University of Oxford

Cancer Epidemiology Unit (CEU)

Policy:
De-identification

Version 1.0

Version History

Version	Issue Date	Author	Description
1.0	08-Oct-2014	Lucy Wright	Initial version

Table of Contents

1 Policy management

This is a controlled document with read-only rights for unit staff and administrative rights for the Information Governance Lead.

This document is one of a number that describe the detailed policies and procedures that support the master Cancer Epidemiology Unit Information Governance Policy.

Title:	De-identification
Location:	v:\IG_policies
Owner:	Information Governance Lead
Approver:	Information Governance Committee
Review:	At least annually (and more frequently if required to make improvements in response to audits or incident management findings)
Applicability:	All Unit employees. All activities performed by or on behalf of the Unit under contract. This policy is applicable to all CEU endeavours that involve personal or possibly personal data.
Interpretation:	Questions relating to the interpretation of this policy should be directed initially to the Information Governance Lead
Unit:	Cancer Epidemiology Unit (CEU) within the Nuffield Department of Population Health, University of Oxford

2 Organisation Roles and Responsibilities

The Unit Information Governance Policy describes the organisational structure, and defines key roles and responsibilities in relation to information governance, including:

- Unit Management Committee
- Information Governance Committee
- Information Governance Lead
- IT & Information Security Manager
- Senior Information Risk Owner

3 Policy wording

Convention	Description
Must	A policy provision that is mandatory
Should	A policy provision that is strongly encouraged but which may be ignored if there is good reason
May	A policy provision that should generally be followed
[…]	Text in [square brackets] does not form part of the policy but is provided by way of explanation or example

4 Abbreviations and definitions

<table>
<tr><th>Abbreviation</th><th>Description</th></tr>
<tr><td>Information Security Incident</td><td>An Information Security Incident is any event or occurrence that has resulted, or could have resulted, in the disclosure of confidential information to an unauthorised individual, a risk to the integrity of the system or data, or risk to the availability of the system.</td></tr>
<tr><td>Information Assets / Information Asset Owners</td><td>Information Assets are identifiable and definable assets owned or contracted by the Unit and which are ‘valuable’ to the business of the Unit. Information Assets may include the computer systems and network hardware, software and supporting utilities and staff that are required to achieve processing of this data, though Information Assets should not be seen as simply technical.
<p>In general terms, Unit Information Assets fall into one of four categories:</p>
<table>
<tr><th>Information Asset</th><th>Information Asset Owner</th></tr>
<tr><td>Clinical research study[1]</td><td>Principal Investigator</td></tr>
<tr><td>Administrative information[2]</td><td>Unit Administrator</td></tr>
<tr><td>IT infrastructure[3]</td><td>Director of Information Science</td></tr>
<tr><td>All other information</td><td>Unit Director[4]</td></tr>
</table>
[1] Examples include MWS, EPIC, DSW
[2] Examples include personnel files, unit accounts
[3] Examples include servers, firewall, networks
[4] or nominated deputy (e.g. Deputy Director)
<p>The Unit must maintain a register of Information Assets and their Owners.</p>
<p>Further information is provided in the Unit Information Governance Policy.</p></td></tr>
<tr><td>Personal data</td><td>Personal data means data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual. http://ico.org.uk/for_organisations/data_protection/the_guide/key_definitions#personal-data
<p>The Health & Social Care Information Centre have published guidance on the drawing the line between personal and non-personal data.</p>
<p>http://www.isb.nhs.uk/documents/isb-1523/amd-20-2010/1523202010guid.pdf</p></td></tr>
<tr><td>Risk assessment</td><td>Risk is the potential for an unwanted event to have a negative impact as a result of exploiting a weakness. It can be seen as a function of the value of the asset, threats and vulnerabilities. The process of risk assessment is discussed as part of the Oxford University toolkit. http://www.it.ox.ac.uk/policies-and-guidelines/is-toolkit/risk-assessment/</td></tr>
</table>

Abbreviation	Description
Sensitive personal data	Sensitive personal data means personal data consisting of information as to: (a) the racial or ethnic origin of the data subject, (b) his/her political opinions, (c) his/her religious beliefs or other beliefs of a similar nature, (d) whether his/her he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992), (e) his/her physical or mental health or condition, (f) his/her sexual life, (g) the commission or alleged commission by him/her of any offence, or (h) any proceedings for any offence committed or alleged to have been committed by him/her, the disposal of such proceedings or the sentence of any court in such proceedings. http://ico.org.uk/for_organisations/data_protection/the_guide/key_definitions#personal-data

5 Introduction

A fundamental principle of the Data Protection Act 1998 is to use the minimum personal data to satisfy a purpose and to exclude information relating to a data subject that is not necessary for the particular processing being undertaken. This principle is aligned with the Caldicott Principles familiar to NHS and Social Care organisations and is supported by both common law confidentiality obligations and the Human Rights Act 1998 which provides a privacy right for individuals.

The Unit has a duty to protect the privacy of individuals, including (but not limited to) employees and study participants.

The key principle is to ensure that, as far as is practicable, individuals cannot be identified from data, unless there is a legitimate and legal requirement for identification.

Effective de-identification (by anonymisation or pseudo-anonymisation processes) depends upon robust information governance and effectively trained staff who understand the importance of data protection and confidentiality.

6 Approaches to de-identification

6.1 Complete de-identification (“anonymisation”)

The aim of complete de-identification (“anonymisation”) is to obscure the identifiable data items within the person’s records sufficiently that the risk of potential identification of the subject is minimised to acceptable levels.

Complete de-identification is essentially a one way process that renders the individuals in the data unrecognisable, even to the data owner. As such, careful consideration should be given to whether this irreversible operation should be performed.

Complete de-identification can be achieved by:

- Removing patient identifiers [e.g. name];
- The use of identifier ranges rather than exact values [e.g. replacing date of birth with year of birth or age range];
- Partially scrambling the data. [In order to construct a sample set (e.g. to test an algorithm) it may be appropriate to “fuzz” the data (e.g. by applying random shifts to the data of birth in a reasonable range) in order to anonymise it. This is commonly done when a representative sample data set is required and manual construction is not practical.]

6.2 Partial de-identification (“pseudo-anonymisation”)

Partial de-identification (“pseudo-anonymisation”) refers to a process that replaces clear identifiers (e.g. name, subject number) with alternative identifiers that bear no overt relationship to the true values. As a consequence, linkage back to the original data or to another de-identified copy from the same source can be achieved with, and only with, knowledge of the de-identification key or algorithm. This allows legitimate linking of data sets and other information but prevents inappropriate or unauthorised access to the identifiable records.

[For example, subsets of the data may be supplied to researcher A and researcher B with different pseudo ID values. Although the two researchers cannot correlate their data, the Information Asset Owner retains the keys for both pseudo anonymisation schemes.]

To effectively pseudo-anonymise data the following actions must be taken:

- Each record containing participant identifiable data [e.g. name, subject ID] should have a unique pseudo-identifier attached and the original identifiers should be removed or replaced [e.g. removing forename and surname, and replacing date of birth with year of birth or age];
- Consideration should be given as to whether a single pseudo anonymisation scheme is sufficient, or whether, if the data is given to multiple targets, multiple schemes should be employed. [This is to reduce the opportunities for two released datasets to be recombined, producing identifiable data.]
- Pseudonyms should be used in place of NHS Numbers and other fields that might be used to back translate the data by reference to external data sources.
- Consideration should be given to using pseudo anonymisation schemes that have similar formats and lengths to the real data (so as not to break import mechanisms). However, if this is done, care should be taken to avoid confusion between the pseudo ID values and the “real” values.

7 Training

All staff must receive training in Information Governance at induction and annually thereafter. This should include De-identification procedures relevant to their role. Requests for additional training or guidance should be discussed with line managers or addressed to the Information Governance Lead.

8 References

8.1 Unit Policies

Policy
Information Security Policy v:\IG_policies
Unit Information Governance Policy v:\IG_policies

8.2 University policies

Policy
University of Oxford policy on Data Protection http://www.admin.ox.ac.uk/councilsec/compliance/dataprotection/
University of Oxford Information Security Policy http://www.it.ox.ac.uk/policies-and-guidelines/information-security-policy
University of Oxford Guidance on Risk Assessment http://www.it.ox.ac.uk/infosec/istoolkit/riskassessment/

8.3 External Guidance

Guidance
Information Commissioner’s Guide to Anonymisation http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation