

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 General Practice Research Database (GPRD), owned by the Sec. State for Health and managed by the GPRD group at the MHRA, has an interest because although the GPRD research dataset is suitably anonymised all data comes from NHS Primary Care records that are stored at the personal level within the GP IT systems. Recently record-linkage of GPRD to other NHS datasets has been enabled that requires the use of a trusted third party (TTP) as the linkage is based upon personal identifiers. Using this TTP system it ensures, at the research level that all data is in a suitbaly anonymous format. The existing DPA and other associated regulations are, we feel, not explicit enough in enabling the correctly governed research that is in the interest of patients and the NHS.

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

Question 2.

Comments: Although much observational research takes place using suitably anonymised data this only happens because a person or group takes on the act of anonymising the data. Such anonymisation is in the interest of all parties and is a privacy enhancing technology. However, the DPA does not refer to the act of anonymisation as a readily acceptable processing of personal data that has obvious benefits. Record-linkage of clinical data held within the many current IT systems requires the use of personal identifiers, such as NHS number, post code, Date of birth, to enable it to be conducted at a level that provides high quality research data. Patients and society benefit from the use of such data; uses that are as fundamental to the delivery of healthcare as they inform future care as well as safeguard public health and the uses of a system of Trusted Third Parties should be explicitly mentioned.

Question 3.

Comments: Over 20 years of use, there have never been any complaints from members of the public or healthcare professionals about the way GPRD operates. This is perhaps testament to the way the processes are organised to limit the risks. As NHS data becomes more widely available for research it is imperative that this record is maintained.

Question 4.

Comments:

Question 5.

Comments:

Question 6.

Comments:

Question 7.

Comments: The linkage of all healthcare data is important to enabling research that is in the interest of safeguarding public health and ensuring that medicines are effective in real world use. Currently healthcare data "apparent" ownership or what appear as issues raised related to the DPA prevent such linkages either taking place at all or require exhaustive time to arrange. An example relates to the use of central death data in research. GPRD records, via the GP, most deaths but cause of death for "sudden deaths" is often missing. Such data can be fundamental to valid research on drug safety issues.

Question 8.

Comments:

Section 3: The legal framework

Question 9.

Comments: The DPA has much strength but has a weakness that it failed to be more directive about the processing of healthcare data from personal to suitably anonymised formats. There is also an issue related to what is "personal data" even after healthcare data has been essentially anonymised. It needs to be accepted that there will be incidents in which the identity of a record might

become known for rare events; particularly rare events about which there is knowledge in the public domain- a famous person with a rare disease. Such should not prevent research rather require that operating procedures related to preventing small cell size analysis are in place that mean that such situations are avoided to the highest possible level.

Question 10.

Comments: There is an issue related to the processing of personal health data and the second principal, that of the processing to make the data anonymised and into a format that will then see the data used many times and for many different purposes.

Question 11.

Comments: Lack of more explicit direction related to healthcare data and in a societal manner of the benefits of enabling the anonymisation of such data and its subsequent wide spread research use.

Question 12.

Comments: The power to enable anonymisation and pseudonymisation (when access back is required to verify coding as is often required, particularly in drug safety studies).

Question 13.

Comments:

Question 14.

Comments:

Question 15.

Comments:

Section 4: Consent and transparency

Question 16.

Comments: Consent to the use of various formats of healthcare data remains unclear, although the CfH Care Record Guarantee is a strong attempt to ensure that the protection of the person and research uses are handled. Information that patients wish to make available via the concept of sealed envelopes remains in contention. Drug safety research needs to take place on all classes of medication but there are instances where the key data is, under the proposed system, more likely to be sealed away; sealed away even from its use in an anonymised format. This will not serve the best interest of public health or that of certain individuals.

GPRD operates a system, at the practices from which it takes its data, of enabling patient opt-out from the use of their data even though it has been de-identified. Were mass use of this opt-out to take place research would be compromised as biases could be introduced. Our experience over 3 years of this system is that the rate of opt-out is very low and at a level that does not compromise research.

Question 17.

Comments: The use of explicit consent for the use of essentially anonymised data for research would have profound implications on the ability to safeguard public health as well as undertake other important research projects.

Question 18.

Comments: The GPRD system of posters and leaflets is one methodology for ensuring patients are aware of how their data may be used. A Care Record Guarantee that covered off all uses that have public benefit- see Q16 would be helpful.

Question 19.

Comments: A standardised and agreed central governance system or the use of detailed governance rules across all healthcare research would be beneficial.

Section 5: Technology

Question 20.

Comments: The availability of electronic healthcare data has transformed the way much healthcare research and particularly complex drug safety and effectiveness studies can be undertaken. The ease of applying anonymised codes to mass (millions of peoples) datasets is also worthy of comment. For the future the role of CfH in enabling greater access to NHS data for research will be important.

Question 21.

Comments: Where portable devices are concerned, their use should be discouraged as far as is possible. They should never be left unattended and never left in an insecure place overnight. All healthcare data should reside in a hidden and encrypted "partition". This is possible to achieve with in-expensive software. All data, even in the hidden partitions should be removed at every opportunity to a more secure locations and the data area used over-written. Such should be universal practice.

Question 22.

Comments: Most observational research can take place using de-identified data. Much of this research is however best conducted using pseudonymised data to enable a way back to validate coded information. The essential element is that those that hold the key to the way back are at a physically different location from the researchers and are under a differing control structure. For instance in the GPRD system the key is held by each GP practice and cannot be accessed by any GPRD staff or user of GPRD data.

It is felt by some researchers, particularly those using geographic data, that it is not possible to use de-identified data. Even in this arena it is possible to de-identify post-codes and other references so creating a de-identified dataset. However there will be circumstances in such research where the location of an event by geography means that closer identity will be known even if anonymisation takes place.

Currently either anonymisation or pseudonymisation is adopted by the large UK databases which keep control of the uses of data. What will be required if the data use landscape changes is an additional level of PET whereby each

research study, unless it can show why, would have its own research level anonymisation code. This has the benefit that data from different studies can never be linked without separate approval; linkage that could reveal a persons identity and should therefore be the subject of Ethics review.

Running large clinical trials with the need to obtain wide-scale recruitment of specific types of people has its DPA issues that can potentially be avoided by the use of simple PETs. It would be possible to run a national system whereby different parts of the CT recruitment process were handled by differing groups. In this way no one group or person would have enough information to cause a potential breach of confidentiality. Alternatively it could be done by a single IT system organised in a manner that no person saw the content of the patient letters.

Section 6: International comparisons

Question 23.

Comments:

Question 24.

Comments: The comment here relates to the fact that subject to comments already made the UK research system currently provides an international example of good practice. The following references supports this: Legal issues of data anonymisation in research: BMJ, May 2004; 328: 1300 - 1301 ; doi:10.1136/bmj.328.7451.1300 Petra Wilson

Question 25.

Comments: Spanish researchers are surprised to hear of the apparent way UK people are more protective of their health care record. Equally the northern European countries appear to use a common identifier across many types of records without their being a public outcry. However it has to be stated that although they appear to have all data linked there are more restrictions on what can be done than in the UK.

Question 26.

Comments:

Section 7: Additional questions

Question 27.

Comments:

Question 28.

Comments: The DPA refers to Medical Research in Schedule 3 but for Observational Research (use of historical data) Part IV 33 come into play . It would seem appropriate to be far more specific by the use of words and phrases throughout that have an acceptable and understood meaning.

Part IV section 33 (1) (b) also has, we believe, an unintended meaning- Drug safety studies may by their nature produce a result that causes a Medicine Regulator to remove a drug from the market. As a consequence of this there may well be

distress to patients on the drug either because they now feel they might have the adverse event in the future or because the drug suited them well and they would prefer to remain taking it.