Chapter 3: pseudonymisation

Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance

February 2022

ico.
Information Commissioner's Office

Contents

Pseudonymisation

At a glance

- Pseudonymisation refers to techniques that replace, remove or transform information that identifies individuals, and keep that information separate.
- Data that has undergone pseudonymisation remains personal data and is in scope of data protection law.
- Pseudonymisation can bring many benefits. It can help you to:
    - reduce the risks your processing poses;
    - implement data protection by design and ensure appropriate security; and
    - make better use of data (eg for archiving, scientific and historical research, and statistical purposes; other compatible purposes; and general analysis).
- Take care not to confuse pseudonymisation with anonymisation. Ultimately, pseudonymisation is a way of reducing risk and improving security. It is not a way of transforming personal data to the extent the law no longer applies. Our chapter on identifiability provides further detail on assessing the risk of re-identification.
- However, you may be able to disclose a pseudonymised dataset (without the separate identifiers) on the basis that it is effectively anonymised from the recipient's perspective.
- The DPA 2018 contains two specific criminal offences to address the potential for harm resulting from unauthorised reversal of pseudonymisation. This applies to the reversal of pseudonymised data and any further processing of it, without first obtaining consent from the responsible controller.
- To use pseudonymisation effectively, you must:
    - define your goals;
    - detail your risks;
    - decide on the most appropriate technique; and
    - document the outcome.
- There are many pseudonymisation techniques. Some will help you achieve pseudonymisation as defined by the law. Others may not, but can still be useful technical measures from a security perspective. Ultimately, the appropriate technique to use depends on the circumstances of your processing.

In detail

What is pseudonymisation?

It is important to understand that pseudonymisation has a specific meaning in data protection law. This may differ from how the term is used in other circumstances, industries or sectors.

Article 4(5) of the UK GDPR defines pseudonymisation as:

Quote

"…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."

There is no equivalent definition in the law enforcement or intelligence services regimes in the DPA 2018 itself, but similar considerations apply.

At a basic level, pseudonymisation starts with a single input (the original data) and ends with two outputs (the pseudonymised dataset and the additional information). Together, these can reconstruct the original data. However, in relation to the individuals concerned, each output has meaning only in combination with the other.

Pseudonymisation therefore refers to techniques that replace, remove or transform information that identifies an individual. For example, replacing one or more identifiers which are easily attributed to individuals (such as names) with a pseudonym (such as a reference number).

While you can tie that pseudonym back to the individual if you have access to the additional information, your technical and organisational measures should ensure that you hold this information separately.

Data protection law specifically mentions "unauthorised reversal of pseudonymisation" as something that can result in harm. You need to assess

the likelihood and severity of this risk, and mitigate it appropriately. For example, if anyone with access to the pseudonymised data also has access to the additional information they could easily reverse the pseudonymisation process.

Relevant provisions in the legislation

See UK GDPR Article 4(5) and Recitals 26, 28 and 29 (external link)

Further reading – ICO guidance

What is personal data?

Is pseudonymised data still personal data?

Yes. Pseudonymisation can reduce the risks to individuals. It can also help you meet your data protection obligations, including data protection by design and security. However, it does not change the status of the data as personal data when you process it in this way.

This is because data protection law is clear that information is personal data if an individual is identified or identifiable, directly or indirectly. The general processing regime also makes it clear that data that has undergone pseudonymisation remains personal data.

For example, the core definition of pseudonymisation describes it as processing of personal data in a particular manner. Additionally, Recital 26 of the UK GDPR says that:

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…"

Ultimately, this means pseudonymisation is a type of processing applied to personal data, rather than a type of data that the law does not apply to (ie like anonymous information).

What is the difference between pseudonymisation and anonymisation?

There can be confusion between pseudonymisation and anonymisation. 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. For example, a mistaken belief that the processing does not involve personal data could mean that the requirements of UK data protection law are not met. This could result in potential adverse outcomes for individuals.

Data protection law says that:

- anonymous information is information that does not relate to an identified or identifiable individual (and the law does not apply to it); and
- data that has undergone pseudonymisation remains personal data.

It is crucial to understand this difference. With pseudonymisation, the processing reduces the links between individuals and the data that relates to them, but does not remove them entirely.

While individuals may not be identifiable from the pseudonymous data itself, they can be identified by referring to other information held separately. Both the dataset and the additional information are therefore still personal data.

Pseudonymisation has many uses. However, the protections it provides against identifiability risk are more limited and different than anonymisation.

Ultimately, you should consider pseudonymisation as a security and privacy risk management measure.

Does this change when disclosing this data to another organisation?

The status of data can change depending on who holds it. For example, pseudonymous data which you can still identify using a key or other separate identifiers might no longer be identifiable in the hands of a different organisation who does not have access to that key.

However, you cannot automatically assume that the pseudonymised data becomes anonymous information in the other party's hands. In practice, this depends on several factors you need to assess, including:

- the ability of the recipient to use other information to enable identification (either something in their possession, or in the public domain);
- the likelihood of identifiability, considering things like the cost of and time required for identification and the state of technology at the time of the processing; and
- the techniques and controls placed around the data once in the recipient's hands.

If you retain the pseudonymised data, it remains personal data in your hands. This is because you hold both the pseudonymised data and the additional information that allows you to identify the individuals the data relates to.

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 when you process it.

Relevant provisions in the legislation

See UK GDPR Articles 4(1) and 4(5), and Recitals 26, 28, 29 and 75 (external link)

See DPA 2018 Section 3 (external link)

Further reading – ICO guidance

What is personal data? – "Identifiers and related factors"

See the section of this guidance on "How do we ensure effective anonymisation?" for more information on identifiability.

The guidance on identifiers and related factors also discusses the considerations you need to take into account when disclosing data to other organisations, including the status it may have once in their hands.

What are the benefits of pseudonymisation?

When properly applied, pseudonymisation can help to:

- reduce the risk your processing poses to individual rights;
- enhance the security of the personal data you process;
- support re-use of personal data for new purposes;
- support your overall compliance with the data protection principles; and
- build individuals' trust and confidence in how you process their data.

Pseudonymisation can enable greater utility of data than anonymisation. However, you should still consider whether you can meet your objectives by using anonymous information.

As a controller, you are responsible for deciding whether and how to implement pseudonymisation. Defining the scope, parameters and objectives – as well as likely risk scenarios – is therefore very important.

There may be cases where pseudonymisation is effectively a requirement. For example, where it is needed to ensure your processing is proportionate to its purpose(s). In other cases, it can help you adopt a data protection by design approach from the design stage of any product, application or service you make.

Pseudonymisation can help you to reduce risk

Recital 28 of the UK GDPR says that:

Quote

"The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned."

There is no specific definition of risk in data protection law, but a number of provisions make clear that this is about the risks to the rights and freedoms of individuals. The key provision is Recital 75, which links risk to the concept of potential harm or damage to individuals:

Quote

"The risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from data processing which could lead to physical, material or non-material damage"

In this context, pseudonymisation can be relevant for any risk assessment you undertake. For example, detailing specific pseudonymisation techniques can help you demonstrate how you intend to mitigate particular risks that your processing may pose. This may be relevant when doing both data protection impact assessments (DPIAs) and legitimate interests assessments (LIAs).

There are several other parts of the law where pseudonymisation is relevant as a risk reduction measure. These include:

- data protection by design and security; and
- personal data breaches.

Data protection by design and security

The law requires you to put in place appropriate technical and organisational measures to:

- implement the data protection principles effectively and integrate necessary safeguards into the processing – this is "data protection by design"; and
- ensure a level of security appropriate to the risk the processing poses – this is "security of processing".

The law also specifically references pseudonymisation in these contexts. If you apply it properly, it can be a useful mechanism to enhance the security of personal data and support your overall compliance with the data

protection principles. It is particularly relevant in the context of the data minimisation principle. Pseudonymisation can limit the level of identifiability in the data to what is necessary for the purpose and reduces the level of personal data shared with other parties.

When considering pseudonymisation, for both data protection by design and security you need to take into account:

- the state of the art and costs of implementation of any measures;
- the nature, scope, context and purpose(s) of your processing; and
- the risks your processing poses to individuals' rights and freedoms.

In assessing state-of-the-art, you should consider whether the technique is suitably robust, ie, it is resistant to known attacks, scalable, and is not cost-prohibitive to implement. The nature, scope and purposes of the processing will influence the type of pseudonymisation technique you choose. For example, if you are required to track individuals over time, some techniques will not be suitable. You should consider whether your chosen approach is able to fulfil your purposes and also whether the measures you choose can reduce the risks to individuals to an acceptable level.

Not all pseudonymisation techniques are equally effective, and they may not have the same implementation costs or requirements. It is therefore important for you to examine pseudonymisation in the context of finding the optimal approach to achieving data protection by design and security.

Personal data breaches

Pseudonymisation techniques can reduce the risk of harm to individuals that may arise from personal data breaches. This can also form part of your assessment of the likelihood and severity of any impact of a personal data breach.

For example, pseudonymisation may be relevant when you assess whether you need to notify individuals of the personal data breach. Article 34 of the UK GDPR requires you to notify individuals without undue delay, unless you have:

Quote

"…implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the data protection breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption."

Although Article 34 does not specify pseudonymisation, it broadly describes technical and organisational measures. Pseudonymisation can form part of

those measures. It can provide another line of defence in the security context. For example, it may prevent security incidents becoming personal data breaches, even if you may still need to take action to address the incident itself.

Relevant provisions in the legislation

See UK GDPR Articles 25 and 32, and Recitals 28, 78 and 83 (external link)

See DPA 2018 sections 57, 66, 103 and 107 (external link)

Further reading – ICO guidance

DPIAs

Legitimate interests

Data protection by design and by default

Security

Personal data breaches

Data minimisation

An upcoming chapter will provide further guidance on a range of technical solutions that could be used to achieve effective pseudonymisation of personal data.

Pseudonymisation can help you process data for other purposes

Data protection law may allow repurposing of personal data for some types of processing, if appropriate safeguards such as pseudonymisation are in place. For example, for research, further analysis or compatible purposes. This means that pseudonymisation can be a useful tool to enable further processing of personal data beyond its original purpose.

This does not mean pseudonymisation automatically allows you to undertake this further processing in all cases. However, it can be an important way for you to demonstrate how you protect personal data, if you do so. It is therefore a factor whenever you are considering whether you can process personal data for other purposes.

If your purposes change over time or you want to use data for a new purpose which you did not originally anticipate, you can only go ahead if:

- the new purpose is compatible with the original purpose;
- you get an individual's specific consent for the new purpose; or
- you can point to a clear legal provision requiring or allowing the new processing in the public interest.

Pseudonymisation is relevant when you are considering compatibility instead of consent or legal provisions. For example, where you:

- undertake further processing for archiving, scientific or historical research, and statistical purposes, which are automatically considered to be compatible purposes; or
- want to undertake further processing for other purposes, and need to assess whether these are compatible with your initial purpose.

Pseudonymisation also allows you to perform general analysis. This activity may be something that you do as part of the further processing.

Archiving, scientific or historical research, and statistical purposes

The purpose limitation principle specifically says that the following are "compatible" purposes for further processing:

- archiving purposes in the public interest;
- scientific or historical research purposes; and
- statistical purposes.

It is important to note that this compatibility depends on the further processing being:

- necessary for those purposes; and
- carried out in accordance with Article 89(1) of the UK GDPR.

Article 89(1) requires that processing for these purposes:

Quote

"…shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject."

The safeguards you implement must ensure that technical and organisational measures are in place, in particular to comply with the data minimisation principle. Article 89(1) says these measures:

Quote

"…may also include pseudonymisation provided that those purposes can be fulfilled in that manner."

The DPA 2018 is also relevant. Section 19 says that the processing is not regarded as subject to appropriate safeguards where it is:

- likely to cause substantial damage or substantial distress to an individual; or
- carried out for the purposes of measures or decisions with respect to an individual, unless these purposes include approved medical research.

Other compatible purposes

Article 6(4) of the UK GDPR says that when deciding if a new purpose is compatible with your original purpose you should take into account:

- any link between your original purpose and the new purpose;
- the context in which you originally collected the personal data. In particular, your relationship with the individual and what they would reasonably expect;
- the nature of the personal data (eg whether it is special category data, or data relating to criminal convictions or offences);
- the possible consequences for individuals of the new processing; and
- whether there are appropriate safeguards (eg encryption or pseudonymisation).

Pseudonymisation does not necessarily mean that you can decide your new purpose is compatible in all cases. It is one of several factors you must consider in this assessment.

If your new purpose is compatible, you don't need a new lawful basis for the further processing. However, you should remember that if you originally collected the data on the basis of consent, you usually need to get fresh consent to ensure your new processing is fair and lawful.

You also need to update your privacy information so that your processing is still transparent.

General analysis

Recital 29 of the UK GDPR says that

Quote

"In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately. The controller processing the personal data should indicate the authorised persons within the same controller."

This means you can perform this general analysis on pseudonymised data within your organisation, provided that you:

- implement technical and organisational measures necessary to ensure data protection compliance; and
- ensure that additional information for attributing the data to a specific individual is kept separately.

Data protection law does not define general analysis. However, if you intend to analyse data relating to specific individuals (eg their behaviour, location, characteristics etc) for the purposes of taking actions about them, this analysis is not general in nature.

In practice, general analysis may be something you undertake in the context of the two purposes detailed above. It can bring many benefits, depending on the purposes you do it for. For example, pseudonymising data about how individuals use your products and services, and then deriving insights and trends from that data. This may allow you to develop new, innovative services or improve existing ones.

This is particularly the case if anonymous information is less useful. However, you should carefully consider whether you can achieve these objectives using such information first.

When you perform general analysis, you need to indicate the authorised persons within your organisation that have access to the additional information. You should also update this, if you make any personnel changes in the future.

Relevant provisions in the legislation

See UK GDPR Articles 5(1)(b), 6(4), 89(1) and Recitals 29 and 156 (external link)

See DPA 2018 Section 19 (external link)

Further reading – ICO guidance

Purpose limitation

Lawful basis for processing

We will produce separate guidance on the research provisions of data protection law soon. We will follow that guidance with more detail on the role of both anonymisation and pseudonymisation in the research context.

Are there any offences relating to pseudonymisation?

Yes. Section 171 of the DPA 2018 specifies two criminal offences relating to the re-identification of de-identified personal data. These are known as the "re-identification offences". They include data that has undergone pseudonymisation.

The first, at Section 171(1), is about the act of re-identification. It states that:

Quote

"It is 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."

The second, at Section 171(5), is about the processing of the personal data after the act of re-identification takes place. It states that:

Quote

"It is an offence for a person knowingly or recklessly to process personal data that is information that has been re-identified where the person does so –

(a) without the consent of the controller responsible for de-identifying the personal data, and

(b) in circumstances in which the re-identification was an offence under subsection (1)."

Section 171(2) also specifies that:

Quote

"(a) 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;

(b) a person 're-identifies' information if the person takes steps which result in the information no longer being de-identified within the meaning of paragraph (a)"

The DPA 2018's explanatory notes add that Section 171(2):

Quote

"…defines the meaning of 'de-identification' and 're-identification' for the

purposes of the offence and reflects the definition of pseudonymisation in Article 4(5) of the GDPR".

This demonstrates how the re-identification offences include data that has undergone pseudonymisation. In this context, they address one of the key risks to rights and freedoms that data protection law requires you to mitigate – the unauthorised reversal of pseudonymisation, and the harms that may arise from it.

Are there any defences?

Yes. Section 171 includes a series of defences. Generally, these are similar to those found in Section 170 for the unlawful obtaining of personal data.

For Sections 171(1) and 171(5), defences include where the re-identification is:

- necessary for the purposes of preventing or detecting crime;
- required or authorised by law or by a court order; or
- in the circumstances, justified as being in the public interest.

The defences also include where the person charged can prove that they acted:

- in the reasonable belief that they were the individual to whom the information relates, had the consent of that individual, or would have had such consent if the individual had known about the re-identification and its circumstances;
- in the reasonable belief that they are the organisation responsible for the de-identification, had the consent of that organisation, or would have had such consent if the organisation had known about the re-identification and its circumstances; and
- for the "special purposes", with a view to publication of any journalistic, academic, artistic or literary material, and in the reasonable belief that the re-identification was justified as being in the public interest in the circumstances.

What is the "effectiveness testing" defence?

Section 172 also contains a specific effectiveness testing defence. This has particular relevance to security and technology researchers acting in the public interest. For example, to identify poor practices, or potential flaws in certain techniques or their implementation.

The defence applies where the following two conditions are met:

- Condition 1: The person was testing the effectiveness of the de-identification systems an organisation uses, where they reasonably

believe the testing is justified as being in the public interest and where they do not intend to cause or threaten damage or distress; and

- Condition 2: The person notifies either the ICO or the organisation responsible for the de-identification about the re-identification.

For Condition 2 to apply, the notification must happen without undue delay and, where feasible, not later than 72 hours after becoming aware of it. Where there is more than one organisation responsible for de-identification, Condition 2 is met if one or more of those organisations is notified.

The defence does not legitimise unlawful or harmful practices. For example, it does not make the actions of malicious actors lawful.

Relevant provisions in the legislation

DPA 2018 Sections 171 and 172 (external link)

UK GDPR Article 4(5) and Recitals 26 and 75 (external link)

DPA 2018 explanatory notes paragraphs 492 to 504 (external link).

Although the explanatory notes are not part of the law, they explain what each provision of the DPA 2018 means in practice and provide background information on the intended outcomes of the provisions.

Further reading – ICO guidance

Exemptions – the "special purposes"

How should we approach pseudonymisation?

As a controller, you are responsible for deciding whether and how to implement pseudonymisation. It is important you clearly establish what you seek to achieve and the most appropriate technique in that context. An inadequate level of pseudonymisation does not meet the requirements of the law, even if the technique you use may fit under existing technical meanings of the term.

Your core considerations should include:

- defining the goals (eg what does your use of pseudonymisation intend to achieve?)
- detailing the risks (eg, what types of attack are possible, who may attempt them, and what measures do you need to implement as a result?)
- deciding on the technique (ie, which technique (or set of techniques) is most appropriate?)

- deciding who does the pseudonymisation (eg you, a processor); and
- documenting your decisions and risk assessments.

While this is not an exhaustive list of relevant considerations, you should address them together due to how they relate to each other.

Define your goals

Overarching goals of pseudonymisation can include:

- ensuring that parties other than yourself (and any processor you may use) cannot easily re-identify individuals in the context of the processing;
- ensuring that pseudonyms cannot easily be reproduced by those parties;
- enabling data accuracy (eg by assigning a particular pseudonym to an individual that allows you to verify their identity); and
- achieving data minimisation (eg if the purposes of your processing do not require you to identify an individual).

One of your overarching goals should also be that once you do implement pseudonymisation, you mitigate any risk of unauthorised reversal of it. This should include any potential source (ie a malicious attacker or an insider threat).

Usability and scalability are also relevant factors for you to consider, both in terms of what you intend to achieve (the goals) and how you go about doing so (the technique).

Detail the risks

When assessing what pseudonymisation techniques to use and how to implement them you should take into account the type of attacker that may exist. For example, it is good practice to consider:

- insider threats – someone with specific knowledge, capabilities or permissions, either in your organisation, a processor you use, or another entity you engage (eg a trusted third party);
- external threats – someone who may not have direct access to the additional information, but wants to increase their knowledge about the pseudonymised dataset (eg by re-identifying the individuals within the dataset); and
- the likely goals of any attack – an attacker may have different goals they seek to achieve (eg re-identification attacks, where the attacker seeks to re-identify individuals (either a subset, or all of them)).

Further reading – ICO guidance

We discuss the methodology to assess the risk of singling out an individual in Chapter 2: How do we ensure anonymisation is effective?.

We will discuss the strengths and weaknesses covering a range of pseudonymisation schemes in future sections of this guidance.

When we publish these sections, we will update this further reading box.

Decide on the technique

You should carefully consider the implementation of pseudonymisation following a risk-based approach, taking into account:

- the nature, scope, context and purpose of the processing;
- the risk factors you identify; and
- the privacy protection, utility and scalability goals your processing requires. For example, assessing whether the technique achieves the level of pseudonymisation required by data protection law. Or, where it does not, the advantages it may still have in terms of security and data minimisation.

Your decision-making process should explore the availability of existing solutions to meet your goals, together with their strengths and limitations. You should choose the appropriate technique after considering:

- the risk of re-identification for the part of the data that will be transformed by the pseudonymisation technique; and
- the required utility and accuracy of the data for the purposes of the processing.

You should also ensure that you have appropriate processes in place for regularly testing, assessing and evaluating the effectiveness of the pseudonymisation techniques you use.

Further reading – ICO guidance

See our guidance on passwords in online services in the Guide to the UK GDPR for more information on appropriate hash functions in that context.

Our upcoming chapter on technical solutions will explore the pros and cons of various pseudonymisation schemes in more depth, including risk mitigation measures and approaches to choose an appropriate technique.

When we publish these sections, we will update this further reading box.

Further reading outside this guidance

A number of publications from the European Union Agency for Cybersecurity (ENISA) provide more details about pseudonymisation techniques, including additional risks that you may need to consider.

These publications include:

- "Recommendations on shaping technology according to GDPR provisions – an overview on data pseudonymisation" (2019) (external link)
- "Pseudonymisation techniques and best practices" (2019) (external link)
- "Data pseudonymisation: advanced techniques and use cases" (2021) (external link)

Decide who performs the pseudonymisation

Different parties may be involved in any pseudonymisation process. There is no one-size-fits-all approach to this. It is ultimately a decision for you to take based on your specific circumstances.

For example, at a simple level, pseudonymisation may be performed by:

- you (eg if you apply pseudonymisation techniques yourself);
- a processor working on your behalf (eg if they have specialist expertise and resources to help you achieve your goals); or
- another organisation, who may be a processor or a joint controller depending on the circumstances.

More complex scenarios can involve many different organisations at different levels of the processing.

Ultimately, the important factor is that you identify roles and responsibilities of each participant, particularly which entity performs the pseudonymisation and who has overall accountability for the processing.

You should also consider a clear separation of functions. For example, clearing specifying and distinguishing between the individuals that:

- carry out the pseudonymisation processes;
- are authorised to access the additional information; and
- undertake any subsequent processing of the pseudonymised data (eg if you are performing general analysis for certain purposes).

With functional separation, you are still processing personal data but have taken steps to ensure that your implementation of pseudonymisation has operated to protect individual rights.

Further reading – ICO guidance

Controllers and processors

Contracts and liabilities

Document the outcome

The above considerations can form part of any DPIA you perform. They can also be relevant in the context of any LIA. The important point is that you clearly document the decision-making process and detail the steps you take.

You should combine your use of pseudonymisation with a thorough security risk assessment for the protection of personal data.

Finally, you should also monitor the state of the art and ensure that the methods you use are appropriate.

Further reading – ICO guidance

Accountability framework

Guide to the UK GDPR – accountability and governance