

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: Sapior Ltd develops and markets privacy-enhancing solutions to facilitate ethical data sharing. Its Pseudonymisation-based De-Identification software is the de facto data privacy standard for the NHS Spine/SUS project.

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

Question 2.

Comments: One-size-fits-all offerings by definition can't actually fit anyone very well and lead to waste and inefficiencies. The huge benefits that can come from sharing personal information stem mainly from a better understanding of the data subject(s), which enables the provision of tailored services and products. In addition, improved knowledge of service demand and usage can facilitate improved management through cost savings, efficiencies and elimination of waste.

All of these examples are applicable to both the public and private sectors and can lead to improved healthcare and government services, as well as

competitive advantage for commercial organisations. Obviously, the individual benefits from more personal service, but society also benefits from improved use of money and resources to do more good.

Question 3.

Comments: The main risk of sharing personal information is the misuse of that data by reuse in ways the data subject did not intend or want. This basic invasion of privacy and the fear of harm are the main risks to individuals. Society loses because systems will not be built that can improve our lives.

Subsequent distrust of organisations by data subjects and the related harm to the data subjects are the potential risks resulting from any misuse.

An example would be the use of health records to screen insurance applications as a requirement of a non-medical purpose such as a loan. Another is the use of genetic proclivities to make judgments rather than on a personal basis.

Question 4.

Comments: The greatest opportunities for personal information sharing come from a Risk-based Data Sharing Framework which would increase sharing opportunities whilst simultaneously reducing the risks of data privacy breaches. Such a system would either indicate an acceptable level of personal information to be made available for sharing or help choose from people/process risk mitigation techniques based on an analysis of the measured breach risk of a given recipient. The three main components of this system include:

1. Setting acceptable risk appetite levels
2. Measuring the breach risk of the dataset at the intended recipient
3. Reducing the privacy breach risk to acceptable by deploying appropriate risk mitigation (technical, people or process)

Data sharing methods that pose the greatest risks are projects that store large volumes of highly individual or characteristic data (about people) within a single large centralised database. This architecture assumes all the data and intended recipients have the same degree of riskiness. This system would either overly denude the dataset in order to protect data privacy or only provide minimal data privacy protection.

Question 5.

Comments: Examples of where public authorities hold too much data include:

1. Police – juvenile records and genetic databases – Some type of record aging policy is needed to reduce future misuse due to the availability of information and the increased chance of false positives that come from an increasingly large data set.
2. DVLA – address details linked to vehicle information – There is a risk of unexpected crimes of opportunity, revenge, etc.
3. NIR – large scale centralized database with comprehensive biometrics – losing this facilitates electronic impersonation. Also tracking capabilities of individuals across all walks of life, though useful for security forces,

raise significant concerns of future misuse.

Question 6.

Comments: Example of where the private sector holds too much data include:

- Experian, 192, Zoom info – lifelong individual histories including both public and purchased information – There is considerable risk when this information is made available solely upon the exchange of funds and without any analysis of the riskiness of the data set or the data recipient.

Example of where the private sector holds not enough data include:

- Healthcare providers/researchers/entrepreneurs – Current data sharing restrictions significantly restrict innovation.

Question 7.

Comments: The example given in Question 6 of potential data sharing to support health care innovation is a good instance of where beneficial data sharing is not yet taking place in sufficient volumes.

The main barriers to this activity are the privacy concerns of the data subjects which cannot yet be fully met. The legal response to address these concerns has been broad based, as is typical in cases where the activity being restricted is in its infancy. However, this legislation should be refined over time as techniques are developed to enable risk based privacy-enhanced data sharing.

These privacy concerns can be overcome by tailoring both how the shared data is, itself, protected, combined with a risk mitigation programme for the recipient to reduce the breach risk to within acceptable levels.

Question 8.

Comments: Again, the previously mentioned example of where data sharing should not occur:

Experian, 192, Zoom info – lifelong individual histories including both public and purchased information – There is considerable risk when this information is made available solely upon the exchange of funds and without any analysis of the riskiness of the data set or the data recipient. Without some sort of data breach notification requirement or other system that applies some cost, and thus value, to personal data, the burden of the cost of a data breach is borne predominantly by the data subject.

Section 3: The legal framework

Question 9.

Comments:

Question 10.

Comments:

Question 11.

Comments:

Question 12.

Comments:

Question 13.

Comments:

Question 14.

Comments:

Question 15.

Comments:

Section 4: Consent and transparency

Question 16.

Comments:

Question 17.

Comments:

Question 18.

Comments:

Question 19.

Comments:

Section 5: Technology

Question 20.

Comments: Technological advances have had tremendous impact on data sharing: both in positive and negative ways.
--

On the positive side, technology (such as PKI and SSL) has enabled internet banking/shopping and the related efficiencies and market opportunities.

More negatively, technology often gives undue comfort and peace of mind. One example of this is the planned increase in sharing of pseudonymised health data without preventing the increased risk of inference attacks. This is the case in instances where non-NHS organisations receive health data and are not subject to improved levels of organisational data privacy. This ignores the increased privacy breach risk associated with sharing pseudonymised data more widely than before.

Question 21.

Comments:

Question 22.

Comments: Privacy Enhancing Techniques (PETs) such as pseudonymisation can be used on a default basis to de-identify personal data being used for analysis.

Obviously, the NHS Spine/SUS project is an example of this in practice, where the conscientious deployment of the technology has occurred in an

environment focused on reducing privacy breach risks.

Pseudonymisation can also help safeguard privacy for medical research when part of a (per project) tailored privacy breach risk mitigation strategy.

When applied to government inter-departmental sharing we are confident that pseudonymisation can significantly enhance their data privacy. For example when needing to share with the NAO, the NAO would access a de-identified version of the departmental data to draw their sample. This sample (and no more) would then be re-identified for transmission to the NAO.

There is not sufficient advice on deployment of such risk based techniques because the concept is new. We are however confident that breach risk measurement is the only objective basis for navigating this "sharing versus privacy" quagmire.

The barriers to deploying a risk based data sharing strategy are:

1. building a risk measurement technique to quantify breach risk of a dataset at a recipient.
2. existing business activities need a lot of help to break their dependence on identifiable data.
3. existing systems will need technical updates to accommodate de-identified data and then re-identify upon request & permission.

Section 6: International comparisons

Question 23.

Comments:

Question 24.

Comments:

Question 25.

Comments:

Question 26.

Comments:

Section 7: Additional questions

Question 27.

Comments:

Question 28.

Comments:
