Cheltenham Borough Council Anonymisation and Pseudonymisation Policy

Version control

Document name: Anonymisation and Pseudonymisation Policy

Version: 1.0

Responsible officer

- Claire Hughes, Corporate Director and Monitoring Officer

Approved by: Cabinet

Next review date: May 2026

Retention period: Delete one year after new version

Revision history

Revision date	Version	Description
May 2023	1	New Policy

Consultees

Internal

- Corporate Governance Group and Leadership Team

External

- N/A

Distribution

All Staff and Council Website

Contents

1. Introduction and purpose of the policy

1.1. The Data Protection Act 2018 and General Data Protection Regulation (GDPR) requires the Council 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.

1.2. This policy identifies how the Council will use Anonymisation and Pseudonymisation to share information safely. This includes the presentation and publication of statistics relating to individuals.

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

1.4. We share and publish information in order to undertake our functions as a council and through a number of channels we collect customer information such as name, address and 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.

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

1.6. 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. Scope of the policy

2.1. This policy extends to all employees who process information on behalf of the Council to perform its everyday business functions.

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

2.3. This policy does not cover the use of information sharing agreements (ISAs) or other tools used to share personal data safely.

3. What is Anonymisation and Pseudonymisation

3.1. Anonymisation and pseudonymisation both relate to the concealment of an individual's identity

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

3.3. 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. It is the process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity.

4. Why Anonymise

4.1. The primary reason for undertaking Anonymisation is to protect individuals' privacy when making available the data resources that activities such as research and planning rely on. It is legitimate to use personal data for certain purposes, for example where the intention is to inform decisions about particular individuals, or to provide services to them. However, where the use of personal data is not necessary, then the objective should generally be to use anonymised data instead.

4.2. 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.

4.3. Further information on Anonymisation can be found in the ICO's Anonymisation Code of Practice

5. Benefits and Anonymisation

5.1. The DPA requires all organisations that process personal data to protect it from inappropriate use or disclosure. However, the same organisations may want, or be required through the Transparency Code, to publish information derived from the personal data they hold. Anonymisation helps organisations to comply with their data protection obligations whilst enabling them to make information available to the public.

5.2. Any organisation processing personal data has to comply with the data protection principles. The principles regulate the disclosure of personal data, and in some circumstances can prevent this. This means that, in general, it is easier to disclose anonymised data than personal data as fewer legal restrictions will apply.

6. Risk of re-identification of Anonymised Information

6.1. 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.

6.2. 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.

6.3. 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.

6.4. 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, e.g. the electoral register, social media, press archives or local library resources to reveal the identity of individuals.

6.5. Issues to consider are as follows:

- What is the risk of a 'jigsaw attack', piecing different items of information together to create a more complete picture of someone? Does 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 did the motivated intruder test reveal?
- How much weight should be given to individuals' personal knowledge?

6.6. 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.

7. Anonymisation/De-identification

7.1. 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.

7.2. 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.

7.3. 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 or super output area instead of full address, age at activity event instead of date of birth.
- Using pseudonyms.

7.4. 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.

8. Pseudonymisation

8.1. 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.

8.2. 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 e.g. Council Tax Account Numbers and similar fields must be of the same length and formatted on output to ensure readability. For example, in order to replace Council Tax Account 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 (e.g. 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, e.g. adding letters that would not ordinarily appear in Council Tax Account 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 give different pseudonym values in order that internal pseudonyms are not compromised;

- The secondary use output must, where pseudonyms are used, only display the Pseudonymised data items that are required;
- Pseudonymised information should have the same security as PII.

9. Effectiveness of Anonymisation

Freedom of Information and personal data

9.1. The Council has to assess Freedom of Information requests to make a decision on whether personal information can be disclosed or if this would breach the GDPR.

9.2. Anonymised information given to a member of the public could breach the GDPR if other information was then combined to produce information that related to and identified a particular individual. This is now personal information.

9.3. Before releasing information that related at one stage to individuals, the council must assess if an organisation or member of the public could identify any individual from the information being released, either in itself or in combination with other available information (re-identification). The risk involved will vary according to the local information environment and particularly who has access to information.

Risk of Re-identification

9.4. Re-identification is when information does not in itself identify anyone (anonymised information) but by analysing it or combining it with other information an individual is identified.

9.5. There are cases in which it will be difficult to determine whether there is a reasonable likelihood of re-identification taking place. For example, it is difficult to determine the risk of re-identification of Pseudonymised data sets, because even though Pseudonymised information does not identify individuals to those who do not have access to the 'key', the possibility of linking several Pseudonymised datasets to the same individual can be a precursor to identification.

9.6. When sensitive information is involved which could significantly affect an individual's privacy, the information must be released with caution and be risk assessed. In borderline cases where the consequences of re-identification could be significant because they would leave an individual open to damage, distress or financial loss, for example, the approach should be to:

- Adopt a more rigorous form of risk analysis;
- Adopt a more rigorous form of Anonymisation to reduce the likelihood of re-identification to acceptably low levels, e.g. for aggregate data, using 'barnardisation', where small value statistics are manipulated in a random

way, or by changing the level of aggregation e.g. increasing the size of geographical areas or the breadth of age bands.

- Obtain data subject consent for the disclosure of the information, explaining its possible consequences; and/or
- In some scenarios, only disclose within a properly constituted closed community and with specific safeguards in place.

10. Is consent needed to produce or disclose Anonymised Information

10.1 An individual's properly informed consent is needed for the publication of personal data. However, there are obvious problems in this approach particularly where an individual decides to withdraw consent. In reality, it may be impossible to remove the information from the public domain, so that the withdrawal of consent will have no effect. Publishing anonymised information rather than personal data is safer even where consent could be obtained for the disclosure of personal data.

10.2 The 'necessity' rules in the GDPR mean that it could be against the law for the Council to publish personal data where anonymised information could serve the same purpose.

10.3 In the Information Commissioner's view, it is generally acceptable to anonymise personal data and to disclose it without the data subject's consent provided that:

- The Anonymisation will be done effectively, with due regard to any privacy risk posed to individuals – a privacy impact assessment could be used here;
- The purpose for which the Anonymisation takes place is legitimate and has received any necessary ethical approval;
- Neither the Anonymisation process, nor the use of the anonymised information, will have any direct detrimental effect on any particular individual;
- The data Controller's privacy policy/notice – or some other form of notification – explains the Anonymisation process and its consequences for individuals; and
- There is a system for taking individuals' objections to the Anonymisation process or to the release of their anonymised information into account.

11. Personal Information and Spatial Information

11.1 Postcodes and other geographical information will constitute personal data in some circumstances under the GDPR. For example, information about a place or property is, in effect, also information about the individual associated

with it. In other cases, it will not be personal data. The context of the related information and other variables, such as the number of households covered by a postcode, is the key.

11.2 Where postcodes are accessed in full as an interim step, e.g. enabling data about individuals to be aggregated or Pseudonymised by assigning them to particular geographical areas such as school catchments or Childrens Centres, the data that includes full postcodes may be personal data, and should be managed as such.

11.3 The more complete a postcode or the more precise a piece of geographical information, the more possible it becomes to analyse it or combine it with other information to disclose personal data.

11.4 The Council should approach the use of postcodes and other spatial information by the size of the dataset, where necessary considering the position on a postcode by postcode basis. For example, this may be necessary where a Freedom of Information Act (FOI) request is for specific information about small cohorts linked to postcodes.

11.5 It may also be necessary to process postcodes, removing certain of their elements to reduce the risk of identification. When anonymising postcodes, the following average characteristics of postcodes should be considered:

- Full postcode = approximately 15 households (although some postcodes only relate to a single property)
- Postcode minus the last digit = approximately 120/200 households
- Postal sector = 4 outbound (first part of the postcode) digits + 1 inbound = approximately 2,600 households
- Postal district = 4 outbound (first part of the postcode) digits = approximately 8,600 households
- Postal area = 2 outbound (first part of the postcode) digits = approximately 194,000 households

12. Publication and limited disclosure

12.1 The Council must make a decision whether to publish even anonymised information. The open data agenda relies on the public availability of information, and information released in response to a Freedom of Information Act request cannot be restricted to a particular person or group.

12.2 This means of making information, whether anonymised or not, available to third parties or the general public includes the following three approaches. Publication decisions should be informed by the realistic scope to control the use to which information is put following its release.

12.3 **Publication.** This is where information is made publicly available and anyone can see it and, in reality, use it for their own purposes. This can further transparency and deliver other benefits but, once published, no strict controls can be placed on re-identification, although other elements of the law may still apply - for example where information is subject to copyright. However, any third party performing re-identification will take on its own data protection liabilities. In reality, publication under licences such as the Open Government Licence falls into this category, as do disclosures made under Freedom of Information or the transparency agenda. The Open Government Licence does not apply to the use or reuse any personal information contained in a publication.

12.4 **Publication under specific licence terms**. This is an attempt to make information publicly available but to place certain specific restrictions on the way it is used. Whilst this can provide useful protection in respect of recipients that respect the rules, this form of publication can clearly present a privacy risk if the conditions attached to the information are either unlikely to be respected or not enforceable.

12.5 **Access control.** This is where anonymised information or, in some cases, personal data, are disclosed but only to particular recipients, with conditions attached to the disclosure. This is often used between groups of researchers. It is appropriate for handling anonymised information that is particularly sensitive in nature or where there is a significant risk of re-identification. The great advantage of this approach is that disclosure is controlled.

13. Useful Contacts

13.1 For further advice and examples of other techniques please refer to the Information Commissioner's code of practice Anonymisation: managing data protection risk, or contact the Information Governance or Research and Intelligence Teams via www.ico.org.uk

13.2 For the ICO Anonymisation Code of Practice, click here.