Pseudonymisation and Anonymisation of Data Policy – Version 1.0

Version history

Version	Date Issued	Notes
1.0	**11/03/2019**	First issue, building on and replacing the previous Pseudonymisation and Anonymisation Policy specific to CCC and PCC Public Health

1. Introduction

The General Data Protection Regulation (GDPR) and Data Protection Act 2018 require 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 also 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. They are part of the Data Protection by Design approach (Article 25 of GDPR) through which data protection is integrated into processing activities and business practices, from the design stage right through the lifecycle[i]. They will often be relevant to a Data Protection Impact Assessment (DPIA) and form a key technical measure to ensure processing complies with the data protection principles[ii].

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*[iii].

2. Policy scope

Although it references the uses of PID for direct care, and a limited set of uses for other social care, 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. The policy will be most relevant to staff in research and business intelligence roles; however it applies to all staff in respect of work they undertake, or commission, that involves PID for secondary use.

3. Purpose and background

This document sets out Cambridgeshire County Council and Peterborough City Council's policy on using PID for secondary use, via pseudonymisation and anonymisation. It seeks to provide all 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.

4. Definitions

Personal data: any information relating to a natural person who can be identified directly from the information, or could be in combination with other information.

Personal Identifiable Data (PID): any information that can identify an individual (note that this includes people who are deceased, unlike the standard definition of personal data that only applies to living individuals). 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. Anonymisation describes the process of data de-identification, producing de-identified data that cannot be linked to the original source or to an individual.

Anonymised Data: Data in a form that does not identify individuals and where identification through its combination with other data is not likely to take place.

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. 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.

5. Governance and responsibilities for the operation of this policy

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 Information Governance Team, for appropriate corrective action.
- Data breaches, or potential breaches, should be reported to Information Governance via the relevant online reporting form (see CCC and PCC Intranets).

6. Business processes

All business processes using service user/patient level data for primary and secondary uses, should be documented in a 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.

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.

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. Recital 26 of GDPR defines anonymous information, as "…information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". The GDPR does not apply to anonymised information.

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.

Any commissioning and contractual process 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 for secondary uses as identified above in Section 6.

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.

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 and Peterborough 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/). Agreements should either be based on the Framework template or other national standards used by third parties, such as NHS Digital.

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.
- General Data Protection Regulation and Data Protection Act 2018.
- 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 Data Security and Protection Toolkit.

12. Training

All relevant staff will be made aware of their responsibilities relating to this policy through generic and specific training programmes and guidance.

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. References

NHS Digital Guide to Confidentiality in Health and Social Care
https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/codes-of-practice-for-handling-information-in-health-and-care/a-guide-to-confidentiality-in-health-and-social-care

NHS Digital Standards and Collections
https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections

[i] The ICO's guidance on Data Protection by design: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/

[ii] CCC's Data Protection Impact Assessment guidance and templates on CamWeb: https://camweb.cambridgeshire.gov.uk/our-organisation/corporate-and-customer-services/gdpr/data-protection-impact-assessments-dpia/ .

PCC's Data Protection Impact Assessment guidance and templates on Insite https://sites.lumapps.com/a/peterborough-city-council/insite/teams/governance-teams/information-governance/information-governance-data-protection-policies

[iii] The ICO's 'Anonymisation: managing data protection risk code of practice': https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf