NIAA Data Handling Policy

Contents

1. **Scope**
2. **Background**
3. **Ten Basic Principles**
4. **Appendix 1: Relevant legislation**
5. **Appendix 2: Six Caldicott Principles**
6. **Appendix 3: Open data and data privacy**
7. **Abbreviations**

1. Scope

This document outlines data access and governance policy for all datasets coming under the responsibility of the NIAA HSRC.

Specifically, this includes datasets where the HSRC is the data owner/guardian (e.g. National Audit Projects; NAPs) and where the HSRC is a data handler (e.g. National Emergency Laparotomy Audit; NELA).

Where the HSRC is data handler, decisions about access will lie with the data owner.

Where the HSRC is the data handler/guardian, then this policy guides that activity.

2. Background

The Health Services Research Centre of the NIAA was established in March 2011 as a world-class centre of excellence in Health Services Research in anaesthesia, perioperative medicine, pain medicine and the anaesthetic specialties. It now oversees a number of national audits, including the NAPs, National Clinical Audits (e.g. NELA) and the new Sprint National Audit Projects (SNAPs) as well as facilitating clinical research for patient benefit.

This policy arises from an identified need for a consistent rational data governance policy for databases falling under the HSRC responsibilities that is consistent with the relevant regulatory framework (Appendix 1), current custom and practice (e.g. Caldicott principles; Appendix 2) and the on-going drive for open access to research data ("Open data").

3. Ten Basic Principles

1. "Open data" (see Appendix 1) is in the interests of all stakeholders in the research process, including researchers. Open data is increasingly demanded by funders, policy makers, public and patients and it has become a basic principle of data management that data should be made openly available unless there is a compelling reason not to do this.

2. Open data is also in the interests of the primary researchers as it allows secondary verification of interpretation of the data and therefore validation of the original findings.

3. The drive for open data needs to be balanced against the overriding right of patients and the public to confidentiality. Adequate anonymisation/psudonymisation of datasets, the use of "airlocks" to provide a "chinese wall" between researchers and primary data, and analysis rules for minimum analysis sizes will be important in this respect.

4. The drive for open data must also be balanced against the needs of the primary researchers to be able to capitalise, through scientific publications and other research outputs, on the process of project origination, development and delivery that led to the acquisition of the dataset concerned. Researchers may therefore wish for a *reasonable** period of time to deliver on the primary research aims and any subsequent analyses prior to data being made open. This period should be defined for each dataset.

5. It is important to distinguish between datasets where the HSRC are data owners and datasets where they are the data handlers acting for other data owners. Data owners, handlers and other stakeholders should be clearly identified for each dataset. This should ideally be done at project initiation, but for already established projects this should be completed as soon as possible.

6. The HSRC will keep a "Register of Data Holdings" with relevant metadata including the source, size and scope of the dataset, identity of the data owners/guardians, identify of the data handlers, "open data" policy and any special provisions of data governance (e.g. project data governance policy). This resource should itself be open, and will serve to signpost external researchers to datasets and the associated investigators and data owners in order to encourage collaborative research proposals. Decisions about collaborative research lie with the data owners. Where the HSRC is the data owner, or is asked for advice by data owners, the presumption should be that collaborative research is encouraged, unless there is a compelling reason not to collaborate. External collaborators will normally be required to provide any reasonable costs associated with collaboration.

7. All projects will require a data governance policy, consistent with principles laid out in this policy. This policy will be important to inform all elements of the research process including the design of the project, consent process and documentation (e.g. acknowledge "open data" policy), maintenance of confidentiality including anonymisation/pseunomymisation methods, analysis plan including provision to avoid identification of individuals, and reporting. For some projects this may include the provision of a "lock-out" that specifies that some or all of the data will not be "open", for example where there is reason to believe that the ability to collect the primary data may be substantially compromised by a requirement to subsequently meet open-data policies.

8. For aggregated datasets, the interests of primary data collectors, often NHS employees working within NHS trusts, need to be acknowledged and this may be facilitated by providing centre-specific "data dumps" to individual centres. Centre specific data belong to the relevant NHS trust. Aggregated data, whether anonymised or not, belong to the designated project data owner/guardian.

9. This data policy will be consistent with the relevant legislation as well relevant customs and practice (e.g. Six Caldicott Principles).

10. This data policy will be reviewed periodically (suggest annually and as needed) to ensure that it remains up to date with and consistent with relevant legislation.

***reasonable:** the criteria for "reasonable period of time" will vary from project to project. A clear definition should be specified for each project based on a defined duration of time following a clear start-time. The default start-time is likely to be the time the dataset is closed and cleaned and therefore ready to use for analysis. The duration of time will depend on the project but is unlikely to be less than 1 year and a clear justification would be expected for duration greater than 5 years.

4. APPENDIX 1: Relevant Legislation (www.opsi.gov.uk)

Modified from "An Information Governance Guide for Clinical Audit" produced by the Health Quality Improvement Partnership (HQIP).

Data Protection Act 1998

- This act makes provision for the regulation of the processing of information relating to individuals, including the obtaining, holding, use or disclosure of such information.
- All data subjects have a right of access to their health records, therefore all records should be traceable whilst in your care.
- Always bear in mind that the eight Data Protection Act principles require that personal data must:
    - be processed fairly and lawfully
    - be obtained or processed for specific lawful purposes
    - be adequate, relevant and not excessive
    - be accurate and kept up to date
    - not be kept for longer than necessary
    - be processed in accordance with rights of data subjects
    - be kept secure
    - not be transferred outside the European Economic Area (EEA) unless there is adequate protection.

Freedom of Information Act 2000

- This Act makes provision for the disclosure of information held by public authorities or by persons providing services for them.
- Under the terms of the Freedom of Information Act, anyone is entitled to apply for copies of clinical audit reports.
- Organisations may need to respond to requests for information regarding clinical audit projects and should ensure that when reporting clinical audit results, there is no stated link between audit conclusions and patients/clients or clinicians.
- Presentation of audit results should always have the approval of stakeholders.

Access to Health Records 1990

This Act has been superseded by the Data Protection Act but still applies to access to the records of the deceased. An Act to establish a right of access to health records by the individuals to whom they relate and other persons; to provide for the correction of inaccurate health records and for the avoidance of certain contractual obligations; and for connected purposes.

Human Rights Act 1998

This act gives further effect to rights and freedoms guaranteed under the European Convention on Human Rights.

The Human Rights Act requires that any invasion of an individual's private life is first subject to a test of necessity and proportionality. It is also underpinned by the Data Protection Act 1998.

Computer Misuse Act 1990

This act makes provision for securing computer material
against unauthorised access or modification; and for connected purposes.

Criminal Justice and Immigration Act 2008

The Secretary of State may by order provide for a person who is guilty of an offence under section 55 of the Data Protection Act 1998 (c. 29) (unlawful obtaining etc. of personal data) to be liable. If you use, obtain or disclose information recklessly and in contravention of the Data Proctection Act 1998 YOU may receive a fine or prison sentence of up to two years if you are sucessfully prosecuted under this Act.

Section 251 of the NHS Act 2006

Section 251 of the NHS Act 2006 re-enacted Section 60 of the Health and Social Care Act 2001. The terms Section 60 and Section 251, when used in relation to use of patient information, therefore refer to the same powers. These powers allow the Secretary of State for Health to make regulations to set aside the common law duty of confidentiality for medical purposes where it is not possible to use anonymised information and where seeking individual consent is not practicable. It was anticipated when section 251 powers were originally established that the NHS would develop mechanisms to seek, record and implement consent. Also that the NHS would endeavour to improve data quality and develop processes to link data in pseudonymised form, reducing the need for identifiable data to be used. These mechanisms are still being developed.

The Health Service (Control of Patient Information) Regulations 2002 (SI 1438)

These regulations were made under Section 60 of the Health & Social Care Act 2001 and continue to have effect under Section 251 of the NHS Act 2006. These regulations established class support mechanisms that support the use of information, one of which allows the use of patient information under strict controls, 'for the audit, monitoring and analysing the provision made by the health service for patient care and treatment'.

Common Law Duty of Confidentiality

This is a key issue in matters of sharing or using personal and/or sensitive information. For NHS purposes using personal information can be justified where the recipient needs the information because he or she is or may be concerned with the patient's care or treatment; the use of the information can also be justified for wider purposes such as improving quality of treatment, promoting effective healthcare administration or research. Where information is shared, there is an implied understanding that the information will not be used except where it is strictly needed to help the professional provide the service. Each member of the team, and any person who provides administrative or secretarial support, has an obligation to treat the information as confidential. The obligation of confidence owed by a professional covers not only information provided by the patient, but also information relating to the patient which the professional obtains from others.

5. APPENDIX 2: Six Caldicott Principles

1. Justify the purpose(s) of using confidential information.
2. Only use when absolutely necessary.
3. Use the minimum that is required.
4. Access should be on a strict need-to-know basis
5. Everyone must understand their responsibilities.
6. Understand and comply with the law.

6. APPENDIX 3: Open Data and Data Privacy

Prepared by Dr Mike Galsworthy

Open data

There is increasing demand for open data. In fact, in some areas it is becoming compulsory. This trend started in the United States; as part of the general demand for government transparency – which knocked on to openness about the science they fund and then the data produced by that funding. The US National Science Foundation (NSF) were the first and then National Institutes of Health (NIH) also started demanding that scientists they fund publish their data. However, they enforcement of this policy has been inconsistent. Now in the UK, we have followed the model with Research Council UK (RCUK) calling for open data. The NHS are also looking at making more data publically available.

The case for open data in general

Open data is good for the following reasons: scientific validation of previous work, overcoming scientific bias in publishing by having it all out there, efficiency- as all data are available to be used (and re-used), meta-analysis to cumulatively improve power to study phenomena, and innovation through exploitation of the data from creative angles not originally intended (including sometimes linking across databases). I've written a piece on it, if you want to read the summary of why it's important but what's holding it up: http://theconversation.com/funding-bodies-will-have-to-force-scientists-to-share-data-14788

A specific clash between open data and patient privacy

However, there is a particular conundrum with scientific research based on people: data privacy. The public have grown accustomed to their information being used in scientific research, but there is a bond of understanding, usually enshrined in a consent form and a promise of confidentiality. With open data, we do not promise that the data will be kept only in the research team and then destroyed – rather, we say that it's going to be out there for everyone, including companies to use. Generally speaking, the public understand the concept of open data and why it is so beneficial to science, but they often have concerns about commercial exploitation.

The other concern, of course, is being identified as an individual. There are threats such as people being to look up neighbours' medical information or potential employers looking at health information of people or their families.

The promise has always been that the data will be made available anonymously, but that proves tricky to actually do. For example, just removing names and addresses was originally thought to do the trick. However, in a very famous case in 1997, the insurance company for public workers in Massachusetts released the "anonymised" data - for sale. The Governor of

Massachusetts (William Weld) assured the public is was all anonymised. However, Latanya simply purchased access to the database – then took the Governor's zip code, age and sex. With this information and knowledge of one recent hospital visit by the Governor, she quickly found his record in the database and mailed his medical history to his office as a publicity stunt. See full story here: http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/ (includes other descriptions of anonymisation problems).

Latanya went on to study this, and found that 87% of all Americans (in year 2005) could be identified uniquely by the combination of zip code, age and sex. These issues are her full time profession – see her page here:
http://dataprivacylab.org/people/sweeney/

Now many people believe that they have learned from this and it's not really a problem, but we simply don't know the clever methods by which databases will be linked and processed in future. For example there is a recent *Science* paper that showed that "anonymised" genetic data could be identified a proportion of the time simply due to genetic relationships with the extensive named genetic data on line in user-contribution genealogy databases.

There is growing literature on methods to preserve anonymity for public release of data. Here is one article: http://www.nature.com/news/2010/100412/full/news.2010.178.html

But the long and short is that we can only do our best, but not guarantee anything absolutely. One of the best things to do is to make it as hard as possible to casually identify individuals and also get anyone downloading the data to legally agree that they will not use it to identify individuals – then there need to be monitoring groups and strict punishments of anyone using open data to identify individuals for nefarious reasons.

Who are the key players in open data and data privacy in the UK?

The most important source of information is the Information Commissioner's Office (ICO):
http://www.ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
That page has links to the ICO code of practice, some interesting blogs/bits on the topic and a link to The UK Anonymisation Network (UKAN). Well worth a thorough read.

Then you should also be aware that HSCIC are deeply involved in this and do get approached by researchers who need guidance. The HSRC will approach this group to engage in discussions.

Finally, the MRC and Wellcome (amongst others) currently have an Expert Advisory Group on Data Access (EAGDA): http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/index.htm .

7. Abbreviations

HSRC	Health Services Research Centre
HQIP	Health Quality Improvement Partnership
MRC	Medical Research Council
NAPs	National Audit Projects
NELA	National Emergency Laparotomy Audit
NIAA	National Institute of Academic Anaesthesia
NIH	National Institutes of Health (US)
NSF	National Science Foundation (US)
RCUK	Research Councils UK
SNAPs	Sprint National Audit Projects
UKAN	UK Anonymisation Network
EAGDA	Expert Advisory Group on Data Access