ORGANISATIONS

GUIDANCE ON ANONYMISATION AND PSEUDONYMISATION

digital TOOLKIT

Guidance for Organisations

This document is purely for guidance and does not constitute legal advice or legal analysis. It is intended as a starting point only, and organisations may need to seek independent legal advice when renewing, enhancing or developing their own processes and procedures or for specific legal issues and/or questions.

JOIC
JERSEY OFFICE OF THE INFORMATION COMMISSIONER

WWW.JERSEYOIC.ORG

Guidance on Anonymisation and Pseudonymisation

Data subjects have a fundamental right to privacy, it is important for organisations which process personal data to be cognisant of this right. When carried out effectively, anonymisation and pseudonymisation can be used to protect the privacy rights of individual data subjects and allow organisations to balance this right to privacy against their legitimate goals.

The guidance note aims to provide information about using these techniques.

Key points

- Irreversibly and effectively anonymised data is not 'personal data' and the data protection principles do not have to be complied with in respect of such data. Pseudonymised data remains personal data.
- If the source data is not deleted at the same time that the 'anonymised' data is prepared, where the source data could be used to identify an individual from the 'anonymised' data, the data may be considered only 'pseudonymised' and thus still 'personal data', subject to the relevant data protection legislation.
- Data can be considered 'anonymised' from a data protection perspective when data subjects are not identified or identifiable, having regard to all methods reasonably likely to be used by the data controller or any other person to identify the data subject, directly or indirectly.

What is personal data?

Personal data means any data (information) relating to an identified or identifiable individual. This individual is also known as a 'data subject'. Art.2(2) of the Data Protection (Jersey) Law 2018 (**'DPJL'**) says that:

> *(2) A data subject is an identified or identifiable, natural, living person who can be identified, directly or indirectly, by reference to (but not limited to) an identifier such as –*
>
> *(a) A name, an identification number or location data;*
>
> *(b) An online identifier; or*
>
> *(c) One or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the person.*

Accordingly, data about living individuals which has been anonymised such that it is not possible to identify the data subject from the data or from the data together with certain other information, is not governed by the DPJL and is not subject to the same restrictions on processing as personal data.

What is anonymisation?

'Anonymisation' of data means processing it with the aim of irreversibly preventing the identification of the individual to whom it relates. Data can be considered effectivelyand sufficiently anonymised if it does not relate to an identified or identifiable natural person or where it has been rendered anonymous in such a manner that the data subject is not or no longer identifiable.

This note is intended to assist with identifying and minimising the risks to data subjects when anonymising data; it cannot say that the techniques you deploy will be able to anonymise the data that you hold. In the case of anonymisation, by 'identification' we mean the possibility of retrieving a person's name and/or address, but also the potential identifiability by singling out, linkability and inference.

What is pseudonymisation?

'Pseudonymisation' of data means replacing any identifying characteristics of data with a pseudonym, or, in other words, a value which does not allow the data subject to be directly identified.

The DPJL defines pseudonymisation (see Art.3) as 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 it is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (individual).

Although pseudonymisation has many uses, it should be distinguished from anonymisation, as it only provides a limited protection for the identity of data subjects in many cases as it still allows identification using indirect means. Where a pseudonym is used, it is often possible to identify the data subject by analysing the underlying or related data.

Uses of anonymisation and pseudonymisation

Data which has been irreversibly anonymised ceases to be 'personal data' as defined, and processing of such data does not require compliance with the DPJL. In principle, this means that organisations could use it for purposes beyond those for which it was originally obtained, and that it could be kept indefinitely.

In some cases, it is not possible to effectively anonymise data, either because of the nature or context of the data, or because of the use for which the data is collected and retained. Even in these circumstances, organisations might want to use anonymisation or pseudononymisation techniques;

- As part of a 'privacy by design' strategy to provide improved protection for data subjects;
- As part of a risk minimisation strategy when sharing data with data processors or other data controllers;
- To avoid inadvertent data breaches occurring when your staff is accessing personal data;
- As part of a 'data minimisation' strategy aimed at minimising the risks of a data breach for data subjects.

Even where anonymisation is undertaken, it does retain some inherent risk. As mentioned, pseudonymisation is not the same as anonymisation and should not be equated as such – the information remains personal data.

Identification – the test for identifiability

In order to determine whether data has been sufficiently anonymised to bring it outside the scope of data protection law, it is necessary to consider the second element of the definition, relating to the identification of the data subject, in greater detail.

The Article 29 Working Party on Data Protection (now replaced by the European Data Protection board, or 'EDPB') (referring to the General Data Protection Regulation (GDPR)) has previously suggested the following test for when an individual is identified or identifiable:

> *In general terms, a natural person can be considered as 'identified' when, within a group of persons, he or she is 'distinguished' from all other members of the group. Accordingly, the natural person is 'identifiable' when, although the person has not been identified yet, it is possible to do it...*

Thus, a person does not have to be named in order to be identified. If there is other information enabling an individual to be connected to data about them, which could not be about someone else in the group, they may still 'be identified'.

In determining whether a person can be distinguished from others in a group, it is important to consider what 'identifiers' are contained in the information held.

Identifiers are pieces of information which are closely connected with a particular individual, which could be used to single them out. Such identifiers can be 'direct', like the data subject's name or image, or 'indirect', like their phone number, email address or a unique identifier assigned to the data subject by the data controller. As a result, removing direct identifiers does not render data sets anonymous. Data which are not identifiers may also be used to provide context which may lead to identification or distinction between users – e.g. a series of data about their location, or perhaps their shopping or internet search history. Indeed, these kinds of data series on their own may be sufficient to distinguish and identify an individual.

However, just because data about individuals contains identifiers does not mean that the data subjects will be identified or identifiable. This will depend on contextual factors. Information about a child's year of birth might allow them to be singled out in their family, but would probably not allow them to be distinguished from the rest of their school class, if there are a large number of other children with the same year of birth. Similarly, data about the family name of an individual may distinguish them from others in their workplace, but might not allow them to be identified in the general population if the family name is common.

On the other hand, data which appear to be stripped of any personal identifiers can sometimes be linked to an individual when combined with other information, which is available publicly or to a particular individual or organisation. This occurs particularly in cases where there are unique combinations of connected data. In the above case for instance, if there was one child with a particular birthday in the class then having that information alone allows identification.

Identifiability and anonymisation

The concept of 'identifiability' is closely linked with the process of anonymisation. Even if all of the direct identifiers are stripped out of a data set, meaning that individuals are not 'identified' in the data, the data will still be personal data if it is possible to link any data subjects to information in the data set relating to them.

The DPJL does not provide guidance as to when a data subject may still be identifiable, but Recital 26 of the GDPR provides some guidance which although not binding in Jersey is helpful in understanding whether certain information may lead to an individual being identified. It says that when determining whether an individual is identifiable or not '*(...) account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly*' and that when determining whether means are 'reasonably likely to be used' to identify the individual '*(...) account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.*'

Therefore, to determine when data are rendered anonymous for data protection purposes, you have to examine what means and available datasets might be used to re- identify a data subject. Organisations don't have to be able to prove that it is impossible for any data subject to be identified in order for an anonymisation technique to be considered successful. Rather, if it can be shown that it is unlikely that a data subject will be identified given the circumstances of the individual case and the state of technology, the data can be considered anonymous.

Some different ways that re-identification can take place are discussed below.

If the source data is not deleted at the time of the anonymisation, the data controller who retains both the source data and the anonymised data will normally be in a position to identify individuals from the anonymised data. In such cases, the anonymised data must still be considered to be personal data while in the hands of the data controller, unless the anonymisation process would prevent the singling out of an individual data subject, even to someone in possession of the source data.

Identification risks

It is not normally possible to quantify the likelihood of re-identification of individuals from anonymised data. However, thinking about the risks which are present will assist in assessing whether identification of data subjects from anonymised data is likely. An effective anonymisation technique will be able to prevent the singling out of individual data subjects, the linking of records or matching of data between data sets, and inference of any information about individuals from a data set.

Singling out

'Singling out' occurs where it is possible to distinguish the data relating to one individual from all other information in a dataset. This may be because information relating to one individual has a unique value; such in a data set which records the height of individuals, where only one person is 190cm tall, that individual is singled out. It might also occur if different data related to the same individuals is connected in the data set and one individual has a unique combination of values. For example, there might be only one individual in a dataset who is 160cm tall and was born in 1990, even though there are many others who share either the height or year of birth.

Data linking

Any linking of identifiers in a data set will make it more likely that an individual is identifiable. For example, taken individually the first and second name 'John' and 'Smith' might not be capable of distinguishing one of a large company's customers from all other customers, but if the two pieces of information are linked, it is far more likely that 'John Smith' will refer to a unique, identifiable individual. The more identifiers that are linked together in a data set, the more likely it is that the person to whom they relate will be identified or identifiable.

A major risk factor which may lead to the identification of individuals from anonymised data is the risk of data from one or more other sources being combined or matched with the anonymised data. This is particularly relevant where data has been pseudonymised, as a direct comparison can be made between the data masked by a pseudonym and other available data, leading to the identification, or unmasking, of data subjects. Researchers have shown many times that only a few pieces of non-identifying information, when combined, can lead to highly accurate re-identification, especially when information in the public domain is combined with otherwise anonymous data sets.

Data minimisation and collection techniques, which are also part of the principles of data protection are helpful in reducing the risk of data matching being successful. The DPJL specifically sets out that controllers must adhere to the principle of data minimisation, namely that personal data processed should be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

Inference

In some cases, it may be possible to infer a link between two pieces of information in a set of data, even though the information is not expressly linked. This may occur, for example, if a dataset contains statistics regarding the seniority and pay of the employees of a company. Although such data would not point directly to the salaries of individuals in the dataset, an inference might be drawn between the two pieces of information, allowing some individuals to be identified. Where this is possible, data protection law continues to applies, and there remains a risk of re-identification that should be considered by organisations which should be appropriately safeguarded.

When is data 'anonymised'?

As set out above, data can be considered 'anonymised' from a data protection perspective when data subjects are no longer identifiable, having regard to any methods reasonably likely to be used by the data controller - or any other person to identify the data subject. Data controllers need to take full account the latter condition when assessing the effectiveness of their anonymisation technique.

If the data controller retains the raw data, or any key or other information which can be used to reverse the 'anonymisation' process and to identify a data subject, identification by the data controller must still be considered possible in most cases. Therefore, the data may not be considered 'anonymised', but merely 'pseudonymised' and thus remains personal data, and should only be processed in accordance with the DPJL.

Where data has been anonymised to such an extent that it would not be possible to identify an individual in the anonymised data even with the aid of the original data, the data has been fully anonymised and is not considered personal data. This might occur where the data is in an aggregated statistical format, or where random noise added to the data is such as to completely prevent a linkage between the original data and the anonymised data from being made.

It is not possible to say with certainty that an individual will never be identified from a dataset which has been subjected to an anonymisation process. It is likely that more advanced data processing techniques than currently exist will be developed in the future that may diminish any current anonymisation techniques. It is also likely that more data sets will be released into the public domain, allowing for cross comparison between datasets. Both of these developments will make it more likely that individual records can be linked between datasets in spite of any anonymisation techniques employed, and ultimately that individuals can be identified.

However, the duty of organisations is to make all reasonable attempts to limit the risk that a person will be identified. In assessing what level of anonymisation is necessary in a particular case, you should consider all methods reasonably *likely* to be used by someone (either an 'outsider' or an 'insider' to the organisation) to identify an individual data subject given the current state of technology and the information available to such a person at present. An approach to anonymisation which affords a reasonable level of protection today may likely prevent identification into the future, but this will have to be monitored and assessed over time.

In deciding what methods are 'reasonably likely' to be used by an outsider, organisations should first consider who the potential intruder might be. Organisations should also consider the sensitivity of the personal data, as well as its value to a potential outsider or any 3rd party who may gain access to the data. The more motivated an outsider will likely be to identify data subjects, the more an organisation should expect extreme measures to be used for identification. Finally, organisations should consider what other data outsiders might be able to access, to compare with the anonymised data. This data may come from publicly available information, such as the electoral register or phone book, or from data known personally to the outsider.

Who might be an 'outsider'?

The word 'outsider' as referred to above, is not used solely to refer to individuals who are not intended to have access to the anonymised data. It can also refer to individuals who are permitted access to the data, but who might, either intentionally or inadvertently identify a data subject from the anonymised data. When it is intended to publish anonymised data to the world at large, there is a much higher burden on organisations to ensure that the anonymisation is effective, as it may be virtually impossible to retract publication in the event of a later realisation that identification is possible, and the intent and actions of recipients goes beyond the supervision of the original data controller.

In some cases, you may want to anonymise data in order to share it with a defined group, rather than releasing it to the public at large. In such cases, you should have regard to the other information and technical know-how available to that group in deciding whether there is any reasonable likelihood of identification occurring.

In academic or institutional settings, it may be possible to include binding commitments aimed at preventing re-identification in any agreement for the sharing of anonymised data. This would reduce the likelihood of identification of data subjects occurring, and would therefore permit the sharing of more detailed data than would otherwise be the case.

In the case of anonymisation of data for use within an organisation, it may not be necessary to impose as rigorous an identifiability test as would be the case where it is intended to release the anonymised data publicly. This is because the organisation will be more likely to retain control over who is able to access the anonymised data, and the conditions under which they may do so. If these conditions are appropriately designed,they can help to reduce the risk of identification, allowing greater detail to be included in the data while retaining anonymity.

However, even when sharing anonymised data within an organisation or within a defined group with binding restrictions on the use of the data, the organisation should consider the risk of the data being accessed by an intruder from outside these groups, or being shared inappropriately. In all cases where anonymised data is prepared for internal use, regard must be had to the possibility of accidental publication, and the physical and technical security measures preventing the unauthorised access of such data.
Where accidental publication or unauthorised access is more likely, greater care must be taken to limit the likelihood of identification of individuals by an intruder obtaining the anonymised data in this way.

How likely are attempts at identification?

The more likely it is that someone may attempt to identify an individual from anonymised data, the more care has to be taken in anonymising the data. However, that in itself is not a reason to consider that anonymisation or other measures on data processing are not required. A wide range of factors will be relevant to assessing this risk, including the value of the information to any outsider, the range of potential outsiders, and the risk of the data being shared beyond the intended recipient. In cases where financial or health information is anonymised, particular care must be taken as there is likely to be a relatively high incentive for other individuals to attempt to identify individuals from the anonymised data.

A related question is whether there is any likelihood of identification occurring inadvertently. This is most likely in circumstances where an outsider with personal knowledge of the data subject comes into possession of the anonymised data. The risk of identification by someone with personal knowledge is discussed below. It should be considered whether such inadvertent identification is possible, and take steps to minimise the potential for this to happen, either by changing the way that the data is anonymised to minimise the risk of identification by those with personal information, or by restricting circulation of the anonymised data to prevent those with personal information from coming into contact with it.

However, it should be remembered at all times that even where personal knowledge of the data is not a factor, re-identification, re-linking and inference may remain a significant risk depending on the anonymisation techniques used and the context of the data.

What other data might an outsider have access to?

As set out above, identification can occur through the matching of different data sets. In selecting an anonymisation technique, you should consider what other data might be available publicly, or to the groups likely to have access to anonymised data, which might make identification possible. Such information includes;

- Public registers, such as the Public Registry (the publicly available registry of immovable land transactions), Electoral Register, or publicly accessible registries of the members of professions;
- Searchable information contained on the internet or in online databases. This category of information might include newspaper stories, blog posts or online directories, or data published in previous data breaches;
- Statistical data published in an anonymised format, which might be combined with certain anonymised data to identify a data subject. This is a particular concern in the case of research or statistical publications concerning the same data subjects;
- Information available to the particular organisation or individual that is being given access to anonymised data.

Personal knowledge

In some cases, the personal knowledge of someone who comes across the data will allow that person to identify a data subject, even though identification would be impossible for someone without that personal information. For example, a doctor might be able to identify one of their patients when reading an anonymised study in a medical journal, or the residents of a particular parish might be able to identify the individuals to whom anonymised crime figures relate.

As a result, special care should be taken in cases where the personal knowledge of an individual or group might allow that individual or group to discover new information about a data subject by linking their personal information to anonymised information about the data subject even in cases where the professional secrecy of the recipient is a factor.

It should be considered whether someone with personal knowledge is likely to have access to the anonymised data in making the assessment of whether the anonymisation process is robust enough to prevent identification in these circumstances. If the individuals with the relevant personal information are unlikely to use that personal information to attempt to identify any data subjects, because to do so would conflict with their professional obligations for example, then the fact that they will have access to the anonymised data does not necessarily mean that the data cannot be considered anonymised.

What anonymisation techniques should be used?

Deciding on an appropriate anonymisation technique has to be done on a case by case basis, having regard to all of the relevant risk factors outlined above, and to the intended purpose of the anonymised data. Organisations have to balance the need to retain all information necessary for the purpose for which the anonymised data is to be used with the identification risks presented by the inclusion of more detailed information in a dataset. Where personal data cannot be effectively anonymised they must still be regarded and treated as personal data.

The DPJL does not prescribe any particular technique for anonymisation, so it is up to individual data controllers to ensure that whatever anonymisation process they choose is sufficiently robust. This document does not provide a comprehensive overview of all available anonymisation techniques, and cannot give detailed guidance on individual cases. Organisations may wish to consult the previous Article 29 Working Party's opinion on Anonymisation Techniques (Opinion 05/2014), and in particular the technical annex thereto for more detailed information about the anonymisation techniques which may be deployed.

Organisations should also be aware of their obligations regarding data protection by design and by default (per Article 15 of the DPJL) as well as regarding the security of processing of personal data (per Article 21 of the DPJL).

There are, broadly speaking, two different families of anonymisation technique: 'randomisation' and 'generalisation'. Other techniques, such as 'masking' or 'pseudonymisation', which are aimed solely at removing certain identifiers, may also play a role in reducing the risk of identification. In many cases, these techniques work best when used together, so as to combat different types of identification risk.

Randomisation

Randomisation techniques involve the alteration of the data, in order to cut the link between the individual and the data, without losing the value in the data. These types of techniques can be used when precise information is not needed for the intended purpose of the anonymised data. Randomisation techniques may assist in reducing the risk of inference from anonymised data, as well as the risk of data matching between data sets, unless other available data sets use the same randomised values.

Randomisation may include the addition of 'noise', or random small changes, into data, to limit the ability of an intruder to connect the data to an individual. For example, in a database which records the height of individuals, small increases or decreases could be made to the height of each data subject, and the data can be stated to be accurate only within the range of the additions and subtractions. It is important to make sure that the scale of the noise to be added is in line with the scale of raw values, so that this process does not produce results entirely out of line with the actual results. For example, in a database of the height of individuals, adding or subtracting between 1cm and 10cm might achieve an acceptable level of anonymity, but adding or subtracting 1m might not produce useful data, and could in some cases make it obvious who the data refers to.

'Permutation' is another type of randomisation technique. This involves swapping certain data between the records of individuals, making it more difficult to identify individuals by linking together different information relating to them. For example, in the case of the height of individuals, instead of adding random noise to the data, the height values for different individuals is moved around, so that is no longer connected to other information about that individual. This is helpful if you need to retain the precise distribution of height values in the anonymised database, but you do not need to maintain correlations between height values and other information about the data subjects.

Generalisation

Generalisation involves reducing the granularity of data, so that only less precise data is disclosed. This means that it will be less likely that individuals can be singled out, as more people are likely to share the same values. For example, a data base containing the age of data subjects might be adjusted so that it is only recorded what band of ages an individual falls within (e.g. 18-25; 25-35; 35-45; etc.).

This can be done by a process known as 'k-anonymisation', which involves ensuring that each value relating to a data subject is shared by at least a minimum number (k) of others within the data set. This allows you to choose an appropriate size for the bands of information. For example, if you want each value to be shared by at least 5 individuals, you might choose to give their location by province or county rather than town, depending on the geographical distribution of individuals within the data.

However, this technique can be weak if data which is linked to the generalised field allows an individual to be singled out. For example, there might be 5 individuals in your database who live in St Helier but if only one of them is over 1.9m tall, they will be identifiable if only the location data is generalised to the Parish level. There are a number of techniques discussed in the technical annex to the previous Article 29 Working Party's opinion on Anonymisation Techniques which can be used to assist organisations in addressing and reducing this risk.

Masking

Masking is useful in supplementing other anonymisation techniques. It involves removing obvious or direct personal identifiers from data. It is a necessary prerequisite of anonymisation that no direct or obvious identifiers are contained in the anonymised data set. Such information might include names, addresses or images.

Masking alone often allows a very high risk of identification, and so will not normally be considered anonymisation in itself. This is because such a technique would allow all of the original unmasked data to be seen, making it at risk of data matching techniques being used to reveal the identity of data subjects.

Pseudonymisation as an anonymisation technique

When used alone, pseudonymisation carries similar risks to masking, in that much of the original, unaltered data will be contained in the pseudonymised data, and so data matching techniques might be able to identify individual data subjects. It has the further disadvantage that if the pseudonym is reused, it permits the linking together of different records relating to the same individual, which would create further identification risks.

However, pseudonymisation has the advantage of permitting different records relating to the same individual to be linked without storing direct identifiers in the data. This is especially useful in longitudinal studies, or for other purposes where it is necessary to link data collected at different times relating to the same data subject. It can be combined with other techniques in some circumstances to allow for anonymised data to be linked to the same individual, however if doing so, it must be considered with each new set of data to be anonymised whether an identification risk exists, having regard to the existing anonymised data. Given that individuals can still be singled out with pseudonymisation

Pseudonymisation should never be considered an effective means of anonymisation, but can be considered a security enhancing measure to reduce the dataset's 'linkability'.

When can personal data be anonymised?

The process of making data anonymous is itself considered to be 'processing' data, so if an organisation wants to anonymise personal data to bring it outside of the scope of the DPJL it must be done fairly, in accordance with the relevant law.

For example, in anonymising data, organisations are still normally subject to the principle of 'purpose limitation', provided by Article 8(1(b) of the DPJL.

Organisations should inform data subjects when collecting personal data if one of the purposes of data collection is to anonymise the data for future use. If this has not been done, such anonymisation could be considered 'further processing' of data for purposes beyond those for which it was originally obtained, which is subject to a number of limitations under the DPJL. In other cases, the anonymisation of data will be ancillary to one of the stated purposes for the collection of data, and so will not be problematic. For example, if anonymisation is used internally within an organisation when data is being accessed for the purpose for which it was obtained, this anonymisation is not a distinct purpose.

There is an exemption to the purpose limitation provided by Articles 8(1)(b) and para.17 of Schedule 2 of the DPJL for the processing of data for archiving and research purposes. Personal data used for such purposes will not be considered to be incompatible with the original purpose for which the personal data was processed.

If anonymisation of personal data is carried out effectively, it can help to reduce the risk of any harm being suffered by data subjects, so it is not likely that data subjects will have a right to prevent their data from being anonymised, but the effectiveness would have to be evaluated in each case.

Extracting personal data from partially anonymised databases

Taking a partially anonymised dataset and processing it to extract personal data is normally considered to be obtaining personal data. The processing of this personal data will thus fall under the remit of data protection law, and will bring with it various obligations regarding personal data which was not obtained directly from an individual. If an organisation can identify an individual in a partially anonymised dataset, they may be considered a data controller, if the organisation meets other criteria for being a data controller.

Article 12(2) of the DPJL, for example, requires that information in relation to the processing of personal data relating to an individual should be given to them by the data controller, where the personal data are obtained from a source other than the individual themselves, within a reasonable period but at the latest within 4 weeks, depending on the circumstances of the case.

As part of the process of anonymising data, organisations should engage in testing the effectiveness of the anonymisation process on their data, in order to determine its success. This will consider what can be identified once the process is complete, the required effort an attacker or intruder might need to expend in order to re-identify, the overall 'usefulness' of the anonymised data and also to gauge how much an increase in anonymisation effort will lead to improvements in the effectiveness of the anonymisation process.
As organisations will, in most cases, have retained the original data, identifying individuals in the course of pen testing will not normally reveal any new information about those individuals, and so such processing is not considered obtaining personal data.

Anonymisation and data retention

Article 8(1)(e) of the DPJL requires that personal data not be kept in a form which permits identification of individuals for any longer than is necessary for the purposes for which the personal data are processed. The wording 'in a form that permits identification' refers to the possibility of retaining data which has been fully anonymised.

Article 8(2)(b) of the DPJL also sets out that personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving archiving and research purposes or for other legal obligations subject to implementation of the appropriate technical and organisational measures in order to safeguard the rights and freedoms of the data subject.

Data retention

As set out above, data which has been anonymised so as to remove the reasonable possibility of identification of any data subjects is not personal data, and the obligation to retain personal data only for so long as is necessary does not apply. However, if an organisation retains anonymised data on this basis, they should keep its identifiability status under continuous review. In particular, the organisation may come into possession of new information which would allow the anonymised data to be linked to an individual.

Where data about individuals is anonymised on a rolling basis for retention, organisations should be careful not to allow linkages between data which has been anonymised at different times, if those linkages would allow individuals to be identified in any of the anonymised data sets. In particular, where pseudonyms are used, this may allow for the linkage of records between different anonymised datasets unless new pseudonyms are chosen each time. Pseudonyms also present a difficulty in anonymisation if a key is retained, as this would make it possible to identify the data subjects from the anonymised data. Where a key is not retained, pseudonymisation will still allow for singling out, possible linkage and inference, and should not be relied on as effective anonymisation.

Deletion of source data

As set out above, data which has undergone a partial anonymisation process will not cease to be personal data if (i) the source data is retained and (ii) individuals would be identifiable from the partially anonymised data with the help of the source data. If organisations intend to retain anonymised data, they are still required to delete the original data once it is no longer needed. Until such original data is deleted, organisations are bound to treat the partially anonymised, or pseudonymised, data as personal data. Individuals will continue to be able to exercise their rights in respect of this data. Once the source data is destroyed, the organisation should again consider and possibly test the effectiveness of the anonymisation.

Subject access and rectification

Data subjects have various rights under the DPJL, including rights under Article 28 to request details about their personal data which is held by an organisation and to access their personal data.

Data subjects also have rights under Articles 31 and 32 of the DPJL to have a data controller correct any incorrect information or delete any personal data in certain circumstances. The obligations on data controllers in responding to such requests are discussed in more detail in our guidance **Subject Access; A Guide for Data Controllers**. Organisations should consult these guidance pages to find out more about dealing with these requests.