



COUNCIL FOR
SCIENCE AND
TECHNOLOGY

Better use of personal information: opportunities and risks

November 2005

Executive summary

The government – through its White Paper on *Transformational government - enabled by technology* - has made clear its ambitions to deliver more effective and personalised public services. CST believes that the ability to fulfil these ambitions is dependent on the intelligent use of personal information about individual people. Government should provide the focus and drive to improve the linkages between, and access to, personal data, while at the same time recognising that there are significant risks which need to be minimised. This should be an incremental process, with the move towards greater data-sharing accompanied by parallel changes in policies and processes that minimise risks and ensure the potential benefits are realised in practice.

We believe there is the potential for both individual citizens and society to benefit from this approach. Citizens should gain more efficient and convenient access to public services. The collective benefit will result from the ability of government to develop more effective policies based on a stronger evidence base, which in turn will result in better targeted, more personalised and efficient public services - from health and welfare benefits through to law enforcement. We do not favour creating a single database because of the potential difficulties with maintaining the security and integrity of the data. Rather, we envisage a series of federated databases, each with its own identity but with common linkages.

But the risks need to be recognised. Concerns about invasion of privacy, accountability and government's ability to deliver the benefits of data-sharing need to be at the heart of an open dialogue. The key here is trust. Without an open dialogue on all the implications of better linkages between, and access to, personal datasets, there is a risk that the public will be mistrustful of government's actions. At the same time, there is a great risk that technology will drive these developments forward in an uncontrolled way. It is essential to do this properly, with appropriate controls, rather than allow the changes solely to be driven by technology. Through extensive engagement with the public and civil society groups, government needs to put regulatory and governance frameworks in place that minimise and manage the risks, while ensuring that the benefits are delivered. Without this intervention, the risks will not be controlled but nor will the potential benefits be realised.

Government therefore needs to engage with the public and opinion-formers to explore and communicate the benefits and risks. As part of our project, we commissioned OPM Ltd to explore issues around data-sharing, through focus groups. Many participants found it difficult to formulate definitive views on the issues discussed and tended to become more uncertain as they became more aware of the complexities of the issues involved. Wider debate would need to be supported by more concrete models of what greater data-sharing could mean for individual citizens – both in terms of the benefits that might be gained and the practicalities of how the system would work. One particular challenge in managing a wider dialogue is the apparent absence of vocal advocates for data-sharing to balance a minority of vocal objectors.

Central to realising the full range of benefits is the need for safeguards to be in place to protect people's personal information and address concerns over privacy. We live in an information-rich society where technological developments are extending the realm of the possible at an ever-increasing rate. Government needs to be clear where it wants to move to in this whole area, how it proposes to get there, and once there, how it ensures that the systems deliver what they are intended to deliver within a safe and secure environment. It is clear that the more joined-up the service, the greater the need for secure systems. We put

forward principles to achieve the necessary balance between promoting greater access to personal data and protecting the individual citizen. Formal policies between those organisations sharing personal data and those accessing these shared data need to be in place.

We believe a degree of creative thinking is needed here, the cornerstone of which should be the concept of citizens *owning* their own data in order for them to be able to exercise some control over how and when it is used, as well as putting the emphasis on citizens themselves to ensure it is kept up-to-date. While accepting there are circumstances where government must retain the right to use individual citizens' data and that government has a role in preventing malicious exploitation, the basic concept would be that citizens themselves would own their personal data. This concept needs to be explored further – at present it appears to be too remote from people's own experiences for them to engage with it.

While the case is made in principle here, more research is needed into the benefits of linking personal datasets, into the knowledge technologies to facilitate this, and into privacy enhancing technologies – an issue where government and business need to work together. This research needs to be interdisciplinary. Business needs to invest more in IT security to generate a greater level of trust in its systems, and government needs to encourage this.

The legislative regime is critical to this area, but it is complex and not well understood, in particular the Data Protection Act. Greater clarity is needed urgently: the large amount of guidance, often at a Departmental level, serves simply to confuse. In parallel, government should look again at whether, and if so what, legislative changes will be necessary to promote sharing of, and access to, personal data for researchers and statisticians. Governance of data management systems is a central issue: the contractual relationship between those who share data needs to be clear and explicit.

Recommendations

General recommendations

1. We consider that there are major benefits to be delivered from developing better linkages between, and wider access to, personal datasets provided the risks are carefully assessed and managed. In order to realise these benefits, there is a need to:
 - engage in dialogue with the public and stakeholders on the full range of benefits and risks, in particular to individual citizens as well as to society and to government;
 - carry out risk analyses with a balanced approach to the benefit-risk equation to strengthen the evidence base for policy formulation and enable improved service delivery;
 - have a focal point within government to think through, plan and co-ordinate joined-up working across different datasets.

Access to personal data for research and statistical purposes

2. Government should put in place mechanisms to ensure that the linkage and use of personal datasets is achieved in a much more co-ordinated, coherent and transparent way across the public sector. We consider that this should be done on the basis of three linked principles:

- personal data must be anonymised whenever this is possible, or pseudonymised in the case of linked datasets¹;
 - there should be a general presumption that access to data should be facilitated where that access is for research or statistical purposes;
 - appropriate safeguards and transparent governance structures should be in place before personal data can be accessed and used.
3. In order to command public trust, we believe it is critical for government to have open standards and protocols for the conditions under which data can be shared. This should involve:
- conducting risk analyses and establishing risk reduction processes among organisations and individual citizens sharing data;
 - addressing real and potential conflicts of interest, and any specific issues – such as involvement of vulnerable groups;
 - putting in place formal data handling policies for researchers or statisticians accessing the shared data.
4. Government should introduce a quality mark or accreditation scheme for the operation of government databases in order to promote greater trust. Monitoring compliance must be through third-party audit, in which the Information Commissioner or an equivalent person or body, must have a central role with agreed delegated authorities. Non-compliance with statutory safeguards must result in severe penalties.

Technological research needs

5. Government should promote research into knowledge technologies that facilitate the benefits of linking personal datasets. A key focus here is research on privacy-enhancing technologies. Government should work closely with business experts from the private sector (such as the banks) in identifying future needs regarding privacy-enhancing technology. In addition, government should:
- initiate a technology road-mapping exercise to identify what technologies will be available when, and plot these against relevant socio-economic and other drivers;
 - stimulate more interdisciplinary R&D – involving computer scientists, engineers and social scientists - in techniques for anonymising and pseudonymising data, encryption, and anti-virus devices;
 - encourage private sector organisations which are involved in privacy to share R&D ideas on security modelling;
 - develop more explicit and proportional confidentiality requirements in its procurement specifications;
 - promote greater trust through encouraging greater levels of investment by business into IT security.

¹ Anonymised data are information which cannot be traced back to an individual (i.e. the encryption key has been thrown away); pseudonymised data are anonymised data where the encryption key has not been thrown away (i.e. it can be traced back to an individual).

Dialogue and communication

6. Government should engage in dialogue with the public and opinion-formers to inform its policy developments on data-sharing and access to personal datasets, and the privacy safeguards that need to be in place. It should use the principles set out in the CST report *Policy through dialogue* and build on the work done for us by OPM Ltd published as part of this Report.
7. As key components of its public engagement work, the government should:
 - sponsor interactions between different stakeholders and the public; and educate individual citizens – especially young people – about personal data and its use in order to promote understanding on how individual citizens could better take responsibility for managing their personal data. Schools and universities would have an important role here;
 - encourage better articulation of, and debate about, the risk–benefit equation that must be analysed and understood in order to promote or restrict greater sharing of personal data;
 - determine where responsibilities lie, and how rectification and recompense will be provided in cases where the security of personal data held by government is compromised.

Regulatory framework

8. Government needs to ensure there is clarity on how the regulatory regime for data-sharing and data protection operates; in particular it should:
 - consider carefully what legislative changes might be necessary to promote sharing of, and access to, personal datasets between public sector bodies;
 - review the guidance which different parts of government have issued - for example on the operation of the Data Protection Act - to ensure there is the necessary consistency.

Implementing these recommendations

9. As part of the government response, CST would welcome an action plan including timetable on how government proposes to address each of the recommendations in this report, followed by a progress report on achievements against this plan in a further 18 months time.

Better use of personal information: opportunities and risks

Council for Science and Technology, November 2005

Contents

Page 1	Executive summary
Page 2	Recommendations
Page 6	Introduction (paragraphs 1-9)
Page 8	Linkages between personal databases (paragraphs 10-31)
Page 14	Future opportunities for personal data-sharing (paragraphs 32-33)
Page 24	Technological, privacy and regulatory issues (paragraphs 34-55)
Page 29	Public attitudes (paragraphs 56-67)
Page 33	Acknowledgements and contact details
Page 34	Annex A: Risks and risk management

Boxes

Page 9	Box 1: A snapshot of existing database linkages
Page 10	Box 2: Ongoing reviews of issues relating to personal datasets
Page 11	Box 3: Some examples of benefits to the individual of linking personal datasets
Page 15	Box 4: Specific opportunities for early action
Page 16	Box 5: Case Study 1: Using data linkage to develop new approaches to detecting adverse drug reactions (ADRs) in routine clinical practice
Page 17	Box 6: Case study 2: Using data linkage to address health inequalities and develop disease registers to improve targeting
Page 19	Box 7: Case Study 3: Use of data linkage to improve risk assessment for health effects associated with environmental pollution
Page 21	Box 8: Case Study 4: Dynamic Traffic Management
Page 23	Box 9: Case Study 5: The Citizens Information Project
Page 27	Box 10: Access to personal data: three principles
Page 30	Box 11: Assessment against CST's Dialogue criteria

Introduction

Context

1. The public sector has a large number of databases containing personal data – in particular, individual citizens' names, addresses and ages. It is impossible to say how many databases exist², but the figure will be a large one. It is estimated there are 300 million personal contact detail records in the UK - on average five sets of contact details for every citizen in the UK. Many of these records will be out of date – between seven and ten million people change their contact data each year.
2. Currently, the management and use of these data is too fragmented and lacks policy coherence, which means that the full range of benefits - to individual citizens, the community and to government - are not being delivered. New benefits and opportunities from linking and accessing personal datasets are becoming apparent in an ever-increasing number of ways.
3. There is an apparent lack of co-ordination between departments regarding the management of personal datasets. For example, the House of Commons Home Affairs Select Committee has expressed concerns over the proliferation of the number of times that an individual's data was placed on databases, in terms of quality, management and cost issues³.
4. Nevertheless, there is an increasing recognition, both in government and more widely, of the importance and value of personal databases, and the ways in which this information could be used. New benefits and opportunities from linking and accessing personal datasets are becoming apparent, not least in facilitating the necessary research that is at the heart of evidence-based policy and well-targeted public services.
5. The benefits of linking datasets and promoting their use need to be balanced against the risks. In particular, there are serious concerns about the protection of privacy.
6. In 2002, the Cabinet Office's Performance and Innovation Unit published a report entitled *Privacy and data-sharing – the way forward for public services*, that outlined the issues surrounding data-sharing and privacy and made recommendations on ways forward. This report argued it was possible to reconcile the better and more widespread use of personal data to help deliver the government's modernising agenda with the resultant intrusion into the individual's right to privacy.

Objectives and scope

7. In preparing this report, the CST had three primary objectives:
 - to understand what beneficial uses could be made of the personal information stored in electronic databases, both now and in the future;
 - to determine what needs to change to enable the benefits to be achieved;

² House of Lords Written Answers 23 February 2005, column WA207, answer to a Written Question from the Earl of Northesk.

³ "...We are concerned by the proliferation of large scale data bases and card systems, since we have seen little to suggest that they are being approached in a co-ordinated way...We believe that the government must tackle this proliferation of data bases, examining in each case whether the number, identifier or data base is needed, what its relationship is to other existing or planned data bases, how data will be shared or verified and other relevant issues. For this action to be effective it must be co-ordinated at the highest levels of the Civil Service." *Identity cards* (paragraphs 42-43). Home Affairs Committee, July 2004.

- to consider, in the context of a risk analysis, what safeguards need to be in place to mitigate the risks.
8. In developing its recommendations, the CST has considered a wide range of issues associated with the linkage and use of personal datasets – including the existing legislative framework and future regulatory needs; technological issues and future research and development directions; and issues around public engagement.

Process

9. In preparing this report, the CST held two evidence-gathering workshops in November 2004 and January 2005, at which expert participants representing government, public and private sector organisations discussed the issues. Based on the outcomes of the workshops, the CST commissioned a series of in-depth case studies on the potential for linking datasets in a number of areas. We also commissioned research into the use of personal datasets by OPM Ltd⁴.

⁴ *Research into the use of personal datasets held by public sector bodies: report for the Council for Science and Technology.* OPM Ltd, October 2005. Available via www.cst.gov.uk

Linkages between personal databases

The proliferation of personal datasets in the public sector

10. The public sector uses personal data to deliver services, understand problems and design and deliver innovative solutions. But much more could be done.
11. Large numbers of public-sector databases exist, largely home-grown by department or agency⁵. This has resulted in inconsistent data storage regimes and duplication of data, a general inability to link these different databases, and differing regimes of access and control. In addition, personal data stored by government is often of low quality – both inaccurate and incomplete – and very expensive to keep up to date. Processing of personal data amplifies errors, and further contributes to the poor quality. The net result is a largely fragmented, inefficient and sub-optimal system that wastes time and resources.
12. The use of such data is also rather *ad hoc*. For example, the use of longitudinally linked administrative data is an important research tool for government, both in setting policy and delivering services. It often relies on bringing together information from different databases. However, there is no general power to link or combine data, and there is the risk that in some cases, in so doing departments and authorities may be acting *ultra vires*. To create separate ‘gateways’ within government may require explicit legislation or regulation in each case⁶. Clearly this is time-consuming and places extra burdens on the legislative process.
13. The area of law enforcement is one in which there is significant convergence and connectivity between personal databases⁷. While some departments (including the Department for Work and Pensions, the DVLA and the NHS) are actively working to rationalise and link their internal databases (**Box 1**), the overall picture is complex and uncoordinated.
14. Establishing greater linkages between government databases in a more systematic way, could help to:
 - rationalise the number of databases across government to improve efficiency and deliver cost-savings;
 - establish a common scheme for personal data to enable more effective accessibility and management of data;
 - facilitate the creation and use of longitudinal data.

⁵ For example, in the late 1980s DHSS was developing its Departmental Single Index Online database, at a similar time that the NHS, DVLA and the Passport Office were also independently building their databases.

⁶ For example, the Cabinet Office PIU report on data-sharing lists eight separate pieces of legislation between 1999 and 2001 under which the Department of Work and Pensions may share data with other departments.

⁷ For example, through the Police National Computer and the Police Fingerprinting System operated by the Police Information Technology Organisation (PITO).

Box 1 – a snapshot of existing database linkages

A number of information gateways exist to enable data-sharing where *vires* are not thought sufficient; for example: between HM Customs and Excise and the Inland Revenue; between DfES and the General Teaching Council; and between Home Office and the Police and HM Customs and Excise. There is also:

The Police Information Technology Organisation (PITO): which is the guardian of the Police National Computer and the Police Fingerprinting Database which link to other databases to allow the police, Customs and Excise and other law-enforcement agencies to carry out their duties.

Work and Pensions Longitudinal Study (DWP): government has legislated to open up Inland Revenue databases to provide better information on movements from welfare into work. It links Jobcentre data with claims for child benefit, employment and tax credits. Longitudinal data is used to prevent fraud, errors and as a basis for government statistics. It provides the basis for the research agenda into the long-term outcome of DWP policies, and greater efficiency of DWP operations as well as to track leavers from the Armed Forces.

Census and Crime: DWP is also linked with the Inland Revenue and the ONS to allow the necessary management and research in advance of the 2011 census. The Home Office's Crime Analysis Unit (CAU) - and the British Crime Survey - have links with other government Departments such as ONS and the Office of the Deputy Prime Minister to access data and carry out research

Health: the NHS Information Authority has links with ONS to provide information on blood donor and cancer registers, and it has links with Home Office and MoD to provide health records of staff. NHS organisations have links with other organisations to communicate information on national health issues.

Education and Skills: DfES has developed a protocol on data-sharing⁸. It is implemented through DfES's Common Basic Data Set which supports a shared technology infrastructure. DfES has links with Ofsted, the Learning and Skills Councils, QCA etc, and has a National Pupil Database containing information on all pupils in England (but with access safeguards in place).

15. Achieving these linkages should lead to improvements in the quality of policy formulation and better targeted and more personalised delivery of public services. At the same time, however, this rationalisation and linkage will need to be done in ways that ensure privacy is maintained and respected.
16. The 2002 report on data-sharing from the Cabinet Office's Performance and Innovation Unit made a series of recommendations for how more widespread and coordinated use of personal data could be promoted by government, while ensuring appropriate safeguards are in place. We agree with the recommendations of this report, but are concerned by the lack of progress in implementing them.
17. We believe that the government should move quickly to develop a co-ordinated and coherent strategy for maximising the linkages and use of personal datasets, ensuring that there is a clear lead and focus within Whitehall to take these initiatives forward. There is a real risk that unless this happens, technology will continue to drive these resources forward in a haphazard manner and the opportunity for government to develop a co-ordinated approach that protects the interests of the public will be lost. Government needs to be clear where it wants to move to in this whole area, how it proposes to get there, and once there, how it ensures that the systems deliver what they are intended to

⁸ Data should be collected once and used many times; the processes should be fully automated; value of collecting the data should outweigh the costs; and personal data should be fully protected.

deliver within a safe and secure environment. We therefore welcome government's White Paper *Transformational government – enabled by technology*⁹ as a starting point for this journey.

18. In taking forward this work, we became aware of a number of other reviews relating to personal datasets (**Box 2**). We believe that it is vital that these reviews, and the implementation of the resulting recommendations, do not take place in isolation.

Box 2 – Ongoing reviews of issues relating to personal datasets

The Cabinet Office Regulatory Impact Unit is looking at data-sharing for front-line DoH staff and NHS, including both data-sharing and data-protection issues. Its focus is on the changes that are needed, in particular improvements to current guidance and review of Data Protection Act. Cabinet Office has published a strategy for *Transformational government*.

The Academy of Medical Sciences is undertaking a study on the Use of Patient Data in research. It is examining the case for using patient data in medical research against the background on national and international regulations. It will report in 2005. There are two related projects: the **MRC** data-sharing initiative covering guidance/data-sharing tools and consents/confidentiality; and the **UK Clinical Research Collaboration** which is looking to streamline the regulatory and governance environment.

The Office of National Statistics is developing a Citizen Information Project (CIP) – see case study 5 (box 9, page 25). It is very closely linked with the Home Office work on Identity Cards and the National Identity Register.

Benefits of linking personal datasets

19. Fostering more effective use of personal datasets held in the public sector could generate benefits both for individual citizens and for society as a whole. However, to secure these benefits it will be necessary for the databases to be linked.

Individual benefits

20. Individual citizens and organisations increasingly expect the public sector to provide more seamless and personalised services to address their needs and to tackle specific problems. To achieve this requires a greater degree of connectivity between personal data which different public sector organisations hold. There is therefore a balance to be struck between better, more personalised services, and the potential intrusion this might mean for the individual citizen. This is a central theme of our Report.
21. There is an opportunity cost to individual citizens if what is being delivered, electronically or otherwise, is not achieved in an optimum way¹⁰. The efficiencies (particularly time) gained from greater access and use by individual citizens of on-line public services is compromised when each on-line service is delivered in a different way, or requires a different set of actions by the individual concerned in each case. Therefore, there is a convenience benefit for the individual.

⁹ *Transformational government – enabled by technology*. Cabinet Office, November 2005. Available via www.cio.gov.uk

¹⁰ For example, an issue of domestic violence can involve over 30 interactions with various agencies before it reaches the criminal justice system.

22. Better linkages between databases will enable identification of those individual citizens who are entitled to benefits and services and who are not claiming them.
23. The CST has articulated a set of potential benefits to individual citizens under four headings: healthcare; life events; public services; and other services (**Box 3**). Our focus group work through OPM Ltd has indicated strongly that there is a need to articulate and debate the benefits to individual citizens in ways in which they can personally empathise.

Box 3 - Some examples of benefits to the individual of linking personal datasets

Healthcare

- Targeting health screening to allow 'at risk' population to be defined. See Case Study 2 (box 6, page 17)

Life events

- Employment – automatically generating accredited CV/qualifications, medical and any criminal record for prospective employees
- Change of address - automatic notification to utilities and government (for example DWP, Inland Revenue, DVLA, Passport Agency)
- Buying a house – central data-point for searches, Land Registry, planning permissions, services and utilities
- Getting married – answers to questions on rights to marry (citizenship, marital status)
- Birth - access to all necessary services, for example health visitor appointments, injections, birth certificate, baby bond
- Death - single death notification triggering appropriate public and private services and administrative arrangements

Public Services

- Benefits - tackle exclusion by providing entitlements whether or not individual citizens apply
- Tax - simplified returns which are completed by the Inland Revenue and sent to individuals for checking
- Transport - abandoned vehicle removal through a local authority - DVLA link
- Services - easier access via a single telephone call since details are already held in the database
- Emergency Services - better linkages between the three emergency services in terms of the information they hold (for example building plans)
- Reminders – MOT, passport, health checks, inoculations
- Assured information correction/destruction - for example relating to credit approval

Other services

- Data protection – the ability to make consolidated Data Protection Subject access requests to different government departments and public bodies
- Integrated environmental hazard analysis – flood, air, noise pollution
- Consumer Protection

Collective Public Benefit

24. In addition to benefits at the individual level, promoting wider linkage and use of personal datasets promises wider collective benefits to society as a whole.
25. One such benefit is strengthening the evidence-base to improve policy making. For example, allowing researchers and statisticians greater access to personal data in the social and medical sciences in particular would strengthen and deepen the evidence base about health, disease, the environment, and individual and social welfare, and the complex inter-relationships between these¹¹. Also, longitudinal data would provide a better understanding of both the labour and energy markets. The key question is the extent to which researchers and statisticians should have access to personal datasets, and on what terms.
26. A second potential benefit is the delivery of improved public services. Delivery of high quality services, and developing well-targeted and effective policies, depends on the effective use of personal information, for example targeting benefits to those groups most in need. The ability to link personal datasets could also have huge benefits for public agencies in taking forward their responsibilities in relation to crime, fraud, security, law enforcement, or welfare benefits.

Balancing benefit and risk

27. The benefits at both an individual and collective public level need to be balanced against the risks. The introduction of risk assessment processes will be crucial to address the concerns arising from increasing access to personal information – which are described in more detail in **Annex A** (page 34). These risks include:
- loss of confidence and trust in privacy;
 - unauthorised use of personal data;
 - exploitation of individual citizens for commercial gain;
 - statistical discrimination (e.g. creating a sub-culture of non-participation by individuals);
 - technical risks such as database failure or incapacitation (eg by spam or unmanageable volumes of data);
 - poor data quality;
 - cyber-terrorism.
28. We emphasise that improved linkages between personal databases and greater access to data by researchers and statisticians must not occur without ensuring that there is the necessary level of privacy protection for the individual, and that government must ensure that these developments should happen in a staged and co-ordinated way. We also believe that the full range of benefits and risks to all stakeholders of promoting greater linkages between, and more widespread use of, personal datasets need to be clearly articulated and debated.
29. In particular, the focus to date has largely been to articulate the benefits that better sharing of personal datasets will bring to society and government, through better public sector policies. What has been lacking is an articulation of the benefits and risks to the

¹¹ Examples include the *MRC/BHF Heart Protection study on cholesterol-lowering therapy*. Lancet, 2002.

individual citizen, in ways in which he or she can empathise, and the necessary safeguards and governance processes that need to be in place to ensure public confidence and trust.

30. We live in an information-rich society. It is important to note that the public-private sector interface is blurring¹². While this Report does not consider the cross-fertilisation of information from private sector databases to the public sector, there is the paradox that the ever-increasing use of the internet for private transactions and the level of trust that this implies appears much greater than the level of trust that individuals have in equivalent public sector IT systems. One explanation is that internet transactions are usually for immediate personal benefit whereas individual citizens see little direct benefit from government databases, made worse by the possibility that they could act against people's interests.

31. The table below shows a broad overview of the issues raised in the focus groups run for us by OPM.

Public Sector/government	Private sector
Do not sell data	Sell data
Technically inefficient	(In general) more technically efficient
Limited or no redress seen as possible in case of misuse of data – channels known to exist but extremely difficult to use, with a lot of problems encountered	Possible redress if data misused eg withdraw business, write to consumer programmes/pages, circulate bad experiences through word of mouth
Absence of choice about whether or not to provide information	Choice about whether or not to provide information, based on perceived advantages/disadvantages in doing so
Social benefits easy to identify eg fraud prevention. Individual disadvantages easy to identify eg access to personal information by unauthorised individuals; identity fraud.	Benefits and disadvantages to the individual can be identified eg value of purchase/service to individual weighed against cost/risk of providing information

¹² Particularly in the area of crime prevention: for example, the Inland Revenue obtains data from credit reference agencies; and the police national computer is linked to the insurance industry.

Future opportunities for personal data-sharing

32. The 2002 Cabinet Office report on data-sharing identified several opportunities for early action, and in the workshops we held to inform this report, a number of further opportunities were highlighted. Taken together with the examples of benefits to the individual outlined earlier, this represents a formidable list of possible opportunities (**Box 4**). We believe these should form priorities for the type of co-ordinated approach across government we have described above.

33. CST considers that public health and transport each constitute an under-used opportunity for better linkages between, and access to, personal datasets. We emphasise that the case studies that follow are *illustrative* and are designed to demonstrate how better linkages and uses of personal datasets can bring about the sorts of benefits we have identified, and the risks that need to be addressed to realise those benefits (**Boxes 5-9**). Including these case studies should not, however, be read as CST endorsing their every detail since we recognise that they would all require substantial further work if they were to be implemented.

Box 4 - Specific opportunities for early action**More Joined up and Responsive services**

- Identifying and supporting children at risk of social exclusion – sharing information between local agencies of children at risk
- Issuing photo-card driving licences – agreement between DVLA and UK Passport Services
- Fostering better access to health records – sharing information between NHS organisations and local authorities
- Providing services for those in real need – sharing information between DWP and LCD on legal aid eligibility
- Developing a more joined-up response to individual needs of ex-offenders (Social Exclusion Unit survey)
- Modernising civil registration – allowing registration of births and deaths in a variety of ways including by telephone or the Internet
- Improving services for families (i) through Sure Start: ensuring that families with young children receive the services that are right for them by better understanding of each family's social and medical situation; and (ii) using the Child Benefit database to send targeted information to parents on the services offered in their region
- Streamlining services for motorists – improving use of data from the Driver, Vehicle and Operator Group to provide more joined-up services for the motorist
- Improving information on the property market – better linkages between the Land Registry, Valuation Office Agency and Stamp Office to provide aggregate data on the local environment, improve policy making and the information available to homeowners
- Tracking the spread of infectious diseases - such as influenza - and providing real time information to frontline medical staff on the predicted pattern of spread and high-risk groups, and using these data to inform vaccination or quarantine programmes
- Using linked datasets to target public services at individuals – eg non-take-up of social benefits

More Effective and Better Targeted Policy Making

- Helping children in need – better data-sharing between local agencies
- Supporting better use of statistical and management information – better use of anonymised health data would enable Drug Action Teams to evaluate the success of treatment programmes and initiatives
- Getting the best from private providers of education and training – new ways of working between DfES and public service providers
- Improving urban planning and investment – wider use of Valuation Office Agency data
- Using DVLA and transport data to improve vehicle emissions and traffic planning. See Case Study 4 (box 8, page 21)
- Using longitudinal data from a variety of existing databases to study the net benefits of particular policies
- Merging census, survey, tax rewards and commercial data to inform social policy - identify, for example, the connection between education, employment and disadvantage/advantage

Tackling Crime and Fraud

- Providing better authentication – access to data on public and private sector databases to corroborate an individual's identity in a more robust and cost-effective way
- Tackling vehicle crime – data-sharing enabled in legislation to allow officers to have roadside access to the information they need to enforce the law and make roads safer
- Using datasets to tackle tax evasion and fraud in (for example) the benefits system

Tackling Debt

- Moving towards effective enforcement: access of the necessary information by the claimant to enable the right method of enforcement to be taken

Box 5 - Case Study 1: Using data linkage to develop new approaches to detecting adverse drug reactions (ADRs) in routine clinical practice

What is the problem?

The UK has no systematic approach to the detection, verification and quantification of adverse drug reactions. (ADRs) are estimated to affect around 7% of patients or hospital admissions at an annual cost of around £380 million. Deaths from prescription errors and adverse reactions have increased five-fold over the past ten years, resulting in 1,100 deaths in 2002. Failure to detect ADRs early increases the risks associated with drug development, with consequent higher drug costs, lower rates of innovation, and greater NHS healthcare costs.

How can linking personal datasets provide the solution?

Early detection of associations between drugs and adverse events may be improved by systematic searching of all available data. A linked dataset that can be used to compile information on health care events for individual citizens and then be used to pilot new methods of detecting adverse drug reactions in routine clinical practice may enable earlier detection of warning signals.

To carry out this research, access is needed to individuals' information on prescribing and medical events. This could be obtained by linking clinical and prescribing information from computerised medical records in general practice with information from hospital information systems, cancer registries and mortality records.

Who will benefit?

Individuals with rare but serious ADRs could be detected more quickly and action taken – for example taking patients off a particular drug. Patients with ADRs could be identified and included in epidemiological, biochemical, and pharmacogenetic research, thus allowing factors that influence ADRs to be identified. This could potentially lead to more targeted drug development in the future.

For government and the NHS, quicker detection of ADRs would allow preventive action to be taken, reducing the number of adverse drug reactions in the population, and lead to lower healthcare costs. The public health, clinical, academic, and economic benefits to society are potentially considerable: the methods described here could be applied to the entire UK population of 60 million people. Millions of years of patient experience, which are needed to detect rare but potentially life-threatening ADRs, could be available to allow appropriate identification and the treatment/measures to be taken to protect people. The UK could become a world leader.

What needs to changes to make these linkages work?

Access to patient datasets. There are also major statistical and computing challenges. Success would require a multi-disciplinary team including epidemiologists, clinicians, statisticians, computer scientists, and civil servants.

What are the Privacy issues and how might they be tackled?

Access would be needed to individual citizens' information on prescribing and medical events. Privacy issues could be overcome by anonymising or pseudonymising the data.

What concerns are there likely to be?

Focus group work undertaken by OPM indicated that attitudes towards information sharing for health purposes are viewed differently from information sharing for other purposes. In part this is because the NHS is seen as trustworthy in a way that other government agencies are not. The benefits to individual citizens are also clearer. ADRs attract considerable media attention. Methods of detecting them more quickly and improving patient safety are likely to be popular with the public.

Box 6 - Case study 2: Using data linkage to address health inequalities and develop disease registers to improve targeting

What are the problems?

One of the government's main priorities for the health service is narrowing the gap experienced by some of the most vulnerable groups in society – in particular lower socio-economic groups, ethnic minorities, and the elderly.

There are major inequalities between ethnic groups. For example, the prevalence of stroke for Black Caribbean men is about two-thirds higher than in the general population. South Asian men are about 30% more likely to have coronary heart disease than men in the general population. Pakistanis and Bangladeshis of both sexes are more than five times as likely to have diabetes and Indian men and women are almost three times as likely to have diabetes as the general population.

There are also inequalities in mortality rates for specific diseases, for example coronary heart disease which causes 35% of male and 27% of female premature deaths before the age of 75. Health screening has been shown to reduce the risk of coronary heart disease, but its cost effectiveness relies on being able to target accurately the population at most risk.

How can linking personal datasets provide the solution?

Linking personal datasets would mean that health inequalities could be monitored more effectively than at present, and provide improved clinical management. Current NHS datasets do not include socio-economic status and ethnicity and do not take an integrated approach. It would mean that disease registers could be developed for use as sampling frames for clinical trials and epidemiological studies. It would allow government interventions to be targeted at the right groups and evaluated to determine whether they are working well.

Datasets that could be used in record linkage include: General Practice records; Community records; Hospital Episode Statistics; Outpatient clinic records; and data collected by other government departments, for example benefits data such as family credit, income support, jobseekers allowance, attendance allowance, disability living allowance, incapacity benefit, and severe disablement allowance.

By linking the Department of Work and Pensions Longitudinal Database and the Department of Health national linked file of hospital statistics and mortality, it should be possible to determine whether or not particular patterns of benefit claiming, benefit types, employment status, household income levels or types of work sought by jobseekers and others correlate with subsequent health indicators as recorded in the hospital episodes and mortality statistics.

Who will benefit from such linkages?

Vulnerable groups of individuals, such as those from lower socio-economic and minority ethnic groups, through better targeting and delivery of health services.

The NHS will operate more efficiently through better-tailored health services, thus providing more health gain for a given level of health spending. Linked datasets would provide excellent sampling frames for the identification of patients for inclusion in clinical trials and epidemiological studies, as well as for follow-up, for example, by providing information on health care episodes.

Wider benefits include better information for managing the NHS; data for measuring the quality of health care; and comprehensive disease registers with individual level socio-economic data for use in epidemiological studies and as sampling frames for randomised controlled trials. This would support the government's aim of making the UK a leading centre for clinical research.

...continued

...Box 6 continued

What needs to change to make these linkages work?

The data to feed into this programme already exist in computerised form at individual patient level, and potentially many of these linkages within the NHS are already part of the national IT strategy. This strategy would need to be widened to include non-NHS data sources such as benefits data. An initial feasibility stage, followed by a pilot data linkage project in 1-2 primary care trusts could be used to trial the technique. The linked data could then be used to examine health inequalities in a number of clinical areas, starting with coronary heart disease. After successful completion of a pilot project, the project could be rolled out into other geographical and clinical areas. The feasibility of data linkage, using hospital episode statistics and linking these data to other sources of information, has already been proven.

What are the privacy issues and how might they be tackled?

It would mean linking NHS datasets with those on socio-economic status and ethnicity. It would mean that disease registers could be developed for use as sampling frames for clinical trials and epidemiological studies. Privacy issues for the monitoring function could be overcome by anonymising or pseudonymising the data. At individual level, the linked data would be used for individual patient care. Procedures would have to be established to ensure that the confidentiality of patient information and individual economic circumstances is maintained at all times. Data linkage procedures would need to be agreed between DWP and DoH. These would need to specify precisely the data to be linked, how data will be exchanged and where the linked data will be held prior to anonymisation as well as considering all aspects of data security and confidentiality.

What concerns are there likely to be?

Focus group work undertaken by OPM indicated that attitudes towards information sharing for health purposes are viewed differently from information sharing for other purposes. In part this is because the NHS is seen as trustworthy in a way that other government agencies are not. The benefits to individuals are also clearer. Public acceptability is likely to be high as this would lead to a higher quality and more efficient NHS service to the public. However, there may be some concern expressed by the public about the potential misuse of the linked data (e.g. to investigate the medical history of suspected benefit fraudsters). It should be possible to publicise the potential benefits of the data linkage in terms of early identification of life-threatening diseases, while outlining the measures put in place to prevent the linked data being used for any other purpose.

Box 7 - Case Study 3: Use of data linkage to improve risk assessment for health effects associated with environmental pollution

What is the problem?

Routine health data at postcode level are being used to address concerns about potential causes of disease related to environmental pollution, for example, from incinerators, landfill sites, road traffic, power lines, mobile phone masts, as well as for the investigation of disease clusters. However such analyses are complex, not least in terms of the socio-economic variables involved and migration effects, as typically only the address at diagnosis is known.

Routine recording of health events only gives the place of residence (typically postcode) at the time of the event, whereas exposure to environmental pollution in the period preceding the event is the parameter of interest. For cancers and mortality, the critical exposure period may be years or even decades before the clinical event is recorded, whereas for birth events such as congenital anomalies the key exposure period may be early in pregnancy. Thus information on residential history is vital in improving the risk assessment, but currently such data are not readily accessible and are not linked to the individual health records.

How can linking personal datasets provide the solution?

The aim is to improve the ability to carry out epidemiological studies on the effects of environmental pollution on health among the general population, through linkage of health data with environmental data, and other datasets on socio-economic variables and migration. Data on sources of environmental exposure are becoming increasingly available. Monitoring data are also available, for example on air pollution through the national network of monitoring sites. These data can be linked to post-coded health datasets using Geographical Information Systems (GIS) techniques and potentially the health risk of populations living near sources of environmental pollution can be estimated. However this requires accurate estimation of the demography of the underlying population, including migration into and out of areas, as well as information on the socio-economic characteristics of the population.

Datasets that could be used in record linkage include: Death records; Cancer registrations; Birth records; Congenital anomalies; Hospital Episode Statistics; General practice records; and Data on migration at individual level. Data collected by other government departments offer opportunities for linking data on socio-economic variables with NHS data. One example is benefits data, which contain valuable information on socio-economic status and other important variables, for example family credit, income support, jobseekers allowance, attendance allowance, disability living allowance, incapacity benefit, and severe disablement allowance.

Who will benefit from such linkages?

Benefit would be gained by Individual citizens and their families who may have concerns about the effects of environmental pollution and about the occurrence of local clusters of disease.

Society as a whole, and government, would also benefit from better interpretation of environmental epidemiological studies, improved risk assessment and hence better informing of environment and health policy.

What needs to change to make these linkages work?

The health data to feed into this programme already exist in computerised form at individual patient level, and potentially many of these linkages are already part of the national IT strategy within the NHS. The linkages would need to be widened to include non-NHS data sources such as benefits data, as well as data on migration such as driving license records and the GP-registered population.

...continued

...Box 7 continued

What are the privacy issues and how might they be tackled?

Linkages would be needed between a broad range of personal datasets, not only health but also benefits, allowances etc related to socio-economic status. Privacy issues could be overcome by anonymising or pseudonymising the data.

What concerns are there likely to be?

Focus group work undertaken by OPM indicated that there were concerns about possible wider consequences of the discovery of environmental pollution in specific areas being damaging to health: falling property prices and residents being unable to move away from a particular area and other related problems.

Box 8 - Case Study 4: Dynamic Traffic Management**What is the problem?**

Growth in the UK economy also means growth in the movement of people and goods and, consequently, in road use. The rise in journeys by car has created challenges in terms of air pollution and CO2 emissions; pressures on road capacity and the pattern of road use, especially in peak hours. All this results in slow and unreliable journeys, and traffic levels are predicted to increase by 30% over the next decade and by 40% by 2025 – which will exacerbate these problems considerably if the status quo is maintained. What is needed is a road network that provides a more reliable and free-flowing system for users; and one where travellers can make informed choices about how and when they travel so that the adverse impact on the environment and other people is minimised.

How can linking personal datasets provide the solution?

By using real-time data-gathering based on two commercial projects – Autoroute1 and Pay as You Drive.

Autoroute1 uses travellers' mobile phones to gather and transmit data in real time back to road users about route choices, traffic flows and optimal journey times. Its main potential for road users is that it provides real time feedback on the quickest route from A to B, allowing them to make the best choice. At the same time, the system allows traffic flows to spread out and keep moving, minimizing congestion and environmental impact from emissions. It has been developed by DECELL and currently operates in Tel Aviv.

Pay as You Drive uses an electronic black box in vehicles to collect information about where, when, how often and how well a vehicle is driven. It is being piloted by Norwich Union to calculate and assign premiums based on actual vehicle usage and they have also used the technology to successfully track, locate and recover stolen vehicles. Pay As you Drive system could be applied directly to the calculation of personalised road tax. One option currently being considered by the Department for Transport would use distance travelled as the main basis for taxation, with an additional charge – set locally – to reflect travel through local congestion points.

Who will benefit from such linkages?

Individual citizens will benefit through faster journey times owing to reduced congestion, greater personal choice based on cost and reduced likelihood of car theft and/or faster recovery times. There are also potential benefits in terms of convenience and personal empowerment – for example, by combining personal datasets with private sector providers users could access personalised insurance. Full linkage of DVLA datasets with insurance companies could create fully automated taxation and MOT procedures.

Collective public benefits will include fewer vehicles on the roads leading to reduced congestion, and reduced levels of emissions through, for example, managed queuing in congestion hotspots. Automated MOT notifications to owners and the police would result in increased roadworthiness of vehicles. In the longer term there would be the facility to alter road pricing dynamically to reflect actual road conditions and traffic flow.

There is potential for real impact on transport modelling. The DECELL system relies on destination data and the Pay As You Drive model can provide it. Linking datasets could also provide information on levels of pollutants, identifying trends and hotspots.

...Box 8 continued

What needs to change to make these linkages work?

To achieve the benefits of dynamic traffic management and road pricing would require proper systems design and effective network management to ensure that the ability for individual citizens to re-route did not lead to network failure or reduced road safety as drivers moved off primary routes onto secondary ones

Pay as You Drive would require vehicles to be fitted with electronic vehicle identification chips, which could be built into all new cars; or it might be possible to include them in new versions of tax discs or on number plates.

Linkages between DVLA and insurance datasets will depend on persuading insurance companies to share personal data, which may be commercially sensitive – however there is already some sharing of data as part of the Department of Transport's Automatic Number Plate Recognition system.

Institutional changes would be desirable – for example departments and agencies operating within a clear framework for data-sharing.

Legislation may be needed to enact these linkages. At the same time, government would need to take a long-term view on the investment required.

What are the privacy issues and how might they be tackled?

The Pay as You Drive system piloted by Norwich Union uses a strict data protection system – analogous to the framework described above – with information on traffic flow provided at the aggregate - anonymised – level. Trust is earned by overcoming privacy concerns, and the recognition by employees that any breach of customer confidence will damage Norwich Union's reputation and lose them business. Customers also have freedom of choice - whether or not to sign up is a purely voluntary act and customers can leave at any point.

This demonstrates that privacy issues are less of a concern where participants enter a scheme voluntarily – as opposed to one which is imposed - and can see the benefits directly.

What concerns are there likely to be?

Focus group work undertaken by OPM indicated that attitudes towards information sharing for the purposes of traffic management were generally positive. Developing a traffic management and road pricing system using these technologies is likely to be acceptable to the public if it draws on the lessons of experience of those consulted for this case study:

- clearly communicate what is being developed and why;
- clearly demonstrate the use and benefits of the technology;
- protect and audit data use and make the system totally transparent;
- do it on a voluntary basis;
- do it one step at a time – and give people time to realise the benefits, articulate their concerns and address them, before moving on to the next stage.

Box 9 - Case Study 5: The Citizens Information Project**What is the problem?**

There is a limited level of data-sharing across the multiple local and central government systems. One implication of this is that many of these systems maintain local citizen contact details (such as name and address) – it is estimated that there are 300 million personal contact detail records in the UK public sector, on average five sets of contact details for every single person in the UK. This imposes an additional administrative burden to both the citizen and each government department while also reducing the quality of the information held locally.

How can linking personal datasets provide the solution?

Sharing contact details across the public sector through a centralised contact details register – in effect a population database – will provide:

- **Economies** - removing duplication of multiple departments updating the same personal information; stopping citizens having to inform multiple departments of a change in circumstances; and making it easier for departments to access government information (including for new initiatives by a common and centralised data format).
- **Efficiencies** - improving service delivery through having more up to date citizen information available; strengthening verification reducing the potential for fraud; making it easier for a citizen to be verified when interacting with government; and eliminating duplicate and potentially conflicting information across departments.
- **Effectiveness** - increasing right first time service to citizens; facilitating more personalised public services; enabling e-government agenda by delivering a citizen-view across systems; providing more up to date information; greater coverage of citizen information available for government policy.

Who will benefit from such linkages?

There will be significant value to both citizens - one central point of communication, rather than (for a typical family) up to seven (Department of Work and Pensions; Inland Revenue; Local council; Local schools; NHS; and DVLA) - and government - the quality of information will improve because any change in circumstance will only need to be recorded once; communications sent to the wrong address will reduce; and it will be easier to collect fines and identify fraud. Having a single and accurate population register in the UK will improve statistical analysis and the quality of output. This will enable better evidence-based government policies, such as schools and hospital requirements.

What needs to change to make these linkages work?

The National Identity Register could be expanded to become the strategic solution for delivering centralised contact details model across the entire UK government. Bringing in under-16s would need addressing separately, for example by converting the Child Index in England to become a child population register. government will need to ensure there is a de minimis number of unique citizen numbers across the entire public sector. This is a complex project over 10+ years.

What are the privacy issues and how might they be tackled?

Government will need to put in place legislation, procedures and audit processes to underpin the use of administrative data for statistical purposes and the longer-term case for a secure statistical register drawing on the NIR and other indexes and administrative registers.

What concerns are there likely to be?

The Office of National Statistics has canvassed views: the public is generally unaware of the CIP proposal. When asked to express an opinion, 85% were positive or neutral; 35% believed that it already exists. There were some reservations, however, on information-sharing with the private sector.

Technological, privacy and regulatory issues

Technological drivers

34. Developments in technology will enable greater and more effective electronic delivery of public services, and increase the ability of public services to make effective use of large amounts of electronic data¹³. Technology is changing rapidly, and is impacting in two main areas:
- information architecture and software engineering - the ability to link databases together and infrastructure/services to make them work
 - privacy enhancing technologies – information security and assurance, regulation and governance.
35. Computer processor speed, disc capacity and communication speeds are all doubling every 15 to 18 months. Devices and computing infrastructures for data storage are ubiquitous, cheap and powerful. More information is being held about more people in more areas affecting their lives.
36. This is leading to an ambient computer infrastructure where computing surrounds an individual: offering up constant information and entertainment, recording financial transactions, monitoring health, optimising transportation, logging geographical presence, and other services. Just as individual citizens have already surrendered information about their personal buying habits and all kinds of associated information to do with lifestyle and social economics, new information spaces will be rich with personal content, both created by ourselves, and by others about us.
37. The integration of diverse sets of information is happening at an increasing rate and on an ever-growing scale. The UK has been at the forefront of efforts to harness the value of integrating diverse sets of scientific information in its national e-Science programme¹⁴. The Ministry of Defence is basing its whole doctrine of Networked Enabled Capability (NEC) on information integration¹⁵. In every case the distributed infrastructure of the World Wide Web is the key technology and its evolution into a Semantic Web of richly annotated and shared content is gathering pace¹⁶. The Semantic Web relies crucially on agreeing core areas of shared terminology to facilitate the integration of information spaces. The commercial drive towards Service Oriented Architectures is also about connected web services using shared standards about how information is represented and mapped. Semantic Web technologies will enable powerful reasoning services to develop over this shared content.
38. All this can be foreseen from current technology drivers. The rate-limiting step is not therefore the technology: in many cases the technological development is already there. The issue is predicting the rate of development in this area and the socio-economic implications. government needs to be clear where it wants to move to in this whole area, how it proposes to get there, and once there, how it ensures that the systems deliver what they are intended to deliver within a safe and secure environment.

¹³ See for example: *Cyber Trust and Crime Prevention*. Foresight, 2004. Available via www.foresight.gov.uk.

¹⁴ See for example: Hey and Trefethen. *Cyberinfrastructure for e-Science*. Science Vol 308, Issue 5723, 817-821, 6 May 2005.

¹⁵ See *Network Enabled Capability*. Ministry of Defence, 2005. Available at <http://www.mod.uk/issues/nec/>

¹⁶ T. Berners-Lee, J. Hendler, O. Lassila. *The semantic web*. Scientific American, May 2001.

39. In order to help inform these assessments, we recommend a study which draws up a technology road-map and plots it against relevant socio-economic and other drivers to enable better forecasting of what practical, monitoring and protection services need to be in place and by when.

Privacy-enhancing technologies

40. Public services have the responsibility to ensure that they are confident about the identity of the individual receiving a particular service; that they have the correct data about individual citizens; and that they can keep that data secure. Greater joined-up service delivery, together with more remote interactions through new technologies, raises new challenges in these areas¹⁷.
41. The legal framework for human rights and privacy has been evolving rapidly¹⁸ and there are signs that public concern about privacy is increasing. A progressive trend favouring an individualist over a collective view of society has meant that the rights of the individual have been recognised and given more prominence¹⁹. Therefore, individual citizens have formal rights and legitimate expectations that their privacy will be protected.
42. Privacy concerns are fuelled in part by the power of technology - in particular the power of data-combing services on the internet and how instantaneous the processes are - and the ease with which personal data held on a private database can be combined with publicly available datasets. The use of the electoral register as an effective marketing tool is one clear example.
43. Privacy enhancing technologies enable users of information technology to interact with each other and with service providers, revealing only the minimum of identifiable information necessary to perform particular transactions. They may also serve to inhibit or prevent unlawful data processing. The design of privacy enhancing technologies depend on the extent of the cooperation, and goodwill, of the different participants: some assume that service providers can be trusted to respect the rights and preferences of users; other scenarios assume a very hostile environment where an adversary may exploit all available means to breach privacy.
44. We believe that government, as a major user of personal data, should make clear what privacy enhancing technologies it is likely to need in the future so that business can work with government to develop such technologies. Government should work closely with business experts from the private sector (such as the banks) in identifying needs. We also believe that government needs to develop more explicit and proportional confidentiality requirements in its procurement specifications.²⁰

¹⁷ Trust and privacy issues are discussed in *Cyber Trust and Crime Prevention*. Foresight, 2004. Available via www.foresight.gov.uk

¹⁸ The regulatory framework consists of Common Law (which recognises that an obligation of confidence arises between particular relationships eg between doctor and patient); the Data Protection Act 1998 (which sets out eight main principles which apply to the use of personal information. The Information Commissioner is responsible for overseeing and promoting compliance with the Act); the Human Rights Act 1998 (which embodies Article 8 of the European Convention on Human Rights and imposes a test of necessity on the invasion of the private and family life of an individual); and the Freedom of Information Act 2000 (which gives a right of access to any information held by a public authority). The extent to which public bodies can collect, use and share personal data also depends on their administrative powers, or vires.

¹⁹ *Managing challenges of consent and confidentiality in research on personal data on medical research*. MRC Data-sharing and Preservation Initiative, 2005.

²⁰ The public sector, including government, is a significant procurer of IT and accounts for an estimated 55% of all spend in the UK on IT services and systems.

45. Privacy enhancing technologies cease to operate effectively against high-powered and prolonged data-mining techniques – both of which combine to increase the chances that connections between disparate pieces of data could be made, and individual identities revealed. Technologies are needed which work at high speed to allow web-browsing over longer timeframes without the risk of identities being discovered.
46. We believe, therefore, that there is a need for government to secure greater levels of investment by business into IT security, in order to promote greater confidence and trust. It should also encourage private sector organisations which are involved in privacy R&D to share R&D ideas on security modelling.
47. In addition, government needs to widen research into software resilience/dependable systems²¹. This should include, as a priority, initiatives to foster more interdisciplinary research – involving computer scientists, engineers and social scientists – into techniques of anonymising and pseudonymising data; encryption; and the development of anti-virus devices.

The current regulatory framework

48. A strategy for achieving better data use needs to be underpinned by a clear framework for decision-making on privacy and data-sharing issues. There is real confusion about the current regulatory framework – in particular the operation of the Data Protection Act (DPA).
49. While the DPA does not provide powers to process or share data, it does not at the same time prevent the processing of data. However, the popular perception of the DPA is exactly the reverse. Consequently, there is no clear guidance on how to act properly. There is, however, a very substantial range of guidance, protocols, agreements between organisations about data-sharing arrangements in specific circumstances which often complicate rather than simplify the overall situation.
50. For example, in the medical area there is both a need to safeguard the interests (in this case the personal data) of participants in research, while at the same time minimising the management and bureaucratic burdens associated with the use of their personal data so that valuable research can proceed²². There is a fear that widespread uncertainty among medical researchers may lead to a disproportionate effort on risk management in areas where there may not be a genuine problem²³. The public perceives real benefits in the capacity to better understand, predict and control disease through the intelligent mining of health records, but are concerned about the security of data, and the potential of such databases to contain lifestyle information²⁴. On balance, there is a feeling that the

²¹ The EU, through its 6th Framework Programme, is spending around 140 million euros (£94 million) on R&D on all aspects of cybertrust.

²² For example, research in the 1970s linking prescriptions with routine health records and hospital admissions and deaths which used personal datasets to highlight highly significant associations between the use of minor tranquilisers and the risk of serious road traffic accidents would be unlikely to be permitted today.

²³ *Managing challenges of consent and confidentiality in research on personal data on medical research*. MRC Data-sharing and Preservation Initiative, 2005. The Medical Research Council (MRC) policy prompts investigators, after first considering what is appropriate and reasonable, to make explicit provision for the preservation and sharing of data in the planning and execution of their research. It is developing guidance on data management and sharing, as well as conducting a study on issues surrounding consent and confidentiality for use of personal data.

²⁴ *Public dialogue on cybertrust and information security*. Royal Society, October 2004

prospect of better healthcare outweighs potential problems of breaches of data collection and management. There is a need to build on this positive position.

51. CST considers that government needs to review the different sets of guidance which different departments have issued, particularly in relation to the Data Protection Act, and ensure there is consistency between these. Looking ahead, government also needs to consider carefully what legislative changes might be necessary, and whether there is a need for a general purpose power to promote sharing of, and access to, personal datasets between public sector bodies.

Moving forward – three core principles for data-sharing

52. CST believes there is a need to widen access to data for researchers, policy-makers and others while at the same time providing necessary safeguards. CST considers that this should be done on the basis of three linked principles (**Box 10**):

Box 10 - Access to personal data: three principles

Principle 1

Personal data must be **anonymised** whenever this is possible, or alternatively in the case of linked datasets (including databases that are updated with personal data on an ongoing basis) the data must be **pseudonymised**.²⁵

Principle 2

There should be a general presumption that **access to data should be facilitated** where that access is for research or statistical purposes.

Principle 3

Appropriate **safeguards** - statutory and other – and **transparent governance structures** must be in place **before** personal data can be accessed and used. This would involve setting open standards and open protocols (in particular for the conditions under which data can be shared) and drawing up appropriate guidance and codes of conduct.

53. One key safeguard that needs to be place to maintain public trust, is the establishment by data-sharers of formal policies for how such data will be handled during and after the study period. These should specify *who* is permitted to see the information; *what* classes of data may be accessed; *why* and *how* it may be accessed and used; and until *when* it may be accessed. Through these arrangements, individual citizens and organisations sharing data should:

- carry out risk analyses – including consideration of possible unintended consequences - to quantify their own risk from entering into sharing relationships with new partners in order to determine the appropriate level of safeguards²⁶;
- be able to act to prevent or reduce risk and damage in the event of misuse by data by partners;

²⁵ Anonymised data are information that cannot be traced back to an individual (i.e. the encryption key has been thrown away); pseudonymised data are anonymised data where the encryption key has not been thrown away (i.e. it can be traced back to an individual). Pseudonymised data is an important instrument for medical research, and are the type of data for which the need for safeguards is absolutely paramount. It is important to recognise that in certain situations in certain areas of relatively low population, for example Northern Ireland, even anonymised information may be traceable back to a named individual.

²⁶ The Risk Assessment process could cover a number of risks: eg (i) loss of confidence and trust in privacy; (ii) unauthorised use of personal data; (iii) exploitation of individuals for commercial gain; (iv) statistical discrimination eg creating a sub-culture of non-participation by individuals; (v) technical risks such as database failure or incapacitation (eg by spam or unmanageable volumes of data); (vi) poor data quality.

- demonstrate clearly that potential partners conform to laws and/or public expectations, and have a shared understandings about the nature of the data and the attendant scale of risks;
- address real and potential conflicts of interest, and any specific issues, for example involvement of special/vulnerable groups.

54. In general, the contractual relationship between those who share data needs to be clear and explicit. Furthermore, governance of the personal data storage function should be separated from ownership of the process to access the information and those that give the rights of access. Governance of use, sharing and analysis of personal data needs to be transparent and monitored by an independent and trusted third party to ensure the system has both credibility and trust. For example, the Economic and Social Data Services run by the Social Science Data Archive together provide a centre of expertise in data acquisition, preservation, dissemination and promotion in the UK. They have a gatekeeper role in checking licensing arrangements, copyright and access conditions.

55. We also believe that, in order to promote greater trust, government should promote a quality mark or accreditation scheme for the operation of government databases. Monitoring compliance should be through third-party audit, in which the Information Commissioner should have a central role, with agreed delegated authorities. Non-compliance with statutory safeguards should result in severe penalties.

Public attitudes

The case for public engagement

56. Government needs to be confident that its evolving policies on the use of personal data carry broad public consent. As part of our project we commissioned OPM Ltd to carry out a two-stage process to explore public perceptions of the current and future use of personal datasets held by public sector bodies²⁷. The objective of Stage 1 was to understand the views of participants before they were exposed to relevant materials or argument, in particular to identify the key questions people raised in relation to the use of their personal data by government and researchers and explore the benefits, aspirations and concerns people perceive that greater use of their personal data by government and researchers could bring. The objective of Stage 2 was to understand the way that individual citizens made decisions between the benefits of greater use of their personal data and the potential losses of personal privacy, and to identify common principles that might help guide development of future government policies on data linkages and access. Stage 2 also explored how best to engage in discussion with the public on the issues raised, in particular the level of public interest in further engagement, their views on how this might take place and their expectations in terms of the legitimacy of the process.
57. The sixty-seven participants in the seven focus groups fell into three main groups – uninterested, undecided and objectors. Throughout these groups two common themes appeared: first, a general feeling of powerlessness that greater sharing of personal data by government would happen anyway; and second, a view that the government was already able to find out information on the individual if it were inclined to do so. Many participants found it difficult to formulate definitive views on the more specific issues discussed and tended to become more uncertain as they became more aware of the complexities of the issues involved. The exception is the overall positive attitudes to the use of personal information by healthcare professionals and researchers.
58. In its report *Policy through dialogue*²⁸, CST proposed criteria against which government could identify those science and technology issues where an investment in public engagement and dialogue would be likely to bring returns in enabling and easing the process of forming and implementing policies. We have assessed the issues around data-sharing against our criteria, drawing on the findings of the focus groups (**Box 11**).
59. Overall, we judge that the issues raised by the linkage of personal data are of the type that policy development should be informed by dialogue with the public. This needs to be a genuine two-way process and would need to be supported by more concrete models of what greater data-sharing could mean for individual citizens – both in terms of the benefits that might be gained and the practicalities of how the system would work. One particular challenge in managing a wider debate is the apparent absence of vocal advocates for data-sharing to balance the minority of vocal objectors.

²⁷ *Research into the use of personal datasets held by public sector bodies: report for the Council for Science and Technology*. OPM Ltd, October 2005. Available via www.cst.gov.uk

²⁸ *Policy through dialogue: informing policies based on science and technology*. March 2005. Council for Science and Technology. Available via www.cst.gov.uk

Box 11 - Assessment against CST's Dialogue criteria**Core criteria**

The envisaged development in science or technology is feasible or there is a significant societal issue that could be addressed using potentially controversial existing technology.

The technology to link and gain access to personal datasets already exists and is developing rapidly. But extending the use this technology may be controversial because of public concerns (often informed by personal experience of interactions with public and private sector bodies) over issues such as: privacy; data security (heightened by concerns over identity theft); maintenance of accurate records; misuse of data; accountability; government's ability to deliver both the technology and the potential benefits.

Potentially controversial ethical issues arise around the conduct of the scientific research, the use of the technology and/or the wider impacts on society. For example: the benefits and risks to different parties (e.g. individual citizens, society, government, industry) are inequitable; the benefits to individual citizens are unclear; individual citizens may have limited or no choice over their use of the technology; risks fall to particularly vulnerable groups.

The collective public benefits of greater linkage between databases are reasonably easily understood and articulated; it is more difficult to describe the potential benefits to individual citizens. Further, individual citizens are unlikely to have a choice (or be asked for their consent) over how their personal data is used, including whether and how it is shared between government agencies, once it has been entered into a database.

The timetable for the development of policy allows for a dialogue process to inform developments.

There is limited knowledge among the public of the way in which their personal data can be used now; we found that as people learn more, their level of uncertainty rises. We are recommending a step-wise incremental approach to increasing data-sharing, which would enable government to put an integrated public dialogue and policy development framework in place.

Additional criteria

New regulatory or governance procedures may be needed

Greater sharing of personal data is likely to require legislative changes and will require new governance arrangements.

Benefit-risk trade-offs

60. We consider that it is crucial for government to explore the trade-offs that individual citizens are prepared to make between the benefits of greater sharing of personal data resulting from better connectivity between databases and the potential losses of personal privacy. There is evidence that people are content to give up personal information to a wide range of services in return for various personalisation and price incentives²⁹ or where they believe that collective public benefits outweigh potential losses³⁰. For databases managed by government and its agencies, issues around the extent to which people expect or want to see personal benefits may be key to getting buy-in to more

²⁹ Trust and privacy issues are discussed in *Cyber Trust and Crime Prevention*. Foresight, 2004. Available via www.foresight.gov.uk

³⁰ *Public dialogue on cybertrust and information security*. Royal Society, October 2004.

widespread uses of personal data and greater connectivity between datasets. In the OPM study thirty-five out of sixty-seven respondents agreed with the statement “*I don’t trust government agencies enough to want them to share any more of my personal information*”, twenty were neutral and eight disagreed. However, when greater information sharing was couched in terms of more personalized services, the trust issue appears to be less apparent, with neutral respondents moving to a position of support. Forty-two respondents agreed with the statement “*I would support greater use of my personal information by government agencies if it led to services that met my needs better*”, twelve were neutral and nine disagreed. The most frequently identified benefits are social – reducing crime and anti-social behaviour; increased information sharing across public agencies is not seen as individually ‘enabling’. All focus groups acknowledged that a more streamlined interface between the individual and public agencies, with fewer requests for repeat information and a reduced capacity for conflicting information, would have benefits.

Earning public trust

61. The work carried out for us by OPM did not imply a crisis in trust said to characterise the current relationship between citizens and their governments. But there is a need for government to establish institutional and governance arrangements that earn public trust before the benefits of connecting personal datasets can be fully realised. Mechanisms for achieving this need to be explored with the public but, as seen in other areas, there is likely to be a desire for processes that are easily understood, with clear lines of responsibility, and with some oversight functions carried out by bodies that operate at arms length from government.
62. There is also a big question in people’s minds about the ability of government to manage the technical changes and issues involved in better linking between and access to personal data. In recent years, the public sector’s reputation in procuring IT systems has suffered, and this may affect public attitudes to proposals for linking databases. A failure to deliver planned database linkages may not be any more controversial than failures in other public sector IT systems, which tend to focus on wasted public money and, to a lesser extent, on lost services. However, the stakes become higher where people are asked to trust in their personal data being protected by privacy enhancing technologies even where that data will be anonymised or pseudonymised. In these cases, public acceptability may be predicated on a confidence that the technology can deliver. At the same time, no systems operate infallibly all the time. There must be answers to the following questions - where do responsibilities lie; and how should rectification and recompense be provided in cases where the security of personal data held by government is compromised.
63. There appears to be increasing public trust in some uses of personal data, such as in banking systems, but less so in others. For example, there are concerns about data-sharing amongst companies with large personal datasets for the purposes of direct marketing. Government needs to explore these variations in trust, and in particular the low level of trust (20%) by the public in the government’s ability to protect personal information, which appears to be driven by the view that it is insufficiently joined-up and skilled to handle data effectively and securely³¹.

³¹ *Public dialogue on cybertrust and information security*. Royal Society, October 2004. *Research into the use of personal datasets held by public sector bodies: report for the Council for Science and Technology*. OPM Ltd, October 2005.

Personal responsibility

64. The work carried out for us by OPM indicated that people felt they should have access to their own information, to be able to check details were correct and to be able to monitor by whom and how it is being used. They also felt that the responsibility for maintaining accurate records should remain with government and were uncomfortable with the proposition that people could update their own information electronically, largely because of concerns of fraud.
65. However, we feel that in the longer term greater convergence between personal datasets will need to be accompanied by empowerment of individual citizens to have more of a role in the management and use of their own data held by the public sector. This would mean more responsibility on individual citizens to provide accurate data and ensure it remained correct throughout changes in their lives³².
66. We believe there is a need for some creative thinking here which develops the concept of citizens *owning* their own data, which would enable them to be able to exercise some control over how and when their data was used as well as putting more emphasis on individual citizens to ensure their personal data was up-to-date. While accepting there are circumstances where government must retain the right to use individual citizens' data, and that government has a role in preventing malicious exploitation, the basic concept would be that citizens themselves would own their personal data. Ownership would mean, as a minimum: that individual citizens would have the right to know what government and agencies hold on them, with certain explicit exceptions; that there would be a shared responsibility between individual citizens and government to maintain the accuracy of the personal data; and citizens would have the right to know the ways in which government uses personal data. This concept needs to be explored further – at present it appears to be too remote from people's own experiences for them to engage with it. It will mean addressing a range of issues, including whether individual citizens have the necessary skills to ensure their information remains accurate and protected.

Ways forward

67. We believe that government should undertake further public engagement and dialogue with individual citizens and opinion-formers to inform its policy developments on data-linkages and access to personal datasets, and the privacy safeguards that need to be in place. It should use the principles set out in the CST report: *Policy through dialogue*³³.
68. As part of this process government should:
- sponsor interactions between different stakeholders and the public; and educate individual citizens – especially young people – about personal data and its use in order to promote understanding on how individual citizens could better take responsibility for managing their personal data. Schools and universities would have an important role here;

³² Note that individuals already have certain responsibilities already, for example to provide accurate data, not commit fraud or criminal activity, respect civil judgements.

³³ *Policy through dialogue: informing policies based on science and technology*. March 2005. Council for Science and Technology. Available via www.cst.gov.uk

- encourage better articulation of and debate about the risk–benefit equation that must be analysed and understood in order to promote or restrict greater sharing of personal data;
- determine where responsibilities should lie and how rectification and recompense will be provided in cases where the security of personal data held by government is compromised.

Acknowledgements

69. This report was prepared for the Council for Science and Technology by a subgroup comprising Professor Janet Finch, Professor Wendy Hall and Dr Mark Walport (convenor). We are grateful to all the people who participated in our two workshops, and to the authors of the case studies for allowing us to use them.

Contacts

70. For further information contact:

Council for Science and Technology
Bay 307
1 Victoria Street
London SW1P 0ET

Tel: 020 7125 2876
Fax: 020 7215 0313

Email: cstinfo@dti.gsi.gov.uk

Web: www.cst.gov.uk

Annex A: Risks and risk management

Developing opportunities for increased use of personal datasets is not without risk i.e. there may be negative consequences which are not intended and can only be foreseen in a general sense. For example, technical and organisational systems for processing personal data may conform to broadly accepted principles of data protection, but still intrinsically infringe notions of individual privacy. This is therefore a sensitive area, and one where wide-ranging public engagement will be essential.

There are also ethical issues which need to be considered. For example: invasion of privacy may be acceptable when government is tracking behaviour which has a predominant impact on the 'public good' (eg to improve public health, reduce crime), but straying into private areas such as lifestyle may be unacceptable - tracking diet may be broadly acceptable, tracking sexual behaviour may not be.

1. Risks to the Public

Erosion of confidence in privacy with loss of trust

Issues:

- Blurring the public/private divide
- Greater invasion of privacy – tensions between the legal need to provide data and data provided voluntarily. Regulatory environment becomes too relaxed. Government acts unlawfully

Actions Needed:

- Government departments (DCA/RIU/OGC) to develop more explicit and proportional confidentiality requirements in procurement specifications for government IT services
- Government reviews regulatory environment and streamlines data protection protocols
- Public dialogue and consultation
- More work on public perceptions of privacy; the validity of privacy-enhancing technology; and role of audit as a means of scrutinising data-sharing eg by the Information Commissioner (see case studies)

Unauthorised use of personal data

Issues:

- Damage to the individual
- Consequent damage to government's reputation – with political ramifications

Actions Needed:

- There is a view that government should acknowledge that such errors cannot be eliminated entirely and should therefore focus on ensuring that such errors are minimised, as well as designing systems for managing them when they do occur.
- Getting consensus around acknowledgement of the fallibility of systems will require intense public dialogue on how such breaches will be handled eg single point of responsibility; quick rectification; recompense etc.
- Define ownership rights more clearly: for example individual citizens taking responsibility for the accuracy of their own data. Controls on how data can be used for statistical and other purposes. Case studies on how unauthorized uses of data have been handled.

Exploitation of the Individual for commercial gain

Issues:

- Risks to individual citizens – denying a service which could benefit them; offering a service of no value or which is harmful; use of personal information for commercial gain to an organisation which disadvantages the individual or at best does not allow sharing of any benefits eg insurance companies using personal data to decide what terms, if any, to offer to individual citizens; discrimination against individual citizens as a result of statistical analyses
- Damage to the reputation and finances of commercial organisations
- Political risks where government seen as ineffective in regulating use of personal data

Actions Needed:

- Raise awareness of, and debate, these risks before embarking on any changes
- Intense public debate on effects of statistical discrimination to establish a regulatory/legal framework

Creation of a sub-culture of non-participants

Issues:

- Political risks – (i) noisy dissenters reduces credibility (ii) incompleteness of information weakens policy analysis

Actions Needed:

- Intense public dialogue, including communication of the benefits to society as a whole – through the press and by other means. Consultation rather than imposition
- Make participation a requirement for individual citizens' having access to certain benefits

- Educate the next generation on the importance of participation

Statistical discrimination

Issues

- Ethical, moral and social risks that particular individual citizens and sections of society will be discriminated against

Actions Needed

- Public discussion of the extent of the risks and which, if any, forms of statistical discrimination may be beneficial, and which not. Separate the risks (eg medical) into those that individual citizens can influence, and those which they cannot
- Transparency; and building into data systems a recognition that individual citizens' circumstances change
- Look at experiences elsewhere eg in Scandinavia and Iceland

2. Technical risks – systems and security

Datasets in different databases will be of differing quality - which may lead to errors in point to point data and, consequently, to administrative mistakes.

One means of increasing confidence in the security of personal data would be the deployment of technology which would require all data owners to hold their data in encrypted format with the ability selectively to release data elements. Researchers etc. requiring access to a dataset would need to generate a document justifying the need for access – the test would include necessity, proportionality and compliance with Data Protection and Human Rights law. The request and response would provide an audit trail available for review e.g. by the Information Commissioner.

Databases fail leading to failure of services

Issues

- Operational and research risks – loss of data; denial of services to [vulnerable] individual citizens, particularly time-critical services
- Technology failure will have greater impact if there are fewer, larger databases
- May lead to resistance to further integration of databases

Actions Needed

- More robust IT systems

Fallibility of information infrastructure

Issues

- Operational and political risks eg spam crisis renders incapacitates databases

Actions Needed

- Filter out spam

Unmanageable volumes of data

Issues

- Information overload: sharing and communication of information becomes unmanageable
- Over-reliance on data seen as a substitute for other forms of analysis
- Failure to meet raised public expectations that data-sharing automatically will lead to public benefit

Actions Needed

- Reduce volume of data to core levels which are actually needed
- Review regularly the data that is held, and whether it is still needed

Poor quality of the product of integrated datasets

Issues

- Operational risk, but poor data would have political ramifications – badly derived policies

Actions Needed

- Plan and define the specifications and compatibilities carefully at the outset; monitor performance regularly
- Training for staff
- Dialogue with the public