

Data Sharing Review

Richard Thomas and Dr Mark Walport

Consultation paper on the use and sharing of personal information in the public and private sector

List of questions for response

We would welcome responses to the following questions set out in this consultation paper. Please follow the question order as set out in the consultation paper, leaving a blank response box for any questions not answered.

Please email your completed form to contact@datasharingreview.gsi.gov.uk

Alternatively you can send a hard copy response to:

Data Sharing Review Secretariat
5.26 Steel House
11 Tothill Street
London
SW1H 9LJ

Thank you.

Section 1: Background

Question 1.

Comments:

The Enterprise Privacy Group is a think-tank dedicated to developing best practice in the handling of personal information. Private companies and public authorities join the Group to collaborate on solutions for their privacy, data protection and identity management concerns. Data sharing is one of a number of privacy areas of interest to the Group.

This response has been provided by a selection of Members of the Enterprise Privacy Group, but does not necessarily reflect the views of those Member organisations. Our response is limited to those questions that were specifically of interest to the Group.

Section 2: Scope of personal information sharing, including benefits, barriers and risks of data sharing and data protection

Question 2.

Comments:

In the context of data sharing, we consider 'society' and 'government' to be interchangeable, since society's needs in this area are represented by the government.

Key benefits of sharing personal information to individuals arise from convenience and a simplified interface with public authorities: when authorities have the right information at the right time in the right place, they are in a better position to deliver prompt outcomes for the individual. Eligibility for social benefits can be identified; health care services delivered more effectively; cumbersome, repetitive form-filling avoided. Countries such as Finland have created an environment where citizens can register a life event, such as a house move, with a single notification to the authorities. DirectGov's 'Tell Us Once' initiative seeks to achieve the same objective.

Society benefits from data sharing – and in particular the Transformational Government agenda – arise from delivering technology-enabled citizen-centric services that improve customer experience, achieve policy outcomes and enhance efficiency. Government has access to reliable datasets to facilitate long-term planning, to the extent that a data sharing 'panopticon' might even do away with the requirement for a National Census. Service delivery can be managed more efficiently and targeted at those individuals that require it. Such cost reductions are attractive benefits for society.

Data sharing is also a key tool in the identification and investigation of benefit and tax fraud, as well as delivering national security and immigration controls. These are 'secondary' benefits for law-abiding citizens, since they do not observe a tangible benefit on a day-to-day basis.

It should be noted that the majority of benefits in the UK model fall to the government rather than the individual. This is not the case in Estonia, where the balance of data sharing is proportionate: individuals are able to access their own records to see what information is held on them, and this provides proportionality between the data controller and the data subject. Transparency provides a critical and commendable check over government personal data management, and goes a long way towards dispelling citizens' fears about data sharing problems.

Question 3.

Comments:

The key risk to both individuals and society/government arising from data sharing in the way it is understood and implemented in the UK is the aggregation of small datasets into larger, monolithic databases without proper consideration of purpose, or examination of alternative methods to deliver the same desired outcomes. Nearly every example of data sharing in central government involves the aggregation and merger of existing databases, rather than creating cross-referenced indices that permit lookups between separate databases. There are virtually no examples of federated or user-centric schemes in use at the national level. The problem is compounded by databases being allowed to grow larger and more complex than they were ever designed to be. The National Identity Scheme has been criticised for this approach.

Large databases are, by definition, a threat to privacy. In nearly every case, an aggregated database will require more users than separate databases, and this increases security risks exponentially. The larger number of users introduces risks of data quality failures; theft of personal information by authorised users; and audit failures where it becomes impossible to hold authorised or unauthorised data users accountable for their action. The database offers a single point of failure,

where separate databases would compartmentalise failures from each other, thus limiting the impacts of any failure.

A further risk to individuals arising from data sharing is the 'de-anonymisation' of data. The statistical community has tried and trusted methods to anonymise data, but this depends upon end users not having the ability to cross-reference data with other sets; for example, if one database holds postcode data but no names or addresses, and a shared dataset includes disability status but no identifying information, then it would be relatively straightforward to identify a wheelchair user in that postcode. This process is known as 'zippering'. In Germany, which had a troubled history of centralised data processing in the Second World War and subsequently in the Cold War, the constitution forbids the central government from holding such databases, which are instead managed on State level.

Citizens are also at significant risk from derived data and metadata. Usage audit trails can imply significant things about individuals; for example, a personal record may be innocuous, but if accompanied by an audit record that shows the police have accessed it on many occasions in recent months, it implies suspicion about that individual, even though those suspicions may be entirely unfounded. Such data is subject to the controls of the Data Protection Act, but if the data subject is unaware of its existence, then they will have no motivation to look for it until they are adversely affected by it. Failures in the metadata can lead to incorrect derived data that can cause failures in traceability, consent or ownership of the original data.

Finally, risks to the individual arise from a prevailing culture of opaqueness about information sharing in the UK. Public authorities are not accustomed to seeking valid consent for such sharing, and few individuals would have any idea where to start looking for personal information in shared systems.

In summary, risks to the individual arise from:

- confusion of 'data sharing' with 'data aggregation' within public authorities;
- dependence upon large merged databases without consideration for alternative technologies that might deliver the same or better outcome;
- unintentional (or deliberate) 'de-anonymisation' ('zippering') of anonymous data through linked datasets;
- creation of derived data and metadata;
- lack of transparency about the purpose or extent of data sharing.

The risks arising from data sharing are strongly weighted against the citizen rather than society, since the misuse of personal information has a proportionally greater impact on the affected individual than it does upon companies or the government. Individual failures of confidentiality, integrity, availability or accountability may, if infrequent and limited in impact, be overlooked at a societal level, but there is a broader risk that would arise from a general loss of trust, by the public, in government data processing. If confidence breaks down, then delivery of Transformational Government benefits will fail, with a massive impact on current and forecasted expenditure. Furthermore, there would be no opportunity in the foreseeable future to remedy the situation: citizens who have lost faith in the State's ability to protect their data will not provide consent to further processing, and success of such programmes would thereafter depend upon coercion and compulsion.

Question 4.

Comments:

Despite the warnings about database failures in the response to Question 3, it should be noted that there are opportunities to improve the security offered by database systems. Designing a database, or a federation of databases, for purpose, rather than reusing existing systems; incorporating granular access controls to ensure that individual users have access only to the minimum amount of information required for the task in hand; comprehensive audit measures to monitor usage; are just examples of controls that can be used to improve security in these systems. An additional technique to protect centralized database systems is that of introducing 'false' information and watchlists so that inappropriate access or modification of data can easily be identified as it happens, and leaks can be quickly attributed back to their source.

'Federated' or 'distributed' approaches to cross-referencing and sharing data are by no means a panacea from the risk perspective: such approaches, if not properly managed, risk data integrity failures arising from synchronization errors. But the opportunity presented by these schemes is one of truly citizen-centric data management, and the government should be investing in greater research into this area, including the development, deployment and trialing of pilot systems in low-risk applications.

Question 5.

Comments:

There are few, if any, examples of public authorities holding too little personal information. In general, problems arise where public authorities fail to share information between them, and this is being addressed by a number of projects including the DirectGov 'Tell Us Once' initiative.

The key concern in this area is not that of authorities gathering too much information, but obtaining that information through data sharing initiatives, where data is zippered (multiple data sources brought together) to create larger sets of aggregated data. Where such larger databases are created, it is critical that controls are in place to minimise disclosure of that data to authorised users, and to monitor that usage such that both data controller and data subject can inspect audit trails.

Question 6.

Comments:

It is common for private sector organisations to collect personal data that is excessive for provision of the service, but which simplifies marketing services for the organisation. For example, users are asked to provide a date of birth for enrolment into almost any service, such as a loyalty card, and many organisations require 'security' information such as a mother's maiden name for authentication. Because these identifiers are common across many service providers, they in fact create a security vulnerability that renders identifiers such as date of birth or maiden name worthless for authentication; this seems to be an inevitable outcome of the information age, but until a pervasive transition to alternative authentication mechanisms is complete, consumers are at risk. One example of such a failing was the loss of HMRC data, where

association of dates of birth with bank account numbers could provide an attacker with a guess at the data subject's PIN numbers.

Recent headlines have highlighted concerns about social networking services that allow users to provide and publish unlimited amounts of personally identifiable information. Users are often unaware of the potential consequences of publishing this data, and unable to remove the trail of data once it is published, either because of cached data or service providers' retention policies (for example, it is not possible to close a Facebook account).

Question 7.

Comments: No response.

Question 8.

Comments: No response.

Section 3: The legal framework

Question 9.

Comments: No response.

Question 10.

Comments: No response.

Question 11.

Comments: No response.

Question 12.

Comments: No response.

Question 13.

Comments: No response.

Question 14.

Comments: No response.

Question 15.

Comments: No response.

Section 4: Consent and transparency

Question 16.

Comments:

There is a lack of clarity about what constitutes valid consent for data sharing, and how that consent should be managed. Current data collection mechanisms do not provide sufficient granularity for the individual to consent to what information is collected, how long it is held, with whom it is shared, and the purposes of sharing. Generalised 'opt-in / opt-out' notices do not cover a necessary level of detail, and rarely provide a transparent mechanism for data subjects to subsequently change their consent permissions or force the deletion of given personal data from its initial storage location and all other locations to which it

has been transmitted for other purposes.
Much valuable work in this area has been done by the Trustguide project
(www.trustguide.org.uk).

Question 17.

Comments:

The phrasing of this question suggests that it might be generally acceptable to share information without consent, something that would certainly not be desirable. Practical consent mechanisms are an essential component of a data sharing strategy for it to be legally compliant and culturally acceptable.

At an individual level, consent mechanisms can become barriers to data collection; for example, recent media coverage of social networking sites is now impacting willingness of individuals to sign up for those services or share data with them. However, overly-complex consent mechanisms could equally inhibit data sharing. There is unlikely to be a 'one size fits all' approach, and further research is required into consent mechanisms that can take into account the needs of a broad spectrum of citizens.

Question 18.

Comments:

There is a pressing need to introduce a 'universal Subject Access Right,' one which allows individuals to a) identify which organisations hold information on them without having to submit multiple Subject Access Requests, and b) to obtain copies of all data, metadata and derived data such as audit trails. Clearly such a right will only be practical if supported by a regulatory body that monitors data sharing. This role could be one for the Information Commissioner's Office, but only if coupled with appropriate extra resources to manage the workload.

Question 19.

Comments:

Whilst the Framework code of practice for data sharing, and privacy impact assessments, are both welcome initiatives it is too early to assess their effectiveness in public authorities or private organisations.

Section 5: Technology

Question 20.

Comments:

Technological advances have, in general, weakened the protection of personal information. Necessary processes and controls have failed to keep pace with technology developments, and a consequence of this is that data sharing is running away from understanding of its consequences. There is a pressing need for greater research and education of technologists, public authorities and citizens to understand what their needs are, and how best to serve these with technology.

Furthermore, government should take leadership in developing frameworks to protect data sharing both within public authorities and the private sector. Model 'privacy protecting policies' would allow organisations to understand what constitutes an acceptable baseline for the management of data sharing, and enable

citizens to make a more informed judgement of an organisation's sharing policies when being asked to consent.

Question 21.

Comments:

Whilst it would not be practical to create, apply and maintain laws that keep pace with the speed of technological change, there is a need for baseline standards to enforce a common standard for what constitutes protection of personal information. Data processors need more specific rules – and associated penalties – for what is considered to be reasonable protection of information.

Before any modification can be made to the law on technical safeguards, data controllers require a more detailed definition of what constitutes personal information or sensitive personal information. There is confusion about how to identify information types, and whilst clearly there will be 'grey areas' and exceptions to any rule, any changes to the law would be ineffective without a mandate to define types of information before safeguards are defined.

A significant step would be to mandate that any portable storage device containing personal information must be subject to encryption. This would provide an effective level of assurance that where data is lost, it has not been compromised; for example, CESG guidelines allow for laptops with suitable encryption to be handled in accordance with protective marking rules at a lower level – 'confidential' data may be treated as 'restricted' if it is encrypted with an approved system. Modern encryption systems are not onerous to deploy or manage on an enterprise scale.

Furthermore, data protection law should specifically mandate the audit of systems and processes to confirm that safeguards are adequate, and identify incidents promptly so that data controllers and data subjects can take remedial actions as soon as possible. Audits will discover where personally identifiable information is held and used, and commercial products already exist to automate this discovery process. Such an approach would benefit both data controller and data subject.

The law should also provide more comprehensive guidance for data controllers on 'end of life' issues, with definitions of best practice for the disposal of IT equipment, removable media and paper documents which contain personal information.

Where data loss incidents occur, the law needs to provide guidance on redress/restitution for affected data subjects. Data controllers that lose data, and are shown to be culpable for that loss, should be held liable to compensate data subjects if the lost information has had a material impact on those individuals. For such a law to be proportionate, it must also define the window of liability after the incident, such that the data controller can draw a line under liability after a given period.

Question 22.

Comments:

The term 'privacy enhancing' is misleading, since it implies a way to give more privacy, rather than protecting the privacy to which the data subject is already entitled. 'Anonymisation' is also an ambiguous term, since truly anonymised data is generally only useful for statistical purposes. Any intermediate level of anonymisation risks the use of 'zippering' to identify data subjects by cross-

referencing with publicly available information.

There is a need for greater research into 'privacy protecting techniques' that consider 'context' of data usage, since this can be the determining factor for the sensitivity of information from the individual's perspective. For example, address information may not be sensitive at the time an individual gives consent to sharing, but becomes sensitive if that individual is the victim of an abusive partner from whom they are trying to hide. Context is, of course, an extremely difficult concept to recognize in an algorithm or process, and this is why it must be explored in detail.

Greater consideration should be given to how volumes of personal information affect risk management techniques. At what point does a database contain sufficient personal information that it requires greater security controls simply because of the quantity? Is it acceptable to provide less security for smaller amounts of data? How can government incorporate proportionality into its risk management procedures so that individuals do not suffer security lapses such as those in the HMRC CD incident?

Both private organisations and public authorities would benefit from the availability of automatic content-based classification and policy management tools. Systems that could, for example, analyse the content of email messages, and encrypt or even block them if sensitive personal information is found. Such an approach would also work for web transactions, instant messaging and removable media.

There is a need for greater exploration of 'brokerage' ideas – the provision of identification and data sharing services by third parties in a competitive environment. For example, banks, mobile phone providers or supermarkets – those organisations that already have a data-driven relationship with their customers – might also offer data brokerage services such that data processors come to them for data when it is required, and the broker manages the consent process on behalf of the individual. Brokerage is a complex topic, but has the potential to simultaneously deliver greater privacy and reduced cost of operation for data sharing exercises. It may even be possible to mandate that data controllers must prove that aggregation of data into centralized databases is the only practical solution to an information problem, that that user-centric approaches such as OpenID mechanisms are not usable in their given case.

Section 6: International comparisons

Question 23.

Comments: No response.

Question 24.

Comments:

The adoption of 'Binding Corporate Rules' (sometimes referred to as 'Model Contracts') by multinational companies has created an environment in which such organisations commit to compliance with a certain level of data protection, even if that exceeds the regulatory requirements of a number of their 'host' states. This contrasts with a number of key initiatives in the UK that have followed the 'Safe Harbour' approach:

- UK airlines disclosing disproportionate amounts of personal data to US immigration authorities without the knowledge of consent of the data subject, under the 'air passenger data' scheme - subsequently amended slightly in the face of EC protest, but not substantively;
- UK public sector organisations off-shoring data and data processing to US multi-nationals without public due diligence or audit of the subsequent security of that data (for instance, Driving Standards Agency).

Question 25.

Comments: No response.

Question 26.

Comments: No response.

Section 7: Additional questions

Question 27.

Comments:

There is an apparent culture within the UK that personal information, once gathered, can be used for any purpose that is related to the original purpose for collection, even if the link is marginal. It is rare for an organisation to approach data subjects and request permission to use information for a fresh purpose, and when that does happen it is often hidden in small print.

There is a clear need to create a culture, and associated procedural and technical mechanisms, to seek 'opt-in' for fresh usage purposes when required. A rise in the number of contacts from data controllers requesting fresh consent would demonstrate a successful change in the culture.

This culture shift might be compared with that of the recent surge of interest in free-range poultry – data controllers and data subjects need to recognize that obtaining and managing personal data in an ethical manner is not necessarily an easy proposition, and will almost inevitably cost more in the short term, but the long-term benefit will be a much greater level of trust between all parties concerned.

Question 28.

Comments:

There is a general assumption within the questions provided – particularly those referring to the private sector compliance with the Data Protection Act – that data sharing is a 'good thing,' when this case has not in fact been proven. Sharing should have to be demonstrably of benefit to both the individual and society as a whole, rather than just an administrative convenience. There is as yet no sense that personal information belongs to the individual, and that is only used with that individual's consent.

