Anonymisation Policy

London Borough of Barnet

www.barnet.gov.uk
## Document Control

<table>
<thead>
<tr>
<th>Document Description</th>
<th>Anonymisation Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Version</strong></td>
<td>V.1</td>
</tr>
<tr>
<td><strong>Date Created</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>Approved</td>
</tr>
<tr>
<td><strong>Document Owner</strong></td>
<td>IMT</td>
</tr>
<tr>
<td><strong>Document Classification:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Authorisation</strong></td>
<td>Name</td>
</tr>
<tr>
<td><strong>Prepared By:</strong></td>
<td>xxxxxx</td>
</tr>
<tr>
<td><strong>Checked By</strong></td>
<td></td>
</tr>
</tbody>
</table>

## Version Control

<table>
<thead>
<tr>
<th>Version number</th>
<th>Date</th>
<th>Author</th>
<th>Reason for New Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>V.1</td>
<td>February 2014</td>
<td>xxxxxx xxxxxx</td>
<td>Creation of new Policy</td>
</tr>
</tbody>
</table>

Date last reviewed: February 2014  
Date of next review: February 2015
Anonymisation Policy

Contents

1 Introduction .................................................................................................................................................. 1
2 Purpose and Scope ................................................................................................................................. 1
3 What is Anonymisation? ......................................................................................................................... 2
3.1 Benefits of Anonymisation ................................................................................................................. 2
4 When to Anonymise Data ...................................................................................................................... 2
5 Anonymisation Techniques .................................................................................................................. 4
6 Risk of re-identification following anonymisation ............................................................................... 6
7 Using s40 of the Freedom of Information Act as a tool ..................................................................... 7
7.1 What to do in borderline cases ......................................................................................................... 7
8 Associated Policies .................................................................................................................................. 8
9 Review of Policy ...................................................................................................................................... 8
10 Contact Information/ Further Guidance ............................................................................................... 9
1 Introduction

Barnet Council shares and publishes information for a number of reasons in order to undertake its functions as a local authority. The council collects information in the process of delivering and providing services to its residents, businesses and visitors to the borough, through a number of communication channels.

Where legally allowed or required this information is shared with other government agencies, partner organisations, local authorities, police, and the general public. In addition to the requirement to share information the council has a strong obligation to ensure transparency in its processes and as such routinely publishes and distributes information as appropriate.

When information constitutes “Personal Data” as defined by the Data Protection Act 1998 (DPA), there are restrictions on how this information can be used. However, when personal identifiers such as name, address, date of birth, reference number are removed from a data source, the information can then be used for secondary purposes and made available to a wider audience without fear of breaching the DPA.

The process of making data available to a wider audience without identifying individuals is anonymisation. By correctly anonymising data, the council has the ability to share, disseminate or publish a greater amount of data with fewer restrictions.

Anonymised data can be used in a variety of ways including for planning (insight) purposes, it can be provided in response to requests made under the Freedom of Information Act 2000 (FOIA) or more widely published as part of a transparency agenda.

2 Purpose and Scope

The purpose of this policy is to provide guidance on when and how to anonymise data correctly. The purpose for which information is required will determine how it should be anonymised. The information provided in this policy will guide officers towards techniques which will ensure that data is correctly anonymised and mitigate the likelihood of individuals being identified from information the council publishes or shares.

This policy applies to all individuals who process information on behalf of Barnet Council (this includes but is not limited to Barnet, Capita and RE employees (permanent, temporary or agency staff and contractors).
3 What is Anonymisation?

Anonymisation is the process of removing, obscuring, aggregating and/or altering any identifiers in a dataset which can point to the particular person(s) the data relate(s) to. Identifiers are any piece of information, or combination of pieces of information, that makes an individual unique in a dataset and as such vulnerable to re-identification.

The Information Commissioner’s Anonymisation Code of Practice defines anonymisation as .....“the process of rendering data into a form which does not identify individuals and where identification is not likely to take place”.

It is established that personal data is any data which can lead to the identification of a living individual. It follows therefore that information that does not relate to and identify an individual is not personal data. Consequently data which has been correctly anonymised will not be personal and can be used in a variety of ways.

3.1 Benefits of Anonymisation

Anonymised information is easier to use than personal data. Information that is correctly anonymised is not subject to the DPA because the data subject(s) is/are no longer identifiable and therefore the data is no longer deemed to be “personal”. For this reason fewer restrictions will apply when using, sharing or publishing anonymised data.

Anonymisation of personal data can help facilitate the council’s information needs in a privacy-friendly way.

Anonymisation aids compliance with our data protection obligations whilst enabling proactive publication and sharing of information.

4 When to Anonymise Data

As a general rule anonymisation should be considered as a possibility at all times when handling personal data, where its disclosure in a non-anonymised form will be likely to reveal personal data resulting in a breach the DPA.

Whilst there may be legal reasons to justify the sharing and disclosure of personal data, the requirement and necessity of every disclosure must still be considered. Just because it is easier to provide a spread sheet of data containing an individual’s personal data does not mean it is appropriate, or necessary to do so. Before sharing or providing data, you should ensure you are only providing what is necessary to meet the purpose.

The reason for releasing data will affect how you make the disclosure, because the risk and consequences of identification will differ. Publication under the Freedom of Information Act 2000 (FOIA) or the open government licence is to the wider world, and carries more risk. Discretionary disclosures, such as those made to another
A department/organisation for insight or research purposes may be easier to control and assess but still carry some risks, although not as much as disclosure to the wider world.

The below table extracted from the ICO’s Anonymisation code highlights the thought process for deciding when and how to anonymised data.

---

**Is personal data involved? If it’s not clear, you should consider:**
- is it reasonably likely that an individual can be identified from those data and from other data?
- what other data are available, either to the public or to researchers or other organisations?
- how and why could your data be linked to other datasets?

- **No** - the data does not relate to identifiable individuals.
- **Yes** - individuals can be identified from the data.

If the DPA prevents you from disclosing personal data, you will need to consider whether the information can be anonymised.

Establish a process for ensuring an adequate level of anonymisation. Factors to take into account include:
- the likelihood of re-identification being attempted;
- the likelihood the re-identification would be successful;
- the anonymisation techniques which are available to use; and
- the quality of the data after anonymisation has taken place and whether this will meet the needs of the person(s) using the anonymised information.

Test the data according to your level of acceptable risk. You should document this process, for example as part of a privacy impact assessment.

Is the data still reasonably likely to allow an individual to be identified?

- **Yes** - there is still an unacceptable level of risk of identification. Consider making further adjustments and retesting the data.
- **No** - the data no longer allows individuals to be identified. You can now disclose or publish the data to the recipients proposed in your risk assessment.

If it is not possible to reduce the risk of identification to an acceptable level, do not publish unless the processing complies with the DPA and other relevant law.
5 Anonymisation Techniques

There are a number of anonymisation techniques that can be used when handling personal data. The technique to use will differ depending on the circumstance of the disclosure being made ie if you are disclosing to the wider world for instance in response to an FOI request or proactively, more rigorous anonymisation would be required than when disclosing internally for insight purposes. Factors to consider when deciding which technique to use will include the sensitivity of the data, the audience it is being released to and the usefulness of the data when it is aggregated.

If in doubt about which technique is appropriate to the data you are dealing with, you should seek further guidance from the corporate Information Management Team, to ensure disclosure will not breach confidentiality or privacy laws.

5.1 Aggregation

This involves publishing / providing the data as a summary, therefore ensuring no data relating to or identifying any individual is shown.

Aggregation is a relatively low risk technique because it will generally be difficult to find anything out about a particular individual by using aggregated data. The resulting data cannot support individual-level research but can be sufficient to analyse social trends on a regional basis, for example.

Small numbers in totals are often suppressed through blurring (disguising or altering the data to reduce the likelihood of identifying the individuals it relates to) or by being omitted altogether. There are different ways of aggregating data; the more common ways are discussed below:

5.1.1 Inference Control – Some cell values (eg small ones such as 1-5) in statistical data can present a greater risk of re-identification. Depending on the circumstances, small numbers can either be suppressed, or the values manipulated. If a large number of cells are affected, the level of aggregation could be changed. For example, the data could be linked to wider geographical areas or age-bands could be widened.

5.1.2 Rounding – rounding a figure up or down to disguise precise statistics. For example if one table may have a cell with value of 10,000 for all people doing some activity up to the present date and the following month, the figure in that cell rises to 10,001. It is easy to deduce the cell of 1. Rounding would prevent this.

5.1.3 Sampling - in some cases, when very large numbers of records are available, it can be adequate for statistical purposes to release a sample of records, selected through some stated randomized procedure. By not releasing specific details of the sample, data holders can minimise the risk of re-identification.
5.1.4 Synthetic data - mixing up the elements of a dataset – or creating new values based on the original data - so that all of the overall totals and values of the set are preserved but do not relate to any particular individual.

5.1.5 Geospatial aggregation – This usually involves aggregating information to a recognised statistical or administrative boundary. Examples could include a ward or a census output area. When this method is used, particular consideration should be given to the precision of the geographic data to ensure that its likely impact on the output is understood. Geospatial boundaries can be very detailed, for example some postcodes cover less than 5 properties, and the wider anonymisation point should still be reviewed. Guidance on some commonly used geospatial types is available from the Office for National Statistics - http://www.ons.gov.uk/ons/guidance-method/geography/ons-geography/index.html

5.2 Pseudonymisation

This involves de-identifying data so that a unique identifier or pseudonym is attached to a record to allow the data to be associated with a particular individual without the individual being identified. The unique identifier can be a randomly generated number or code, or simply a replacement of the individuals name with Customer A, Customer B. For example, Customer A is 25 years old and lives in Edgware. Customer B is 35 and lives in Hendon.

However, simply replacing an individual’s name with a unique identifier doesn’t guarantee anonymisation. Each field of personal data needs to be considered as a separate entity. Think about what information needs to be disclosed and consider replacing fields which won’t obscure the meaning of the data itself. For example, if the council is interested in finding out the average age of individuals taking up a particular type of service, we do not need to provide a report detailing full dates of births. We can instead remove the name and address as standard, but also replace the full date of birth with either a name, or a partial date of birth as an alternative. This will still provide meaningful data, but also provide a level of anonymity to the individuals concerned.

5.3 Data masking

This involves removing all the obvious personal identifiers from the data such as name, address, date of birth etc., to create a data set in which no personal identifiers are present. There are two types of this technique:

5.3.1 Partial data removal – this results in data where some personal identifiers, eg name and address have been removed but others such as month or year of birth remain, so that individuals cannot be identified.

5.3.2 Data quarantining - The technique of only supplying data to a recipient who is unlikely or unable to have access to the other data needed to facilitate re-identification. It can involve disclosing unique personal identifiers – eg reference numbers – but not the ‘key’ needed to link these to particular individuals.
These are relatively high risk techniques because the anonymised data still exists in an individual-level form. Electoral roll data, for example, could be used to fairly easily reintroduce names that have been removed to the dataset.

When anonymising post codes, consider how much of the postcode information is required for the purpose. This is necessary because full post codes when put together with other information could still identify individuals. So instead of providing the full postcodes consider using partial postcodes e.g if the postcode is AB1 0XY use AB, AB1 or AB1 0 depending on what the outcome would yield.

6 Risk of re-identification following anonymisation

In some cases the risk of anonymised data being combined with other data to result in personal data being created will be high. An obvious example is where publicly available data – such as the Electoral Roll or data easily retrievable from a web-search – can be combined with the ‘anonymised’ data, allowing an individual to be identified. Note that ‘identified’ does not necessarily mean ‘named’. It can be enough to be able to establish a reliable connection between particular data and a known individual.

However, in some circumstances it can be difficult to establish the likelihood of re-identification, particularly where complex statistical methods might be used to match various pieces of anonymised data. This can be a particular vulnerability where pseudonymised data sets are concerned, because even though pseudonymised data does not identify an individual, in the hands of those who do not have access to the ‘key’, the possibility of linking several anonymised datasets to the same individual can be a first step to identification. This does not however mean that effective anonymisation through pseudonymisation becomes impossible as it is recognised that some forms of research can only take place where different pieces of data can be linked reliably to the same individual. The DPA allows this provided that:

a) identification does not take place, or
b) if identification does take place, this does not constitute a breach of the data protection principles.

However the DPA will be breached if individuals were assured that only anonymised data would be published but in fact their personal data was disclosed. This is because the reasonable expectation of the individual was that their personal data would not be shared.

You must be aware of the risk of re-identification and also that this risk can change over time, eg powerful data analysis techniques that were once rare are now common-place. However, if anonymisation is carried out effectively in the present it is likely to protect personal data from future re-identification.
7 Using s40 of the Freedom of Information Act as a tool

The DPA is primarily concerned with the risks associated with the identification of individuals by data controllers. However, section 40 of FOIA, the personal data exemption, introduces a broader concept of risk because the test for deciding if personal data can be disclosed is whether disclosure to a member of the public would breach the data protection principles. Please see the FOI Policy for more information about this exemption.

Considering section 40 when anonymising data will ensure we take into account the additional information that a particular member of the public might have that could allow data to be combined to produce information that relates to and identifies a particular individual – and that is therefore personal data.

The test in FOIA can be difficult to apply in practice because different members of the public may have different degrees of access to the ‘other information’ needed for re-identification to take place. It is therefore considered good practice to look at identification ‘in the round’, i.e. before disclosing anonymised data assess whether any organisation or member of the public could identify any individual from the data being released – either in itself or in combination with other available information. The risk involved will of course vary according to factors like, time, sensitivity of the information, recipients of the information and mode of communicating the information.

Anonymised data disclosed within a controlled environment, eg a discrete disclosure to an authority/organisation could remain anonymous even though if published more widely, the likelihood of re-identification would mean that the anonymised data would become personal data. When contemplating discrete disclosure however, we should still consider other limitations on disclosure that may apply eg duty of confidence or statutory prohibitions on disclosure.

7.1 What to do in borderline cases

There will be borderline cases where, in reality, it will be difficult, or even impossible, to determine whether it is likely that re-identification will take place. In situations like this, the risk posed to individuals by disclosure, or the public benefit of the information must be assessed and this is what will inform the decision to disclose or not disclose.

Clearly the identification of an individual can have a range of consequences depending on the nature of the data, the context in which it is disclosed and who it relates to.

In borderline cases where the consequences of re-identification could be significant eg because they would leave an individual open to damage, distress or financial loss, it is worth considering:

- seeking the data subject’s consent for the disclosure of the data, explaining its possible consequences;
- adopting a more rigorous form of risk analysis and anonymisation
• disclosing within a properly constituted closed community eg a specific council team and with safeguards in place.

Acceptable safeguards to reduce the likelihood of re-identification include:
• purpose limitation, ie telling the recipient that the data can only be used for an agreed purpose or set of purposes;
• restricting the recipient from further disclosing the data to another party
• ensuring that the recipients’ staff with access to data have been properly trained, especially on security and data minimisation principles;
• limiting the use of the data to a particular project or projects;
• prohibiting any attempt at re-identification and putting in place measures for the destroying of any accidentally re-identified personal data;
• arrangements for technical and organisational security, eg staff confidentiality agreements;
• encryption and key management to restrict access to data;
• limiting the copying of, or the number of copies of the data;
• making arrangements for the destruction or return of the data on completion of the project; and
• penalties, such as contractual ones that can be imposed on the recipients if they breach the conditions placed on them.

Please note that in some particularly high-risk situations, it may not even be possible to share anonymised data within a closed community.

If you are finding it difficult to determine whether it is likely that an individual can be re-identified from anonymised data, please seek further guidance from the corporate Information Management Team.

8 Associated Policies

Data Protection Policy
Data Protection Act Compliance Toolkit
FOI and EIR Policy

9 Review of Policy

This policy will be reviewed every two years and the suitability of the techniques recommended assessed based on current and foreseeable future threats.
Further advice and guidance is available from Information Management Team

Address: London Borough of Barnet
         Building 2
         North London Business Park
         Oakleigh Road South
         London N11 1NP

Tel No:  (020) 8359 7080
E-mail:  foi@barnet.gov.uk