

# 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](mailto: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:

- We acquire patient-identifiable healthcare data, both NHS (Scotland) routine datasets and study-specific datasets
- Our data is collected directly by in-house systems (often Windows or web-based), and indirectly through NHS partner and other institutions. NHS acquisitions are via secure transfers over NHSnet.
- Identifiable data is held in a central, secure facility with named staff access
- The principle uses of the data are: epidemiological studies, genetic studies, drug and patient safety studies, audit, management, etc.
- Data extracts for studies are released with Caldicott Guardian approval to internal and external users as minimal, anonymous, encrypted datasets and delivered via networks only

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

#### Question 2.

##### Comments:

- Within the healthcare setting, data sharing offers improvement of health outcomes for both individuals and society. For example, the care of diabetics within Scotland has been considerably enhanced through the introduction of the SCI-DC diabetes register, constructed by pooling relevant information from primary and secondary care sources and others. This has also permitted substantial research underpinning future care strategies. See Appendix A for a wider selection of work using our data.

#### Question 3.

##### Comments:

- The sharing of information either between individuals and institutions, or between institutions, can suffer from various modes of failure. For example: loss of data (and its subsequent misuse); misuse of data through poor internal institutional controls; failure to share data accurately or completely; differences in strength of security and privacy mechanisms between two data-sharing institutions. For the individual, this failure to share data appropriately and carefully can result in: (a) a loss of autonomy, (b) misrepresentation (through erroneous or false data), (c) exploitation, (d) increased liability, (e) may simply result in a failure to act. Society may lose confidence that the data-sharing mechanisms are viable and that the risks outweigh the benefits. It is therefore important to improve transparency of decision and data sharing processes to prevent a loss of confidence in those institutions that are vital to our lives today.

#### Question 4.

##### Comments:

- There should be no general restrictions regarding scope as the opportunities and risks of sharing data are dependent on the details and should be assessed on a case-by-case basis. This is particularly true of healthcare and related sectors. For example, the value of linking data between the food retail sector and routine healthcare datasets is inestimable, but the perceived privacy risks are high. That is not to say this shouldn't happen if satisfactory methodologies could be developed.
- Any data-sharing architecture has an opportunity and risk profile. Centralised sharing of existing datasets has the greatest risks with respect to unauthorised use of data, access control failures and inappropriate exposure of data through cultural or security failings. Opportunities for sharing centralised data will be complex because of the above risks and the controlling organisation's perceived remit. Virtualisation of distributed databases reduces only some of these risks. The main feature here is a single entity operating as a conduit for sharing data.
- Other distributed architectures incorporating multiple middlemen have much-improved risk and opportunity characteristics but these need carefully designed, and require supportive political and economic models. A key feature of these approaches is a clear separation of activity by participating institutions. (See response to Q28)
- Many organizations share information by copying data from one system to another sometimes incorporating the data in the receiving system. This can create a chain of dependency if this data is 'live' which can cause a variety of problems. Specifically, the ability to correct data is much reduced if not impossible. These sharing methods, which create multiple copies of data, are not desirable and their use should be minimised.

Question 5.

Comments:

- When we talk about public authorities holding 'too much data' we are talking about data available within one institution or readily available to one institution (through linkage say). The NHS holds a great deal of information on most of the population, but complete data is generally held within a few small systems, such as primary care, while incomplete data is dispersed among a few larger systems such as hospital PAS. This configuration is generally not viewed as posing high risk. However, the consolidation of complete data in a single large system is viewed as being high risk. The proposal for a consolidated EHR within the NHS therefore presents a challenge and may well represent a case of the public authorities holding too much information. Similar concerns surround complete 'linkage' databases such as the proposed UK National Identity Register where many personal identifiers are held together within one institution for the entire population. We suggest that the issue of 'too much data' is one of configuration, rather than quantity.

Question 6.

Comments:

- We are all familiar with centralised services such as Google, Facebook and Amazon. Amazon is considered low risk because it holds a limited type of data on any given individual and this is sufficient for its purpose. Amazon is not accused of holding too much data. On the other hand Google holds very large amounts of unstructured data on a substantial fraction of the population. Because Google's business is advertising, almost any data about an individual is useful and necessary to its purpose. Google's purpose undermines the DPA's principle of data holdings being adequate, relevant and not excessive. Google is regularly cited as holding too much information.

Question 7.

Comments:

- Medical research is almost universally regarded as beneficial to society. However, the sharing of healthcare data for medical research has proved difficult over the last decade with great uncertainty in the availability and timeliness of access to data by researchers. Many studies directly collect further patient data raising the additional problem of how such data is linked with routine healthcare data and by whom. The barriers to sharing of these data are numerous: legal responsibility, authority and institutional culture, competition, and lack of agreed technical mechanisms. The choice of technical architecture for the sharing of data – not the specific technologies - is important to the removal of all these barriers. (See Q28 for further discussion)

Question 8.

Comments:

### **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:

- Explicit informed consent should never be a substitute for other measures ensuring safe linkage and sharing of data.
- The sharing of identifiable information between distinct institutions should always require explicit informed consent for the precise content to be shared. Intra-institutional sharing of identifiable information should not require consent. Voluntary arrangements between individuals and institutions formed from a recruitment process should also require explicit informed consent based on knowledge of the recruitment process and purpose. The concept of informed consent is obviously problematic. However, when the use of data does not require identification then anonymisation has often been considered sufficient to negate the requirement of consent as is the case in Scotland.
- Example initiative requiring informed consent: Scottish Family Health Study (<http://www.ncbi.nlm.nih.gov/pubmed/17014726>)

Question 17.

Comments:

- There are few barriers to gaining consent for use of identifiable data in the health sector or where patients are recruited directly by the research community. However, the form of the consent may still be problematic.
- The sharing in bulk of anonymous healthcare data is an increasing requirement in research. There are many proposed options for handling consent in such circumstances but none address the fundamental issue: for a specific research question for which specific data is required, would a patient choose not to take part, even though at a general level they are supportive?

Question 18.

Comments:

- Transparency in data sharing can only arise from the introduction of appropriate mechanisms covering data-holding institutions, data users and the development of political bodies and economic incentives necessary to support it. All requests to link data from distinct institutions should be submitted to a legitimate independent authority. A

formal review process should be followed to decide on whether the linkage should be permitted. The review should cover individual dataset characteristics, linked dataset characteristics, re-identification possibilities in relation to the recipient of the linked dataset, ethics, etc., and result in a formal, public process specification for the construction of the linked dataset. The process is then executed by the relevant institutions. (See Q28 for further discussion)

- Offering individuals access to their data does not help them or others assess what is being done with these data. Access to the process descriptions and output descriptions of the data sharing mechanism would allow the public to directly experience how organizations use and share their data.

Question 19.

Comments:

- 

### **Section 5: Technology**

Question 20.

Comments:

- The following technologies have advanced the ability to share and protect data:
  - Availability of high bandwidth networks (share)
  - Secure network transmission protocols (protect)
  - Platform independent formatting such as XML (share)
  - Platform independent programming methods such as SOA (share)
  - Coding systems and ontologies (share)

The maintenance of privacy cannot be solved by technology; this is a matter of system architectures and institutional organisation and authority.

Question 21.

Comments:

- 

Question 22.

Comments:

- Anonymisation / pseudonymisation does not guarantee privacy of data, and data so prepared should still be viewed as personal information. The use of anonymisation techniques does however make the re-identification of a dataset non-trivial. It therefore helps to enforce formal rules forbidding re-identification, and these should be developed. The preparation of anonymous or pseudonymous data should not take place in a stand-alone manner, but must be part of a wider data-sharing mechanism.

### **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:
-----------

### **Appendix A - References**

#### **Genetics**

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes. **Science**. 2007 Apr 26

#### **Drug Safety Studies**

McMahon AD, Evans JMM, MacDonald TM. Pharmacoepidemiology Report. Hypersensitivity reactions associated with exposure to naproxen and ibuprofen: A cohort study. **Journal of Clinical Epidemiology**. 2001 2001;54:1271-4.

Donnan PT, Libby G, Boyter AC, Thompson P. The population risk of fractures attributable to oral corticosteroids. **PharmacoepidemiolDrug Saf**. 2005 3/2005;14(3):177-86.

#### **Diabetes epidemiology**

Tan HH, McAlpine RR, James P, Thompson P, McMurdo ME, Morris AD, Evans JM. Diagnosis of type 2 diabetes at an older age: effect on mortality in men and women. **Diabetes Care**. 2004 12/2004;27(12):2797-9.

Evans JMM, Wang J, Morris AD. Comparison of cardiovascular risk between patients with type 2 diabetes and those who had had a myocardial infarction: cross sectional and cohort studies. **British Medical Journal**. 2002 2002;324:939-42.

Evans JM, Donnelly LA, Emslie-Smith AM, Alessi DR, Morris AD. Metformin and reduced risk of cancer in diabetic patients. **British Medical Journal**. 2005 4/22/2005

Libby G, Murphy DJ, McEwan NF, Greene SA, Forsyth JS, Chien PW, Morris AD. Pre-eclampsia and the later development of type 2 diabetes in mothers and their children: an intergenerational study from the Walker cohort. **Diabetologia**. 2006 Dec 23

Laing SP, Swerdlow AJ, Slater SD, Burden AC, Morris A, Waugh NR, Gatling W, Bingley PJ, Patterson CC. Mortality from heart disease in a cohort of 23,000 patients with insulin-treated diabetes. **Diabetologia**. 2003 Jun;46(6):760-5.

### **Birth injury and epilepsy**

Murphy DJ, Libby G, Chien P, Forsyth S, Greene S, Morris A. Cohort study of forceps delivery and the risk of epilepsy in adulthood. *AmJObstetGynecol*. 2004 8/2004;191(2):392-7.

### **Demonstrating the link between antibiotic use in primary care and infection with antibiotic resistant bacteria**

P. DONNAN, L. WEI, D. STEINKE, G. PHILLIPS, R. CLARKE, A. NOONE, F. SULLIVAN, T. MACDONALD, P. DAVEY, (2004). Presence of bacteriuria caused by trimethoprim resistant bacteria in patients prescribed antibiotics: multilevel model with practice and individual patient data. *British Medical Journal*, 328, pp.1-5.

### **Using routine data to facilitate recruitment and outcome measures in clinical trials**

M. WELLS, A. HARROW, P. DONNAN, P. DAVEY, S. DEVEREUX, G. LITTLE, E. MCKENNA, R. WOOD, R. CHEN, A. THOMPSON, (2004). Patient, carer and health service outcomes of nurse-led early discharge after breast cancer surgery: a randomised controlled trial. *Br.J Cancer*, 91 (4), pp.651-658.

### **Measurement of the economic cost of adverse drug events**

S. V. MORANT, A. D. MCMAHON, J. G. CLELAND, P. G. DAVEY, T. M. MACDONALD, (2004). Cardiovascular prophylaxis with aspirin: costs of supply and management of upper gastrointestinal and renal toxicity. *British Journal of Clinical Pharmacology*, 57 (2), pp.188-198.

### **Improving quality of antimicrobial management for hospital inpatients:**

G. BARLOW, D. NATHWANI, F. WILLIAMS, S. OGSTON, J. WINTER, M. JONES, P. SLANE, E. MYERS, F. SULLIVAN, N. STEVENS, R. DUFFEY, K. LOWDEN, P. DAVEY, (2007). Reducing door-to-antibiotic time in community-acquired pneumonia: Controlled before-and-after evaluation and cost-effectiveness analysis. *Thorax*, 62 (1), pp.67-74.

### **Studies in orphan diseases**

Sullivan FM, Swan IR, Donnan PT, Morrison JM, Smith BH, McKinstry B, Davenport RJ, Vale LD, Clarkson JE, Hammersley V, Hayavi S, McAteer A, Stewart K, Daly F. Early treatment with prednisolone or acyclovir in Bell's palsy. **N Engl J Med**. 2007 Oct 18;357(16):1598-607.

Flynn RW, Macdonald TM, Jung RT, Morris AD, Leese GP. Mortality and Vascular Outcomes in Patients Treated for Thyroid Dysfunction. **J Clin Endocrinol Metab.** 2006 Mar 14