Public Health Directorate

Pseudonymisation and Anonymisation of Data Policy – Version 3.1 (FINAL)

Version history:
Published: March 2016
Reviewed with edits: 16 February 2017
Reviewed with edits: 05 March 2018
Approved (PH DMT):   15 March 2018

1. Introduction

Data protection legislation[1] requires organisations to process the minimum personal data needed for the purpose at hand and to not use information that identifies individuals unless necessary. This is formally recognised in the second Caldicott principle, which states: "Don't use personal confidential data unless it is absolutely necessary".

In practice, this means that personal identifiable data (PID) can be used to support the direct care of individuals and/or the service they have consented to receive, or to quality assure the care provided. As far as is practicable, any other purposes should be supported by using data where individual service users cannot be identified.

Where this is not practicable, data should flow through business processes that minimise the risk of inappropriate use or disclosure. In many circumstances this requires data to be received by a part of the organisation designated as a 'safe haven', or operating as a virtual safe haven, where it can be processed securely and only used in an identifiable form for specific authorised procedures within the safe haven boundary. Appropriate information barriers, contractual agreements and appropriate business processes effectively support this. Onward disclosure, for purposes other than direct care or documented and agreed uses, should be limited to pseudonymised or anonymised data.

Effective pseudonymisation and/or anonymisation processes depend upon robust information governance and effectively trained staff who understand the importance of data protection and confidentiality.

The Information Commissioner's Office (ICO) lays out the good practice recommendations for effective anonymisation of data in the ICO code *Anonymisation: managing data protection risk code of practice* (https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/).

---

[1] Data Protection 1998, from 25 May 2018 General Data Protection Regulation (GDPR) and Data Protection Bill as implemented into UK law.

2. Policy scope

Although it references the uses of PID for direct care, and a limited set of uses for other public health and healthcare purposes, this policy is specifically concerned with the security of service user/patient identifiable information when it is being used for purposes other than direct care, known as secondary use. Direct care involves the primary use of these data.

3. Purpose and background

This document lays out Cambridgeshire County Council (CCC) Public Health Directorate's policy on dealing with PID for secondary use, via pseudonymisation and anonymisation. In doing that, it seeks to provide all Public Health staff who use service user/patient identifiable data with guidance to safeguard confidentiality when the data is used for purposes other than direct care and guidelines for those staff using pseudonymised, anonymised or de-identified data for secondary use.

While describing Public Health's policy for pseudonymisation and anonymisation, where appropriate, the policy does recognise the current position and the steps required to achieve the policy's full intent and provides an update on significant changes during 2017/18.

The first Cambridgeshire County Council Public Health Directorate Pseudonymisation and Anonymisation of Data Policy was published in March 2016. It was first reviewed and revised during January and February 2017 and has been again in February and March 2018. This revised policy reflects the results of the February 2018 review and, in particular, takes into account the new data protection legislation, known as the General Data Protection Regulations (GDPR) that come into force on 25 May 2018.

4. Definitions

Personal Identifiable Data (PID): any information that can identify an individual. This could be one piece of data for example a person's name or a collection of information such as name, address and date of birth.

Primary use: information is used for direct care and service provision purposes. This would directly contribute to the treatment, diagnosis or the care of the individual. This also includes relevant supporting administrative and service management processes and audit/assurance of the quality of the service provided.

Secondary use: information that is used is for non-care and non-medical purposes. Generally this could be for health improvement, population health surveillance, audit, commissioning, target-setting, evaluation, contract monitoring and associated reporting. When data derived from PID are utilised for secondary use this should, where appropriate, be limited and de-identified so that the secondary uses process is confidential.

Anonymisation: is the process of turning data into a form which does not identify individuals and where identification is not likely to take place (see https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/). Anonymisation describes the

process of data de-identification, producing de-identified data that cannot be linked to the original source or to an individual.

De-identification: the de-identification of data refers to the process of removing or obscuring any personally identifiable information from service user / patient records in a way that minimises the risk of unintended disclosure of the identity of individuals and information about them. Methods used to de-identify information may vary depending on the circumstances, but should be appropriate to protect the confidentiality of the individuals and the intended secondary use of the data.

Pseudonymisation: is a method of anonymisation and is the process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity. In order to de-identify the data a coded reference or pseudonym is attached to a record to allow the data to be associated with a particular individual without the individual being identified. The ICO draws a distinction between anonymisation techniques used to produce aggregated information and those such as pseudonymisation that produce anonymised data but on an individual-level basis (see https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/). A coded key (pseudonym) is used to re-identify the data if necessary and permitted.

Data Protection legislation supports the use of pseudonymisation as an appropriate safeguard where anonymisation is not practical, yet it should be recognised that data that has undergone pseudonymisation can still be considered information about an identifiable natural person.

Aggregation: is where data are displayed as totals, so no data relating to or identifying any individual is shown. Small numbers in totals are often suppressed through 'blurring' or by being omitted altogether (see https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/).

5. Governance and responsibilities for the operation of this policy

This policy was approved by Cambridgeshire County Council's Public Health Directorate Management Team on 15 March 2018.

Operationally, overall responsibility rests with the Public Health Directorate Management Team (PH DMT), with day-to-day responsibility for managing and implementing the Policy resting with the Business Intelligence Team (Information and Records), the Assistant Director for Public Health Intelligence and the relevant service and commissioning leads. Implementation of the Policy will include having effective training in place for staff to access, as required, or will be covered by standard NHS or local information governance training.

All staff who create, receive and use service user/patient records have pseudonymisation and anonymisation responsibilities under the Data Protection Act and supporting information governance policies. All staff will be responsible for:

- Ensuring they understand the requirements of this policy and any supporting standards and guidelines.
- Ensuring that this policy and its supporting standards and guidelines are built into local processes and that there is ongoing compliance on a day to day basis.

- Ensuring all breaches or suspected breaches of confidentiality or information security are reported for immediate investigation. In particular, this includes the unauthorised reversal of pseudonymisation.
- Immediately highlighting areas of potential weakness to their line managers, service/commissioning leads and the Business Intelligence Team (Information and Records), for appropriate corrective action.

6. Business processes

All business processes using service user/patient level data for primary and secondary uses, should be documented in the Public Health Directorate Data Flows Register and Information Asset Register. Business processes can include, but are not limited to:

- Processes using patient data involved in the direct care of service users/patients (primary use).
- Processes using patient data not involved in the direct care of service users/patients (secondary use).
- A combination of primary and secondary uses for direct care.

Internal to the Public Health Department itself, **primary use** includes, but is not restricted to, the delivery and management of public health services and duties, including:

- Public health services for gypsies and travellers.
- Public health support to NHS Cambridgeshire and Peterborough Clinical Commissioning Group, under the mandated Healthcare Public Advice Service covering applications for exceptional treatments.
- Placement information from providers to support invoicing for drug and alcohol services (currently under review at 05/03/2018).
- Internal evaluation of the winter Stay Well public health campaign.

Other direct care services that use PID are delivered externally and appropriate contractual measures are in place to ensure the appropriate use of PID and the safe receipt of anonymised or de-identified data for secondary use.

All information recorded about a person should be recorded in line with data protection legislation.

Secondary use business processes should be initially documented and then reviewed regularly to assess any requirement to use de-identified data. Following assessment any processes that require de-identified data must be modified in line with this policy. All onward disclosure should be limited to pseudonymised or anonymised / de-identified data.

Secondary use of primary data and personally identifiable data includes, but is not wholly restricted to:

- Analysis of anonymised data for commissioner and provider service based target setting, performance management and contract monitoring.
- Analysis of anonymised data for public health purposes, health needs analysis and support to the local Clinical Commissioning Group.

The following public health commissioned services receive anonymised data for secondary use, in line with the purposes stated above:

- Adult lifestyle's services (including services delivered by general practitioners and in workplaces): weight management, physical activity, behaviour change, smoking cessation, NHS health checks and associated services.
- Sexual health and contraceptive services.
- Services for children and young people: weight management, school nursing services, vision screening and health visiting.
- Mental health services.
- Healthy Fenland Fund – community based health improvement projects.

7. Anonymisation / de-identification

Staff only have access to the data that is necessary for the completion of the business activity which they are involved in. This is reflected in Caldicott Principles; *access should be on a need to know basis*. This principle applies to the use of PID for secondary or non-direct care purposes. By de-identification users are able to make use of patient level clinical data for a range of secondary purposes without having to access the identifiable data.

The aim of de-identification and anonymisation is to obscure or remove the identifiable data items within a person's records sufficiently that the risk of potential identification of the subject or a person's record is minimised to acceptable levels, so as to provide effective anonymisation.

De-identified and anonymised data should still be used within a secure environment with staff accessing it on a need to know basis. De-identification and anonymisation can be achieved by:

- Removing patient identifiers, sensitive and personal data.
- The use of identifier ranges, for example; value ranges instead of age.
- Aggregation.
- Using a pseudonym (although, as covered further below, pseudonymising data will not necessarily completely ensure that re-identification is impossible).

If patient data is required the NHS Number is the most secure form of identifiable data.

For all Public Health commissioner secondary use of identifiable data are de-identified before use and there are clear information barriers in terms of access to systems and access to network locations for PID held under those services or functions that require its use.

Any commissioning and contractual process the PH commissioner undertakes should include assurances that the Provider's processes are robust in respect of the supply of data. This will include the Provider ensuring that non-identifiable information is supplied to Public Health commissioners for secondary uses as identified above in Section 6.

Public Health commissioner analytical services have access to some identifiable data, including the Primary Care Mortality Database (PCMD), having completed the appropriate

approvals with NHS Digital. There are clear documented role based measures to separate the processing of these data from the secondary analysis in order to appropriately limit access to PID for PCMD and ONS births.

The PH commissioner's analytical team and the Drug and Alcohol Action Team (DAAT) also receives some local mortality data for the purposes of suicide and drug related deaths audits and real-time monitoring, in line with national public health priorities and policy. These data contain details of deceased people and next of kin in some cases and are sent securely, are stored in restricted network locations, with access restricted to those service and analytical leads who are directly involved in the work. All published audit results and other analyses are anonymised and/or aggregated.

8. Pseudonymisation

When pseudonymisation techniques are consistently applied, the same pseudonym is provided for individual patients across different data sets and over time. This allows the linking of data sets and other information, which would not be available if the PID were removed completely. The techniques when applied correctly can provide an effective safeguard to allow information on an identifiable natural person to be shared, linked or used where appropriate.

To effectively pseudonymise data the following actions must be taken:

- Each field of PID must have a unique pseudonym;

- Pseudonyms to be used in place of NHS Numbers and other fields must be of the same length and formatted on output to ensure readability. For example, in order to replace NHS Numbers in existing report formats, then the output pseudonym should generally be of the same field length, but not of the same characters. Letters should be used within the pseudonym for an NHS number to avoid confusion with original NHS numbers;

- Consideration needs to be given to the impact on existing systems both in terms of the maintenance of internal values and the formatting of reports;

- Where used pseudonyms for external use must be generated to give different pseudonym values in order that internal pseudonyms are not compromised;

- The secondary use output must, where pseudonyms used, only display the pseudonymised data items that are required. This is in accordance with the Caldicott Guidelines;

- Pseudonymised data should have the same security as PID.

Having reviewed the need to adopt pseudonymisation within the Public Health Directorate, we have determined that, at this time, the direct use pseudonymisation by Public Health Directorate staff is not required or appropriate for the uses and treatment of data within the Directorate for the reasons detailed below. However, this will be kept under review.

- We are not permitted to, or have no business need to, link data.

- De-identification of PID and limiting publication to aggregated results for secondary analysis are more appropriate, reasonable and proportionate steps in the current context in which the Directorate uses data.
- Internally, clear information barriers are in place between all PID data and those members of the Directorate who are not engaged in either direct care or the use of those data for other public health purposes.
- For external providers, contractual arrangements ensure that no PID or sensitive data are submitted to the Public Health commissioner.

The Directorate's analytical team does use record-level pseudonymised data in receiving the Hospital Episode Statistics (HES) data supplied by NHS Digital. The HES extracts provided are pseudonymised at source by NHS Digital and contain full individual HES records but with all identifiable information removed and certain fields replaced with derived data

9. Use of identifiable data

If service user/patient records are viewed in an identifiable form beyond the standard primary use requirements and documented secondary use agreements, then the reasons and usage of the data should be fully documented and approval is required by the appropriate data owner. This auditable trail of access to patient's records supports the Care Record Guarantee where patients are to be informed as to who has accessed/seen their data and the audit will provide accurate data in the event of untoward incidents.

The key items to be documented are:

- Who has accessed each data base containing identifiable data;
- Date and time of access;
- The reason for the access;
- The output from the access.

This audit should be kept within a separate structured database to enable queries and audit. The log of accesses must be regularly audited via sampling of users or subject matter to check for unusual patterns of access. If any unusual patterns of access are noted this should be reported.

10. Transferring information

Appropriate data sharing agreements should be in place when information is to be transferred to another organisation. If the transfer of information is required for secondary use then a form of anonymised or pseudonymised data should be sent. The Cambridgeshire Information Sharing Framework should be used to inform any data sharing (see https://www.cambridgeshire.gov.uk/data-protection-and-foi/information-and-data-sharing/information-sharing-framework/).

Secure methods of data transfer and storage must should be used, as per the relevant data sharing agreement.

An annual external data flows audit will be carried out to ensure compliance and to address any unknown issues and a register of external data flows will be maintained.

11. Legal and Professional Obligations

We will take actions to comply with the relevant legal and professional obligations, in particular:

- The Caldicott Principles.
- Data Protection Act 1998, and as superseded from 25 May 2018 by General Data Protection Regulation and Data Protection Bill as implemented into UK law.
- Human Rights Act 1998.
- Common Law Duty of Confidentiality.
- The Information Commissioner's Office (ICO) code: anonymisation: managing data protection risk code of practice.
- NHS Digital Information Toolkit.
- General Data Protection Regulation Order (GDPR), approved 14 April 2016; Enforcement date: 25 May 2018.

12. Training

All relevant staff will be made aware of their responsibilities relating to this policy through generic and specific training programmes and guidance. The Policy is on the Public Health Directorate website at https://www.cambridgeshire.gov.uk/council/council-structure/senior-managers-and-structure-charts/public-health-directorate/ and has been emailed to the Directorate during March 2018.

We have not offered specific training covering this policy. Rather, we have determined that the NHS IG Toolkit training and CCC's internal information security provide adequate training.

13. Validity of this Policy

This Policy is designed to avoid discrimination and be in accordance with the Human Rights Act 1998 and its underlying principles. This Policy should be reviewed annually. Anonymisation and Pseudonymisation standards should be subject to an ongoing development and review programme.

14. Policy contacts and key points of contact for incident reporting

- Matthew Smith, Senior Information Management Analyst, Business Intelligence Team – Information and Records. matthew.smith@cambridgeshire.gov.uk or 01223 714694.
- David Lea, Assistant Director, Public Health Intelligence. david.lea@cambridgeshire.gov.uk or 01480 379494.

15. Key references

- Information Commissioner's Office, Anonymisation Code of Practice: https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/
- NHS IG Toolkit: https://www.igt.hscic.gov.uk/ (and in particular the 'Pseudonymisation Implementation Project' documentation in the Knowledge Base).

Authors
David Lea, Assistant Director, Public Health Intelligence, Public Health.
Matthew Smith, Senior Information Management Analyst, Business Intelligence Team – Information and Records

Date
05 March 2018
Version 3.1