A guide to the **ANONYMISATION OF PERSONAL DATA**

2015

Datatilsynet

Contents

Introduction

This guide is intended to help organisations in their efforts to anonymise the personal data they have collected in a robust and secure manner. Anonymisation is about removing the possibility of identifying individuals in a data set. Anonymisation is an important means of enabling the extraction of valuable insights through data analysis, while reducing the risks for those concerned. When personal data are anonymised, they are no longer deemed to constitute personal data. The processing of such data therefore falls outside the scope of the Data Processing Act.

Why is this guide necessary?

Anonymising data is challenging – and it is more challenging today than it was before. The vast reservoir of publicly accessible data, combined with the availability of ever cheaper and more powerful analysis technology, has increased the risk of re-identification.

The increased risk of re-identification makes it even more important to perform thorough risk assessments before publishing anonymised data, and to use robust anonymisation techniques. In this guide, we will review key legal provisions, point out risk factors it is important to take into consideration, and discuss the strengths and weaknesses of various different anonymisation techniques.

Who is the guide intended for?

This guide is intended for all those wishing to anonymise personal data in the public and private sectors. It applies irrespective of the purpose of such anonymisation. There may be many reasons why an organisation wishes to anonymise the personal data it has collected. It may, for example:

- have been ordered to publish data in anonymised form.
- be obliged to disclose information to a third party and wish to protect the identities of those concerned.
- wish to publish data in order to be open and transparent about its own operations.
- wish to use already collected data for new purposes, such as building personas in connection with target marketing, or to identify trends and patterns.
- wish to release data for statistical analysis or scientific purposes

In brief

- Privacy legislation does not apply to anonymous data. Data is anonymous if it is no longer possible, with the tools that can reasonably be expected to be used, to identify individuals in a data set.
- The anonymisation of data makes it possible to exploit the value inherent in data analysis in a way not injurious to privacy.
- This guide will help organisations wishing to anonymise personal data, irrespective of their reason for doing so.
- This guide will help organisations to identify the challenges and risks associated with the anonymisation of personal data, in order for the outcome to be as secure as possible.
- This guide provides an introduction to key aspects of the Personal Data Act.

Anonymisation and personal data

When any person or organisation domiciled in Norway processes digital data, they must take into consideration that such processing may trigger obligations and rights under to the Norwegian Personal Data Act. Anyone processing the data must comply with the Act's provisions, or risk incurring a financial penalty, civil liability or even criminal liability.

However, this presumes that the data being processed are *personal data*, since the Personal Data Act applies only to data that relate to specific individuals. The limits of what may be defined as personal data are therefore crucial to the law's applicability.

More simply put, the processing of personal data is covered by the Act, while the processing of anonymous or anonymised data is not. The same applies to data that is not linked to individuals at all. It is the distinction between personal data and anonymous data that is the topic for this guide.

Personal data and anonymous data

It can be tricky to decide where the line between personal data and anonymous data should be drawn. The starting point for any such assessment is the legislation's definition of the term "personal data"

Personal data comprise information and assessments that may be linked to an individual (natural) person (Section 2 of the Personal Data Act).

This definition has three main components:

1. It can relate to any form of information.
2. The phrase *that may be linked to* is the bridge between 1 and 3.
3. An identifiable or identified natural person.

We believe a certain understanding of what is deemed to constitute personal data is necessary in order to understand anonymisation. For a more thorough analysis of the personal data concept, see section 4.2 of the report Big Data – principles of personal data under pressure (2013, pdf), available from datatilsynet.no.

Here follows a brief discussion of the three elements:

1. Any form of information

All types of information are encompassed by the definition. Firstly, it means *objective* information, such as a person's age, address or annual income. Secondly, it can include *subjective* impressions, such as a person's assessments or characterisation of another individual. The veracity of the information is unimportant. It is an item of personal data irrespective of whether it is an assertion, verifiable fact or pure invention.

Nor is the term "personal data" restricted to matters traditionally associated with an individual's private life. Other, more prosaic matters, such as where one works or what one is studying, also fall within the definition of personal data.

The question of how worthy of protection the information is only arises at a later point in time, often in connection with an assessment of whether the way in which the data are processed complies with the law or not.

Nor is the format in which the data are held of any significance. Personal data can be expressed verbally, numerically, in drawings, photos, sound or as biometric characteristics. Furthermore, the data may be found in emails, in public case documents, on social media, in apps, text messages, online, etc.

2. The linking element

It must be possible to link the information to a physical (natural) person. Sometimes, this link is easy to recognise, sometimes not. For example, information on the condition of a vehicle will probably be associated primarily with the object itself. Nevertheless, that same information could also reveal matters relating to people who have had to do with the object, such as the vehicle's owners. In certain circumstances, information on one person may, at the same time, constitute information about one or more others. This could be the case, for example, in a medical or genetic context.

Thus, the link between the information and the person may also be *indirect*. Such an indirect link is sufficient for the Act to be applicable. This follows directly from the wording of Section 2 of the Personal Data Act.

3. Identifiable natural person

The information must also be linked to an individual (natural) person, and that person must be *identifiable*. That a person has been *identified* means that he or she has been distinguished from a group of people. That the person is identifi*able* means that such identification is *possible*. That such identification could feasibly occur at some point in the future is sufficient.

Information may, at first glance, appear to be anonymous, but nevertheless constitute personal data in the eyes of the law. This is because it may be possible to identify one or more people indirectly. Examples include a vehicle's registration number or a smart phone's IMEI number. These data can, in certain cases, be linked to other data sets or other databases, thereby revealing the identity of the vehicle or phone's owner. Other information that appears together with such numbers, such as where the

vehicle or phone has been, will therefore also be considered to be personal data.

Anonymous data

In the above, we have attempted to explain the meaning of the term "personal data". It is important to have a certain understanding of what personal data are in order to be able to determine what is required for an item of personal data to be deemed anonymised.

As previously mentioned, the Personal Data Act has no provisions with respect to the processing of anonymous or anonymised data, and the correct identification of where the line is drawn can therefore be of great significance.

Data of the type defined in point 1 above can be said to be anonymous when it is not possible to find any such linking element as stated in point 2, or the individual in point 3 is not identifiable.

Anonymous data can be defined as data that are impossible to link to an identifiable individual, taking account of all the means that may reasonably be envisaged used to identify the person concerned, either by the data controller or any other third party (See recital 26 of the EU Data Protection Directive's preamble.)

Definitions

Anonymisation is the act of rendering personal data anonymous.

Pseudonymisation is the replacement of directly identifiable parameters with pseudonyms, which will still constitute unique identifying indicators.

De-identification is the removal of all uniquely personal characteristics from the data, so that they can no longer be linked to a specific individual.

Anonymisation

Anonymisation is the act of rendering personal data anonymous. In other words, data sets that can be linked to an identifiable person are prepared in such a way as to make it impossible to link the data to a specific person. Several techniques can be used to achieve this aim. The various techniques' strengths and weaknesses are described in the appendix to this guide (see page 16).

When the anonymising process is finished, it is important to realise that true anonymisation has been achieved only if the process is *irreversible*. In other words, it must not be possible to re-establish the link between the data and the specific individual, taking account of the means which may reasonably be envisaged used to identify the person concerned, as mentioned earlier.

Determining whether the data make it possible to identify a person or whether the data may be considered anonymous or not depends on the actual circumstances. The assessment must rest on the likelihood of re-identification. Each individual case must be assessed and analysed not only on the basis of the means available today, but also with an eye on tomorrow's technology – within reasonable limits, naturally. The benchmark is the extent to which such means can be envisaged used to discover the identities of the people concerned.

Sometime, anonymisation is confused with two similar phenomena, *pseudonymisation* and *de-identification*. Such confusion may be unfortunate. At worst, it could result in the commission of a criminal offence, with all the consequences that could entail.

Pseudonymisation

Pseudonymisation is the replacement of directly identifiable parameters with pseudonyms, which will still constitute unique identifying indicators. A likelihood therefore exists that the specific individual may be indirectly identified. Indeed, it is often the point that the same (pseudonymised) person can be tracked over a certain period of time, in connection with research studies, for example. We therefore find ourselves within the scope of the Personal Data Act's definition of personal data, with the consequence that the Act's provisions must be respected.

In other words, it is extremely important to be aware of this distinction, since pseudonymised data are subject to the provisions of the Personal Data Act, while the opposite is the case with respect to anonymous data.

However, this does not mean that pseudonymisation is without merit. Pseudonymisation can make it more difficult to link a specific data set to the data subject's identity. It can therefore be seen as a useful technique for promoting privacy. Pseudonymisation may protect the individual to which the data are linked, and it may be easier to justify the processing of such pseudonymised data in relation to one or more of the lawful grounds provided in the Act.

The terms we use in this guide are based on shared European assessments (see, for example, the Article 29 Working Party's opinion/recommendations on anonymisation techniques (pdf). They may deviate from the way in which such terms are understood in Norway,

particularly in the health sector. In the Personal Health Data Filing Systems Act, which came into force on 1 January 2015, the definitions of pseudonymised and de-identified health data were replaced by the broader term "indirectly identifiable health data". (Further information on the term pseudonymisation as it was understood prior to the new legislation, can be found in Circular I-8/2005 (regjeringen.no, pdf). This also applies to the legal sense of the term "encryption", which deviates from how the term is used in this guide, where it denotes a technique).

Advantages of anonymised data

If you have a sufficiently robust and securely anonymised data set, you can make use of the information without any risk of contravening the Personal Data Act. You do not need to take account of the duties applying to the *data controller*, and further use and analysis of this type of data is not subject to any notification or licensing requirement.

Nor do you need to make sure that there are lawful grounds for processing the data, or comply with requirements relating to relevance or purpose. Furthermore, the data holder has no obligation to delete the data, etc.

Anonymisation could be the solution in cases where there are doubts about whether the law permits personal data to be processed in a certain way. In cases where the law specifically precludes the processing of personal data, the answer could be to render the data anonymous, since anonymous data fall outside the scope of the Personal Data Act.

Anonymisation and the concept of data processing

It is also a prerequisite that the data to be rendered anonymous have been collected and processed in accordance with the Personal Data Act's provisions. In theory, the very act of anonymisation must be deemed to constitute the processing of personal data. In consequence, therefore, anyone undertaking the anonymisation of the data must respect the requirements set out in Section 11 of the Personal Data Act during the anonymisation process. (See section 2.2.1 in the Article 29 Working Party's opinion/recommendations on anonymisation techniques (pdf) for further details.) The restrictions relating to purpose set out in Section 11(1)(c) must, for example, be respected.

Grounds for anonymisation will probably often be found in the so-called balancing of interests stipulated in Section 8(f) of the Personal Data Act. The provision states that personal data may be processed only if the processing thereof enables the data controller, or third parties to whom the data are disclosed, to protect a legitimate interest, *except* where such interest is overridden by the interests of the data subject.

In other words, a legitimate interest must exist, and the Act's stipulation of necessity must have been met. A key issue, however, is that the data subject's privacy can be said to have been infringed to only a minor degree by the anonymisation of data that can be linked to him or her. This will naturally play an important role in the balancing of each party's interests.

If anonymisation is deemed to constitute the processing of data in the legal sense, it is clear that anonymisation cannot "repair" a lack of lawfulness or legitimacy in the original data collection. In other words, it is not possible to first collect personal data in contravention of the law and then render it anonymous.

Challenges linked to anonymisation

Risk of re-identification

Access to a vast pool of publicly accessible data, combined with the availability of ever more powerful analysis technology, has increased the risk of re-identification. Re-identification means that it is possible to identify specific individuals from what, at the outset, is presumed to be an anonymous data set. Studies have shown that by collating data from several sources, it is possible to re-identify a person from only two known attributes in an anonymised data set, such as postcode and birthdate.

Re-identification can occur when someone takes personal data they already have on an individual, and searches for hits in an anonymised data set, or when a hit in an anonymised data set is used to search for hits in publicly accessible data. Examples of such publicly accessible data include data from public records (eg tax lists, the Brønnøysund Register Centre, the vehicle registration database, the electronic public records system (OEP)), social media, local and national media archives and genealogy websites.

There have been several known cases of re-identification, the majority performed by researchers on real data sets. In the majority of these cases the data had been poorly anonymised to begin with. For example, the organisations had retained too many identifying elements in the data set, or, in connection with the publication of the data, had not analysed which other accessible data sets existed "out there" that could be used to deduce information from their own data set.

Example: Netflix

A well-known example of the publication of poorly anonymised data was when Netflix announced a competition for developers. The prize was USD 1 million. The aim was for someone to develop a solution that improved their recommendation module by 10 per cent.

Netflix made a "trial data set" available to the competing developers, which they could use to train their systems. A disclaimer was issued along with this data set: "To protect our customers' privacy, all personal data that identifies individual customers has been removed, and all customer IDs have been replaced by randomly generated ID nos."

There are several online film-rating portals, including IMDb. Individuals can register with IMDb and rank films under their full name.

Researchers Narayanan and Shmatikov ran Netflix's de-identified trial database against IMDb's database (based on the date of a user's assessment post) and managed in this way to partially re-identify customers included in Netflix's trial database.

Anonymisation in brief

- Personal data are made up of information and assessments that can be linked to a specific person.
- It can be difficult to draw the line between personal data and anonymous data. It is therefore important to have some understanding of what personal data are.
- Anonymous data fall outside the scope of the Personal Data Act. For this Act not to apply, it is crucial that the anonymisation of data is real. In other words, it must be impossible to recreate any link between the data and the individual concerned, taking into account the means that may reasonably be envisaged used.
- The advantage of anonymisation is that the further processing of the data can take place without incurring any form of processing liability.
- Anonymisation will not always be necessary. In many cases, the data will be processed in accordance with one or more of the lawful grounds provided in the Personal Data Act.

Example: AOL

The AOL case is a typical example of mistaken pseudonymisation.

In 2006, a database containing 20 million search words for over 650,000 users was published. The only thing AOL had done to protect the privacy of its users was to replace their AOL username with a serial number. This resulted in some of the people being identified and localised.

Pseudonymised search strings for search engines have an extremely high identification rate, particularly if they are linked to other attributes like IP addresses or other client configuration parameters.

Certain types of data are more difficult to render anonymous than others. This applies to localisation and genetic data, for example. People's patterns of movement are so unique that the semantic part of the localisation data – the places where the data subject has been at a certain point in time – can reveal a lot about the data subject, even without other known attribute values. This has been demonstrated in many representative academic studies (de Montjoye et al: «Unique in the Crowd: The privacy bounds of human mobility», *Nature*, 3, 1376 (2013)).

Genetic data profiles are another example of personal data that risks being re-identified if the only anonymisation technique used is to remove the data subject's identity. This is because the genes are inherently unique. Studies have shown that the combination of publicly available genetic resources (such as genealogy databases, obituaries, search engine results) and metadata about DNA donors (donation time, age, address) can reveal certain people's identity, even though they have provided the DNA sample "anonymously".

Pseudonymised data is not the same as anonymous data

Pseudonymised data is not synonymous with anonymised data, as pointed out in Chapter 2. Pseudonymisation does allow individuals to be identified. Data controllers who choose to pseudonymise data, rather than anonymise them, must be aware that the data will still be defined as personal data, and must therefore be processed in accordance with the Personal Data Act's provisions.

There are several examples of data controllers believing that they have anonymised data, while – in realty – they have merely pseudonymised them by, for example, replacing the subject's name with a serial number.

Challenges in brief

- The danger of re-identification: When data from several sources are compared against each other, there is a risk that individuals in what are nominally anonymous data sets may be identified.
- Two known data points, e.g. postcode and birthdate, could be enough to re-identify individuals from a data set.
- Pseudonymous data must not be confused with anonymous data. Pseudonymous data are personal data, and their processing falls within the scope of the Personal Data Act.
- Encryption is not the same as anonymisation. The objective of encryption is to protect the data, not make them unidentifiable.

Encryption is not anonymisation

Another widely held misconception is that encrypted or single-coded (hashed) data is the same as anonymised data. This misconception rests on the following two assumptions: a) that when an element in a database (eg name, address, birthdate) has been encrypted or replaced by a randomised code with the help of encryption technology (eg a hash function with a key), this element has been anonymised; and b) that anonymisation is more effective if the key's length is correct and an advanced encryption algorithm has been used.

It is important to understand that the objectives of encryption and anonymisation are different. Encryption is a security method intended to protect confidentiality in a communication channel between two identified parties (people, entities or software programs) to prevent eavesdropping or unintended publication. The purpose of

anonymisation, on the other hand, is to avoid individuals being identified by preventing hidden connections between attributes linked to the subject of the data.

Neither encryption nor key coding as such helps to make the data subject unidentifiable, since the original data are still accessible or can, at the very least, be deduced by the data controller. Being able to perform a semantic translation of the personal data, as in the case of key coding, does not eliminate the possibility of recreating the data in their original structure.

Advanced encryption can help to provide better data protection by making them incomprehensible to parties that do not have access to the encryption key, but this does not necessarily render the data anonymous. As long as the key to the original data is accessible (even though it is kept by a trusted third party), the possibility of the data subject being identified is not eliminated. Encrypted and single-coded data must therefore be deemed to constitute personal data, and must be treated as such.

Encryption vs anonymisation

Encryption is used to protect confidentiality in a communication channel, and has an entirely different purpose to anonymisation. The purpose of anonymisation is to avoid individuals being identified.

(This chapter is based partly on the Article 29 Working Party's opinion/recommendation on anonymisation techniques (pdf))

Recommendations for secure anonymisation

Should data be anonymised, pseudonymised or left identifiable?

At an early stage, the data controller must decide whether the personal data to be processed should be anonymised, de-identified or left identifiable. This choice will affect what the organisation must do in relation to the Personal Data Act when it processes the data concerned. If anonymisation is chosen, any further processing will, as previously mentioned, fall outside the scope of the Personal Data Act.

Describe the purpose of anonymisation

It is important to be clear about the purpose of anonymisation. The way in which the data set will be used plays a crucial role in determining the risk of re-identification. If the data are to be published online, this represents a bigger risk than restricted release of the data for research purposes or for use in one's own organisation, for example. Even though restricted release is easier to control and assess, this too is not without risk.

Several factors must be taken into account when data are to be anonymised. For example:

- What type of data is to be anonymised?
- What control mechanisms are linked to the data set? What security precautions will limit access to the data?
- How big is the data set? (What quantitative properties does it have?)
- Which other publicly accessible information resources exist, which could potentially be used to deduce information about individuals in a data set that is to be anonymised?
- Will the data set be released to third parties? (If so, will access be limited, or will it be made publicly accessible online, for example?)
- Might unauthorised third parties wish to stage a targeted attack on the data in an attempt to identify individual people? (In this kind of risk assessment, the data's sensitivity and type is particularly important.)

Perform a risk assessment both before and after publication

It is almost impossible to assess the risk of re-identification with absolute certainty. To obtain an absolute overview of all the other information that is "*out there*", who it is accessible to and how it may be used in a re-identification attempt, is extremely challenging. Such "other information" could be information that is accessible to another organisation or particular person, or it may be generally available online.

Since you can never be entirely certain about which data are available at any given time, or which data will be made accessible at some point in the future, it is necessary to perform as thorough a risk assessment as possible, early in the anonymisation process.

Furthermore, it is important that the data controller also remembers to consider the risks surrounding a data set *after* it has been published. Due to the risk of re-identification, the data controller should *regularly* investigate whether *new* risk factors have cropped up, and *re-evaluate* the risk factors already identified. It is also important to assess whether *control* of the identified risk factors is adequate or needs to be adjusted.

If someone should succeed in re-identifying the data, and this results in personal data being processed, the organisation responsible for the data must assume the role of data controller for them, in accordance with the Personal Data Act.

Use of the re-identification test

One test that can be useful to perform to examine the risk of re-identification, is what is known as the "motivated intruder test". This involves testing whether the data can be re-identified *if* an intruder should attempt to do so.

The motivated intruder must be considered to be a person/organisation which, with no prior knowledge, attempts to identify individuals in an anonymised data set. Even though the intruder has no prior knowledge, they can be considered as sufficiently competent. In other words, they have access to resources such as the internet, public databases and libraries. It must also be presumed that they will be able to use various investigative techniques to communicate with people who know the identities of individuals in the data set. However, the motivated intruder should not be deemed to have specialist knowledge, such as expertise in data hacking, or access to specialist equipment to force access to data that is securely stored.

Some types of data are obviously more attractive for a motivated intruder than others. There may be many reasons why someone would attempt to re-identify individuals in an anonymised data set. These include:

- Disclosure of newsworthy information about public figures.
- Political or social activism, e.g. as part of a campaign against a particular organisation or person.
- Inquisitiveness – a desire to find out who is involved in a local planning application, for example, or simply to see whether it is possible to deduce information linked to an individual from the data set.

However, this does not mean that data which, at first glance, appears to be "ordinary" "innocuous" or without value can be published without a careful assessment of the threat associated with re-identification.

The motivated intruder test is useful because it sets a threshold for the risk of re-identification higher than that at which a normally inexpert citizen could manage to re-identify the data, but lower than that at which an expert with access to specialist competence, analytical resources and prior knowledge could do so. It is good practice to perform a motivated intruder test as part of any risk assessment. In practice, the performance of a motivated intruder test could, for example, comprise:

- Performing an online search to find out whether a combination of birthdate and postcode could be used to reveal an individual's identity.
- Searching in the archives of a national or local newspaper to see whether it is possible to link the victim's name to data about where crimes have been committed.
- Using social media communities to see whether it is possible to link anonymised data to a user's profile.
- Using data from various public databases to try and link anonymised data to someone's identity.

Since access to available information and computing power is increasing all the time, it is important to carry out motivated intruder tests regularly to re-evaluate the risk of re-identification.

Motivated intruder test – a checklist

- What is the risk of a so-called "jigsaw attack", i.e. that various bits of information are put together to form a complete picture of one person? Are the data of such a nature as to enable them to be linked together? For example, is the same code used to refer to the same individual in several different data sets?
- What other type of "linkable" information exists that is easily or publicly accessible?
- What technical measures can be deployed to succeed in re-identifying the data?
- How much weight should be accorded to individuals' possible knowledge of the data subject in an anonymised data set?
- If an intruder test has been performed, what weaknesses did it reveal?

Sources of information include public libraries, public databases (tax lists, the Brønnøysund Register Centre, the vehicle licensing database), the electronic public records system (OEP) in which correspondence to/from the public administration is listed, parish records, genealogy websites, social media, search engines, local and national media archives, and anonymised data published by other organisations.

Anonymisation techniques must be determined on a case-by-case basis

There are various anonymisation techniques, which are more or less robust. The data controller must consider what guarantee against re-identification can be achieved through application of a specific technique, in light of three key risks associated with anonymisation:

- Is it still possible to *distinguish* one individual person in a data set?
- Is it possible to *link together* various data sets associated with one and the same person?
- Is it possible to *deduce* information associated with an individual person from the data set?

No anonymisation techniques fully meet the requirements for effective anonymisation. How a data set can best be anonymised must therefore be determined on a case-by-case basis, where account is taken of the purpose of anonymisation and the context involved. It is necessary to work consciously and meticulously to determine which technique best suits a particular situation, as well as how several techniques may possibly be combined to make the outcome even more robust.

	Is it still possible to *distinguish* one individual person in a data set?	**Is it still possible to *link together* various data sets associated with one and the same person?**	**Is it still possible to *deduce* information associated with an individual person?**
Addition of noise	Yes	Probably not	Probably not
Substitution	Yes	Yes	Probably not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	Probably not
Differential privacy	Perhaps/probably not	Probably not	Probably not
Hashing (tokenisation)	Yes	Yes	Probably not
Pseudonymisation	Yes	Yes	Yes

Knowing the strengths and weaknesses of the various techniques makes it easier to decide how the data can be anonymised in the most secure way possible. The strengths and weaknesses of the various techniques are examined in detail in Appendix 2.

The table above provides an overview of the strengths and weaknesses of the various models, as they relate to the fundamental criteria mentioned above. A solution that meets all three criteria will be resilient against identification. If a proposed solution does not meet one of the criteria, a thorough assessment of the identification risks relating to the data set should be carried out.

(This chapter is based on two publications: The Article 29 Working Party's opinion/recommendations on anonymisation techniques and a guide to the anonymisation of personal data published by the Information Commissioner's Office, the Norwegian Data Protection Authority's British sister organisation: Anonymisation: managing data protection risk code of practice, ICO (2012). We reuse the text with the permission of the ICO.)

Quick guide

- Choose one or more anonymisation techniques on the basis of the context and purpose concerned.
- The optimal anonymisation solution must be chosen on a case-by-case basis.
- All techniques have their strengths and weaknesses. So use a combination to achieve the best possible degree of anonymisation.
- Risk assessments must be performed not only before the data are published, but also afterwards. New risks may emerge.
- Use tests to assess the probability of re-identification.
- The data controller should disclose which anonymisation techniques or combination of techniques have been used when publishing anonymised data sets.
- Easily identifiable (i.e. rare) attributes should be removed from the data set.

Glossary

Aggregated data	Statistical data about several individuals, which has been combined to show general trends or values without identifying individuals within the data set. Aggregated data is not necessarily anonymous data.
Attacker	A third party (i.e. neither the data controller nor the data processor), who gains access to the original data entries, either intentionally or by accident.
Anonymous data	Data which it is impossible, using all reasonable technical means, to link back to an individual person.
Anonymisation	The process of rendering data into a form in which it is no longer possible to identify individuals and where the risk of identification does not exist, using all reasonable technical means.
Attribute	In a relational database, an attribute is a property or characteristic belonging to a table (component). For example, in a customer database, there is a table (component) called Customer. Attributes ascribed to this table may include customer no., first name, surname, address, postcode, etc. These attributes will form the headings of the table's various columns. Values accorded to the attributes might be: 123456, Per, Hansen, Storgata 1, 0123.
De-identified	Data from which the directly identifying attribute values have been removed and replaced by a unique identifier to conceal the identity of the data subject.
Limited access	The release of data to a limited and narrowly defined group of users, e.g. researchers or an institution.
Data controller	The organisation or person who determines the purpose for which the personal data are to be processed, and the manner in which this is to be done. The data controller is responsible for ensuring that the data are processed in accordance with the provisions of the Personal Data Act.

Data processor	The organisation or person who processes personal data on behalf of the data controller. The data processor must process the personal data only as specifically agreed with the data controller.
Data set	A data set comprises different entries relating to individual people (the data subjects). Each individual entry is related to a single data subject and is made up of a set of values (e.g. 2013) for each attribute (e.g. Year). A data set is a collection of entries, which can take the form of a table (or a series of tables) or an annotated/weighted graph.
Quasi-identifiers	Combinations of entries relating to a data subject or group of data subjects. In some cases, a data set may contain several entries on the same person.
Entry	The contents of a table column (e.g. Per, Hansen, Rødveien 9, etc.). In other words, the values recorded under each attribute heading, in this case: first name, surname, address, etc.
Personal data	A piece of information or an assessment which may be linked to an individual person.
Pseudonymisation	A process whereby directly identifiable parameters are replaced by unique indicators to conceal the real identities of the data subjects.
Publishing	The act of making data publicly available, e.g. by uploading them to the internet.
Data subject	An individual person about whom information is recorded.
Sensitive personal data	Information regarding a person's racial or ethnic origin; political, philosophical or religious views; health issues; sexual relations; membership of a trade union; or that a person has been suspected, charged, indicted or convicted of a criminal offence.

Office address:
Tollbugata 3, 0152 Oslo

Postal address:
PO Box 8177 Dep., 0034 Oslo

postkasse@datatilsynet.no
Tel: +47 22 39 69 00

datatilsynet.no
personvernbloggen.no