Introduction to anonymisation

Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance

May 2021

ico.

Information Commissioner's Office

Contents

About this guidance

At a glance

- Anonymisation is a privacy-friendly way to harness the potential of data, including when developing new and innovative products and services.
- Effective anonymisation of personal data is possible in many circumstances. It depends on the techniques you use. You need to reduce the risks of identifying individuals to a sufficiently remote level so that the information is effectively anonymised.
- This guidance will help all organisations that seek to anonymise personal data, for whatever purpose.
- It will help you identify the issues you need to consider to use anonymisation techniques effectively.

In detail

Why have you produced this guidance?

Data is the lifeblood of the digital economy, and data sharing is key to opening up new opportunities.

We understand the benefits that data sharing can bring to organisations, individuals and society as a whole, but there are risks too. However, effective anonymisation techniques provide a privacy-friendly alternative to sharing personal data.

This guidance sits alongside our data sharing code of practice, which gives practical guidance on how to share personal data in line with data protection law. Anonymisation offers an alternative way to use or share data by making sure that individuals are not identifiable.

You need to have a reasonable degree of confidence that disclosing or sharing apparently anonymous information will not lead to an inappropriate disclosure of personal data, eg through 're-identification'.

Determining the status of information in different circumstances is therefore a key challenge. For example, you may hold information that is clearly

personal data, but its status when processed by another organisation or by the world at large may be unclear.

Anonymisation safeguards individuals' privacy and is a practical example of the data protection by design approach that the law requires.

Effective anonymisation of personal data is possible, desirable and can help society to make rich data resources available whilst protecting individuals' privacy.

Further reading outside this guidance

Visit our data sharing information hub for more information about the data sharing code.

What is this guidance about?

This guidance:

- explains what we mean by anonymisation and pseudonymisation;
- details how this affects your data protection obligations and responsibilities;
- discusses what you should consider when anonymising personal data;
- provides good practice advice for when you seek to anonymise this data; and
- discusses technical and organisational measures to mitigate the risks to individuals when you do so.

This guidance deals with the role that anonymisation plays in the context of data protection law:

- the Data Protection Act 2018; and
- the three data protection regimes:
  - general processing under Part 2 of the DPA 2018 and the UK GDPR;
  - law enforcement processing under Part 3; and
  - intelligence services processing under Part 4.

Where relevant, the guidance highlights and explains any differences between the regimes.

This guidance does not generally consider the impacts of anonymisation on areas of ICO competence outside data protection. However, some sections are relevant under other laws such as the Freedom of Information Act 2000 (FOIA).

This guidance is not a statutory code. It contains advice on how to interpret relevant law in the context of anonymous information, and recommendations on good practice. There is no penalty if you fail to adopt good practice recommendations, as long as you find another way to comply with the law.

This guidance does not describe every possible anonymisation technique in detail, but includes case studies and good practice recommendations.

Who is this guidance for?

You should use this guidance if you are considering turning personal data into anonymous information. For example, this guidance is relevant if you:

- are required by law to publish anonymous information, eg some health service bodies;
- are looking to use data in new and innovative ways, eg to improve services or design new products or collect large volumes of data to train AI models;
- need to deal with a request for information under FOIA, and it includes personal data;
- want to become more transparent and accountable to the public; or
- want to provide anonymous information for research purposes, or to enable wider societal benefits.

This guidance describes ways you can assess and mitigate the risks that may arise, particularly in terms of how to assess whether other data is available that may make re-identification likely. It also helps you to assess other risks (eg those involved with producing and publishing anonymous information). These may include:

- information about an individual's private life ending up in the public domain;
- a supposedly anonymous dataset being 'cracked' so personal data about individuals is compromised;
- re-identification causing harms to individuals, such as a loss of control of their personal data causing harms such as damage, embarrassment anxiety or financial loss;
- reduced trust and confidence if you disclose information unsafely; and
- legal problems where insufficiently redacted qualitative data is disclosed, eg under FOIA.

Anonymisation can help you to mitigate these risks and share information fairly and proportionately. This guidance will help you develop your understanding of anonymisation techniques, their strengths and weaknesses, and the suitability of their use in particular situations.

This guidance is for a general readership and does not provide in-depth knowledge of security engineering or statistical methodology. However, it does assume a level of familiarity with key data protection terms and concepts, and related terms about anonymisation. We discuss these in more detail where it helps to explain the risks that anonymisation may create or exacerbate.

It will also help technical experts understand how the data protection framework applies to (and facilitates) their activities.

The guidance focuses on specific risks and controls to ensure your approach to anonymisation provides safeguards for individuals' rights and freedoms. However, it is not intended as an exhaustive guide to data protection compliance and it is not prescriptive. It gives you the flexibility to implement anonymisation techniques in your own way, taking a proportionate and risk-based approach.

When the ICO investigates an issue about the effectiveness of anonymisation, we will take the advice in this guidance into consideration. It will stand you in good stead if you can demonstrate that your approach to producing and disclosing anonymous information takes account of these good practice recommendations. However, you may find alternative ways of complying with your data protection obligations and may choose to adopt good practice measures above and beyond those we have set out.

You need to make sure you are aware of all your obligations and you should read this guidance alongside our other guidance in the Guide to data protection.

Further reading outside this guidance

Read the Guide to Data Protection for more information.

How is this guidance structured?

This guidance is divided into several parts covering different aspects of anonymisation in the context of data protection law.

Firstly, it introduces the key concepts of anonymisation and pseudonymisation, places them in the context of the UK legal framework and explains the role they play.

Secondly, the guidance covers the concept of identifiability, including approaches such as the 'spectrum of identifiability' and how these can apply in data sharing scenarios. It also looks at how you can manage re-identification risk, and covers established concepts like the 'reasonably likely' and 'motivated intruder' tests.

Thirdly, we look at relevant accountability and governance requirements in the context of anonymisation, including data protection by design and DPIAs and the use of trusted third parties.

Fourthly, the guidance looks at the role anonymisation can play in the context of research.

Finally, it covers key anonymisation techniques, technological solutions, case studies and practical examples that illustrate how you can ensure anonymisation is effective.

In each section, we discuss what you must do to comply with data protection law as well as what you should do as good practice.

Introduction to anonymisation

At a glance

- In order to understand anonymisation, you first need to understand personal data.
- Anonymous information is data which does not relate to an identified or identifiable individual (ie data that is not personal data).
- Anonymisation is the process of turning personal data into anonymous information so that an individual is not (or is no longer) identifiable.
- Data protection law does not apply to truly anonymous information.
- Effective anonymisation is possible. However, the techniques you use must reduce the risks of identifying individuals to a sufficiently remote level that they **effectively** anonymise the information.
- Pseudonymisation is a type of processing designed to reduce data protection risk, but not eliminate it. You should think of it as a security and risk mitigation measure, not as an anonymisation technique by itself.

In detail

What is personal data?

Data protection law regulates the processing of personal data. Effective anonymisation therefore depends on a sound understanding of what constitutes this data.

Section 3(2) of the DPA 2018 says that personal data means:

Quote

'any information relating to an identified or identifiable living individual'

Section 3(3) defines an 'identifiable living individual' as:

Quote

'a living individual who can be identified, directly or indirectly, in particular by reference to—

(a) an identifier such as a name, an identification number, location data or an online identifier, or

(b) one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual.'

Clearly, information (or a combination of information) is not personal data if it does not relate to an identifiable individual. Data protection law does not apply to truly anonymous information.

The definition of personal data in Section 3 applies to the UK data protection framework as a whole. Article 4(1) of the UK GDPR also defines personal data for the purposes of the UK's 'general processing' regime, and this definition is not materially different.

As personal data has to be about living individuals, data protection law does not apply to information about the deceased. However, you should note that this data may still be protected by confidentiality or other legal rules.

Relevant provisions in the legislation

Sections 3(2) and (3) of the DPA 2018 (external link)

Article 4(1) of the UK GDPR (external link) and the Keeling Schedule (external link)

Further reading

Read our guidance on 'What is personal data?' in the Guide to the UK GDPR.

For more information on the UK data protection framework and its three regimes, see 'About the DPA 2018' in the Guide to data protection.

What is anonymous information?

Data protection law does not explicitly define 'anonymous information'.

However, in data protection terms, you should understand it as the end result of a process that converts personal data into information that the data protection legislation no longer applies to. Essentially it is information that is then out of scope of the DPA 2018.

The data protection framework's general processing regime provides further guidance on what the term 'anonymous information' means. Recital 26 of the UK GDPR says this is:

Quote

'…information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.'

As anonymous information is therefore no longer personal data, the principles of data protection law do not apply when you process it.

In the ICO's view, the same information can be personal data to one organisation, but anonymous information in the hands of another organisation. Its status depends greatly on its circumstances, both from your perspective and in the context of its disclosure.

You need to take into account all the means reasonably likely to be used, by yourself or a third party, to identify an individual that the information relates to. This will determine whether the data is anonymous information. We refer to this as the 'reasonably likely' test.

Further reading

We will discuss the concepts of 'identifiability' and the 'reasonably likely' test in more detail in future sections of this guidance. When we publish these sections, we will update this further reading box.

What is anonymisation?

Data protection law also does not specifically define 'anonymisation'.

However, its meaning for the purposes of the UK data protection framework is clear from the wording of Recital 26 of the UK GDPR. It is the **way** in which you turn personal data into anonymous information, so that it then falls outside the scope of data protection law.

You can consider data to be effectively anonymised when it:

- does not relate to an identified or identifiable individual; or
- is rendered anonymous in such a way that individuals are not (or are no longer) identifiable.

We use the broad term 'anonymisation' to cover the techniques and approaches you can use in the pursuit of these aims – ie, of preventing the identification of the individuals the data relates to, taking into account all relevant factors.

It is important to note that you must carefully assess each case individually based on the specific circumstances. This will help you to decide the effectiveness of an anonymisation technique and therefore whether the data is effectively rendered anonymous. Clearly, 100% or 'absolute' anonymisation is the most desirable position. At the same time, you will not always be able to state that a specific technique or set of controls will achieve these aims, particularly as technology changes over time.

This means that even where you use anonymisation techniques, a level of inherent identification risk may still exist. However, this residual risk does not mean that particular technique is ineffective. Nor does it mean that the resulting data is not effectively anonymised for the purposes of data protection law when you consider the context.

Also, data protection law does not require anonymisation to be completely risk-free. You must be able to mitigate the risk of re-identification until it is sufficiently remote that the information is 'effectively anonymised'.

Some information can present a need for caution (eg datasets that contain special category data). Anonymisation issues may also be more complex where you have large datasets that contain a wide range of personal data. You may therefore need specialist expertise and input beyond this guidance.

You should also be clear that other laws may still apply, even where effective anonymisation takes place. For example, certain aspects of the Privacy and Electronic Communications Regulations 2003 (PECR) apply to 'information', not just personal data (such as the provisions on terminal equipment, also known as the 'cookie law').

Further reading – ICO guidance

See the Guide to PECR's sections on traffic data, location data, and cookies and similar technologies.

Is anonymisation always necessary?

No. It is legitimate to use personal data for certain purposes. Indeed, the outcome you seek to achieve may mean it is necessary for you to process this data – eg, if you provide services to individuals or use their data to inform any decisions you make about them.

For example, much medical research involves access to patients' personal data, and is carried out on the basis of their involvement and agreement.

Data protection law provides a framework to enable this processing to be fair, lawful and transparent. However, if you don't need to use personal data to achieve your objectives, then in general you should seek to use anonymous information instead.

This is because anonymous information is no longer 'personal data' and is not subject to data protection requirements. Therefore, you may see it as desirable to achieve effective anonymisation of the data you hold.

Is anonymisation always possible?

In reality it can be difficult to determine whether the information you hold is personal data or anonymous information. This requires measured judgement based on the circumstances.

In some instances effective anonymisation may not be possible due to the nature or context of the data, or the purpose(s) for which you collect, use and retain it. For example, we recognise that some collections of personal data do not lend themselves well to anonymisation. Although the sensitivity of data will generally decrease with the passage of time, this is not always the case. The inappropriate release of records many decades old (eg criminal records), could still have a severely detrimental effect on an individual. That is why the security of data that cannot be anonymised is paramount.

What are the benefits of anonymisation?

Anonymisation limits your data protection risks, and can enable you to make information available to other organisations or to the public.

It also supports the principle of data minimisation. If you process personal data, you have to comply with the data protection principles and be able to demonstrate how you do so. The principles regulate the disclosure of personal data and establish a framework through which you can do this fairly, lawfully and transparently.

In general, it is easier to disclose anonymous information rather than personal data as fewer legal restrictions apply. It is also easier to use anonymous information in new and different ways, as the data protection rules on purpose limitation do not apply.

Implementing effective anonymisation can therefore help you to:

- better understand the legal requirements about the information you hold and intend to share or disclose;
- improve your decision-making and risk reduction and management processes;
- adopt a data protection by design approach;
- protect individuals' identities;
- reduce reputational risks caused by inappropriate or insecure disclosure or publication of personal data;
- reduce questions, complaints or disputes about your disclosure of information derived from personal data;
- develop greater confidence in publishing anonymous information in rich, re-useable formats; and
- navigate potentially challenging issues such as when handing FOI requests involving personal data.

Wider benefits of anonymisation include:

- developing greater public trust and confidence that data is being used for the public good, while privacy is protected (ie by ensuring legally required safeguards are in place and being complied with);
- greater transparency as a result of organisations being able to make anonymous information more widely available;
- incentivising researchers and others to use anonymous information instead of personal data, wherever this is possible;
- economic and societal benefits deriving from the availability of rich data sources; and
- improved public authority accountability through better availability of information about service outcomes and improvements.

If we anonymise personal data, does this count as processing?

Yes. For the purposes of data protection law, applying anonymisation techniques to turn personal data into anonymous information counts as processing.

Section 3(4) of the DPA 2018 defines processing as:

Quote

'…an operation or set of operations performed on information or on sets of information, such as—

a) collection, recording, structuring or storage,

b) adaptation or alteration,
c) retrieval, consultation or use,
d) disclosure by transmission, dissemination or otherwise making available,
e) alignment or combination, or
f) restriction, erasure or destruction'

This is relevant for all three regimes. For example, the definition in Article 4(2) of the UK GDPR is substantially the same, but contains the term 'personal data' instead of 'information' (the DPA 2018 defines 'personal data' as 'information' that relates to an identified or identifiable living individual).

Techniques and approaches that are designed to turn personal data into anonymous information constitute processing operations performed on that data. For example, when you create aggregate statistical information from personal data, that data is 'adapted' or 'altered' within the meaning of Section 3(4)(b) (or, for example, within the meaning of Article 4(2) of the UK GDPR where the general processing regime applies to you).

This means that you need to comply with data protection requirements for this processing. This includes ensuring you have a lawful basis for it and you clearly define your purpose(s).

In general it is likely that applying anonymisation techniques to the personal data you hold will be fair and lawful. However, it is still necessary for you to clearly define your purpose and detail the technical and organisational measures you intend to implement to achieve it.

Relevant provisions in the legislation

Section 3 of the DPA 2018 (external link)

Article 4(2) of the UK GDPR (external link) and the Keeling Schedule (external link)

What is pseudonymisation?

It is important to understand what data protection law means by the term 'pseudonymisation', and how this may differ from its use in particular circumstances, industries or sectors.

The DPA 2018 does not define pseudonymisation for the UK data protection regime as a whole. However, the general processing regime defines it at Article 4(5) of the UK GDPR as:

Quote

'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, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'

Pseudonymisation is therefore a technique that replaces or removes information that identifies an individual. For example, it may involve replacing names or other identifiers (which are easily attributed to individuals) with a reference number. This is similar to how the term 'de-identified' is used in other contexts, for example the removal or masking of direct identifiers within a dataset.

You must also ensure that you keep the additional information separately and appropriate technical and organisational controls are in place. This is so you can ensure that it is not possible to re-identify an individual from use of the separately held additional information, or indeed any other information.

This guidance uses the term 'pseudonymous data' to describe personal data that has undergone pseudonymisation in line with the above definition.

Relevant provisions in the legislation

Article 4(5) and Recital 26 of the UK GDPR (external link) and the Keeling Schedule (external link)

What about 'de-identified' personal data?

While the term 'de-identified' is widely used, its meaning may differ depending on the circumstances. For the purposes of data protection law, it is important to note that Section 171 of the DPA 2018 refers to 'de-identified personal data' in the context of the re-identification offence.

Section 171(1) states:

Quote

'It an offence for a person knowingly or recklessly to re-identify information that is de-identified personal data without the consent of the controller responsible for de-identifying the personal data'.

Section 171(2)(a) then says:

Quote

'personal data is "de-identified" if it has been processed in such a manner that it can no longer be attributed, without more, to a specific data subject.'

The DPA 2018's explanatory notes clarify that this provision:

Quote

'…reflects the definition of pseudonymisation in Article 4(5) of the GDPR.'

Therefore, for the purposes of the re-identification offence, the DPA 2018 refers to 'de-identified' personal data as personal data that has undergone pseudonymisation as defined in the UK GDPR rather than (for example) anonymous information.

Relevant provisions in the legislation

Section 171 of the DPA 2018 (external link)

DPA 2018 explanatory notes (external link) – while Explanatory Notes are not part of the law, they are intended to assist the reader in understanding the DPA 2018.

Article 4(5) of the UK GDPR (external link) and the Keeling Schedule (external link)

Further reading

We will discuss the re-identification offence and the concept of 'de-identified personal data' in more detail in future sections of this guidance. When we publish these sections, we will update this further reading box.

What is the difference between anonymisation and pseudonymisation?

Anonymisation means that individuals are not identifiable and cannot be re-identified by any means reasonably likely to be used (ie, the risk of re-identification is sufficiently remote). Anonymous information is not personal data and data protection law does not apply.

Pseudonymisation means that individuals are not identifiable from the dataset itself, but can be identified by referring to other information held

separately. Pseudonymous data is therefore still personal data and data protection law applies.

For example, Recital 26 of the UK GDPR makes it clear that personal data which has undergone pseudonymisation remains in scope of the law:

Quote

'…Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person…'

You should therefore exercise a level of caution. For example, it is common to refer to datasets as 'anonymised' when in fact they still contain personal data, just in pseudonymised form. This poses a clear risk that the requirements of UK data protection law may be disregarded in the mistaken belief that the processing does not involve personal data. If there are reasonably available means that could be used to re-identify individuals, then the data in question is not effectively anonymised.

However, it is also important to consider the processing's context. For example, whether a dataset that is pseudonymised from your perspective has the same status from the perspective of another organisation you share it with.

Example

An organisation applies a pseudonymisation technique that divides personal data into two parts – a dataset that by itself does not identify individuals, and 'additional information' such as a key that enables re-identification.

The organisation may refer to the first set as 'anonymous information'. This may indeed be the case in the hands of a third party that has no means reasonably likely to be used to re-identify individuals within that dataset.

However, if the first organisation continues to hold both the 'treated' data and the 'additional information', the data overall remains personal data in their hands. From its perspective, the data has undergone pseudonymisation (ie it is pseudonymised personal data).

Both parties need to carefully assess the status of the data in the hands of the second organisation (ie whether from their perspective they could regard it as anonymous information).

The key point is that if you apply a pseudonymisation technique, this does not necessarily change the status of the treated data from your perspective. The data you hold may still be personal data. Additionally, depending on the

circumstances, the data may still be personal data even when you disclose it to another organisation.

Ultimately, you should consider pseudonymisation techniques as ways you can reduce the risks your processing may pose to individuals (ie they may act as security measures).

Further reading

We will discuss the disclosure of pseudonymised datasets in more detail in future sections of this guidance. When we publish these sections, we will update this further reading box.

Further reading outside this guidance

See our detailed guidance on 'What is personal data?'.

What are the benefits of pseudonymisation?

An overarching benefit of pseudonymisation is that it can make your data protection compliance simpler in a number of areas.

The general processing regime in the UK GDPR provides a number of examples, such as:

- **general analysis** – Recital 29 of the UK GDPR incentivises you to adopt pseudonymisation not just as a security measure. This is because it enables you to undertake 'general analysis' of pseudonymised datasets that you hold, provided you put in place appropriate technical and organisational measures;
- **purpose limitation** – pseudonymisation is one of the factors you should take into account when deciding if further processing for a new purpose is compatible with your original purpose. This is also one of the important safeguards for processing personal data for scientific, historical and statistical purposes;
- **data protection by design** – pseudonymisation is one of the key ways in which you can implement appropriate safeguards for the personal data you process, both at the design stage and throughout any project lifecycle;
- **security** – pseudonymisation is referenced as one of the 'appropriate technical and organisational measures' in both the security principle and the specific provisions on security of processing;
- **personal data breach notifications** – pseudonymisation techniques can reduce the risk of harm to individuals that may arise from personal data breaches. This will assist you in assessing when you need to

notify individuals (both anonymisation and pseudonymisation techniques have application here); and

- **individual rights** – employing pseudonymisation techniques may reduce the amount of data you have to consider when responding to requests from individuals. For example, if your purposes for processing do not or no longer require identification of individuals, you are not required to process additional information in order to do so (or to comply with other requirements of data protection law). So, where you can demonstrate you are not in a position to identify individuals, the rights of access, rectification, erasure and data portability do not apply. However, you need to be able to respond to these requests if individuals provide you with additional information that enables their identification.

Relevant provisions in the legislation

See UK GDPR Articles 5, 6, 11, 25, 32, 33, 34 and 89 and Recital 29 (external link).