1. INTRODUCTION

As technology advances, and with it capability in the areas of sharing and using data to its full effect, the privacy issues inherent in personal data become more and more apparent. Nowhere is this more acute than with medical data.

The potential benefits of providing shared access to personal digital information are becoming increasingly evident throughout the medical research and health service provision communities. Repository systems for medical data offer enormous potential for large scale research by a wide range of researchers which often extends well beyond the original purposes of data collection. However, such repositories are not without their risks.

“A national repository of electronic health records … in personally identifiable form containing highly sensitive personal information, such as chronic health problems, sexually transmitted diseases, neurological, genetic and psychiatric conditions can potentially give rise to gross privacy invasion.”

Livia Iacovino, Monash University

As such, it is increasingly vital that an appropriate balance is realised between maximising the potential advantages to the research community and to the general public while ensuring that individuals’ rights are respected with regards to their personal data. In the UK the protection of personal data against unlawful or unethical disclosure to third parties is covered by statute, the common law and non-statutory guidance.

2. THE DATA PROTECTION ACT 1998

The Data Protection Act 1998 (the DPA) (see our DCC Briefing Paper on Data Protection) places obligations on people and organisations that process personal data and in addition gives individuals certain positive rights in relation to data pertaining to them.

Personal data is defined as “data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.” This can include things such as an individual's personnel file, their medical records or home phone number. A further tier of personal data is defined in the DPA: sensitive personal data. This is personal data relating to more typically sensitive areas of an individual's life such as religion, race and political beliefs and any data relating to "physical or mental health or condition".

The primary implication of being covered by the legislation is that anyone processing personal data must comply with eight principles of good information handling. Research is given special exemptions in respect of the second and fifth principles. This means that, subject to certain conditions, further processing of personal data for research purposes is not to be regarded as incompatible with the purposes for which they were originally obtained and personal data which are processed only for research purposes may be kept indefinitely.

Fair and Lawful

The first principle in the DPA calls for data to be "fairly and lawfully processed". While it is not immediately apparent what is intended by this principle, further guidance is provided in Schedule 2 of the DPA which stipulates that for processing to be considered fair and lawful one of the six conditions must be satisfied. The conditions most likely to be relevant are:

1. The data subject has consented to the processing.
4. The processing is necessary in order to protect the vital interest of the data subject.
5. The Processing is necessary — (a) for the administration of justice […] or (d) for the exercise of any other functions of a public nature exercised in the public interest by any person.
6. (1) The processing is necessary for the purposes of legitimate interest pursued by the data controller or by the third party or parties to whom the data are disclosed, except where the processing is unwarranted in any particular case by reason of prejudice to the rights and freedoms or legitimate interests of the data subject. (2) The Secretary of State may by order specify particular circumstances in which this condition is, or is not, to be taken to be satisfied.
Where sensitive personal data is involved, Schedule 3 provides additional and more exacting conditions. Again, only one of these has to be satisfied to comply with the Act. The most relevant for medical data are likely to be:

1. The data subject has given his explicit consent to the processing of the personal data.
2. (1) The processing is necessary for medical purposes and is undertaken by — (a) a health professional, or (b) a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional. (2) In this paragraph "medical purposes" includes the purposes of preventative medicine, medical diagnosis, medical research, the provision of care and treatment and the management of healthcare services.

A further provision to note in the DPA is Section 10(1) which provides that an individual can require a data controller to stop or not to begin processing their personal data at any time in writing on the ground that:

(a) the processing … is causing or is likely to cause substantial damage or substantial distress to him or to another, and (b) that damage or distress is or would be unwarranted. This is quite a high hurdle for the data subject given that the request must be in writing and must show actual substantial damage or distress or likelihood thereof.

3. CONSENT

It is important to understand the role of consent in relation to medical data. Whilst consent is one of the conditions listed in each of the DPA Schedules, only one of those conditions in each Schedule need be satisfied to justify the use of personal data. As such, contrary to common belief it is not always necessary to obtain consent to process personal data. It should be noted however that, in addition to these Schedules, there is a general requirement, within the DPA's first principle that all processing be lawful. This includes meeting common law (i.e. judge made law as opposed to statute) confidentiality obligations, which are likely themselves to require consent to be obtained. So, especially with regard to sensitive personal data, best practice would be to seek explicit, positive consent to process the data in the majority of cases. This is only legally valid if properly informed (although as mentioned below the standards to be reached vary). Written consent is not legally necessary but is practically prudent.

In Scotland, if planning to use identifiable data without consent, approval from the Privacy Advisory Committee is required. In England and Wales, if it is not feasible to obtain consent from the study participant then you should gain approval from the Patient Information Advisory Group (PIAG) to access identifiable patient information for research purposes.

There is still a lack of clarity surrounding the concept of consent. A number of issues remain unresolved, such as:

- Can a patient agree to a general unspecified consent?
- If not, does the information that the patient needs to be given in order to consent differ depending on the use to be made of the data?
- Is consent time limited?
- Is consent required in any circumstance for access to a patient's de-identified data?
- If a patient revokes consent, what should happen to their data that has been used for secondary uses/research prior to the revocation. Should and, indeed, could further use be prevented?

4. ANONYMISATION AND CODING

If data is anonymised so as to prevent subject identification it is not covered by the DPA, thereby allowing for wider use. Practitioners may wish to consider whether any personal details in data being curated or preserved add to its usefulness or could be removed.

There are different ways that personal data can be modified to conceal identities. The definitions used by the Medical Research Council (MRC) are:
Coded information. This contains information which could readily identify people, but conceals identity by using a code, the key to which is held by members of the research team. This helps to meet legal and ethical obligations to protect personal information, but the identifiable data is still held by the research team. In this way coded data still falls within the scope of the DPA.

Linked anonymised data. This is anonymous to the research team that holds it, but contains coded info which could be used to identify people. The key in this case may be held by a larger research database or register.

Unlinked anonymised data. This contains nothing that has reasonable potential to be used by anyone to identify individuals. The link to the individuals has been irreversibly broken.

However it is important to realise that even with linked and unlinked anonymised data there may still be the potential to deduce individual’s identities through combinations of information, held either by people handling the research data or by those who see the results. This problem of connectivity applies where one piece of data on its own may not provide identification, but used together with another it may do so. It requires a practitioner to ask “does the person I am passing the data to hold data which may allow them to make the connection?” In many situations it will not be possible to answer this question with any degree of certainty. Exactly how much of this potentially identifying information can be safely included in data that is assumed to be ‘unidentifiable’ can only be judged on a case by case basis.

Further issues are raised where the anonymised data concerned might of itself uniquely identify a person. For example when researchers acquire MRI brain scans their anonymisation procedures strip off images of the face and ears, before they are analysed for psychiatric research purposes (the topic of a DCC SCARP Project case study). Other physical characteristics of brain scans may prove to be unique, so in this and other medical imaging areas the consequences of data sharing, and the limits placed on that, need continual review. Given the potential for legal ambiguity NHS research ethics committees and the research councils have a key role in regulating practice and providing guidance.

5. NON-STATUTORY GUIDANCE

It is important to acknowledge the role of medical research ethics committees and data access policies of the MRC and the Wellcome Trust. For example the MRC provides advice on how medical research data should be used (see Additional Resources). All researchers supported by MRC funds are expected to adhere to key principles. Such principles include for example:

- All medical research using identifiable personal information, or using anonymised data from the NHS, which is not already in the public domain, must be approved by a research ethics committee.
- All personal information must be coded or anonymised as far as possible and be consistent with the needs of the study. This should be done as early as possible in the data processing.
- At the outset, researchers must decide what information about the results should be available to the people involved in the study once it is complete, and agree these plans with the research ethics committee.
- Hospitals, and practices involved in research, must develop procedures for making patients aware that their information may sometimes be used for research, and explaining the reasons and safeguards.
- When consent is impracticable confidential information can be disclosed without consent only if:
  - The likely benefits to society outweigh the implications of the loss of confidentiality, so that it is clearly in the public interest for the research to be done
  - There is no intention to feed information back to the individuals involved or take decisions that affect them, and
  - There are no practicable alternatives of equal effectiveness

From MRC Executive Summary — Personal Information in Medical Research.
6. INTELLECTUAL PROPERTY

Intellectual property rights also play a part here. For instance, image data will attract copyright, and a database of medical data may attract the database right. This is not the focus of this paper but for more information see our DCC Legal Watch Paper on IPR in Databases.

7. FURTHER CONSIDERATIONS

- Technology that may protect privacy is not always compatible with the retention of records in the long term. For example, without adequate curation activity, encryption will be unsuitable for records that need to be accessed over time. Encryption adds a layer of complexity for long term preservation as the algorithms and software used to encrypt data are likely to become obsolete over time, while the access keys for decryption may become lost.
- The DPA refers only to living individuals and does not apply to the records of those who have died. The Access to Health Records Act 1990 deals with requests for access to records relating to the deceased.
- There is a crossover between the DPA and the more recently introduced Freedom of Information Act 2000 and Freedom of Information (Scotland) Act 2002. Personal data is one of the absolute exemptions in the Acts. This relationship can sometimes be unclear.
- As discussed above, gaining express consent from each patient whose data is used in research would be ideal. But is this an overly idealistic approach for large scale research databases? In practice it would be extremely complicated.
- Other statutory law that will be relevant in some circumstances is:
  - the Human Rights Act 1988
  - the NHS (Venereal Diseases) Regulations 1974 and the NHS Trusts (Venereal Diseases) Directions 1991
  - the Human Fertilisation and Embryology Act 1990 (as amended)
  - the Abortion Regulations 1991

8. ADDITIONAL RESOURCES

- DCC Briefing Paper on Data Protection — Mags McGinley, February 2007
- DCC Legal Watch Paper on IPR in Databases — Mags McGinley, October 2007
- data-protection@JISCMAIL.AC.UK — a JISCmail-based mailing list for those interested in data protection and related topics
- The Data Protection Act 1998
- Peter Carey, "Data Protection Handbook" (The Law Society) Edited by Peter Carey
- The Information Commissioners Office
- Common Services Agency (ISD) v Collie
- MRC Regulatory Support Centre
- MRC Guidance — Personal Information in Medical Research, specifically the simplified decision tree on using personal information in medical research (Figure 1, page 11) and the table describing controls on the use of information in medical research (Table 1, page 15)
- MRC Guidance — Executive Summary on Personal Information in Medical Research
- **Access to Collections of Data and Materials for Health Research** — a report by William W Lowrance to the Medical Research Council and the Wellcome Trust
- **Beyond the Tomb: Privacy and the long term preservation of EHRs in national systems: a case study of Australia’s HealthConnect project**, Livia Iacovino, Monash University
- **Records Management: NHS Code of Practice**
- **The NHS Confidentiality Code of Practice**
- **NHS Records Management Roadmap**
- Kalra, D et al. (2003). "**Security and confidentiality approach for the Clinical E-Science Framework (CLEF)**"
- CHERRI Project Report: **Clinical Recordings for Academic Non-Clinical Settings**
- Julian Peto, Olivia Fletcher, and Clare Gilham (2004). "**Data protection, informed consent, and research**". British Medical Journal 2004; 328: 1029-1030 (plus the interesting Rapid Responses to the article)