

OP111

De-identification and Pseudonymisation Policy

Contents

Policy Sections	Page
1 Policy Statement	2
2 Definitions	2
3 Accountabilities	3
4 Policy Detail	5
5 Financial Risk Assessment	6
6 Equality Impact Assessment	6
7 Maintenance	6
8 Communication and Training	6
9 Audit	7
10 References	7
11 Document Control Sheet	9
12 Implementation Plan	11

Attachments:

[Attachment 1 – What is anonymisation and pseudonymisation?](#)

[Attachment 2 – Deciding when and how to release anonymised data](#)

[Attachment 3 – Anonymisation techniques](#)

[Attachment 4 – Pseudonymisation techniques](#)

[Attachment 5 – Examples of unsafe/ineffective techniques not to be used or be used with caution](#)

1.0 Policy Statement

This document directs the measures to be taken to safeguard patient confidentiality when data is used or shared for purposes other than direct patient care. It instructs all Trust personnel who use patient data how to de-identify data for secondary use.

This policy explains anonymisation and pseudonymisation techniques and the reasons they are required, also outlining recommended techniques and methods to avoid. This is to ensure that information reporting and datasets are processed and shared, where appropriate, using the correct techniques to preserve the confidentiality of the patient.

The aim of de-identification is to obscure the identifiable data items within the person's records sufficiently that the risk of potential identification of the subject or a person's record is minimised to acceptable levels. Although the risk of identification cannot be fully removed, this can be minimised with the use of multiple pseudonyms and data masking.

2.0 Definitions

Aggregated data - Statistical data about several individuals that has been combined to show general trends or values without identifying individuals within the data.

Anonymisation - The process of rendering data into a form which does not identify individuals and where identification is not likely to take place.

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.

Information or Data Sharing - The disclosure of data from one or more organisations to a third-party organisation or many organisations

Personal Data or Personal Identifiable Data (PID) - any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Personal Confidential Data (PCD) categories - Personal confidential data is information (an identifier) about a person e.g., a patient, client, service user or staff, from which the individual could be singled out from others. It may be a single or combination of two or more identifiers such as:

- Name,
- Address (home or business),
- Postcode (e.g., a house in rural area),
- NHS number,
- Email address,

- Date of birth,
- Driving licence number (date of birth and first part of surname),
- Telephone numbers,
- Local Patient Identifier and
- National Insurance number.

A single identifier, such as an unusual surname or an isolated postcode, may be fairly explicit, as may a combination of identifiers, such as of postcode and telephone number.

Primary Uses - Purposes that directly contribute to the safe care of the patient are classified as primary uses. They include care, diagnosis, referral and treatment processes together with relevant supporting administrative processes, such as clinical letters and patient administration, patient management on a ward, and managing appointments for care. Audits and methods to provide assurance of the quality of the healthcare are also primary uses.

Pseudonymisation - The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. If that additional information is kept separately and is subject to technical and organisational measures, pseudonymisation ensures that the personal data are not attributed to an identified or identifiable natural person. However, the UK GDPR classifies pseudonymised data as identifiable data.

Re-identification - The process of analysing data or combining it with other data with the result that individuals become identifiable.

Secondary Uses - Secondary use of patient data is the use for purposes that do not directly contribute to the safe care of the individual concerned e.g, research purposes, service management and planning, commissioning, performance management, contract monitoring and reporting.

Sensitive information or data - This can be described as any category of information legally defined as particularly sensitive relating to racial or ethnic origin, political or religious beliefs, sexual orientation, any offences or ongoing investigations relating to the individual, trade union membership, safeguarding information and physical, mental or sexual health

3.0 Accountabilities

This policy applies to all staff and external contractors who are employed to carry out work on behalf of the Trust, whether it is a temporary or time limited capacity. Each individual is responsible for ensuring that they comply with the relevant legislation, Trust policy and guidance for sharing information safely and effectively.

Caldicott Guardian

The Caldicott Guardian will be responsible for ensuring robust policies are in place to ensure that patient information will remain confidential, that information is shared

appropriately and in line with the law and NHS guidance, and that the best interests of patients are maintained.

Senior Information Risk Owner (SIRO)

The role of the SIRO is to take responsibility for key decisions and to inform an organisation's general corporate approach to anonymisation and pseudonymisation.

Data Protection Officer (DPO)

The statutory role of the DPO is to monitor compliance with data protection law, provide advice taking into consideration the associated risks and acting as the contact point for the Information Commissioner.

Caldicott Guardian/Nominated Lead(s) for Information Sharing

The Caldicott Guardian may delegate operational responsibility for the policy, in cases of specific sharing activities to senior managers (i.e., Heads of Service or Departments) who will be the Nominated Lead(s) for Information Sharing. These staff will ensure dissemination and the use of this policy and associated techniques and monitor the implementation of and compliance with this policy within their own departments.

Managers

Managers have a responsibility to ensure that they are aware of the policy detail and its associated Trust policies and that all members of staff that they manage are aware of the policy content and recommended techniques.

All staff/parties involved in information sharing/data processing

- All staff and other parties who are involved in information sharing and, or data processing have a duty of confidentiality and must ensure that individual rights in relation to the disclosure and use of personal information are understood and upheld. Please see [Confidentiality Code of Practice](#) for more detail on maintaining confidentiality.
- They also have the responsibility to ensure that requests for information are specific, recorded and provided on a 'need to know' basis in line with the Caldicott Principles and the General Data Protection Regulation.
- They are responsible for using the correct techniques when sharing data
- If there is any doubt about whether information should be shared or disclosed, staff must refer to the Trust's Information Sharing Policy [OP85](#), or speak to their line manager, the Trust's Information Governance Department and, or Caldicott Guardian.
- If there is any doubt on applying the techniques in this policy, staff must speak to the Trust's Information Department [REDACTED]

***IMPORTANT* - Requirements for data leaving Trust control**

Irrespective of the method used to de-identify data, there must be data governance arrangements in place to ensure de-identified data leaving the Trust is not subject to re-identification outside of Trust control.

The requirements for de-identification must be taken into consideration at the initiation stage of any project and be included within the impact assessment exercise, as a requirement and a risk. Processing requiring pseudonymised data must be treated at the same level of risk with identifiable data.

If it is discovered that re-identification is possible for data that has left Trust control, incident reporting arrangements must be initiated as soon as possible. Consideration must be given to statutory requirements to report serious data breaches to the ICO within 72 hours. Therefore, time taken for impact assessment should be maximised.

4.0 Policy Detail

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 PCD for secondary or non-direct care purposes. By using de-identification techniques, users can make use of patient level clinical data for a range of secondary purposes without having to access the identifiable data items.

The aim of de-identification is to obscure the identifiable data items within the person's records sufficiently that the risk of potential identification of the subject or a person's record is minimised to acceptable levels. Although the risk of identification cannot be fully removed, this can be minimised with the use of multiple pseudonyms and data masking.

De-identified data should still be used within a secure environment with staff access on a need-to-know basis.

De-identification can be achieved by:

- Removing direct patient identifiers (anonymisation),
- The use of identifier ranges, for example, value ranges instead of age (masking), or
- By using a pseudonym.

4.1 What are anonymisation and pseudonymisation?

[Attachment 1 - What are anonymisation and pseudonymisation](#) provides an explanation of what anonymisation and pseudonymisation are and why they are important.

4.2 Deciding when and how to release anonymised data

[Attachment 2 - Deciding when and how to release anonymised data](#) provides a flow diagram to understand the steps to follow to determine how data should be released.

4.3 Anonymisation techniques

[Attachment 3 - Anonymisation techniques](#) outlines approved techniques to anonymise data for data sharing and reporting.

4.4 Pseudonymisation techniques

[Attachment 4 - Pseudonymisation techniques](#) outlines further details on approved tools to use for pseudonymisation. However, it is not recommended that you create your own algorithm for creating a pseudonym.

The Trust's Information Department have built a bespoke pseudonymisation and masking tool within the Data Warehouse, or can recommend an accredited tool for individual use, if needed. Please contact them via [REDACTED] for any further information or support.

4.5 Unsafe and ineffective techniques not to be used or be used with caution

[Attachment 5 – Unsafe ineffective techniques not to be used or be used with caution](#) provides examples of common unsafe/ineffective techniques which are not to be used or be used with caution.

5.0 Financial Risk Assessment

A financial risk assessment has been undertaken and no financial risks have been identified as a result of implementing this Policy.

6.0 Equality Impact Assessment

An assessment has been undertaken, no adverse effects have been identified for staff, patients or the public as a result of implementing this Policy.

7.0 Maintenance

This policy will be reviewed every three years but sooner if changes in legislation or guidance require, or there are changes which arise from overarching area- or region-wide protocols. Responsibility lies with the Head of Information and will be overseen by the Information Governance Steering Group.

8.0 Communication and Training

Approved Trust policies will be made available to staff via the Trust's intranet page. This policy will be implemented and communicated through the work of the Information Governance Action Group and Information Governance Steering Group. Advice and support will be provided by the Information Department

9.0 Audit Process

Criterion	Lead	Monitoring method	Frequency	Committee
A peer review of policy contents to compare with other Trusts and to ensure up to date techniques are being used	Head of Information	Peer review of policy or reviewing examples from other NHS Trusts.	Every 3 years in line with policy refresh	IGSG
Audit of application of this policy	Head of Information	Report number of requests for pseudonymisation to Information Department	Annually	IGSG

10.0 References

The Royal Wolverhampton NHS Trust Policies and Strategies:

[OP85 Information Sharing Policy](#)

[OP13 Information Governance Policy](#)

[OP97 Confidentiality Code of Conduct for Staff](#)

Other sources

The Information Commissioners Office. (2012). ICO: Anonymisation: managing data protection risk code of practice

<https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

The Information Commissioners Office. (2021). ICO Data sharing code of practice

[data-sharing-a-code-of-practice-1-0.pdf \(ico.org.uk\)](https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf)

Information Commissioners Office website – What is personal data?

[What is personal data? | ICO](#)

The Information Commissioners Office. (2012). ICO What is personal data? – A quick reference guide

https://ico.org.uk/media/for-organisations/documents/1549/determining_what_is_personal_data_quick_reference_guide.pdf

The Information Commissioners Office. (2012). ICO Determining what is personal data

<https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf>

11. Document Control

Reference Number and Policy name: OP111 De-identification and Pseudonymisation Policy for Data Sharing	Version: Version 2.0	Status: Final	Author: Head of Information Director Sponsor: Chief Finance Officer/SIRO	
Version / Amendment History	Version	Date	Author	Reason
	V 1.0	Jan 2018	Head of Information	Creation new policy
	V 1.1	Aug 2022	Head of Information	Extension until November 2022
	V2.0	Sep 2022	Head of Information / Interim Head of Data Security & Protection/DPO	Review, restructure content and ensure up to date techniques included Additional statements and definitions added in line with GDPR legislation
Intended Recipients: Staff groups who process patient/staff datasets or reports and distribute internally or to third parties.				
Consultation Group / Role Titles and Date: Chief Finance Officer, Executive lead/SIRO – August 2022 Information Governance Lead/Data Protection Officer – August 2022 Policy Review Group – November 2022				
Name and date of Trust level group where reviewed			Trust Policy Group – November 2022	
Name and date of final approval committee			Trust Management Committee – November 2022	
Date of Policy issue			December 2022	
Review Date and Frequency (standard review frequency is 3 yearly unless otherwise indicated)			November 2025	

Training and Dissemination: Policy will be made available to all staff on Trust intranet page with a communicated release. It will also be taken to IG Steering Group and IG action group. Guidance and support will be provided by the Trust information and Information Governance Departments.

To be read in conjunction with:

[OP85 Information Sharing Policy](#)

[OP13 Information Governance Policy](#)

[OP97 Confidentiality Code of Conduct for Staff](#)

[Wolverhampton Overarching Information Sharing Protocol](#)

[ICO Anonymisation Code of Practice](#)

[ICO Data Sharing Code of Practice](#)

[ICO How to disclose information safely - Removing personal data from information requests and datasets](#)

[ICO Determining what is personal data quick reference guide](#)

Initial Equality Impact Assessment (all policies): Completed ~~Yes~~ / ~~No~~

Full Equality Impact assessment (as required): Completed ~~Yes~~ / ~~No~~ / ~~NA~~

If you require this document in an alternative format e.g., larger print please contact policy administrator

Monitoring arrangements and Committee

Information Governance Steering Group

Document summary / key issues covered:

What is anonymisation?

Spectrum of identifiability

What is pseudonymisation?

Deciding when and how to release anonymised data – flow diagram

Techniques for anonymising reports and datasets

Recommended approach for pseudonymising datasets.

Examples of unsafe/ineffective techniques not to be used or used with caution

Key words for intranet searching purposes

OP111, Policy, Procedure, Anonymise, Anonymisation, Pseudo, Datasets

12.0 IMPLEMENTATION PLAN

To be completed when submitted to the appropriate committee for consideration/approval

Policy number and policy version OP111 Version 2.0	Policy Title De-identification and Pseudonymisation Policy	
Reviewing Group	Trust Policy Group	Date reviewed: November 2022
Implementation lead: [REDACTED], Tel: 01902 307999 ext [REDACTED] or mobile: [REDACTED]		
Implementation Issue to be considered (add additional issues where necessary)	Action Summary	Action lead / s (Timescale for completion)
Strategy; Consider (if appropriate) 1. Development of a pocket guide of strategy aims for staff 2. Include responsibilities of staff in relation to strategy in pocket guide.	N/A	
Training; Consider 1. Mandatory training approval process 2. Completion of mandatory training form	N/A	
Development of Forms, leaflets etc; Consider 1. Any forms developed for use and retention within the clinical record MUST be approved by Health Records Group prior to roll out. 2. Type, quantity required, where they will be kept / accessed/stored when completed	N/A	
Strategy / Policy / Procedure communication; Consider 1. Key communication messages from the policy / procedure, who to and how?	To all staff via intranet and through IG groups & committees. Also, via Trust Brief article including references to policy and pitfalls attachment 5	Two months from approval
Financial cost implementation Consider Business case development	N/A	
Other specific Policy issues / actions as required e.g. Risks of failure to implement, gaps or barriers to implementation	N/A	

What are anonymisation and pseudonymisation?

4.1.1. Anonymisation

Anonymisation is a very valuable tool that allows data to be shared, whilst preserving confidentiality. It should be used when individual level records are needed, but there is no legal basis for sharing Personal Confidential Data items and it does not need to be linked to other datasets.

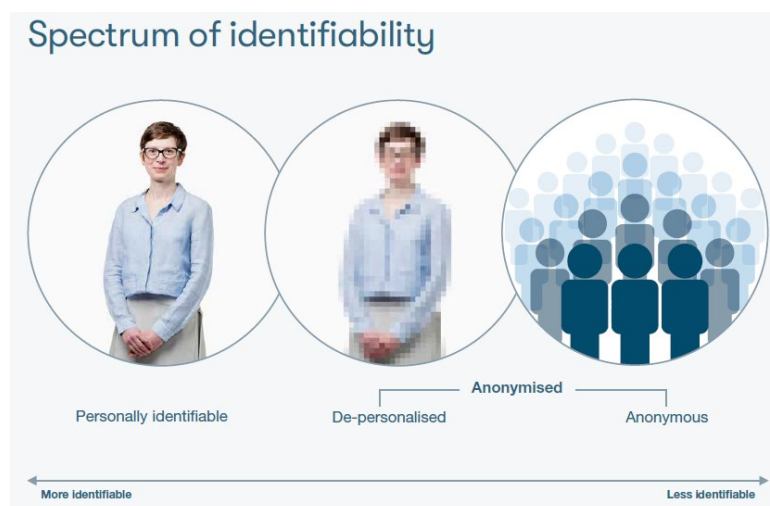
All organisations collect some information from their service users as part of their organisational activities, and, increasingly, they are sharing (at least some of) the data they collect. The information they collect on their service users, for example their names, addresses, dates of birth, health status etc., is what is termed personal data.

Health service organisations are required to protect the identities of individual patients but are also required to produce national and local datasets, data analysis, and publish statistics on patient care and outcomes. Anonymisation helps organisations to comply with their data protection obligations whilst enabling them to make information available to stakeholders and the public.

De-identifying data requires that identifiers are removed, obscured, aggregated and, or altered in some way. The term 'identifiers' is often misunderstood to simply mean formal identifiers such as the data subject's name, address and unique identification numbers e.g., NHS number. However, identifiers could include 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.

4.1.2 Identifiability spectrum

The picture below gives a great visual representation of the levels of identifiability and illustrates the difference between being identifiable and anonymised.



At one end of the spectrum, a person is fully identifiable. As you remove or encrypt information, you blur the image more and more, and it becomes more difficult to identify who that person is. At the other end of the spectrum, it is not possible to identify who someone is — they are effectively anonymous.

Different controls are needed at different points along the spectrum depending on the risk of re-identification, which is why you may hear people talking about the ‘environment’ or ‘context’ in which data is used. The controls that are taken to protect the data are just as important as the data itself. It may also be possible to work out who someone is by joining together information from different sources — like joining together different pieces of a jigsaw puzzle.



Personally identifiable

What is it?

This is information that identifies a specific person. Identifiers include: name, address, full postcode, date of birth or NHS number.

How is it protected?

Personally identifiable information will always be stored in a highly secure way. There are strict laws that safeguard how personally identifiable information can be used if you are not asked for consent. There are also sanctions under the Data Protection Act if personally identifiable data is misused.

Example

A patient's medication history, including their NHS number (but no contact details).

Other words that you may see

Personal data, confidential information, patient identifiable information, confidential personal information.

De-personalised

What is it?

This is information that does not identify an individual, because identifiers have been removed or encrypted. However, the information is still about an individual person and so needs to be handled with care. It might, in theory, be possible to re-identify the individual if the data was not adequately protected, for example if it was combined with different sources of information.

How is it protected?

There are strict safeguards on how de-personalised information can be used, because there is the potential that it might be possible to re-identify someone. The higher the possibility of re-identification, the greater the level of control needed.

Example

A report that someone has suffered side-effects from a common medicine, including the patient's age and gender but with name, NHS number and date of birth removed.

Other words that you may see

De-identified, pseudonymised, key-coded, masked, anonymised in context, effectively anonymised, non-disclosive, non-identifiable, de-identified data for limited access.

Anonymous

What is it?

This is information from many people combined together, so that it would not be possible to identify an individual from the data. It may be presented as general trends or statistics. Information about small groups or people with rare conditions could potentially allow someone to be identified and so would not be considered anonymous.

How is it protected?

Because it would not be possible to identify someone, this information does not need special protection and can be published openly.

Example

The number of people who have been prescribed a certain medicine over ten years in five cities.

Other words that you may see

Aggregated data, grouped data, pooled data, statistics.

¹ These excerpts have been taken from national guidance: <https://understandingpatientdata.org.uk/sites/default/files/201704/Identifiability%20briefing%205%20April.pdf>

Below are two examples which outline how any piece or combination of pieces of information could be identifying to reveal the individual.

Example 1: using the attributes age and marital status. At first glance these attributes are not obvious identifiers but let us imagine a case where one of the patients in our dataset is a sixteen-year-old widow. Our implicit demographic knowledge tells us that this is a rare combination which means that if we were to publish this information for this patient then she could potentially be re-identified by, for example, someone spontaneously recognising that the data corresponds to their friend or colleague or neighbour.

Example 2: let us consider the attribute gender. Again, this attribute is not an obvious identifier but let us imagine a case where we have a dataset in which there is only one female patient who has received treatment for a particular condition within a certain month. The gender attribute then would be identifying for this female patient.

As with any security measure, de-identification is not fool proof. Although a rare event, a de-identified dataset could potentially be re-identified by somebody who has sufficient auxiliary information. Re-identification is variously known as data intrusion, the mosaic effect and jigsaw identification. The idea is this: if one can bring extra information to the (released) de-identified data, then one might be able to piece together enough evidence to identify specific respondents, and, or to disclose certain attributes about specific respondents. Let us illustrate this point by returning to example 2, where our dataset contained data about only one woman. If a data intruder possessed the extra information that the woman was the shortest of all the respondents, then they would be able to re-identify the woman from the 'height' attribute even if the gender attribute has been removed from the data.

In some cases, there may be a high level of risk to individuals should re-identification occur. For example, health data, where although there may be no obvious motivation for trying to identify the individual that a particular patient 'episode' relates to, the degree of embarrassment or anxiety that re-identification could cause could be very high. Therefore, the de-identified techniques used to protect data should reflect this.

What other information might conceivably be available, and linked to a de-identified dataset, will be defined by how the data is shared. There are a range of ways in which data can be shared such as through secure databases, secure network access, network folders (which restrict who has access to the data and, or how the data can be used) or openly on the internet. Of course, there is a huge difference between, at one end of the spectrum, making data available to a small number of vetted individuals in a secure folder and at the other end of the spectrum publishing that data as open data on the Internet. In the former case, opportunities for re-identification can, to a large degree, be controlled and limited simply because the data environment is being managed in terms of who can access the data and how. In the latter case there is much less opportunity to control the data environment. As such, the potential for any data anywhere in the world to be used by a determined data intruder to re-identify data is much greater. Therefore, it is very important to ensure that data is shared in a way that is appropriate to the re-identification risk associated with it. So, for example, very sensitive and very detailed de-identified data should only be shared in a secure and controlled environment whilst non-sensitive and less detailed anonymised data can be shared in less controlled environments.

It is also very important to recognise when anonymising data that the process of anonymisation may impact on the usefulness of data. For instance, in example 2, removing the gender of the respondents from the data could result in important generalisations being missed or incorrect inferences being made. So, one should be mindful that the overprotection of data is undesirable since there is little point in sharing and, or disseminating data that does not represent whatever it is that it is meant to represent.

4.1.3 Pseudonymisation

The General Data Protection Regulation defines pseudonymisation as:

“Pseudonymisation - the processing of personal data such that it can no longer be attributed to a single data subject without the use of additional data, so long as said additional data stays separate to ensure non-attribution”.

It is the term used for the process of de-identifying data, so that a coded reference or pseudonym is attached to a record, allowing the data to be associated with a particular individual, but prevents the individual being identified without the pseudonymised key.

Pseudonymisation works by taking a unique identifier (like an NHS number) and completing an algorithm to convert that number into another one, which cannot be traced back to the patient without access to the ‘key’ (a reference table of the original v new values). Providing you always use the same original value, the pseudonymised output will always be the same too, meaning you can link different datasets together to analyse the data without ever being able to reverse engineer the code to identify the patient.

An NHS number works on the same principle as it is a unique identifier relating to all records for that patient, but it is important to remember that this is **not** an adequate pseudonymised ID number. This is because the ‘key’ to trace the patient’s details is widely available to staff with permission to access the NHS spine, the Trust’s PAS system, and the Trust’s Clinical Web Portal. The NHS number should **never** be used as a pseudonymised reference in its own right for this reason. However, it is recommended that the NHS number is used as the original unique identifier which is used when converting to a new pseudonymised version.

As an example, imagine that when you put NHS number 123456789 through the pseudonymisation algorithm creates a new unique number of 5L7TWX619Z to replace the existing one. If you looked on the PAS system, Clinical Web Portal or the NHS spine you would not be able to identify this new number, as they are unique outputs from the algorithm. Without the encryption salt key, you would not have any way of identifying this patient. 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 is not available if the PID is removed completely.

If it is imperative that the data request enables you to link multiple data items together; pseudonymisation techniques should be used if the recipient does not have a legal basis for identifying the patient.

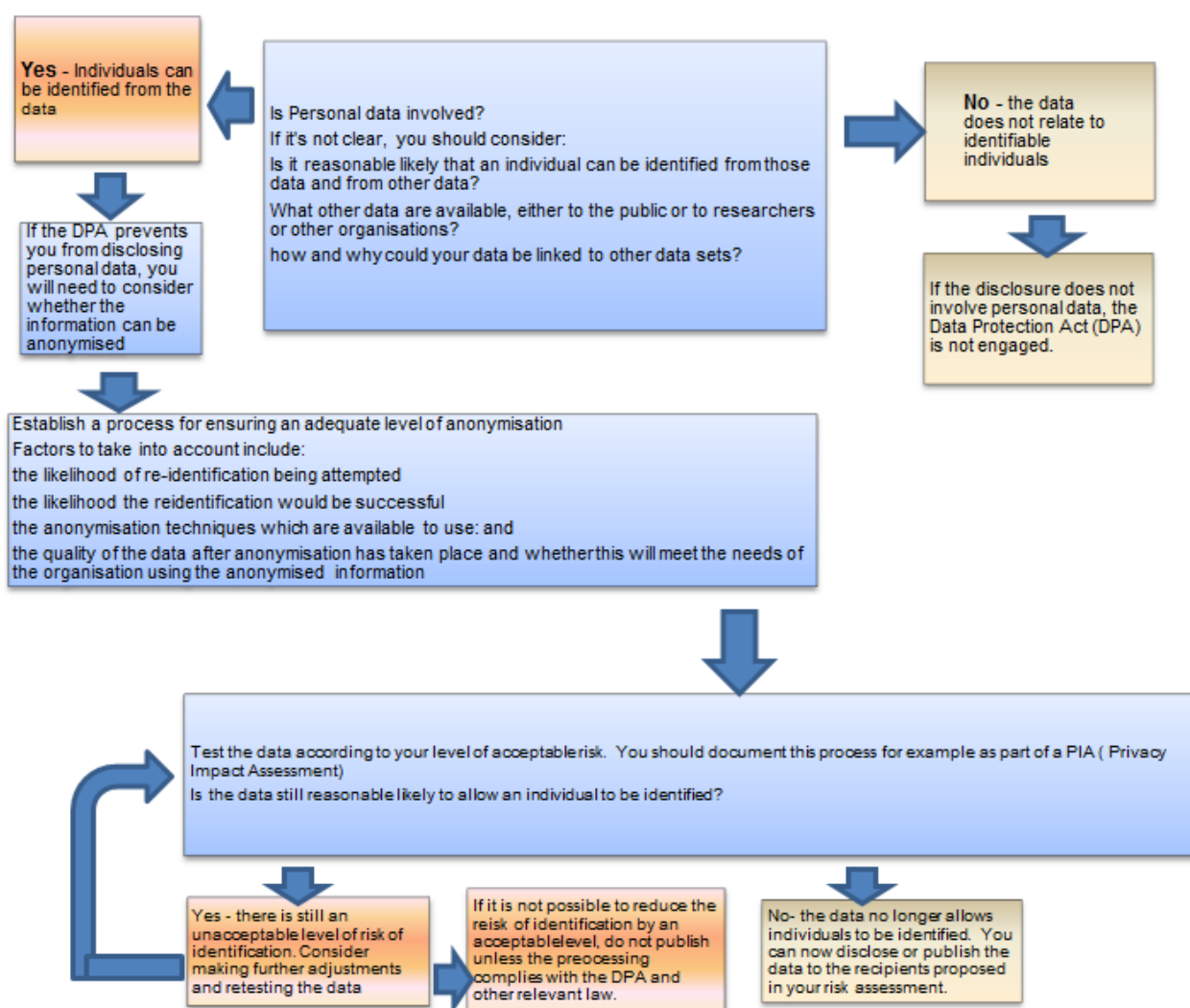
The recommended approach to pseudonymisation is outlined in [Attachment 4](#) of this policy document.

The aim of de-identification is to obscure the identifiable data items within the person's records sufficiently that the risk of potential identification of the subject or a person's record is minimised to acceptable levels. Although the risk of identification cannot be fully removed this can be minimised with the use of multiple pseudonyms.

Deciding when and how to release anonymised data

The reason for releasing data will affect how you make the disclosure, because the risk and consequences of identification will differ.

- Publication under freedom of information is to the wider world and carries more risk because of the audience being bigger.
- Discretionary disclosures, such as those made for research purposes or in your own commercial interests, can be easier to control and assess but are not without risks.



Please use the diagram above in conjunction with [OP85 Information Sharing Policy](#)

¹ Diagram taken from ICO Anonymisation: Managing data protection risk code of practice <https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

Anonymisation techniques

4.3.1. Data masking

This involves stripping out obvious personal identifiers, such as names, from a piece of information to create a dataset in which no person identifiers are present.

- **Partial data removal** – results in data where some personal identifiers, e.g., name and address have been removed but others, such as date of birth, remain.
- **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 reference numbers (e.g., PAS ID) but not the ‘key’ needed to link them to particular individuals (e.g., no access to PAS).

In practical terms, when producing a dataset or using a dataset for reporting purposes, the data can be extracted at patient level in order to have the correct number of records at a granular level, but the personal and, or sensitive items contained within it can be deleted before sending elsewhere by removing the fields completely. This preserves the usefulness of the dataset but ensures confidentiality for the patient as no identifiable items are present.

4.3.2. Derived data items and banding

Derived data is a set of values that reflect the character of the source data, but which hide the exact original values. This is usually done by banding or derivation techniques.

Examples

- Replacing date of birth with age or banding into years (e.g., 1-10, 11-20, 75+ etc).
- Replacing post codes with areas of residence codes e.g., Local Super Output Area codes (LSOA codes) or reducing to partial postcode (i.e., first 4 characters).
- Replacing NHS number with system ID e.g., PAS ID (which is not identifiable to those who do not have access to the system) or replacing with a system generated ID number which is only available in the back-end SQL tables and not visible to anyone who does not have agreed permissions to the SQL tables.

NB: an LSOA lookup table is available from the Trust’s Information Department. Please contact [REDACTED] for more information or support.

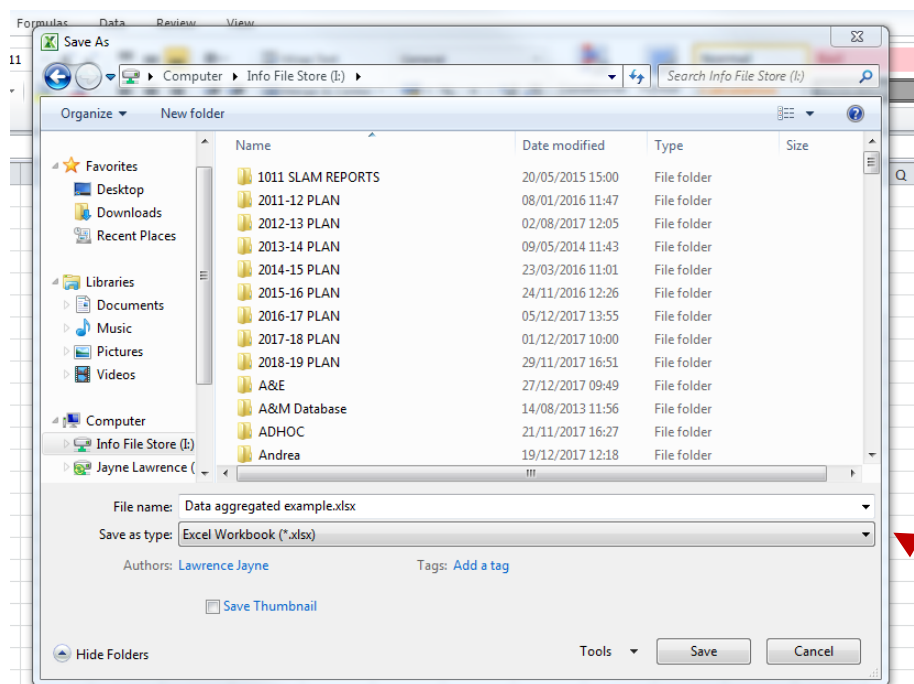
4.3.3. Aggregation

Data is grouped and displayed as totals, so no data relating to or identifying any individual is shown. Small numbers can be identifiable, so should be suppressed through blurring or being omitted altogether.

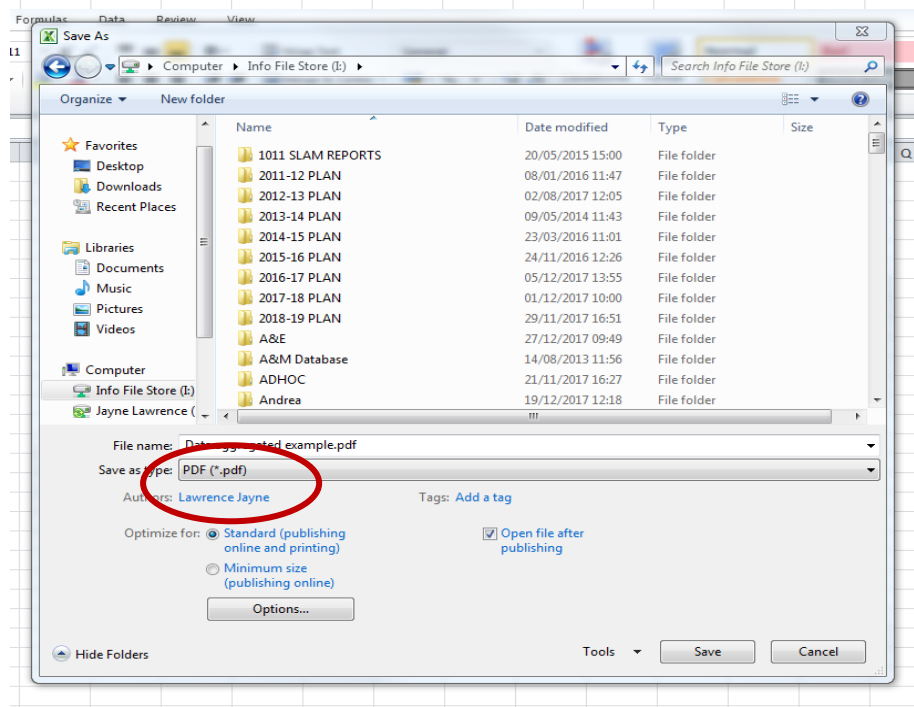
Tabular reporting is a means of producing tabular (summarised or aggregated) data, which protects against re-identification. However, caution must be taken if the pivot table function

within Microsoft Excel is being used to create the table as underlying data can be hidden within the table itself, so you would need to copy and paste values in circumstances where the base dataset includes patient personal/sensitive data fields (for further details see [Attachment 5](#)).

Another method would be to copy and paste the table into a word document, or to save the document as a PDF file by using 'save as' function and selecting PDF (*.pdf) as the preferred file type (shown below):



Save in preferred location but click here and select PDF (*.pdf)



***IMPORTANT* - Requirements for data leaving Trust control**

Irrespective of the method used to de-identify data, there must be data governance arrangements in place to ensure de-identified data leaving the Trust is not subject to re-identification outside of Trust control.

The requirements for de-identification must be taken into consideration at the initiation stage of any project and be included within the impact assessment exercise, as a requirement and a risk. Processing requiring pseudonymised data must be treated at the same level of risk with identifiable data.

If it is discovered that re-identification is possible for data that has left Trust control, incident reporting arrangements must be initiated as soon as possible. Consideration must be given to statutory requirements to report serious data breaches to the ICO within 72 hours. Therefore, time taken for impact assessment should be maximised.

Pseudonymisation techniques

4.4.1. Background

Pseudonymisation is the term used for the process of de-identifying data, so that a coded reference or pseudonym is attached to a record, allowing the data to be associated with a particular individual, but prevents the individual being identified without the pseudonymised key.

Pseudonymisation works by taking a unique identifier (like an NHS number) and completing an algorithm to convert that number into another one, which cannot be traced back to the patient without access to the 'key' (a reference table of the original v new values). Providing you always use the same original value, the pseudonymised output will always be the same too, meaning you can link different datasets together to analyse the data without ever being able to reverse engineer the code to identify the patient.

An NHS number works on the same principle as it is a unique identifier relating to all records for that patient, but it is important to remember that this is **not** an adequate pseudonymised ID number. This is because the 'key' to trace the patient's details is widely available to staff with permission to access the NHS spine, the Trust's PAS system, and the Trust's Clinical Web Portal. The NHS number should **never** be used as a pseudonymised reference in its own right for this reason. However, it is recommended that the NHS number is used as the original unique identifier which is used when converting to a new pseudonymised version.

As an example, imagine that when you put NHS number 123456789 through the pseudonymisation algorithm creates a new unique number of 5L7TWX619Z to replace the existing one. If you looked on the PAS system, Clinical Web Portal, or the NHS spine you would not be able to identify this new number, as they are unique outputs from the algorithm. Without the encryption salt key, you would not have any way of identifying this patient.

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 is not available if the PCD fields are removed completely.

4.4.2. Recommended pseudonymisation techniques/tools

There are tools available to create a pseudonym identifier across datasets. The free tool which is used by the RWT Information Department is called Open Pseudonymiser. This tool is available at the following website: <https://www.openpseudonymiser.org/Default.aspx>.

It contains a guidance document explaining the process to follow. This can be found here: https://www.openpseudonymiser.org/resources/2.0.2/docs/OpenPseudonymiser-User_Guide_v2.0.2b.docx.

It is imperative that the salt key table is stored securely once the process has been undertaken. It should be saved somewhere secure on the Trust's network and can only be

accessed by authorised staff members. If this key is divulged at any point, the data can be de-pseudonymised and re-identification of patients can occur. It may be that this has been agreed as part of a data sharing agreement where data is sent pseudonymised for encrypted data transfer only, but you must make sure necessary agreements are in place with the third party to also keep the key locked down and accessible by authorised personnel only. This should be done in line with [OP85 Information Sharing Policy](#).

It is not recommended that you create your own algorithm for creating a pseudonym, but if this is required, it is essential that you always use the same original identifier (e.g., NHS number) and the same algorithm for one work output if it has multiple datasets. It is also vital that the algorithm is thoroughly tested to check that it cannot be reverse engineered and cracked to reveal the initial identifier e.g., by rounding the number in the middle of a string of calculations, it would be impossible to work the calculation backwards to arrive at the original number. However, you will need to create a separate algorithm for each individual work item or output, to prevent the risk of re-identification across different work outputs.

The aim of de-identification is to obscure the identifiable data items within the person's records sufficiently that the risk of potential identification of the subject or a person's record is minimised to acceptable levels, this will provide effective anonymisation. Although the risk of identification cannot be fully removed this can be minimised with the use of multiple pseudonyms.

The Trust Information Department have experience in using Open Pseudonymiser and have also developed a tool in the Trust Data Warehouse that also masks additional items such as postcode, ethnicity, gender etc and can offset dates for additional security where needed too. For any technical help and support, please contact the Information Department via their team email address: [REDACTED].

4.4.3. Pseudonymisation considerations

To effectively pseudonymise data the following actions must be taken.

- Each identifying field of PCD must have a unique pseudonym or mask.
- Pseudonyms being 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 i.e., 5L7 TWX 619Z. 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 Personal or PID and the encryption salt key must be kept secure.

***IMPORTANT* - Requirements for data leaving Trust control**

Irrespective of the method used to de-identify data, there must be data governance arrangements in place, to ensure de-identified data leaving the Trust is not subject to re-identification outside of Trust control.

The requirements for de-identification must be taken into consideration at the initiation stage of any project and be included within the impact assessment exercise, as a requirement and a risk. Processing requiring pseudonymised data must be treated at the same level of risk with identifiable data.

If it is discovered that re-identification is possible for data that has left Trust control, incident reporting arrangements must be initiated as soon as possible. Consideration must be given to statutory requirements to report serious data breaches to the ICO within 72 hours. Therefore, time taken for impact assessment should be maximised.

Unsafe/ineffective techniques not to be used or be used with caution

The following examples have been taken from the ICO guidance “How to Disclose information safely - Removing personal data from information requests and datasets”. A link is provided at the bottom of the document for further reference. All of these techniques are acceptable for use when no patient identifiable data items are included within the file. If the report or dataset does include personal data items, these techniques are not recommended for anonymising the file, or additional steps are to be considered as outlined below.

Hidden data - Hiding in plain sight

The simplest case of data being disclosed in error can occur when data is not immediately visible on the screen but elsewhere within the file. This can be due to a range of design choices or the rendering of certain formatting styles. For example, when setting up a template, a user might have chosen to ‘hide’ certain data by setting the font colour to be the same as the background (e.g. white on white or black on black).

Whilst hiding data in this manner prevents personal data being disclosed on a printed version of the file, it will still remain within the source file. This personal data is at risk of unintended disclosure if the electronic version is distributed. Highlighting the text or changing the font colour will expose it (see Figure 1).

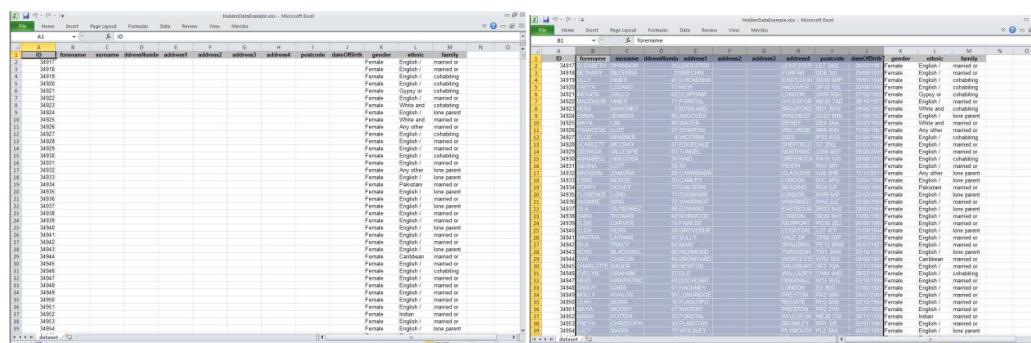


Figure 1: The left-hand side shows the identifying elements hidden from immediate view as the font colour is set to white. The right-hand side shows the text easily revealed simply by selecting the relevant cells.

Another example of where data might be hidden from obvious view is when it is placed in the fringes of a file where it is not expected to be found.

As an example, Microsoft Excel 2007 and upwards support up to 16,384 columns and 1,048,576 rows of data. A user might place data outside of the normal visible area with the aim to hide it from being displayed on a standard sized monitor. For example, personal data being moved to columns AA to AI.

The data is still available within the original file and accessible to anyone who accesses it, so this is not an effective technique in anonymising data.

Hidden rows and columns

A common method of 'hiding' data within a spreadsheet is through the use of hidden rows or columns.

Figure 2 shows part of a Microsoft Excel spreadsheet that contains a short log of property rentals and associated payment information.

	A	C	D	E	F
1	Property	Tenancy start date	Payment day	Monthly rent	Comments
2	1	05/03/2013	15	£ 495.00	
3	2	09/11/2009	15	£ 575.00	
4	3	15/06/2014	1	£ 495.00	
5	4	01/07/2012	1	£ 525.00	
6	5	01/09/2010	5	£ 400.00	
7					
8					

Figure 2: A log of property rentals and associated payment information

Column B is not visible in the column headings but that does not mean that there is no Column B within the file or that it does not contain any personal data. Column B has been set to be hidden, meaning hidden from view. Selecting 'Unhide' from the appropriate submenu is a trivial series of clicks to return the data to full view (shown in Figure 3).

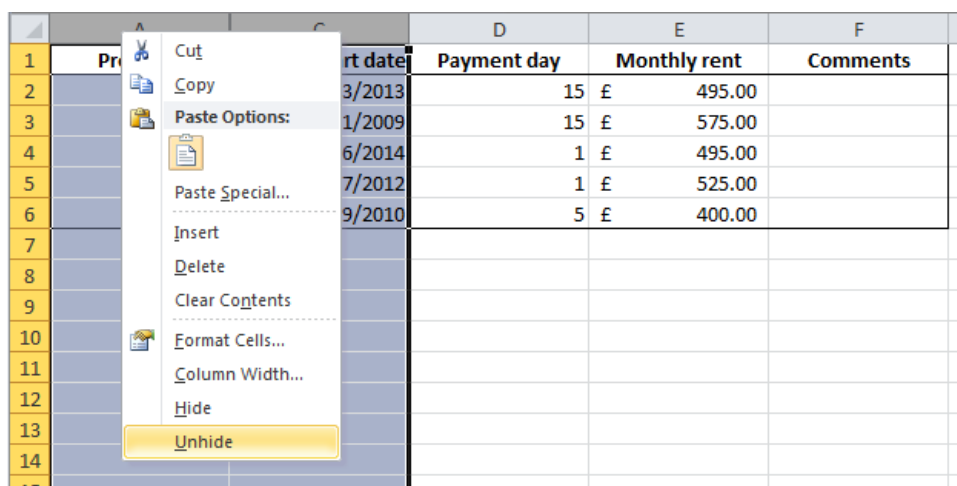


Figure 3: Menu selection required to unhide Column B

Once the column has been unhidden, personal data contained within the column is accessible to anyone who has access to the file (Figure 4). Therefore, the hiding of data in this manner is not an effective or appropriate means to protect against unauthorised access. It is also obvious, from the column headings themselves when hidden columns exist.

	A	B	C	D	E	F
1	Property	Tenant	Tenancy start date	Payment day	Monthly rent	Comments
2	1	Mr A	05/03/2013	15	£ 495.00	
3	2	Mrs B	09/11/2009	15	£ 575.00	
4	3	Mr C	15/06/2014	1	£ 495.00	
5	4	Ms D	01/07/2012	1	£ 525.00	
6	5	Dr E	01/09/2010	5	£ 400.00	
7						

Figure 4: Personal data contained within the previously hidden Column B is now visible

Whilst hidden rows and columns might seem a trivial example, it is important to remember that the decision to release a file may have been taken by an individual who has no working knowledge of the file or its original purpose. The original author of the file may even have left the organisation many years previously, and it may not be immediately apparent that the data fields were hidden from view.

Hidden rows and columns can also be identified from the fact that the (column and row) headings do not flow in a consecutive order.

Hidden sheets and tabs

An entire worksheet can also be hidden from view. It is less obvious that a worksheet has been hidden as they can be renamed so it is more difficult to notice that a sequential number or letter is missing. However, relying on obscurity as a security measure is poor practice, and is not to be considered as an appropriate measure to prevent unauthorised access (as hidden sheets can be easily unhidden).

Pivot tables

A pivot table can be used to summarise a large set of data. This can create an automatic summary of the underlying data. Figure 8 shows an example extract of the dataset created in Microsoft Excel 2010.

ID	forename	surname	address1	address2	address3	address4	postcode	dateOfBirth	gender	ethnic	family
2	34917	ELIZABETH	SHANNON	35	LEICESTER ROAD		LEICESTER LE7 9AQ	06/03/1987	Female	English /	married or civil partner couple family without dependent childr
3	34918	BETHANY	BECKERA	2	BRECKIN ROAD		FORFAR DD8 3JX	29/09/1937	Female	English /	married or civil partner couple family without dependent childr
4	34919	LILLY	HINES	41	CHICKENHALL LANE		EASTLEIGH SO50 6RP	19/01/1946	Female	English /	cohabiting couple family without dependent family
5	34920	FREYA	LOZANO	77	NEW STREET		ANDOVER SP10 1EL	20/08/1990	Female	English /	cohabiting couple family without dependent family
6	34921	IMOGEN	GALLO	62	CLAPHAM ROAD		LONDON SW9 0QH	07/02/1925	Female	Gypsy or I	cohabiting couple family without dependent family
7	34922	MADDISO	HINES	71	FORSTAL ROAD		AYLESFORD ME20 7AD	28/10/1977	Female	English /	married or civil partner couple family without dependent childr
8	34923	HEIDI	MAHONEY	7	BOWLAND STREET		BRADFORD BD1 3BW	09/06/1952	Female	White anc	cohabiting couple family without dependent family
9	34924	EMMA	JENKINS	95	ANDOVER ROAD NORTH		WINCHES' SO22 6NN	01/06/1927	Female	English /	lone parent family without dependent family
10	34925	MAYA	LIM	86	WATER LANE		DERBY DE4 4AA	25/03/1956	Female	White anc	married or civil partner couple family with dependent children
11	34926	FRANCES	LUTZ	57	STANTON ROAD		WELLINGE NN8 4HN	13/06/1967	Female	Any other	married or civil partner couple family without dependent childr
12	34927	ELLIE	WARNER	8	VICTORIA ROAD		DISS IP22 4GS	01/08/1964	Female	English /	cohabiting couple family without dependent family
13	34928	SCARLETT	MCCRAY	67	EDGEDALE ROAD		SHEFFIELD S7 2BQ	02/03/1926	Female	English /	married or civil partner couple family without dependent childr
14	34929	GEORGIA	GILLESPIE	83	TUNNEL ROAD		NORTHWY CW8 4EP	09/08/2005	Female	English /	married or civil partner couple family without dependent childr
15	34930	ANNABEL	HINOJOSA	34	HAIQ STREET		GREENOCIA PA15 1JG	09/08/1931	Female	English /	cohabiting couple family with dependent family
16	34931	SIENNA	LOTT	20	ST CATHERINES ROAD		PERTH PH1 5RY	20/06/2007	Female	English /	married or civil partner couple family without dependent childr
17	34932	MADISON	ZAMORA	68	CARNWADRIC ROAD		GLASGOW G46 8HR	13/11/2011	Female	Any other	lone parent family with dependent family
18	34933	ESME	MORSE	28	CAMLEY STREET		LONDON N1C 4PG	12/04/1998	Female	English /	lone parent family without dependent family
19	34934	POPPY	DICKEY	77	CHILTERN ROAD		READING RG4 5JF	15/03/1995	Female	Pakistani	married or civil partner couple family with dependent children
20	34935	FLORENCE	LUND	17	EDWARE ROAD		LONDON NW9 6AF	22/02/2006	Female	English /	lone parent family without dependent family
21	34936	JASMINE	KINGS	72	WARRINGTON ROAD		WARRING WAS 2JZ	22/08/2009	Female	English /	married or civil partner couple family without dependent childr
22	34937	ISLA	GUTIERRE	96	EDWARD ROAD		GASTROU BN23 8AS	28/03/1959	Female	English /	lone parent family without dependent family
23	34938	SARA	THOMAS	62	NORWOOD ROAD		LONDON SE24 5AY	13/05/1983	Female	English /	married or civil partner couple family with dependent children
24	34939	ELSIE	CARVER	74	FAIRLEE ROAD		NEWPORT PO30 2EJ	06/08/1985	Female	English /	married or civil partner couple family with dependent children
25	34940	ELIZA	KERR	56	GROVEBURY ROAD		LEIGHTON LU7 4FF	21/06/1944	Female	English /	lone parent family with dependent family
26	34941	MARTHA	LATHAM	61	SULLY MOORS ROAD		VALE OF G CF64 5RP	24/05/2013	Female	English /	married or civil partner couple family with dependent children
27	34942	ISLA	TRACY	64	MAIN ATRUNK ROAD		SPALDING PE12 0BW	26/01/1921	Female	English /	married or civil partner couple family without dependent childr
28	34943	ROSE	BLACKWEI	66	RICHMOND ROAD		TWICKEN TW1 3AW	22/10/1970	Female	English /	lone parent family without dependent family
29	34944	AIVA	CHACON	94	BROMYARD ROAD		WORCEST WR2 5EE	08/04/1941	Female	Caribbean	married or civil partner couple family without dependent childr
30	34945	CHARLOTTE	BAUER	99	NEWTON ROAD		SALISBURY SP2 7QA	13/11/1985	Female	English /	married or civil partner couple family with dependent children
31	34946	EVELYN	GRAHAM	57	OLD GORSLEY LANE		WALLASEY CH44 4HD	28/03/1933	Female	English /	cohabiting couple family with dependent family
32	34947	HEIDI	KIRKPATR	69	DOCHARTY ROAD		DINGWAL IV15 9UG	03/10/1986	Female	English /	married or civil partner couple family without dependent childr
33	34948	DARCY	CARR	61	HACKNEY ROAD		LONDON E2 9ED	01/02/1925	Female	English /	married or civil partner couple family without dependent childr

Figure 8: An extract of the example dataset displayed in Microsoft Excel 2010

A summary of a dataset can be created using the pivot table feature. Figure 9 shows the pivot table created to summarise the number of individuals by gender and family type.

Row Labels	Count of ID
Female	5000
cohabiting couple family with dependent family	277
cohabiting couple family without dependent family	456
lone parent family with dependent family	518
lone parent family without dependent family	500
married or civil partner couple family with dependent children	1238
married or civil partner couple family without dependent children	2011
Male	5000
cohabiting couple family with dependent family	300
cohabiting couple family without dependent family	442
lone parent family with dependent family	507
lone parent family without dependent family	506
married or civil partner couple family with dependent children	1216
married or civil partner couple family without dependent children	2029
Grand Total	10000

Figure 9: A pivot table displaying a summary of individuals by Gender and Family type

As with hidden data fields, even though the underlying data is not immediately visible on the screen, it can still be accessed. A double-click on the pivot table can signal to the software to automatically extract the data used to calculate the clicked data and display this in a new worksheet.

Even if the worksheet containing the original data is deleted from the workbook or if the pivot table is copied into a new workbook, the underlying data may be copied across with it, making the data accessible to the user. Figure 10 shows the result of double-clicking on the count of 'Female / Lone parent family with dependent family'. The personal data of the 518 individuals which were summarised in that row are extracted by the software and displayed in a new worksheet.

ID	forename	surname	addressNumber	address1	address2	address3	address4	postcode	dateOfBirth	gender	ethnic	family
1	EMILIA	HENSON	58	HIGH STREET				UTTOXETER ST14 7HT		Female	English	\ lone parent family with dependent fa
2	SKYE	HOWE	1	COLHAM GR				HILLINGDON UB8 3JY		Female	English	\ lone parent family with dependent fa
3	ISABELLE	VELAZQUEZ	85	HIGH STREET				BRIDGWATAS 1TB		Female	Caribbean	lone parent family with dependent fa
4	ELEANOR	EMERY	7	NEWPORT R				GNOSALL ST20 0BN		Female	English	\ lone parent family with dependent fa
5	ANNA	STEELE	51	BLACKSTOCI				LIVERPOOL L3 6EP		Female	Any other	lone parent family with dependent fa
6	LUCY	ZAVALA	54	LONDON RC				ALTON GU34 4HA		Female	English	\ lone parent family with dependent fa
7	ZARA	SOSA	57	RIGNALL RO				GREAT MISS HP16 9AN		Female	English	\ lone parent family with dependent fa
8	ELLIE	BRIDGES	73	STOCKPORT				HYDE SK14 3QT		Female	Pakistani	lone parent family with dependent fa
9	ANNA	RAINEY	21	PORTWAY R				WARLEY B65 9BY		Female	Indian	lone parent family with dependent fa
10	BETHANY	LARSEN	4	RINGWOOD				POOLE BH12 3JN		Female	Banglades	lone parent family with dependent fa
11	CHLOE	LARKIN	76	OAKLAND R				LEICESTER LE2 6AN		Female	Arab	lone parent family with dependent fa
12	ANNABELLE	STOUT	66	WALLINGFO				WANTAGE OX12 8BB		Female	English	\ lone parent family with dependent fa
13	VIOLET	LAKE	71	PORTWAY R				OLDBURY B69 2BT		Female	Caribbean	lone parent family with dependent fa
14	EMILY	OTTO	37	BEAU STREE				LIVERPOOL L3 3JE		Female	English	\ lone parent family with dependent fa
15	ELIZA	CONTRERAS	56	HUMBERSTC				LEICESTER LE5 3AP		Female	English	\ lone parent family with dependent fa
16	MADISON	ZAMORA	68	CARNWADR				GLASGOW G46 8HR		Female	Any other	lone parent family with dependent fa
17	BELLA	CORDOVA	29	STATION RC				PRESTON PR4 2HD		Female	Pakistani	lone parent family with dependent fa
18	ABIGAIL	JENNINGS	65	HIGH ROAD				WOODFORD IG8 0PR		Female	English	\ lone parent family with dependent fa
19	GEORGIA	WIGGINS	100	BURY NEW F				SALFORD M7 2YJ		Female	English	\ lone parent family with dependent fa
20	ELIZABETH	STROUD	68	GOWER STR				BOLTON BL4 7EY		Female	English	\ lone parent family with dependent fa
21	ROSE	REECE	76	MAYFIELD R				ASHBOURNI DE6 1AR		Female	English	\ lone parent family with dependent fa
22	WILLOW	COTE	37	UPWELL STR				SHEFFIELD S4 8AJ		Female	African	lone parent family with dependent fa
23	ALICE	CHAMBERL	11	BLAGUEGAT				SKELMERSD WN8 8TY		Female	English	\ lone parent family with dependent fa
24	ELIZA	KERR	56	GROVEBURY				LEIGHTON B LU7 4FF		Female	English	\ lone parent family with dependent fa
25	ELIZABETH	HAAS	77	GREAT PERC				LONDON WC1X 9QU		Female	English	\ lone parent family with dependent fa
26	WILLOW	CHILDERS	40	HOLES BAY f				POOLE BH15 2BD		Female	English	\ lone parent family with dependent fa
27	MADISON	HUTCHISON	45	LANGLANDS				GLASGOW G51 4AW		Female	English	\ lone parent family with dependent fa
28	PAIGE	RAY	60	LAWFORD R				RUGBY CV21 2EA		Female	Pakistani	lone parent family with dependent fa
29	BELLA	HOWELL	71	PARKFIELD f				BIRMINGHAM B8 3AY		Female	Any other	lone parent family with dependent fa
30	EMMA	WISE	29	ACADEMY S				DUMFRIES DG1 1DA		Female	English	\ lone parent family with dependent fa
31	EVA	RODRIGUES	12	BUTTERLEY S				LEEDS LS10 1AW		Female	English	\ lone parent family with dependent fa
32	DAISY	WINKLER	47	DRUMMONI				STAFFORD ST16 3HU		Female	English	\ lone parent family with dependent fa

Figure 10: An extract of the personal data of the 518 female individuals within the lone parent family with dependent family category

Pivot tables can be used to create data tables but are not safe for transferring information if it contains personal identifiable data. However, you can copy the pivot table and paste only the values to a new workbook. This is sometimes referred to as a 'paste special' operation. The copied data can also be checked by double-clicking cells within the copied pivot table to ensure there is no link back to the source data.

Charts

As with pivot tables, charts can also contain a copy of the source data embedded in the background. A further risk could arise when a chart is embedded into a document or presentation as the embedded chart could also contain a copy of the source data.

Figure 12 shows an example of a chart being created from a pivot table summarising the count of family types in the dataset.

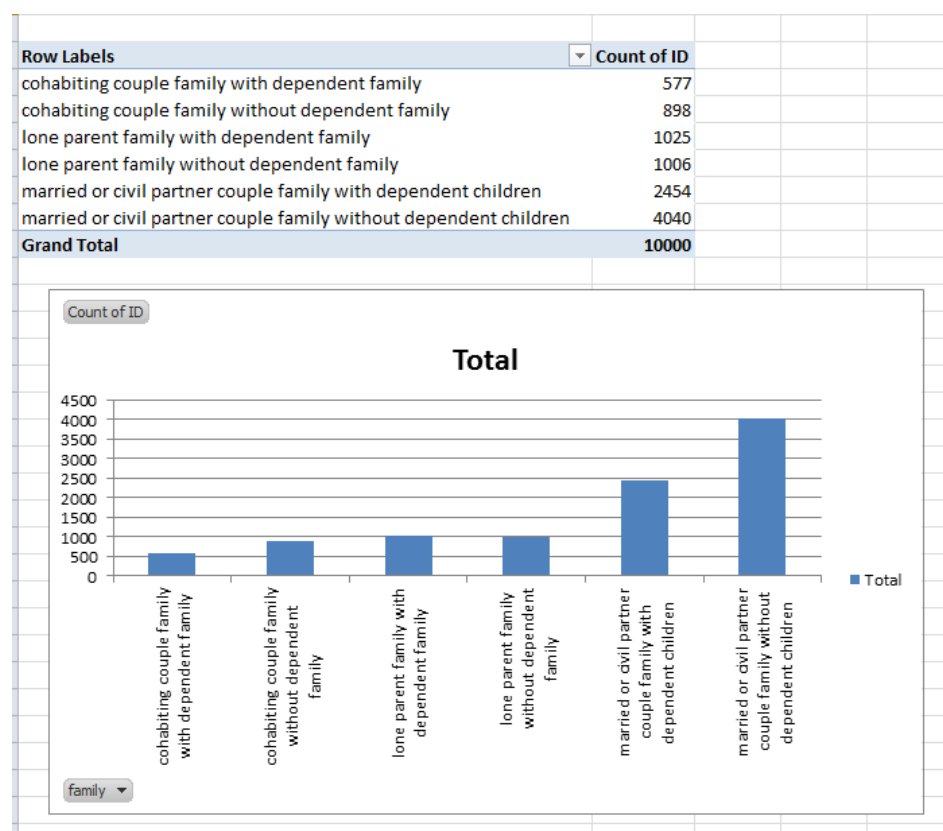


Figure 12: A chart and pivot table summarising the count of family types in the dataset

If the chart is embedded within a Microsoft Office Word document, then a copy of the underlying data is also copied across and embedded within the document. Simply double-clicking on the chart and selecting the data worksheet can reveal the underlying data (Figure 13).

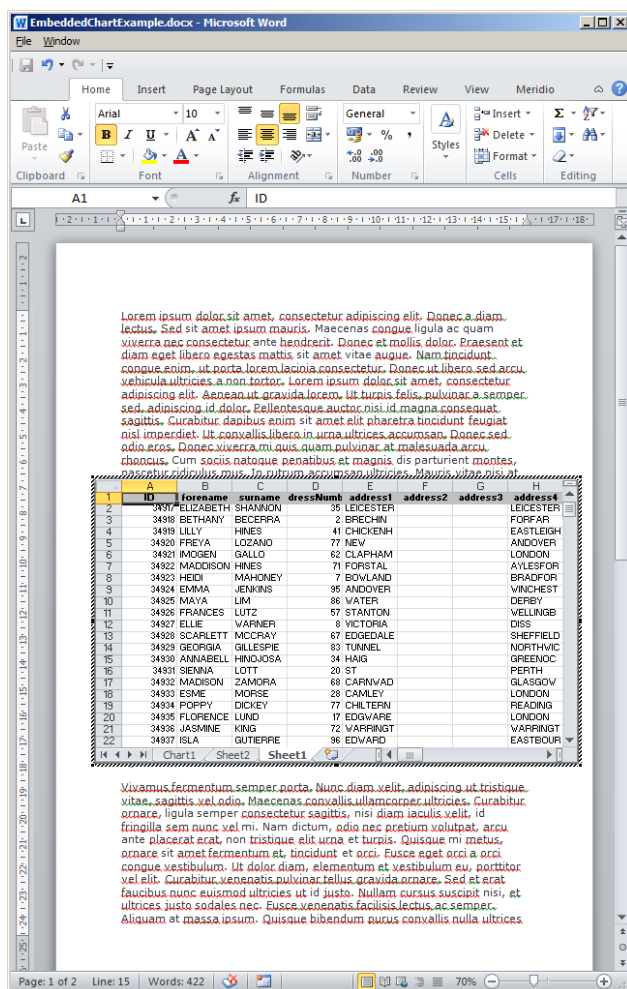


Figure 13: After double-clicking the embedded graph object the underlying data can be revealed by selecting the data worksheet.

It is recommended that the chart is copied and pasted as an image file (e.g., jpg or png) into the destination file. Exporting the document into a format such as PDF will also remove the underlying source data from the graph.

Functions

Functions such as LOOKUP and VLOOKUP also create and store a cache of the source data that can be exposed through careful manipulation of the function. Figure 14 shows an example of the VLOOKUP function. The source data was located in a different file but a cache of the dataset is also stored within the current file.

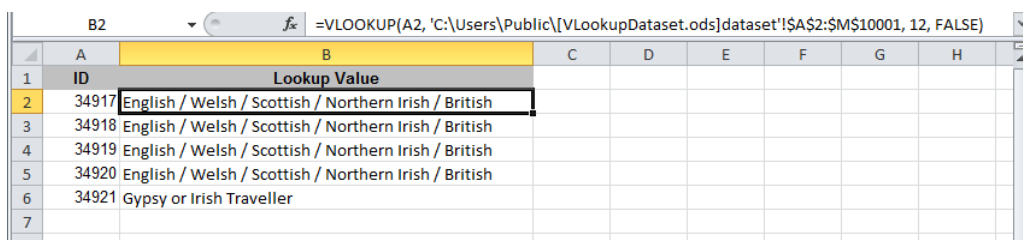
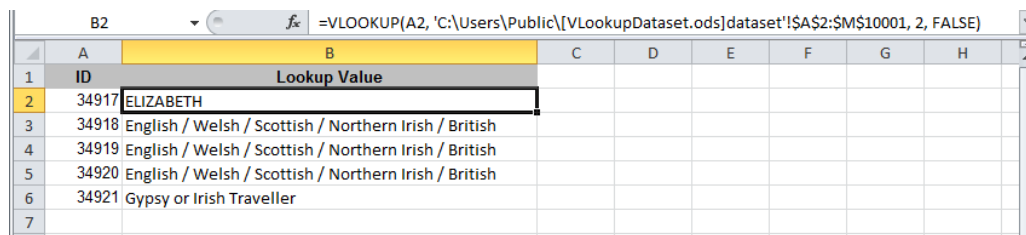


Figure 14: A VLOOKUP function referencing a dataset in a different file

This can be demonstrated by editing the column index in the VLOOKUP formula. By manually changing the column index from 12 to 2 the formula will return data from the second column in the cache without access to the source file as can be seen in Figure 15.



	A	B	C	D	E	F	G	H
1	ID	Lookup Value						
2	34917	ELIZABETH						
3	34918	English / Welsh / Scottish / Northern Irish / British						
4	34919	English / Welsh / Scottish / Northern Irish / British						
5	34920	English / Welsh / Scottish / Northern Irish / British						
6	34921	Gypsy or Irish Traveller						
7								

Figure 15: Manually editing the VLOOKUP function can return a different value from the cache

Therefore, caution should be applied when using VLOOKUP functions when the data file could be shared, and it contains personal identifiable data. It is recommended that a copy of the file is made, and the VLOOKUP formula is copied and pasted over as text.

[How to disclose information safely \(ico.org.uk\)](http://ico.org.uk)