Luton

Anonymisation and pseudonymisation policy

Sharing information safely

Author: Yvonne Salvin, Data Protection Officer

Contact: Yvonne.Salvin@luton.gov.uk

Version: 1.0 (published)

Last updated: May 2019

Document history

Version	Date	Notes	Prepared by
0.1	22/3/19	First draft	Yvonne Salvin
0.1	23/4/19	Approved by Information Governance Steering Group	Yvonne Salvin
1.0	9/5/2019	Submitted to CLMT for approval	Yvonne Salvin
Final	9/5/2019	Approved by CLMT – change of name	Yvonne Salvin

Contents

1 Introduction

1.1 The Data Protection Act 2018 and General Data Protection Regulation (GDPR) requires us to use the minimum personal data necessary for a purpose. Secondary uses of personal information must not breach our obligations of confidentiality and respect for private and family life. This guidance identifies how we will use anonymisation and pseudonymisation in share information safely. This includes the use of storyboards for training and publicity purposes and the presentation and publication of statistics relating to individuals.

1.2 Anonymisation and pseudonymisation enables the council to undertake secondary use of personal data in a safe, secure and legal way.

1.3 We share and publish information in order to undertake our functions as a council, through a number of channels we collect customer information such as name, address, date of birth, however if we remove these identifiable details, information can then be used for secondary purposes without fear of breaching the General Data Protection Regulation.

1.4 This process is called anonymisation. By removing the personal information it allows the council to share or publish more data with fewer restrictions.

2 Purpose

2.1 The purpose of this policy is to ensure a standardised approach to enable consistency throughout the council, with regard to how and when to anonymise information correctly.

2.2 This policy is part of a suite of Information Governance policies.

3 Scope

3.1 This policy extends to all employees of the council who process information on behalf of the council when used for non-healthcare medical purposes while balancing the needs of the council to perform its everyday business functions.

3.2 All must comply with this policy where anonymised information is to be produced or published from individual level data

3.3 This policy does not cover the use of information sharing agreements (ISAs) or other tools used to share personal data safely. For further information on Information Sharing Agreements please refer to the Information Sharing Framework.

4 What is anonymisation and pseudonymisation?

4.1 Anonymisation and pseudonymisation both relate to the concealment of an individual’s identity.

4.2 Anonymisation is the process of removing, replacing and / or altering any identifiable information (identifiers) that can point to the person(s) it relates to.

4.3 Pseudonymisation is the technical process of replacing the identifying information to protect the individual’s identity whilst allowing the recipients to link different pieces of information together. A nickname is an example of pseudonymisation, although other identifying information such as age, ethnicity, gender or specific medical condition may also be changed to prevent a person being identified.

5 Definitions

5.1 Personal Identifiable Information (PII) is any information that can identify an individual. This could be one piece of information, or a collection of information, for example a name, address and date of birth.

5.2 Primary use refers to the use of information for the purpose of delivering council services to individuals. This also includes relevant supporting administrative processes and audit/assurance of the quality of services provided. Primary use requires information at the person identifiable level.

5.3 Secondary use refers to the use of information about individuals for research purposes, audits, service management, commissioning, and contract monitoring and reporting. When PII is used for secondary uses the information should, where appropriate be limited and de-identified so that the secondary use process does not enable individuals to be identified.

5.4 Anonymisation is a term for a variety of statistical and other techniques that depersonalise information about people so that the specific data subjects cannot be identified, including via aggregation and pseudonymisation.

5.5 Aggregation is an anonymisation technique in which information is only presented as totals, so that no information identifying individuals are shown. Small numbers in total are a risk here and may need to be omitted or ‘blurred’ through random addition and subtraction.

5.6 Pseudonymisation is the de-identification of individual level information by attaching a coded reference or pseudonym to each record that allows the information to be associated with a particular individual without the individual being otherwise identified. If the same system of pseudonyms is used across different datasets, then these datasets can be combined for analytical purposes without revealing the identities of individuals. Again, care needs to be taken if combining datasets, for example, could lead to individuals being identifiable via a combination of their circumstances.

5.7 Re-identification or de-anonymisation is where anonymised information is turned back into personal information through the use of for example data matching or combining. Where anonymisation is being undertaken, the process must be designed to minimise the risk of re-identification

6 Why anonymise

6.1 Anonymisation is undertaken to protect the privacy of individuals, whilst still making data available for statistical or analytical purposes. Personal data does have to be used directly where the intention is to inform decisions about particular individuals, or to provide services to them. Where this information is not needed at this level and for these purposes, however, it should be anonymised.

6.2 The GDPR is concerned with ‘personal data’ which relates to living individuals who can be identified from such data. Anonymised data where the prospect of identifying individuals is remote is not seen as personal data. The GDPR is therefore not applicable.

7 Benefits of anonymisation

7.1 All organisations that process personal information are required by the GDPR to protect it from inappropriate disclosure.

7.2 Where the council wants to or is required to publish information derived from such personal information, for example for analytical or statistical purposes, anonymisation techniques enable this information to be made available to the public and others without revealing any person identifiable information, so complying with Data Protection obligations.

8 Risk of re-identification of anonymised information

8.1 When anonymising information, the council must be sure that information is assessed and risks mitigated. This includes assessing whether other information is available that is likely to facilitate re-identification of the anonymised information.

8.2 The GDPR states that personal information is information which relates to a living individual who can be identified from that information, or from those information and other information which is in the possession of, or is likely to come into the possession of, the data controller.

8.3 When assessing whether information has been anonymised effectively, it is necessary to consider whether other information is available that, in combination with the anonymised information, would result in a disclosure of personal information. This is most likely where the circumstances described by the combined information are unusual or where population sizes are small.

8.4 Anyone considering anonymisation should carry out a 'motivated intruder' test, recommended by the Information Commissioner's Office as a means to check whether information has been effectively anonymised. This checks whether a reasonably competent individual who wished to de-anonymise information could successfully do so.

The test involves finding out whether information in the anonymised dataset could be combined with searches of easily available online or other information, such as the electoral register, social media, press archives or local library resources to reveal the identity of individuals.

8.5 Issues to consider are as follows:

- the risk of a 'jigsaw attack', piecing different items of information together to create a more complete picture of someone
- whether the information have characteristics which facilitate information linkage
- what other 'linkable' information is easily available
- what technical measures might be used to achieve re-identification
- what re-identification vulnerabilities the motivated intruder test revealed
- how much weight should be given to individuals' personal knowledge

8.6 Re-identification would lead to the unintentional disclosure of personal or sensitive personal information and would therefore be an information security incident. This should be reported as soon as possible using the council's information security incident process.

9 Anonymisation / de-identification

9.1 Staff should only have access to the information that is necessary for the completion of the business activity they are involved in. This principle applies to the use of PII for secondary or non-direct purposes. Through de-identification, users are able to make use of individual information for a range of secondary purposes without having to access the identifiable information items.

9.2 The aim of de-identification or anonymisation is to obscure the identifiable information items within the person's records sufficiently that the risk of potential identification of the information subject is minimised to acceptable levels: this will provide effective anonymisation.

9.3 De-identification can be achieved via a range of techniques. Whether de-identification is achieved depends on the fit of the technique with the specific dataset. Techniques include:

- aggregation so that information is only viewed as totals
- removing person identifiers
- using identifier ranges, for example:
  - age ranges instead of age
  - full or partial postcode
  - super output area instead of full address
  - age at activity event instead of date of birth
- using pseudonyms

9.4 De-identified information that goes down to the level of the individual should still be used within a secure environment with staff access on a need to know basis.

10 Pseudonymisation

10.1 When pseudonymisation techniques are consistently applied, the same pseudonym is provided for individuals across different datasets and over time. This allows datasets and other information to be linked in ways that would not be possible if person identifiable information was removed completely.

10.2 To effectively pseudonymise information, the following actions must be taken:

- each field of PII must have a unique pseudonym
- pseudonyms to be used in place of NHS numbers and similar 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, the output pseudonym should generally be of the same field length, but not of the same characters.
- other identifiable fields should be replaced by alternatives which render the information less specific (such as age at activity event replacing date of birth, lower super output area replacing postcode)
- it should be clear from the format of pseudonym information that it is not 'real' information to avoid confusion, such as adding letters that would not ordinarily appear in 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

11 Related documents

- Information sharing agreement - guidance
- Data sharing agreements - guidance
- Information security policy
- Data protection policy
- Information Governance Strategy
- Freedom of Information Policy
- Environmental Information Regulations Policy