relevant legislation.

Government departments have since presented ever more detailed ideas about sharing data. The Cabinet Office followed up its 2002 report in 2005 with a white paper called 'Transformational Government Enabled by Technology'. Earlier this year, the Cabinet Office's Ministerial Committee on Data Sharing provoked outrage from privacy campaigners by releasing a position statement saying that "information will normally be shared in the public sector, provided it is in the public interest". In September, the committee followed up with an 'Information Sharing Vision Statement' describing current projects and future plans.

The government's desire to share data is beset with technical, legal, political and social issues. It also goes to the heart of a society's sense of privacy. UK citizens tend to believe they have a right to privacy, in their personal and family lives, their homes and their correspondence, without interference from a public authority, unless life and limb are threatened. The new plans for personal data subvert these beliefs, outlined in the Human Rights Act, in favour of active risk reduction for a common good defined by the state. The arguments for data sharing have become easier to make in the light of child abuse and terrorism cases, but implementation problems remain.

LEGAL ISSUES

At the same time that the police's ability to share data has been hampered by the lack of a well supported national IT strategy, other public sector bodies worry about legal constraints. Data protection law is often blamed for the problem. However, the real issue is that there are often specific statutory restrictions on data sharing, such as restrictions on the ways local authorities use council tax data. The principles of administrative law are another hurdle, since they say that public bodies cannot do things

outside their express statutory, or implied, powers.

"The common law of confidence is also a barrier, especially in the health service, where there have been doubts about whether medical data can be shared for services in the public good, such as public health surveillance and cancer registries," says Ruth Boardman, co-head of the international data privacy practice at law firm Bird & Bird.

Yet public data sharing continues to spread. The Cabinet Office's 2002 Privacy and Data Sharing report paved the way for a big change in government powers. By adding data sharing gateway clauses to a number of pieces of UK legislation, data such as health records, tax returns, welfare benefits, law enforcement records, driving licence information, passport data, and even car insurance details are now available for various agencies to share.

One of the most controversial projects is the Children's Index, which springs from a proposal in the Children Act 2004 to list every child with details of their health, education, learning difficulties and youth offences, and links to local authority and social services. Data on the children of public figures will have special protection status, raising concerns about data security, because all children should have this level of protection but, as the proposal stands, will not. This is already an issue with the ID and passport databases, which have suffered five security breaches in five years by staff using their access privileges to conduct unauthorised checks. There is already a significant black market for personal information, according to a recent report by the Information Commissioner.

Why is the government so keen to share data among its departments? In part it is because powerful business-oriented 'intelligence' software is now available to sift, analyse, and match vast numbers of electronic records, to categorise groups of people and their behaviours, and then forecast or intervene on the basis of this information. This software is already used by supermarkets, online bookseller Amazon, and by the Universities & Colleges Admissions Service to help universities pick candidates. For a government, the software opens up the prospect that the information it gathers on its citizens can be used to target policy.

ACCURACY AND ANALYSIS

There is, of course, the problem that if you put garbage into such a system you'll get garbage out. Pooling, sharing and matching data from diverse sources is an enormous and costly task: the NHS's data-sharing project will cost an estimated £12.4bn, according to the National Audit Office. The problem is well illustrated by the 43 police forces across England and Wales, which typically have separate databases for crime reporting, custody, case preparation, and fixed-penalty offences. The Metropolitan Police has separate databases for the capital's 32 boroughs, and other relevant data is held

continues p34

DATA-SHARING PROJECTS

National Programme for IT (NPfIT)

The biggest data sharing project in the UK connects GPs to hospitals and allows electronic care records to be shared among health professionals. NPfIT uses a data sharing gateway clause in the Health and Social Care Act 2001. The £2.3bn initial cost estimate is expected to reach £12.4bn. Accenture, a main contractor, walked out of the project in September, blaming iSoft's late delivery of the central information management software. Earlier in the year Accenture wrote off $450m because of 'significant delays' in the programme.

The National Automatic Number Plate Recognition Network

A network of 3000 cameras that now photographs all vehicle number plates and compares them with multiple databases, including the Police National Computer, Revenue & Customs, licensing, and commercial motor insurance databases. This exists because of gateway clauses in the Anti-Terrorism Act of 2001, and the Serious Organised Crime and Police Act of 2005, which allow for the disclosure of vehicle insurance.

The National Identity Register/ Citizen Information Project

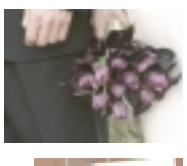
A one-stop shop for basic details. When moving house, a citizen would register the change online and all records, local authority, central government, DVLA and Inland Revenue would be updated. Each record will potentially include an audit trail, accessible for law enforcement agencies, recording all transactions with public and private services. This is allowed through the ID Cards Act 2006.

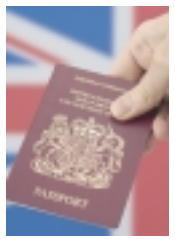
IMPACT

The £367m IMPACT project will link police information nationally, and connect with the Criminal Justice Information Technology organisation's shared services network, CJS Exchange.

Your digital footprint, from cradle to grave

This diagram, based on one produced as planning for a Citizen Information Project, shows how many government agencies hold data on us. It also illustrates the concept of data creep: when the original was drawn, the Citizen Information Project and the National ID Register were separate functions (as shown). The two projects were merged as part of the 2006 ID Cards Act, although this wasn't disclosed to Parliament until afterwards.





Register birth

Register adoption

Register death

Apply for state pension

Register emigration

Registered on police intelligence

Convicted of criminal offence

Convicted of civil offence

CRB check (via Police / LA)

Register divorce, name change

Register marriage

GRO

DWP

?

Police

?

Police

PNC

CRB

?

GRO

GRO

Allocate NHS No

NHS

DWP

?

Allocate Child Reference Vo. (Dormant NINO)

BABY ON BOARD!

Add to, transfer or remove from Child Register

Register with doctor

NHS

Start in state education

DfES

LA

Child at risk report

LA

Apply for passport
Renew passport
Replace passport

UKPA

NIR

Apply for Identity Card
Renew Identity Card
Replace Identity Card

Citizen Information Project data spine and population register

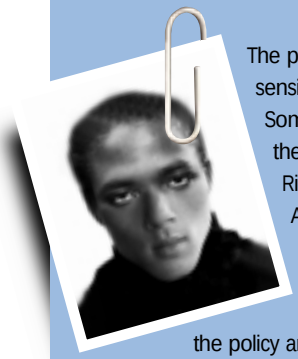
[Alamy]

Register to vote
LA
CORE
Register for local services
LA
Register as disabled
Register as unemployed
DWP
Register as carer / guardian
Claim maternity / child benefit
Apply for NINO
IND
Apply for asylum
Apply for permit to work
Direct Gov
Register on Govt Gateway
NLPG
LLPG
Create / demolish property
NIR
LA
DWP
DVLA
NHS
Advise 'known as' name change
Change Address
DVLA
Apply for driving licence
Renew driving licence
Replace driving licence
DWP
Convert Child Ref No to NINO
DfES
Register with Connexions
LA
Apply for student grant
SLC
LA
Apply for student loan
Employer
Start / leave job
IR
Complete tax return
Claim tax credits
LA
LR
Buy / sell house (council tax)

KEY

CRB
Criminal Records Bureau
DfES
Department for Education and Skills
Direct Gov
Portal to UK government information
DVLA
Driver and Vehicle Licensing Agency
DWP
Department for Work and Pensions
GRO
General Register Office
IND
Immigration and Nationality Directorate
IR
Inland Revenue
LA
Local Authority
LLPG
Local Land and Property Gazetteer
LR
Land Registry
NHS
National Health Service
NINO
National Insurance Number
NIR
National Identity Register
NLPG
National Land and Property Gazetteer
PNC
Police National Computer
SLC
Student Loans Company
UKPA
UK Passport Agency

PROTECTING PRIVATE DATA THROUGH TECHNOLOGY



The proliferation, use and potential misuse of sensitive data is clearly a subject of global interest. Some of the most useful research going on is in the US within PORTIA (Privacy, Obligations, and Rights in Technologies of Information Assessment), a five-year project centred on Yale and Stanford Universities. The project is looking at how to solve the technical challenges of handling sensitive data, and the policy and legal issues facing data subjects, data owners, and data users.

"The cryptology and security research communities have always concentrated on preventing the transmission of sensitive information. They encrypt it, put up access controls and other preventive approaches," said Joan Feigenbaum, the Henry Ford II Professor of Computer Science at Yale University, and Yale principal investigator of the PORTIA project. "In many senses, they are looking at the wrong end of the problem. Many people are willing to accept that sensitive information about them is transmitted from one machine to another but they want some assurance, legal, technological and social, that the information will be used in a legitimate way."

One of the project's major research themes is data-analysis methods that preserve privacy.

"It means you can use a data item in a calculation without having to reveal it. You split up the calculation in such a way that no one party who is doing the computation ever sees the raw data in their entirety. You could apply that to a distributed database, a set of data that is owned by disparate parties, for very specific applications," said Feigenbaum.

The project is also working on formal definitions of privacy, focusing on the idea of 'contextual integrity' developed by PORTIA investigator Helen Nissenbaum, an associate professor at New York University.

Contextual integrity is a way of thinking about privacy, which has now been turned into a formal privacy language by computer scientists at Stanford. The World Wide Web Consortium already has a Platform for Privacy Preferences Project (P3P), to enable websites to express their privacy practices in a standard format. Unlike P3P, though, the Contextual Integrity language can express privacy legislation and enforce compliance with it. The team at Stanford has managed to express, mostly successfully, several US privacy laws, including those covering healthcare, financial and children's online activities.

"Contextual Integrity says there are norms of information flow and problems occur when you violate these norms," said Nissenbaum. Key to the new definition is the 'principle of transmission', the basis under which information is given. In other words, was the information bought, shared voluntarily, taken by legal force, or handed over confidentially.

"With a formal privacy description, we can evaluate laws and practices and data arrangements in a rigorous way and assess whether the norms of flow have been violated. Then you can ask, is it acceptable or not, or what harm might be caused?" she said.

The privacy language uses temporal logic, which can express detailed constraints on the past and the future. Adam Barth, a computer scientist from Stanford who has worked on the language, said: "With temporal logic we can express that a certain communication is permitted if the data subject has previously consented and the recipient of the information keeps the information in confidence – in other words, that, in the future, he or she must not transmit the data to another person. These detailed temporal conditions are essential in correctly expressing privacy laws."

in the Police National Computer and national databases of DNA and violent offenders.

"For any government organisation or department, historically you'll find the situation is similar," says John Tidy, an IT strategy consultant who works with central government and spent many years working in the police sector. "For various reasons, common infrastructures and common architectures – all the things that would make it easier to transfer data from one department to another or make data available in a meaningful manner – usually don't get built into the original specification."

Often the only way to share records is to employ expensive software tools to transform one dataset into a format readable by another.

"You have to do that for every different type of data user who doesn't have your data. It adds immensely to the technological complexity," adds Tidy. Inaccurate, out-of-

date and missing information has to be sifted out and all the records that might refer to the same person matched and resolved.

"Not only do you need to find technical methods of understanding that you're talking about the same subject or person or activity, but you need to be able to add some level of sophistication and nuance. What is information that has been written down – for example a date of birth – and what is simply intelligence?" says Peter Dorrington, head of fraud solutions in the UK for the US firm SAS, which sells tools for building data warehouses and data analysis.

When trying to match records from many different databases, perhaps as part of a fraud prevention exercise, bad data will

quickly propagate; so the more data sources being compared, the greater the ramifications of any error. Some years ago, for example, a bank changed its customer questionnaire so that the query 'Have you ever been bankrupt?' became 'Have you ever had financial difficulties?' The bank forgot to tell the companies receiving the data that its meaning had changed, so many customers found that they were marked as bankrupts in other systems that used the data.

It may be possible to take a bank to task, but it's more difficult to do the same to a government.

"Even on something you think would be as robust as ID, it's usually a sliding scale of confidence between various systems that this is the same identity and it's not being shared," says Dorrington. This is particularly problematic for fail-safe systems, such as the proposed national ID register, which are designed to assume the worst. Innocent people can get flagged as criminals, as happened earlier this year when several hundred people were denied jobs and university places because the Home Office's Criminal Record Bureau database incorrectly labelled them.

"The challenge, once you go to a national ID register, is that even a small percentage of error results in thousands of people having problems dealing with it on a day-to-day basis. And if it isn't failsafe, then why have it?" says Dorrington.

DATA CREEP

You don't need identifiable personal information to understand trends and patterns, but British government data sharing focuses on pinpointing individuals. Some government departments are already planning to analyse public and private-sector databases for suspicious activity. The new Serious Organised Crime Agency (SOCA) is reviewing public and private-sector databases, to find data-matching opportunities that could highlight suspicious behaviour by individuals that implies they are involved in organised or financial crime. The SOCA consultation paper 'New Powers Against Organised and Financial Crime', says the public sector could share private-sector suspicions of fraud by joining CIFAS, the UK's fraud-prevention service. It also proposes matching suspicious activity reports with data from Revenue & Customs, the Department for Work and Pensions, the Passport Office and Driver and Vehicle Licensing Authority (DVLA) databases. This, it says, would be quite legal.

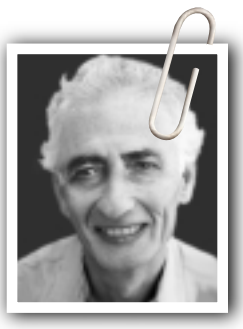
"Once you can share the data, then the temptation is to do more with it. Then, potentially, you end up with a surveillance society," said Boardman. "This kind of 'data creep' should be prohibited by the Data Protection Act, which requires data to be used fairly, in accordance with our reasonable expectations, and prohibits the use of data obtained for one purpose for other, incompatible purposes. However, the Data Protection Act is largely self-policing. There is no requirement in the UK, unlike some other countries, to submit intrusive schemes to the Information Commissioner for authorisation."

One striking example of data creep has been the merger of the planned functions of the ID cards database (the National Identity Register) with a proposed Civil Service programme to build a population register called the Citizen Information Project (CIP). The two projects differed mainly in how the personal data were to be used, accessed and disclosed.

The idea of the CIP project was to create a central repository of basic details, such as name, sex, address history, place and date of birth, and unique identifying numbers, to ease public sector administration through widespread data sharing. This would mean that all government records relating to an individual could be updated at once if their circumstances changed. The ID Card, in contrast, was promoted for purposes relating to counter-terrorism, crime, immigration control and illegal employment, with limited data sharing. Parliament wasn't notified of the merger of the two schemes until after it passed the ID Cards Act, despite civil servants advising ministers to do so.

To borrow from the conclusion of the academics Charles Raab and Christine Bellamy, in a paper entitled 'Joined-up government and privacy in the United Kingdom: managing tensions between data protection and social policy': "Today in the UK there is no general balance between privacy and data-sharing. Nor is there any single agency responsible for striking a balance."

Nevertheless, the government is planning to build a universal information system whose scope and reach, many fear, has not been clearly set out for public scrutiny. ■



Christine Evans-Pughe is a freelance journalist