apdcat

Autoritat Catalana de Protecció de Dades

Privacy by design and privacy by default

A guide for developers

February 2023

Guides collection. No. 7

Generalitat de Catalunya

Index

1. Introduction

The concept of privacy by design, which has been developed since the late 1990s, largely through the work of the Information and Privacy Commissioner of Ontario, refers to the need to consider the impact in terms of privacy of products and services, especially technological ones, from the design stage.

The concept of privacy by default is closely related, and involves taking appropriate technical and organisational measures to ensure that only the personal data that is essential for each specific processing purpose is collected and that the user having to take any steps (by default).

With the passing of Regulation (EU) 2016/679 of the European Parliament and of the Council, of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation or GDPR), data protection by design and data protection by default both went from being recommendations or good practice to being obligations.

In particular, article 25 of the GDPR states:

> 1. the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects [**data protection by design**].
>
> 2. The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed [**data protection by default**].
>
> That obligation applies to
>
> - the amount of data collected,
> - the extent of their processing,
> - the period of their storage
> - and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

Identifying these needs and providing solutions for them from the very moment of the design of technological tools makes it possible to save time, resources, harm for the persons concerned and the obvious reputational costs that incorporating these requirements at a later stage might cause.

Furthermore, as it is an obligation, failure to fulfil it could be an infringement of the GDPR, which could result in a penalty of up to 10,000,000 euros or of 2% of the total worldwide annual turnover of the preceding financial year if this figure is more than 10,000,000 euros.

In any case, beyond this obligation, data protection is a qualitative factor that is increasingly valued by businesses and institutions that acquire a given product or service and also by users.

Ultimately, data protection is a competitive advantage and this guide aims to be a useful tool to make the most of it.

That the different agents in the digital ecosystem see it this way is a strategic priority for the Catalan Data Protection Authority (APDCAT). Therefore, we have prepared this guide to enable developers, and controllers who commission them to develop applications, to identify the different important elements for personal data protection, and the steps that can be taken to deal with it right from the moment of design.

2. Roles linked to data protection by design and by default

Data protection by design and by default, as envisaged by the GDPR, directly addresses the data controller, the person who has the obligation to implement and ensure implementation of the appropriate technical and organisational measures.

To identify who the controller is, it is necessary to consider the definition in article 4(7) of the GDPR, which defines the controller as the person who determines the purposes of the processing of data and the means used.

In any case, the GDPR allows the controller to delegate data processing to a third party or simply allow a third party to access the data to provide a service on behalf of the controller. This person is called the “processor”. This figure can be an entity that provides the service itself (such as a concession holder for a public service), one that collaborates with the controller to provide the service (for example, a hosting service) or a company that carries out the development or maintenance of an application or platform that involves accessing or processing personal data). In any case, if the developer needs to access personal data on behalf of the controller, even if only at the development stage, a contract must be a signed that sets out the content that the data protection regulations envisage [1].

It is important to underline that the controller must choose a processor that offers sufficient guarantees regarding the application of appropriate technical and organisational measures. One factor to take into account when making this choice is whether the processor has stamps and certificates or procedures and action protocols that include data protection by design and by default.

[1] See article 28 of the GDPR.

It is especially important to note that personal data can only be transferred outside the European Economic Area if the destination country has an adequacy decision issued by the European Commission[2] or if there are appropriate safeguards in accordance with one of the mechanisms set out in article 46 of the GDPR.

In any case, the measures that the controller must comply with in accordance with articles 25 and 32 of the GDPR must also be required of the processor and of any products and services that are acquired or commissioned.

In other words, any new products and services developed, internally (by the controller or processor itself) or externally, must comply with data protection by design and by default. Otherwise, the controller will be unable to fulfil its obligations.

Furthermore, developers that are classed as processors might subcontract activities to third parties (hosting, use of tools provided by third parties, etc.), meaning that these third parties will have to access personal data. These third parties are classed as sub-processors. In this case, a developer who is classed as a processor must select a sub-processor that provides appropriate guarantees and must have the authorisation of the controller. The same obligations and guarantees apply to the sub-processor as to the processor.

In summary, the principal roles in the development of a technological solution that is entrusted to a third party which, at the same time, entrusts the hosting at this stage to another third party, are set out in the table below:

Roles	**Controller**	**Processor**	**Sub-processor**
Task	Determining the purposes and means of the processing of personal data	Providing a service to the controller that involves accessing personal data	Providing a service to the processor that involves accessing personal data
Example	An entity that initiates the technological solution that involves processing personal data	Developers	Hosting service

It should be noted that, for data protection by design and by default to be effective, it is especially necessary to guarantee that developers have taken into account the context in which the product they develop is used. To configure the product with all appropriate guarantees, it is vital to know the circumstances in which the application will be used.

[2] The list of countries with an adequacy decision can be consulted here.

This knowledge and the implementation of the corresponding guaranties is a clear competitive advantage, as the controller has to prioritise the most appropriate solutions to guarantee the rights of users.

3. Effective application of data protection by design and by default

The rules governing data protection by design and by default do not determine which specific technical and organisational measures have to be implemented. Identifying the necessary measures must be the result of a prior analysis carried out by the controller and, by extension, by the developers of the technological solutions that the controller must use.

Necessary and proportionate measures must be applied. In any case, at the moment of design, it is necessary to take into account:

- The nature, scope, context and purposes of the processing.
- The risks processing presents for the rights and freedoms of people.
- The state of the art.
- The cost of the application.

In principle, it is the controller's responsibility to define the nature, scope, context and purposes of the processing. The selection of a particular technological solution must take these elements into account, as well as the risks inherent to each available technology. For this reason, it is essential to collaborate with developers at this stage.

Once these aspects are known, the risks processing presents to people's rights and freedoms must be assessed. If it is expected that they might be high, a data protection impact assessment (DPIA) must be completed.[3] It is important that the developer is involved in this assessment, as this is probably the best placed person to evaluate the technologies that might be used and that might involve a high risk for the rights and freedoms of people that makes the DPIA necessary.[4] Also, to help define the measures that can be put in place to reduce this risk, taking into account the state of the art and the cost of applying them. In summary, the data protection impact assessment is a systematic process that, as well as requiring a systematic description of the envisaged processing and of its necessity and proportionality, requires (i) assessment of the risks deriving from the processing and (ii) determining the measures to mitigate these risks.[5]

Similarly, in cases where performing a DPIA is not a requirement, the available technical options must be assessed, taking into account the current state of the art and the cost of implementing them.

[3] Article 35 of the GDPR contains a non-exhaustive list of cases. The list in section 28.2 of Spain's Organic Law 3/2018 is also useful, as is the list published by the APDCAT.
[4] The APDCAT has materials and an application to help carry out this data protection impact assessment.
[5] If the risks cannot be mitigated enough, before starting the processing the supervisory authority must be consulted.

Privacy enhancing technologies (PET) will be of use for this purpose. These make it possible to minimise risks without losing the functionality of the application or data system. Section 4 of this guide lists some measures that are considered key for protecting personal data and which form part of this more general set known as PET.

Ultimately, the design of technological solutions must take into account the risks deriving from the processing to determine the technical measures that must be applied.

Data protection by design and by default must be planned in the different phases of the processing of the data:

- Design.
- Development and testing.
- Collection of data.
- Use of the data.
- Communication or disclosure of data.
- Keeping and storing data.

Next, without setting out to be exhaustive, we identify some aspects that are considered vital to assess in relation to each of these stages, and which, from the very moment of the design, have to enable compliance with the principles of personal data protection.

In order to facilitate the systematic review of the elements that it is suggested should be assessed in relation with the different stages that comprise processing of personal data, a specific checklist has been prepared that is included as an annexe to this guide. As a summary, this list includes the principal aspects to consider, which are covered in more detail in sections 3 and 4 of this guide, and it will make it possible to assess the degree of incorporation of data protection in the design of the solutions.

3.1 Design stage

In the design stage of technological solutions, the different components in which the software will be structured and their interactions are defined, with the objective of fulfilling a series of functional and non-functional requirements. Data protection by design and by default introduces data protection among these requirements.

The inherent complexity of software development means that it is essential to use a development methodology to manage projects and ensure their success.

While data protection by design and by default does not involve using new design methods, it does require the tasks or analyses that are performed to be adapted. In this sense, different design strategies are mentioned:[6]

[6] Privacy Design Strategies (The Little Blue Book), JAAP-HENK HOEPMAN (2022)

- **Minimisation**: restricting the processing of personal data to the minimum possible. Processing the minimum amount of personal data, limiting the impact that the system might have on people.

- **Hiding**: hiding personal data from people who do not need to access them, making it difficult for them to be misused. There are many ways of hiding data, and their usefulness depends on the specific situation: cryptography, access control, etc.

- **Segregation:** distributed processing of data in compartments that are kept as separate as possible. Segregating data into different leak-proof compartments prevents easy access to people's complete profiles.

- **Aggregating:** processing the data in the most aggregated way possible, whenever this permits achieving the purpose pursued. Aggregating data in groups of people, if they are sufficiently large and diverse, means that the data cannot be associated with a specific person.

- **Informing:** appropriately informing people about the processing of their personal data.

- **Control:** people have to be able to make decisions about the processing of their data.

- **Enforcement:** there must be a privacy policy that is compatible with the legal requirements, and the means to fulfil it must be provided.

- **Demonstrating:** It is necessary to be able to provide evidence that the processing of personal data is carried out in a "friendly" way in terms of privacy.

At a more practical level, there are privacy patterns that provide design solutions to common problems in data protection.[7] That is to say, they are a pre-validated way of applying design strategies to specific problems.

When designing a solution, the identification of flows of personal information that will be processed must be included as essential elements (where the data are collected, how they are collected, who needs to have access to them and for what purpose, etc.) as well as the security measures that are enforceable in each specific case.

These measures will depend on the result of the risk analysis that must be performed in all cases. The following sections of this guide analyse many of these measures, grouped by the processing phase in which they are most relevant. However, it is at the moment of the development of the application that it is necessary to take them into account to incorporate them into the design.

[7] Privacy patterns.

3.2 Development and testing stage

Development and testing software usually involves using data. When these data are personal, the data protection rules are fully applicable.

Factors to take into account regarding the use of data during the design and testing stages

- If access to personal data is necessary in the development and testing stage, the development team or company and anyone the developer contracts will be processors (see part 2 of this guide on “Roles linked to data protection by design and by default”).

- To safeguard the confidentiality and integrity of the information, it is necessary to separate appropriately the production environment from the development and testing environment and ensure the testing environment is secure (if possible, isolated from external connections until the moment that it is necessary).

- If development and testing can be done without having to use real data, you should opt for this possibility. It is generally possible to use synthetic data (see section 4 of this guide).

  If this is not possible, minimal data must be used. By minimisation, we do not just mean the amount, but also the quality of the data (presence of identifiers and pseudoidentifiers, level of detail, etc.).

  If using real data is necessary, we recommend using them in an anonymised or pseudoanonymised form if possible.

  If use of data is necessary, data that the controller already has can be reused or new data can be collected. If new data has to be collected, the recommendations in section 3.3 of this guide must be taken into account.

- Special attention must be paid to the location of the server where the information collected will be stored. This matter could compromise the confidentiality of the information, as not all countries apply the same guarantees. Furthermore, using a server outside the European Economic Area involves an international transfer of data, which must fulfil certain requirements in order to be valid (Chapter V of the GDPR).

- Mechanisms to control access must be implemented to ensure that unauthorised users do not access the data. A user account must be created for each person who needs to access the data and access must be restricted to the necessary data to carry out their duties.

- Backup copies of the development and testing infrastructure are essential for the success of the project, but it is necessary to consider

that if these copies contain personal data, access to them must be controlled.

- Where Wi-Fi networks are used, the protocol that provides the highest degree of security must be used[8].
- Software packages and libraries that are used in the development must be up to date. Failure to update these might give rise to vulnerabilities in the software that is developed.
- The developing organisation must train its staff in data protection and must establish the appropriate confidentiality commitments.

Other factors to consider in the design stage

- Use a method to guarantee the quality of the source code. Whether manually, through code reviews carried out by different people, or automatically using static or dynamic analysis tools. Poor-quality code can give rise to vulnerabilities, which can put personal data at risk.
- Follow the recommendations of secure programming guides.

In the launch stage of the technological solution, and also when new users join the organisation, it is essential to provide training for staff who will use it that is appropriate for their different profiles. This involves providing training sessions so they understand how the solution works, the risks deriving from its use, the organisational and the technical measures that must be adopted to minimise them, as well as providing exhaustive user guides.

3.3 Data collection

3.3. Data minimisation

Only data that is appropriate and necessary to achieve the purpose pursued should be collected.

[8] In this sense, it is necessary to bear in mind that significant vulnerabilities have been found in the WEP, WPA and WPA2 protocols. At the time of writing this guide, the most recent protocol is WPA3. It must also be noted that some older routers might not support it.

From the start, this means that it is necessary to consider whether the purpose can be achieved without using personal data. This would simplify the controller's obligations, as personal data protection rules would not be applicable.

If it is necessary to use personal data, only the minimum amount required to achieve the specific purpose established by the controller and which the persons concerned have been informed of can be collected. Furthermore, it is necessary to be much more restrictive when special categories of data are collected (data that reveal ethnic or racial origin, political opinions, religious or philosophical beliefs or trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation), and so, whenever possible, collecting this type of data should be avoided.

The data collected can be very varied. They can include, among others, the following:

- Identifying data, contact details, employment data, data about education, socioeconomic data, data about interests, lifestyles, etc., alongside the special categories of data already mentioned.
- Activity logs, that are used in all types of services (web servers, email servers, in an online shop, etc.).
- Geolocation data.
- Data from wearable devices, such as pulse rate, blood pressure, blood oxygen level, etc.) that are collected automatically.
- Transactions (payments using mobile phones, card, etc.).
- Identifying data associated with devices or communication protocols (IP address, MAC address, telephone number, IMEI, ICC SIM, etc.).
- Metadata associated with files.

Given that most digital products or services constantly gather data, the need to obtain personal data must be verified continuously. In addition, the first access (often involving filling in a form or providing access to a previously configured profile) is an especially sensitive moment in this sense.

Verification of the need for the personal data must cover *all* of the data obtained, regardless of the means by which they are obtained. In this sense, it is necessary to pay special attention to data obtained from multiple sources – not just forms –, including collection of data about users' activity or cookies, whether first or third party, which, in addition, can have various purposes (behaviour patterns, demographic data of users, etc.).

Collection can be direct or indirect. So, for example, when a developer uses particular third-party libraries, this might involve processing personal information, even sometimes inadvertently. For example, when a Javascript library is included on a website, this library has access to the personal information contained on the website. It is even more worrying when third-party libraries used in the development of mobile applications have the same permissions as the developed application. This would mean, for example, that if the application can access the geolocation or the microphone, the external libraries used also could also do so.

On the other hand, the principle of data minimisation is not complied with adequately if an application requests permission to access and collect information that is not relevant to the purpose pursued. For example, collecting information about geolocation when the service offered is exactly the same independently of the location.

It is also important to consider that a collection method might involve an international data transfer outside the European Economic Area, something that requires appropriate safeguards in accordance with the GDPR. For example, a service for obtaining statistical data about visitors to a website might automatically send information to servers located outside the European Economic Area, and so it will be necessary to analyse whether the destination country of the data has equivalent safeguards to those that there would be within the European area.[9]

Important aspects

- It is necessary to consider whether the purpose pursued can be achieved without collecting personal data, or by collecting less data than originally envisaged. In other words, whether there is an alternative that makes it possible to achieve the purpose with less data. One option might be to generate artificial data, that simply replicate the behaviour of the real data, and to work exclusively with these data, known as synthetic data. It is also possible to modify data so that they cannot be associated with one person (pseudonymisation, anonymisation), reduce the amount of data collected or the level of detail, restrict to the maximum access to the data to the parts of the system that need them, etc.

- It is advisable to avoid processing data from special categories if this is not strictly necessary. For example, instead of configuring access to a particular digital service using biometric data, this can be done through a password and, perhaps, implementing two factor authentication.[10]

- Special care must be taken when designing forms and, in particular, when deciding on obligatory fields. It is necessary to distinguish clearly between information that it is compulsory to provide and that which is optional.

- The application must only request permission to access and collect information that is relevant to the purpose pursued.

[9] Specifically, in relation to the use of "Google Analytics", the French data protection authority – CNIL – in February 2022 ordered a website to stop using this service and had general information published in relation to the use of Google Analytics. At the start of the year, the Austrian Data Protection Authority ruled on these same lines.

[10] It is important to recall in this sense that two factor authentication involves combining two means of verification comprising: a piece of information known by the user (for example a password), something the user has (for example, a mobile phone) and something the user is (for example a piece of biometric data).

- It is necessary to avoid storing by default technical data that are not strictly necessary, such as MAC address, the IP address, the name of the device or the advertising ID.
- Collection of unnecessary connected information (such as metadata, data linked to activity, etc.) must be avoided.
- The possibility of certain information being processed on the user's own device must be assessed. One clear example of this course of action is the system of decentralised contact tracing, during the COVID-19 pandemic.
- The possibility of **anonymising or pseudonymising** data must be assessed, when knowing the identity of the user that is connected to it is not essential.
- It is necessary to take care with the use of external software components, as they could have access to personal data, above all without this access being documented. It is advisable to review the conditions of use of these components and, above all, monitor how they access personal data.

3.3.2 Lawfulness of data collection and processing

To be able to collect and process personal information, one of the legal bases set out in article 6 of the GDPR must apply.

Furthermore, in the case of special category data (data that reveal ethnic or racial origin, political opinions, religious or philosophical beliefs or trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation), one of the exceptions set out in article 9.2 of the GDPR must apply.

The consent of the person concerned is one of the legal bases set out in article 6 of the GDPR and can often be used as a legal basis for collecting data through applications or web services that users can install, or which they can voluntarily and freely access. Consent must be unambiguous, specific, informed and freely given, and choosing not to give it must not have any consequences other than those strictly linked to the impossibility of processing that information.

Factors to take into account regarding the management of consent in the design of technological solutions

- Consent must be a clear statement or affirmative action and must require action by the user. For example, it can be a box that has to be ticked, an electronic signature or an action that makes it possible to understand unambiguously that consent is being given (such as

following a link after having been clearly informed of it). It cannot be a pre-selected option.

- When consent is given to process special categories of data, or to receive commercial communications, it must be explicit. In this case, it is not enough to infer consent from the user's activity (for example, continuing to browse or accessing certain sections); instead an express statement is needed.

- Consent must be specific. When it is requested for various purposes, the user must be able to choose separately for each one (granular consent), for example with boxes or option buttons.

- Before requesting consent, it is necessary to inform the persons concerned of the aspects mentioned in the "Transparency and fairness for the user" section of this guide.

- Consent must be freely given. It is not enough to configure different consents for different uses if this is done in such a way that selecting a particular option also requires giving consent in another field, or the offer of the service is conditional on consenting to another purpose that does not necessarily need to be linked to it.

- The logs or means of accrediting the granting of consent must be stored through all of the processing of the data, and until the term of limitation of any potential infringements deriving from the processing has passed (the term of limitations for very serious infringements is three years after they occur).

- Consent must be revocable at any moment. Mechanisms for withdrawing consent must be provided separately for each of the intended purposes. These mechanisms must be accessible in a similar way to when obtaining consent.

- Secure mechanisms must be established to identify people who give consent and those who revoke it.

- Consent given by children is only valid if they are older than 14.[11] In this case, session cookies might be of use, despite their limited effectiveness, that contain the age initially entered by the child so that, for example, when children aged under 14 years enter their date of birth and realise that they cannot access the service, it is not particularly easy for them to change their age.

[11] Unless the rules applicable to the sector in question establish a different minimum age.

3.3.3 Transparency and fairness for the user

To comply with the principles of transparency and fairness, it is necessary to guarantee that users can have all of the information necessary about how their data are processed so that they can take the decisions that correspond to them or exercise their rights, and that the processing is in accordance with the expectations that the person concerned has been able to generate on the basis of this information.

It is the responsibility of the controller to determine the content of the information clauses. Despite this, when designing the application, it is advisable to consider certain aspects regarding how this information is provided.

The principles of transparency and fairness also involve an obligation to refrain from practices known as dark patterns or misleading design. These patterns consist of interfaces and the implementation of experiences that lead users to take unintended, unwanted and potentially harmful decisions relating to the processing of their personal data.[12]

The European Data Protection Board lists various types of dark patterns:

- **Overloading**: excessive requests, information or options, so that users share more data or involuntarily allow processing of their personal data against their wishes. The following three dark patterns are part of this category: **continuous prompting**, **privacy maze**, **too many options**.
- **Skipping**: designing the user experience interface so that users forget or do not reflect on aspects linked to data protection. The following two dark patterns are part of this category: **deceptive snugness** and **look over there**.
- **Stirring**: the users' capacity for choice is affected, because their emotions are played on or visual nudges are used. The following two darks patterns are part of this category: **emotional steering** and **hidden in plain sight**.
- **Hindering**: obstructing or restricting users in the process of obtaining information or managing their data, so that this task becomes especially difficult or impossible to

[12] EDPB Guidelines 3/2022 on Dark patterns in social media platform interfaces: How to recognise and avoid them. Version 1.0. Adopted on 14th March 2022.
On this matter, see also the contributions of Harry Brignull at https://www.deceptive.design/, where he describes various types of deceptive designs

achieve. The following three dark patterns are part of this category: **dead end**, **longer than necessary** and **misleading information**.

- **Fickle design**: designing the interface so that it is inconsistent and unclear. As a result, the user finds it hard to browse between the different data protection control tools and understand the purpose of the processing are. The following two dark patterns are part of this category: **lacking hierarchy** and **decontextualising**.
- **Left in the dark**: this means that the interface is designed to hide information or tools for managing data protection, or to leave users unaware of how their data are processed and how they can control them by exercising their rights. The following three dark patterns are part of this category: **language discontinuity**, **conflicting information** and **ambiguous wording or information**.

Important aspects

- Information must be provided before data is collected and, where applicable, before consent is given.
- When the data are not obtained from the person affected, it is also necessary to inform him or her. This must be done within one month of obtaining the data, or if it is anticipated that the data will be used in communications, at the latest when the first communication to the affected person or to a third party is made.
- The information must be complete, simple, understandable, visually clear and, where appropriate, adapted for people with functional difficulties.[13]
- In the case of services aimed at children, the information must be provided in language that is adapted to this group's level of knowledge.
- When the data are collected from the affected person, he or she must be informed about the aspects set out in paragraphs 1 and 2 of article 13 of the GDPR. If the data is provided by a third person, the information set out in paragraphs 1 and 2 of section 14 of the GDPR must be provided.

[13] The rules regarding accessibility must be taken into account: Directive (EU) 2016/2102 of the European Parliament and of The Council, of 26 October 2016, on the accessibility of the websites and mobile applications of public sector bodies; Spanish Royal Decree 1112/2018, of 7 September, on the accessibility of the websites and mobile applications of public sector bodies. The following may also be of interest: Accessibility requirements for ICT products and services and Web Content Accessibility Guidelines.

- Information can be provided in layers. So, information is initially given about the purpose of the processing, the identity of the controller and the possibility of exercising the rights of information self-determination (access, rectification, erasure, objection, restriction of processing and portability) as well as any other information that is considered essential. Users are then also offered the possibility of consulting the rest of the information if they want to know about the implications of the service in more detail.

- The information provided must be accurate. Users must be clearly and promptly notified of any change so that they can take the appropriate decisions.

- It is necessary to allow for copies to be made of the websites and applications, to be able to verify the mechanisms and the content of the information clauses existing at each moment by applying a verifiable timestamp.

- It is necessary to refrain from implementing dark patterns.

3.4 Use of data

The personal data that are collected cannot be used for any unrestricted purposes. The purpose must be specific (it cannot be unclear or excessively generic), explicit (it must be included in the information provided to the person concerned and in the record of processing activities that the controller has to keep) and legitimate (see the "Lawfulness of the collection and processing of data" section of this guide).

The actual use made of it must always be the same as the purpose that the person concerned was informed of when the data was collected. It can only be used for other purposes with the consent of the person concerned, when authorised by law or in the case of an activity that can be considered compatible, in accordance with the criteria established by the GDPR.[14]

Important aspects

- It is advisable to implement technical self-control mechanisms that ensure that use of the information is at all times restricted to the stated purpose. There must be an appropriate system for classifying information, that guarantees that the way it is used is limited to the legitimate use that can be made of it.

[14] Articles 5(1)(b) and 6(4) of the GDPR.

- Appropriate protection mechanisms must be put in place to avoid misuse by agents outside the organisation – for example processors – or internal ones (it is important to bear in mind that the most significant risk often comes from the organisation's own staff). On this matter, see the section on confidentiality in this guide.

- It is recommended that technological solutions are designed so that the exercise of the rights of the persons concerned is facilitated (rights of access, rectification, erasure, objection, restriction of processing and portability), if possible through the application itself, and that it also enables the controller to make them effective easily.

3.5 Communication and disclosure of data

It is necessary to ensure that communications of data to third persons only occur when there is a legal basis to do so and appropriate security measures have been put in place.

Important aspects

- When the use of an application or a technological solution is based on the consent of the person concerned, mechanisms must be established so that, by default, the data are not accessible to an indefinite number of people without the intervention of the person concerned.

- When information is published through the technological solution, it is advisable to provide mechanisms that enable its automated depublication, once the established period of publication has expired.

- It is recommended that communication uses **end to end encryption.** This means that the data can only be decrypted by the device of the person who receives the communications – using its private key – and not for example by the providers of the communication service.

In web services, the **HTTPS protocol** should be used and the server configured so that it cannot be accessed through other protocols.
This communication protocol, based on the use of certificates by the server, is of particular interest as it not only guarantees confidentiality and integrity (encryption), but also the authenticity of the provider of services. This can be complemented with the HSTS protocol, which instructs the browser to use only HTTPS; in this way, the risk of MitM attacks is avoided, especially if the HSTS header is preloaded.

- It is necessary to analyse whether the recipient of the data needs to access it in plain text. If this is not the case (such as a processor that only has to host the information), the data must be encrypted so that the processor cannot access it.

- If the recipient needs access to do calculations, the possibility of access to data only being possible with homomorphic encryption must be assessed. There are situations in which this is of particular interest, such as when wishing to use the cloud to store data and make calculations.

On the same lines, it is important to take into account that there are a large number of cryptographic protocols that enable a variety of tasks to be done while only revealing minimal information. For example, secret sharing schemes, zero-knowledge proofs, etc.

- It is also advisable to explore the possibility of using techniques that enable calculations to be done in a distributed way so that each agent involved in the calculation has its own data (secure multi-party computation – SMPC). On a similar line, when developing artificial intelligence models, it is advisable to use federated learning mechanisms so that the agents who have the data (individuals, bodies, etc.) can train the model in a distributed way, without having to give the data to a body that centralises the training.

3.6 Keeping and storing data

In this section, we refer both to the measures that must be taken to guarantee the confidentiality, integrity and availability of the information and the limitation of the period of storage.

3.6.1 Confidentiality, integrity and availability of the information

The security of information must be guaranteed. This essentially means that personal information must not be accessible to unauthorised persons (**confidentiality**), it must not be improperly altered (**integrity**) and it must be available when needed (**availability**).

These three characteristics can be protected with different measures, which should be determined in accordance with the probability and severity of the existing risks.

Confidentiality, integrity and availability are different aspects of security that, despite being linked, can require different measures. Therefore, different measures are presented below, grouped according to the different security dimensions, although it is important to take into account that some measures might simultaneously cover more than one of these dimensions. For example, preventing improper access to information (confidentiality) also helps to protect its integrity and availability (for example, by making hijacking it more difficult). Ultimately, although we have chosen this separate structure by way of example, it is necessary to bear in mind that security is a holistic concept.

Confidentiality and integrity

To guarantee confidentiality and integrity, it is essential to establish access controls for the information.

Important aspects

- A **permissions policy** should be implemented that only grants the essential operational permissions. Roles must be defined so that individual users only have the essential permissions to exercise their own functions (need-to-know principle). Even senior managers of the organisation that controls the processing should not have unlimited permissions if their ordinary duties do not require it, although they can be granted on a one-off basis if needed at any time.

- To identify users, electronic **certificates** can be requested, which offer a high degree of robustness. However, when the requirement this means of identification is not proportionate or appropriate for other reasons, other identification mechanisms must be used, such as concerted key ones.[15]

- If it is possible and proportionate in terms of security versus usability, it is advisable to establish a system based on **multi-factor authentication**. Systems based on a single factor (such as a password) can sometimes be easily breached (for example, by poor storage, lack of robustness or shoulder surfing). Accordingly, the use of double factor systems is recommended. These have to combine different aspects, from (i) something I know (password), (ii) something I have (a token or a mobile phone) or (iii) something I am (biometric aspects).

Regarding the possibility of using biometric data, it is important to bear in mind that these are special category data and so can only be used if any of the exceptions provided for by the GDPR[16] apply and when their use is proportionate. Therefore, we recommended that you do not choose this option unless it is strictly necessary.

- **Managing access keys**: issuing access keys to people who have to access the information, whether they are the end users of the service or the people entrusted with managing the application, must take into account various aspects that affect both the characteristics of the key and the creation, management and retrieval process.

[15] In the case of public administrations, section 9 of Spain's Law 39/2015, of 1 October, on the common administrative procedure of the public administrations governs electronic means of identification of members of the public.
[16] See article 9.2 of the GDPR.

- **Secure credentials:** requiring strong passwords protects against password guessing[17] or dictionary and brute force [18] attacks: and so, it is advisable that the application itself requires a minimum number of characters, upper case, lower case and special characters.

Given the importance of having sufficiently strong passwords, it is advisable that when the user chooses a password, the system should not only indicate its level of strength but also prevent the use of ones that do not fulfil certain strength requirements.

Using default passwords in goods or services is not advisable.

To deal with these potential attacks, it is also possible to implement measures such as:

1. Restricting users in the case of multiple failed access attempts: the user's profile or IP is blocked after a certain number of failed attempts in a set time period.

2. Including a CAPTCHA in the validation process.

- The **process of generating and**, where applicable, **communicating** the password must guarantee that only the user knows the keys.

- The technological solution must be configured so that the controller has to store the **users' access credentials** through an electronic hash of the password.

- Establishing a mechanism that requires it to be **renewed periodically** is advisable.

- The application or website should be configured so that **the user session closes automatically, after a certain period** of inactivity, and so that **the cookies that store the session are invalidated,** so that to access it again it is necessary to identify yourself.

- It is necessary to have an **access and activity log** that makes it possible to analyse in detail and retrospectively any circumstance that might affect the information. There are multiple benefits to having an access log, as it is a deterrent in itself, it can help detect anomalous or suspicious actions and, if there has been any improper action, it makes it possible to collect evidence about what has actually happened. This evidence can help with taking internal or external corrective measures (including reporting a security breach, so that

[17] Password guessing: this can be done by people who know potential victims and so can guess their passwords, or by professionals who use Open Source Intelligence (OSINT) techniques to discover information such as date of birth, identity of children, etc., that is often used to create passwords.
[18] "Dictionary" or brute force attacks comprising multiple consecutive attempts.

users can protect themselves appropriately), as well as pursuing relevant liabilities.

As well as access control, it is also advisable to consider other aspects that make a decisive contribution to preserving confidentiality and integrity and that refer to “how” information is stored.

Important aspects

- Information must be stored in an **encrypted** format whenever possible.
- An electronic signature must be used when possible, to guarantee the authorship, date and time and integrity of the document.
- It is important to ensure that the anonymisation and pseudonymisation processes used have not become ineffective.

In addition, it is also advisable to carry out penetration testing to identify potential vulnerabilities, as well as to include an incident monitoring system that makes it possible to detect any anomaly quickly.

Important aspects

- It is necessary to carry out **periodic tests** for possible vulnerabilities in the protection of the confidentiality of the information against external attacks. There are three main types of external security audit: (i) black box, when the auditor attempts to gain access without prior knowledge of the system or application; (ii) white box, when the auditor starts with full knowledge of the system or service to be analysed; and (iii) grey box, when some type of information is provided to the auditor who is to attempt an attack. Black-box testing is generally recommended as it simulates a scenario that is as realistic as possible of someone external who does not have prior knowledge of the system that is being verified. Despite this, it might sometimes be of value to provide auditors with some information so they can simulate attacks from different levels of user privilege.

- It might be especially appropriate to **audit your own code**, to detect possible security risks.[19]

- Appropriate mechanisms, automatic if possible, must be implemented to **identify abnormal behaviour or data leaks** that enable a fast and appropriate reaction. For example, through measures that control the transit of information and make it possible to detect unusual behaviour and alert the controllers of the device or service.

Finally, it is necessary to note that it is important to adopt any other security measures that are **appropriate to the existing risks.**[20]

Availability

Important aspects

- Automated mechanisms must be put in place to create **backup copies** periodically in a different location. Although knowledge of particular circumstances that could affect availability in a specific processing (power cuts, lack of connectivity, fire, flooding, obsolescence of certain parts of pre-existing IT systems, etc.) might go beyond the area that corresponds to the developers, they must have an exhaustive knowledge of it when establishing an application's system for making copies and the retrieval system. The mechanism (periodical copies, mirror server, etc.) and frequency of the copies will depend on the probability and severity of a loss of availability. The 3-2-1 rule is often mentioned: three backup copies on at least two different media and with one of the copies outside the working environment (isolated).

- For this approach to be effective, it is necessary to configure **different accesses to the backup copies**. Otherwise, if a system administrator's credentials are compromised, the attacker could access the live information and the information in the different copies and corrupt all of it at the same time.

[19] As an example, there are products that make it possible to automate this type of audit. Specifically, it would be a case of verifying, for example, that access to the databases has been designed to ensure that it cannot be improperly accessed through XSS attacks (validating input fields before they are executed in the database itself).

[20] In the case of the public sector, in Spain, it is necessary to consider the National Security Scheme (ENS), approved by Royal Decree 311/2022, of 3 May. However, it is important to bear in mind that section 3 of the ENS expressly states that measures deriving from a relative risk analysis specifically carried out taking into account the protection of personal data take precedence.

- It is necessary to plan periodical tests of the recovery procedure, to safeguard the availability of the information and the continuity of the services.
- As availability also depends on factors such as (i) machinery (ii) supplies such as electricity and (iii) connectivity, it is advisable to asses the possibility of having redundancy these critical factors.

3.6.2 Limitation of the storage period

Data must only be stored for the time necessary to achieve the purpose pursued.

After this time, they must be erased.

Erasure is not the same as deletion, but rather it results in restricting[21] during the period for which the data must be stored to respond to potential liabilities. Once the restricting period has ended, the data must be destroyed or, if appropriate, anonymised.

Data that are to be stored for statistical or similar purposes must not be stored in a way that makes it possible to identify the persons concerned for longer than is strictly necessary to fulfil the purpose of the processing.

Measures

- Before starting to collect personal data, it is vital to decide on its storage period (including the restricting period) and inform the persons concerned of it.[22]
- Automated mechanisms to erase unnecessary data must be established.
- It is advisable to establish mechanisms that make it possible to block the erased or rectified data, for as long as they are needed to respond to potential liabilities, and facilitate their destruction once the restricting period has ended.

[21] Restricting involves identifying and setting aside data, with technical and organisational measures to prevent its processing, including viewing, apart from making the data available to judges and tribunals, the public prosecutors or the competent public authorities. In particular, to the data protection authorities, to pursue potential liabilities deriving from the processing so long as they are not barred by time limits (sec. 32 LOPDGDD).

[22] In the scope of the Catalan public administrations, the storage periods established in the document evaluation tables must be taken into account.

- If you wish to store data for statistical purposes beyond the established storage period, appropriate measures such as anonymisation of the information must be taken.

- Methods must be used that ensure effective elimination of information that is not being kept. In particular, it is necessary to erase files and units securely (overwriting the data multiple times) or, in the case of devices that are no longer used (hard discs, USB memory, etc.), through physical destruction, especially when they leave the control of the organisation.

- The time that the browsing data is stored on the user's device should be minimised. It is better to use session cookies rather than persistent cookies and, if persistent cookies are used, an expiry date should be set for them. Similarly, it is necessary to avoid the use of storage mechanisms that seek to store data without the user having any control of it. For example, zombie cookies, which use various techniques to regenerate when the user erases them.

4. Key measures to protect personal data

This guide identifies a range of measures that help improve the protection of personal information in one or more stages of processing.

This section covers three measures that the European data protection regulation specifically mentions because of their importance. These are encryption, anonymisation and pseudonymisation.

As these measures involve a clear benefit for the protection of personal data, they form part of the broader group that in the field of data protection is known as PET (privacy enhancing technologies).

4.1 Encryption

Encryption is a process that transforms information (plain text), so that it is not comprehensible (encrypted text). Encryption is carried out using an encryption key and only the person with the decryption key can reverse the process and access the information. In any case, the encrypted personal data are still personal data subject to the personal data protection rules.

Encryption systems are usually classified by the method used to encrypt and decrypt the information:

- **Symmetric or private key cryptosystems**: these use the same key for encryption and decryption.

- **Asymmetric or public key cryptosystems**: these use different keys for encryption and decryption. In this case, the user has two keys (public and private). When someone wants to send a message or communication to another user, the message is encrypted using the recipient's public key, so that only the recipient (who is the only person with the private key) can decrypt it.

The major advantage of public key systems is that people who want to communicate information in secret do not have to share their private keys at any moment. This system does require a mechanism so that the different parties can share their public key securely. This is done using digital certificates, in which a recognised entity attests to the validity of a key. This is known as public key infrastructure (PKI).

Given the high computational cost of public key cryptosystems (resulting from the complexity of the mathematical operations that are necessary), hybrid cryptosystems in which the plain text is encrypted with a randomly generated key are also an option. In turn, this key is encrypted with a public key algorithm and the message is attached. The recipient decrypts the key using his or her private key and, so, will be able to decrypt the text.

Encryption is especially recommended for protecting personal data, both from the point of view of confidentiality and of integrity. So much so that the data protection rules state that there is no need to report a breach of data security to the supervisory authority or inform the persons concerned of it if measures have been taken that make the data unintelligible to any unauthorised parties, such as encryption.

Notwithstanding what the risk analysis might reveal, it is especially advisable to encrypt information when processing a large amount of especially sensitive information or data, whether of special category or other data (such as economic data). In any case, as with any protection measure, it has to be proportionate with the existing risks, so that the benefits of implementing this measure are greater than the costs of doing so.

If the encrypted data have to be subjected to operations or calculations, homomorphic encryption can be especially useful. This makes it possible to perform certain algebraic operations on encrypted data. This can, for example, be useful when sensitive data are stored in the cloud and working with them is planned.

It is important to consider that encryption processes always involve some risk of decryption. These risks can be computational (encryption algorithms become obsolete over time) or associated with the management and storage of the keys used to encrypt and decrypt the information. Accordingly, it is advisable to review[23] the robustness of the algorithm over time and also ensure that the keys are generated and stored securely.

[23] For example, the Centro Criptológico Nacional (CCN) frequently publishes information and guides that might be of use. In May 2022 it published the Guía de seguridad de las TIC CCN-STIC 807 Criptología de empleo en el Esquema Nacional de Seguridad.

There are other cryptographic mechanisms that make it possible to obtain different features and so should also be assessed when developing an application, such as electronic signatures, hash functions, secret sharing schemes, etc.

4.2 Anonymisation

This process aims to prevent natural persons from being identified from a data set, without disproportionate efforts, whether directly or indirectly, and as such it is an irreversible process. As the data can no longer be attributed to natural persons, they are no longer classed as personal data. Consequently, the risks for the persons concerned reduce and the personal data protection rules no longer apply.

> Before starting any processing, it is advisable to analyse whether this can be done without using personal data or using anonymised data. If so, this option must be selected.

It is necessary to bear in mind that the process of anonymisation is in itself processing of personal data. Therefore, the personal data protection rules must be taken into account in this phase.

Likewise, in the anonymisation process, it is necessary to establish a separation of functions so that the people involved do not overlap with those who are linked to the processing of the data once they have been anonymised.

Anonymisation techniques alter the data to protect privacy. This inevitably has an impact on the usefulness of the data.

4.2.1 Anonymisation techniques

Anonymisation techniques are classified in two large categories, according to the procedure used: masking and creating synthetic data.

Masking techniques

Masking techniques start with the original data and modify them. A relationship is maintained between the records of the original data and the records of the masked data. This relationship means that the risk of reidentification has to be taken into account.

The range of masking techniques is very extensive.[24] Depending on how they affect the accuracy of the data, these techniques are classified as perturbative and non-perturbative. Perturbative techniques change its accuracy. For example, the following are perturbative masking techniques:

- **Adding noise**. A certain level of random noise is added to the original data so that the values are not exact. The risk associated with these data will depend on the level of noise added. The more noise is added, the more uncertainty there will be about the original data and, therefore, the risk will be lower.

- **Microaggregation**. This involves aggregating records in groups, with a minimum fixed cardinality, and replacing each group with a new record that is representative of the group. In this way, the records of the anonymised data no longer correspond to a specific person, but instead refer to a group of people. The larger the group, the lower the risk.

- **Rank swapping**. This involves replacing the value of an attribute from a record with the value of another record that is within a range of the initial value. Unlike when adding noise, the distribution of data in each attribute is preserved in rank swapping.

The properties of each masking technique can make it more suitable than another, in a specific situation. However, we must always bear in mind that these techniques alter the accuracy of the data. From the point of view of privacy, this is positive because, even if a record is reidentified, there would be uncertainty about the accuracy of the data it contains. However, this loss of accuracy in the data, means that these techniques cannot be used in certain situations.

Non-perturbative masking techniques can be used to maintain the reliability of the data. These reduce the granularity of the information so that a record cannot be reidentified. For example, non-perturbative masking techniques include:

- **Erasure**. This involves eliminating certain data, so that the person concerned is no longer identifiable. Erasure is very common with identifying variables; these variables do not generally provide much statistical value and so erasing them is not usually problematic. On the other hand, reidentification can also result from the combination of different attributes that are not identifiers on their own (these combinations of attributes are known as quasi-identifiers). To avoid reidentification, it is necessary to erase quasi-identifiers that can make it possible to identify specific people.

- **Generalisation (or global recoding)**. This involves replacing the information in a variable or attribute so that the new value corresponds with a broader category. For example, age can be replaced with age ranges.

Generating synthetic data

[24]For more information: A Network of Excellence in the European Statistical System in the field of Statistical Disclosure Control.

Techniques for generating synthetic data derive new data from a model of the original data. There is no direct relationship between the original data and the synthetic data.

The idea is that the model has to preserve the statistical properties of the data we want to analyse. It is necessary to bear in mind that the synthetic data only contain certain properties. Therefore, using this type of data can limit the types of analyses that can be done.

The fact that synthetic data do not reproduce the original data considered in isolation means that this technique is regarded as secure against the risk of reidentification. However, this fact, as well as the risk of disclosure, depend on the model that has been used. A model that is too precise can entail elevated risks. If the model only includes some statistical properties of the original data that we wish to preserve, the risk of reidentification can be controlled. However, the use of very complex models can give rise to overfitting; that is to say, that the model is not a representation of the statistical properties of the original data, but instead represents specific data. In this case, the risk of reidentification can be high.

Therefore, it is critical to determine the properties of the data that that are to be processed. For example, in the development and testing stage, it may be enough to preserve the syntactic and semantic validation of the data (that is to say, that they have the appropriate types and that there are no combinations without meaning). Data generated in this way avoid any risk.

4.2.2 Risks of anonymisation

Anonymisation processes do not fully guarantee that it will be impossible to discover personal information from the anonymised data. Once the techniques described above have been applied, it is necessary to assess what the level of risk is and adjust the anonymisation.[25]

Many anonymisation techniques are based on assumptions about the information available to a potential attacker. The more information is available, the easier it will be to reidentify a record. It can be very hard to determine what information is available externally and this information can also change over time, and so it is necessary to assess the risks that exist at each moment.

In the anonymisation processes that have just been described, there are privacy models that seek to give guarantees of privacy at the moment of carrying out the anonymisation. In this way, the risk is first identified and then an anonymisation technique is used to deal with it.

The best known privacy models include *k*-anonymity (and its related models, such as *l*-diversity and *t*-closeness) and differential privacy.

[25] Regarding the effectiveness of anonymisation, see chapter 2 "How do we ensure anonymisation is effective" of the guide to anonymisation, pseudonymisation and privacy enhancing technologies (PET) that is being prepared by the *Information Commissioner's Office*.

k-anonymity

k-anonymity deals with the risk of reidentification by making it so that each record can be associated with a group of *k* people. *k*-anonymity assumes that reidentification occurs through a particular set of attributes (quasi-identifiers) and requires each combination of these quasi-identifiers that appears in the anonymised data to be repeated at least *k* times.

The most common ways of obtaining *k*-anonymity are:

- **Generalisation and erasure**. The granularity of the information in the quasi-identifiers is reduced, so that each combination of values present in the anonymised data is repeated *k* times.

- **Microaggregation**. Microaggregation is applied to quasi-identifiers with groups of size *k*.

Specific people can be identified in *k*-anonymity when the variability of an attribute in a group of *k*-anonymous records is small. To avoid this problem, other privacy models have been developed, such as *l*-diversity and *t*-closeness, that require minimal variability.

Differential privacy

Differential privacy is a privacy model for database consultation. In other words, it does not generate a new anonymised database that can be analysed, but instead the consultations are done on the original database and the value of the answer is altered to protect privacy.

Differential privacy is recognised for giving strong guarantees of protection of privacy which, furthermore, are independent of the externally available information. However, strict application of differential privacy has many limitations owing to its significant impact on the utility of the data. For example, differential privacy is used in the census in the USA, but to maintain the utility of the data, such extreme parameters have to be applied that all of the guarantees are lost[26].

4.3 Pseudonymisation

Pseudonymisation is a process that makes it impossible to identify the natural person to whom data correspond without using additional information that has to be stored separately and subjected to technical and organisational measures to avoid the reidentification of the data subjects[27]. In other words, only the person who has carried out the pseudonymisation can link information to identified or identifiable people. Third persons cannot establish this relationship.

---

[26] Differential Privacy for census data explained.

[27] For this reason, the data resulting from pseudonymisation are still classed as personal data and so are still subject to the requirements of the GDPR. Similarly, the GDPR encourages the use of pseudonyms in the processing of personal data. The GDPR also considers that pseudonymisation makes it possible to reduce the risks for data subjects and contributes to compliance with the rules

Unlike anonymisation, pseudonymisation is a reversible process. Pseudonymised personal data are still personal data and so the data protection rules apply to them.

Some techniques for making pseudonymisation effective are:

- Use of pseudonyms: only the person who has the correspondence between the pseudonym and the real identity should be able to attribute the information to the individual person.
- Use of random codes that cannot be predicted by third persons.
- Encryption is also commonly regarded as a pseudonymisation technique, given that the process is reversible through the use of a key.

As with anonymisation, for pseudonymisation to be fully effective, it is advisable to establish a separation of roles so that the people involved in the pseudonymisation process do not overlap with those who are linked to the processing of the data after they have been pseudonymised. It is also necessary to bear in mind that pseudonymisation processes do not 100% guarantee that third parties cannot reidentify the people affected, and so it is necessary to analyse this risk and review it periodically.

It is also necessary to consider the possibility in certain contexts (for example, social networks, metaverse, etc.) of using digital identities that are different from the user's real identity. Although the use of aliases or avatars is not strictly a pseudonymisation technique, it can be especially useful for preserving a degree of privacy, as it still a layer of protection between users' real identities and their actions in the digital ecosystem with these other identities. However, this mechanism does not guarantee anonymisation or even pseudonymisation, given that depending on the context and the associated information, the person concerned might still be identified. Therefore, they are classed as personal data.

5. Data protection legislation

General legislation

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).

Organic Law 3/2018, of 5 December, on personal data protection and guaranteeing digital rights (LOPGDD)

Organic Law 7/2021, of 26 May, on the protection of personal data processed for reasons of prevention, detection, investigation and prosecution of criminal offences and execution of criminal sanctions (LOPDSPJP).

Royal Decree 311/2022, of 3 May, regulating the National Security Scheme (ENS).

Other legislation

Law 34/2002, of 11 July, of information society services and electronic commerce (LSSICE) (sections 21, 22, 38.3 *c*, *d* and *i*, 38.4 *d, g* and *h*, 43.1.

6. References

Privacy by design and by default

Privacy by Design. The 7 Foundational Principles. Ann Cavoukian. Ph.D. Information and Privacy Commissioner, Ontario, Canada.

Privacy Design Strategies (The Little Blue Book), JAAP-HENK HOEPMAN (2022).

Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices. (2012) Ann Cavoukian, Ph.D.

Guidelines 4/2019 on Article 25. Data Protection by Design and by Default. (version 2.0, 2020). European Data Protection Board.

Data protection by design and by default. Information Commissioner's Office (ICO).

Guide GDPR CNIL de l'equip de development. Commission Nationale d'Informatique et des Libertés (CNIL).

Software Development with Data Protection by Design and by Default, (2017). Norwegian Data Protection Authority.
Guía de privacidad desde el diseño. Agencia Española de Protección de Datos (AEPD).

Guía de protección de datos por defecto. Agencia Española de Protección de Datos (AEPD).

Data protection by design e data protection by default. Garante per la Protezioni dei Dati Personali (GPDP).

Report "Data protection engineering: From theory to practice" (2022). ENISA.

Report Privacy and data protection by design: from policy to engineering (2014). ENISA.

Other resources

Dictamen 02/2013 sobre las aplicaciones de los dispositivos inteligentes, (2013). Grupo de trabajo «artículo 29 sobre protección de datos».

Guidelines 3/2022 on Dark patterns in social media platform interfaces: How to recognise and avoid them (v. 1.0). European Data Protection Board.

Deceptive design. Harry Brignull.

PETs Maturity Assessment Repository (2019). ENISA.

Protecting privacy in practice. (2019). The Royal Society.

Privacy patterns. Privacy Patterns.org.

Web Security Testing Guide. Open Web Application Security Project (OWASP).

Cross Site Scripting Prevention Cheat Sheet. Open Web Application Security Project (OWASP).

Dictamen 05/2014, sobre técnicas de anonimización. Grupo de trabajo sobre protección de datos del artículo 29.

Data Pseudonymisation: Advanced Techniques and Use Cases (2021). ENISA.

Anonymisation, pseudonymisation and privacy enhancing technologies guidance. (2021). Information Commissioner's Office (ICO).

Orientaciones y garantías en los procedimientos de anonimización de datos personales. Agencia Española de Protección de Datos (AEPD).

Introducción al Hash como técnica de seudonimización de datos personales. Agencia Española de Protección de Datos (AEPD).

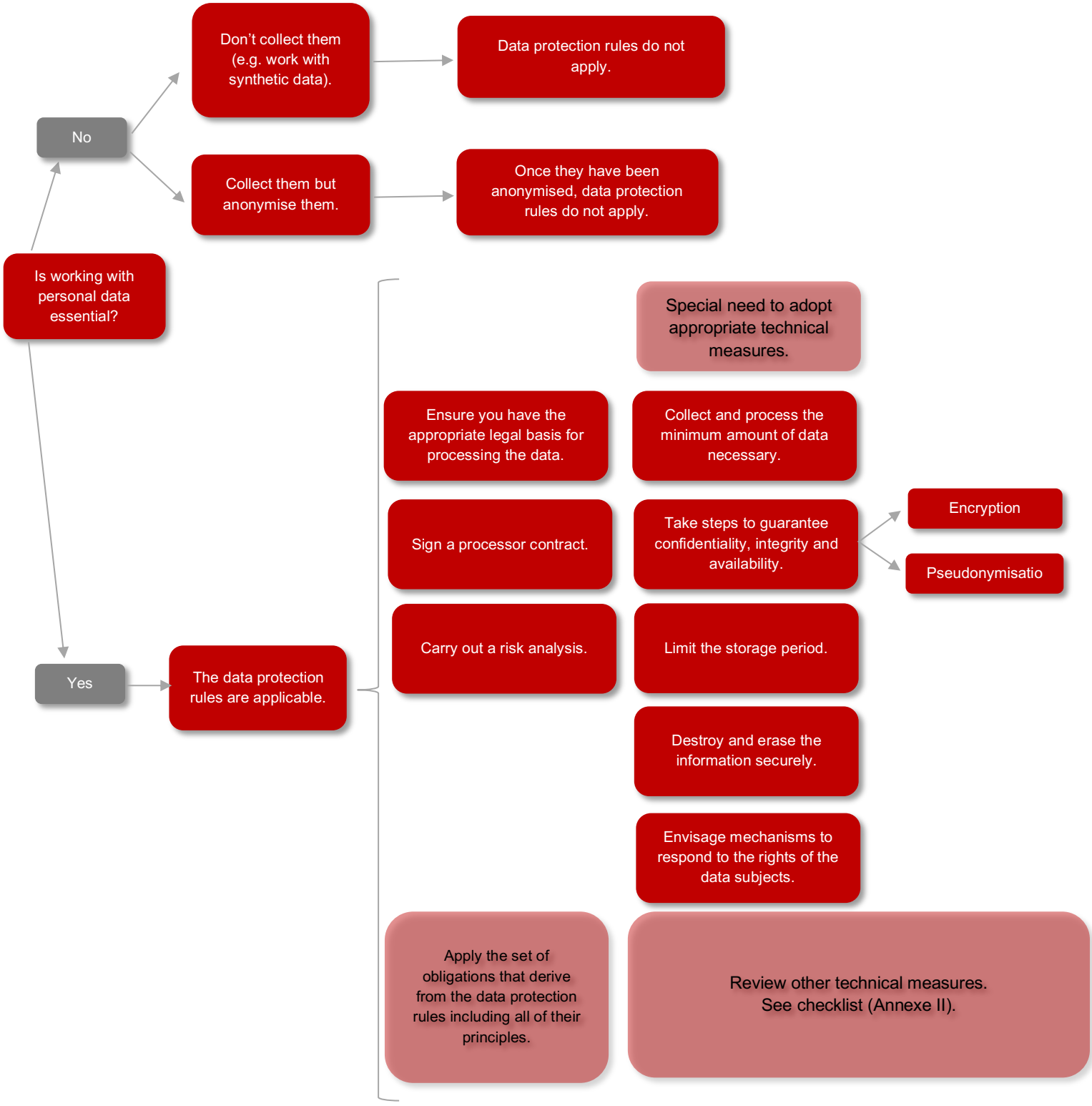
10 malentendidos relacionados con la anonimización. Agencia Española de Protección de Datos (AEPD).

Cifrado y Privacidad III: Cifrado Homomórfico (2020). Agencia Española de Protección de Datos (AEPD).

Guia pràctica Avaluació d'impacte relativa a la protecció de dades (2019). Catalan Data Protection Authority (APDCAT).

Guidance on AI and data protection. Information Commissioner's Office (ICO)

Annexe I. Prior analysis



Annexe II Checklist

Design stage

☐ Introduce design strategies to guarantee data protection
☐ Apply privacy patterns

Development and testing stage

☐ Sign a processing contract if the developer or a third party contracted by the developed have to access personal data.
☐ Adequately separate the production and development and testing environments.
☐ Assess the possibility of working with fictitious data, in testing environments.
☐ Minimise the amount of data collected in the testing stage.
☐ Assess the possibility of anonymising or pseudonymising the data collected in the testing stage or using synthetic data.
☐ Ensure the information is hosted or processed is in a country that offers appropriate guarantees.
☐ Use the Wi-Fi protocol that provides the best degree of security.
☐ Review the quality of the code used and follow the recommendations of safe programming guides.
☐ Carry out a risk analysis to determine the security measures.
☐ Anticipate staff training to guarantee minimisation of risks at all times.

Data collection

☐ Only collect essential data. This affects: :
- Data collected using forms.
- Technical data that are collected (IP, MAC, geolocation, etc.).
- Inserted code that might transfer information.
- Other ways of collecting data.

☐ Assess the possibility of anonymising or pseudonymising the personal data collected.
☐ In particular minimise the collection of special categories of data.
☐ Distinguish clearly between information that it is compulsory to provide and optional information.
☐ Only require permissions to access and collect information that is relevant to the purpose pursued.
☐ Assess the possibility of doing processing directly on users' devices.
☐ Appropriately inform the persons concerned about how their data will be processed.
☐ Make copies of the information clauses existing at each moment and apply a verifiable timestamp.
☐ Refrain from implementing dark patterns.
☐ Verify that the requirements for consent to be valid, where necessary, have been fulfilled:
- Boxes cannot be pre-ticked.
- Consent must be different for each purpose.
- Consents cannot be connected or conditional on each other.

- Establish mechanisms that make it possible to verify the identity of the person giving consent.
- Verify the identity and age of the person who gives consent.
- It must be possible to withdraw consent with mechanisms of equivalent complexity to those used to grant it.

☐ Conserve proof of the consent obtained.

Use of data

☐ Implement mechanisms that make it possible to classify information appropriately, in accordance with the purposes and processing to which it will be subjected.
☐ Incorporate measures that facilitate the exercise of and attention to rights.

Communication or disclosure of data

☐ When dissemination is based on the consent of the person concerned, mechanisms must be put in place so that, by default, data are not accessible to third persons.
☐ Establish controls on who can access information.
☐ Allow for automatic depublication of the information once its period of use has ended.
☐ Include end-to-end encryption in communications and, if necessary, the possibility of using VPN connections.
☐ Use https protocols for online services.
☐ Analyse the possibility of disclosing the encrypted data.
☐ In the development and learning stage of artificial intelligence models, apply federated learning systems.

Keeping and storing data

Confidentiality and integrity

☐ Establish an appropriate permissions policy.
☐ Explore the option of requiring the following identification and authentication mechanisms:

- Use of electronic certificates
- Systems based on multi-factor authentication
- Concerted key systems

☐ Guarantee the security and the process of management and safekeeping of the access keys:

- Guarantee the robustness of credentials.
- Implement technical mechanisms to prevent dictionary or brute force attacks.
- Establish a secure procedure for password management and recovery.
- Guarantee that only the user can know his or her credentials.
- Encrypt keys.
- Oblige users to change their passwords periodically.

☐ Set a session time-out that clears the cache memory.
☐ Establish an access and activity log.
☐ Keep information encrypted whenever possible (using homomorphic encryption if necessary) .
☐ Use electronic signatures, if possible.
☐ Ensure that the anonymisation or pseudonymisation processes used do not lose effectiveness.
☐ Carry out tests to check for vulnerabilities to attacks and schedule them periodically.
☐ Audit code periodically.
☐ Establish automatic mechanisms to detect intrusions and data leaks.
☐ Adopt any other security measure that is appropriate to the existing risks.

Availability

☐ Configure making back-up copies.
☐ Establish different permissions for accessing back-up copies.
☐ Plan information recovery tests based on copies.
☐ Assess redundancy in machinery, electricity supply and connectivity networks.

Limit the period of storage

☐ Define the storage period for the information and inform the persons concerned of it.
☐ Implement automated mechanisms to facilitate compliance with the erasure and restricting periods established.
☐ Apply measures such as anonymisation if data are to be stored for statistical purposes beyond the established storage period.
☐ Destroy hardware effectively in the case of machinery that is no longer used.
☐ Check you are not using persistent cookies or, alternatively, set an expiry date for this type of cookie.