Leicester
City Council

Information Assurance
Policies and Guidance

Policy and Guidance on Anonymising Personal Data for Secondary Uses

January 2016

Document Version: V1.1
Review Date: January 2017

Owner: Information Governance Manager

Document History

Revision Date	Version Number	Summary of Changes
2014 01 24	V0.1	Policy draft.
2014 02 10	V0.2	Draft incorporating comments from Public Health.
2014 02 10	V0.3	Draft incorporating comments from Information Assurance and Governance
2014 02 14	V1.0	Approved by 14 February 2014 Information Management Programme Board
2016 01 04	V1.1	Reviewed by Information Governance Manager. Removal of Head of Information Assurance post.

Audience: Staff handling person identifiable and/or anonymised information.

Acknowledgement: this policy was developed from a template provided by the HSCIC.

Users of this document are responsible for familiarising themselves with the latest version on a regular basis. You should be aware that a physical copy might not be the latest available version. The latest version, which supersedes all previous versions, can be found at http://interface.lcc.local.

Contents

1. Introduction

The principles of the Data Protection Act regulate the disclosure of personal data and, in some circumstances, prevent its disclosure. This Policy aims to ensure that, where personal data is used for secondary purposes by Leicester City Council, this is done using anonymised, aggregate or pseudonymised data. This is to protect the privacy of 'data subjects' – the individuals the data relate to – and so ensure compliance with the Data Protection Act 1998.

Anonymised information, where the prospect of identifying individuals is remote, can be used in numerous ways: the Data Protection rules do not apply to such information as it does not enable living individuals to be identified.

2. Scope

This Policy applies to all Leicester City Council employees, agency workers, external contractors, casual workers, volunteers, employees from other organisations using Leicester City Council facilities and equipment, elected members and those working on secondment (referred to below as "workers"). All of them must comply with this policy where anonymised information is to be produced or published from individual level data.

3. Roles and Responsibilities

All people to whom the policy applies are responsible for ensuring that they manage personal data appropriately and anonymise it when it is to be used for secondary purposes, using appropriate pseudonymisation or aggregation techniques.

The **Information Governance Manager** is responsible for the production, review and maintenance of the Council's anonymisation policy.

The **Information Governance Manager** is responsible for the Council's compliance with the Data Protection Act and will provide guidance on where anonymisation would be appropriate and promote anonymisation practices where appropriate as part of information sharing agreements.

Where relevant, and particularly where access to data is conditional on the Council holding a current NHS Information Governance Toolkit accreditation, all **Directors and Managers** will implement this policy within their business areas and ensure awareness of it and adherence to it by their staff. Specifically, they will ensure that:

- Where secondary uses are being made of personal information, the information is appropriately anonymised, via pseudonymisation or aggregation techniques.
- The risk of re-identification of anonymised data is actively considered prior to releasing anonymised datasets directly to third parties or publishing them.

4. Definitions

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

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 data at the person identifiable level.

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

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

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

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.

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

5. Anonymisation of personal data

5.1 Why anonymise?

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.

The Data Protection Act 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 Data Protection Act is therefore not applicable.

5.2 Benefits

All organisations that process personal data are required by the Data Protection Act to protect it from inappropriate disclosure.

Where Leicester City Council wants to or is required to publish information derived from such personal data, 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.

5.3 Risk of re-identification of anonymised data

When anonymising data, Leicester City 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 data.

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

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 data. This is most likely where the circumstances described by the combined data are unusual or where population sizes are small.

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

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 data 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?

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.

6. Anonymisation / de-identification

Staff should only have access to the data that is necessary for the completion of the business activity they are involved in. This is reflected in the Caldicott Principles ('need to know' access). This principle applies to the use of PID for secondary or non-direct purposes. Through de-identification, users are able to make use of individual data for a range of secondary purposes without having to access the identifiable data items.

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

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 data 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.

De-identified data 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.

7. Pseudonymisation

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

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 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 data 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 data that it is not 'real' data to avoid confusion, e.g. 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;
- 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. This is in accordance with the Caldicott Principles;
- Pseudonymised data should have the same security as PID.

8. Use of identifiable data

If records are viewed in an identifiable form for other purposes than normal service delivery, then the reasons and usage of the data should be fully documented and approval is required from the appropriate data owner.

Relevant services should set up an appropriate tracking tool, eg. an Excel spreadsheet, to capture this activity. 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.

A structured log of accesses should be kept 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.

9. Transferring Information

Appropriate data sharing agreements should be in place when information is to be transferred to or from another organisation.

If the transfer of information is required for secondary use then a form of anonymised or pseudonymised data should be sent.

10. Effectiveness of anonymisation

10.1 Freedom of Information and personal data

Leicester City Council has to assess Freedom of Information requests to make a decision on whether personal data can be disclosed or if this would breach the Data Protection Act.

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

Before releasing information that related at one stage to individuals, Leicester City 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 data environment and particularly who has access to information.

10.2 Risk of re-identification

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.

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.

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, eg. 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.

11. Is consent needed to produce or disclose anonymised information?

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.

The 'necessity' rules in the Data Protection Act mean that it could be against the law for Leicester City Council to publish personal data where anonymised information could serve the same purpose.

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 – 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.

12. Personal data and spatial information

Postcodes and other geographical information will constitute personal data in some circumstances under the Data Protection Act. 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.

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 Sure Start Centres, the data that includes full postcodes may be personal data, and should be managed as such.

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.

Leicester City 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 (FOA) request is for specific information about small cohorts linked to postcodes.

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

13. Publication and limited disclosure

Leicester City 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.

The 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.

- **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.

- **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.

- **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.

14. Further Information

For further advice and examples of anonymisation through aggregation, pseudonymisation and 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.