

# Guidelines



## **Guidelines 01/2025 on Pseudonymisation**

**Adopted on 16 January 2025**



## EXECUTIVE SUMMARY

The GDPR defines the term ‘pseudonymisation’ for the first time in EU law and refers to it several times as a safeguard that may be appropriate and effective for the fulfilment of certain data protection obligations.

As per that definition, pseudonymisation can reduce the risks to the data subjects by preventing the attribution of personal data to natural persons<sup>1</sup> in the course of the processing of the data, and in the event of unauthorised access or use.

Applying pseudonymisation, controllers can thus retain the option to analyse the data, and, optionally, to merge different records relating to the same person. Pseudonymisation can also and often will be set up so that it is possible to revert to the original data. Thus, controllers can process personal data in original form in some stages of the processing, and in pseudonymised form in others.

Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person,<sup>2</sup> and is therefore personal. This statement also holds true if pseudonymised data and additional information are not in the hands of the same person. Even if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data can be considered anonymous only if the conditions for anonymity are met.

The GDPR does not impose a general obligation to use pseudonymisation. The explicit introduction of pseudonymisation is not intended to preclude any other measures of data protection (Rec. 28 GDPR). It is the responsibility of the controller to decide on the choice of means for meeting its obligations having regard to the accountability principle. Depending on the nature, scope, context and purposes of processing, and the risks involved in it, controllers may need to apply pseudonymisation in order to meet the requirements of EU data protection law, in particular in order to adhere to the data minimisation principle, to implement data protection by design and by default, or to ensure a level of security appropriate to the risk. In some specific situations, Union or Member State law may mandate pseudonymisation.

The risk reduction resulting from pseudonymisation may enable controllers to rely on legitimate interests under Art. 6(1)(f) GDPR as the legal basis for their processing provided they meet the other requirements of that subparagraph; contribute to establishing compatibility of further processing according to Art. 6(4) GDPR; or help guarantee an essentially equivalent level of protection for data they intend to export.

Finally, the contribution of pseudonymisation to data protection by design and default, and the assurance of a level of security appropriate to risk may make other measures redundant – even though pseudonymisation alone will normally not be a sufficient measure for either.

Controllers should establish and precisely define the risks they intend to address with pseudonymisation. The intended reduction of those risks constitutes the *objective* of pseudonymisation within the concrete processing activity. Controllers should shape pseudonymisation in a way that guarantees that it is *effective* in reaching this objective.

---

<sup>1</sup> For a definition of what it means to attribute data to a natural person see paragraph 17. Prevention of attribution does not imply anonymity of the data.

<sup>2</sup> Rec. 26 GDPR.

Controllers may define the context in which pseudonymisation is to preclude attribution of data to specific data subjects. This context will be called the *pseudonymisation domain* in these guidelines. The pseudonymisation domain does not have to be all-encompassing, but may be restricted to defined entities, most often to the set of all authorised recipients of the personal data that will process the data for a given purpose. The effectiveness of pseudonymisation in the implementation of data-protection principles or in the assurance of a level of security appropriate to the risk is highly dependent on the choice of the pseudonymisation domain and its isolation from additional information that allows the attribution of pseudonymised data to specific individuals.

Thus, pseudonymisation is a safeguard that can be applied by controllers to meet the requirements of data protection law and, in particular, to demonstrate compliance with the data protection principles in accordance with Art 5(2) GDPR. These guidelines will help controllers to choose effective techniques for the modification of original data, to protect pseudonymised data from unauthorised attribution, and to manage user rights when processing pseudonymised data.

Controllers must always bear in mind that pseudonymised data, which could be attributed to a natural person by the use of additional information, remains information related to an identifiable natural person, and thus is personal data (Rec. 26 GDPR). Therefore, the processing of such data needs to comply with the GDPR, including the principles of lawfulness, transparency, and confidentiality under Art. 5 GDPR, and the requirements of Art. 6 GDPR. Controllers must maintain an appropriate level of security by implementing further technical and organisational measures. Finally, controllers must ensure transparency, and need to facilitate the exercise of the data subject rights set out in Chapter III of the GDPR, unless the exception provided for in Art. 11(2) and 12(2) GDPR applies.

## Table of Contents

|                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------|----|
| Executive summary .....                                                                                                  | 3  |
| 1 Introduction.....                                                                                                      | 7  |
| 2 Definitions and legal analysis .....                                                                                   | 9  |
| 2.1 Legal definition of pseudonymisation .....                                                                           | 9  |
| 2.2 Objectives and advantages of pseudonymisation .....                                                                  | 10 |
| 2.2.1 Risk reduction .....                                                                                               | 10 |
| 2.2.2 Analysis of pseudonymised data and planned attribution.....                                                        | 11 |
| 2.3 Pseudonymisation domain and available means for attribution.....                                                     | 12 |
| 2.4 Meeting data-protection requirements using pseudonymisation.....                                                     | 13 |
| 2.4.1 Pseudonymisation as an effective measure for data protection by design and by default.....                         | 13 |
| 2.4.2 Ensuring a level of security appropriate to the risk.....                                                          | 15 |
| 2.4.3 Pseudonymisation as a supplementary measure for third country data transfers .....                                 | 16 |
| 2.5 Transmission of pseudonymised data to third parties .....                                                            | 17 |
| 2.6 Implications for the rights of the data subjects .....                                                               | 19 |
| 2.7 Unauthorised reversal of pseudonymisation .....                                                                      | 19 |
| 3 Technical measures and safeguards for pseudonymisation .....                                                           | 20 |
| 3.1 Pseudonymising transformation .....                                                                                  | 20 |
| 3.1.1 Structure of the pseudonymising transformation .....                                                               | 20 |
| 3.1.2 Types of pseudonymising transformations .....                                                                      | 21 |
| 3.1.3 Modification of original data necessary for the objectives of pseudonymisation .....                               | 22 |
| 3.1.4 Pseudonymisation in the course of data collection .....                                                            | 23 |
| 3.2 Technical and organisational measures preventing unauthorised attribution of pseudonymised data to individuals ..... | 24 |
| 3.2.1 Preventing reversal of the pseudonymising transformation.....                                                      | 24 |
| 3.2.2 Securing the pseudonymisation domain .....                                                                         | 25 |
| 3.3 Linking pseudonymised data .....                                                                                     | 25 |
| 3.3.1 Controlling the scope for the linkage of pseudonymised data .....                                                  | 26 |
| 3.3.2 Linking data pseudonymised by different controllers .....                                                          | 27 |
| 3.4 Summary of procedures for pseudonymisation .....                                                                     | 29 |
| Annex – Examples of the Application of Pseudonymisation .....                                                            | 31 |
| Example 1: Data minimisation and confidentiality in internal analysis .....                                              | 31 |
| Example 2: Separation of functions allowing for data minimisation, purpose limitation, and confidentiality.....          | 32 |
| Example 3: Data minimisation and purpose limitation in the course of external analysis .....                             | 34 |
| Example 4: Safeguarding identity – confidentiality and accuracy .....                                                    | 36 |

|                                                                                                                           |    |
|---------------------------------------------------------------------------------------------------------------------------|----|
| Example 5: Secondary use for research .....                                                                               | 37 |
| Example 6: Reduction of confidentiality risks .....                                                                       | 39 |
| Example 7: Risk reduction as a factor in the balancing of interests, and ascertainment of compatibility of purposes ..... | 40 |
| Example 8: Risk reduction justifying further processing.....                                                              | 42 |
| Example 9: Supplementary measure .....                                                                                    | 43 |
| Example 10: Granting access rights to pseudonymised data .....                                                            | 45 |
| Glossary .....                                                                                                            | 45 |

# The European Data Protection Board

Having regard to Article 70(1)(e) of the Regulation 2016/679/EU 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, (hereinafter “GDPR”),

Having regard to the EEA Agreement and in particular to Annex XI and Protocol 37 thereof, as amended by the Decision of the EEA joint Committee No 154/2018 of 6 July 2018<sup>3</sup>,

Having regard to Article 12 and Article 22 of its Rules of Procedure,

**HAS ADOPTED THE FOLLOWING GUIDELINES**

## 1 INTRODUCTION

1. These guidelines intend to clarify the use and benefits of pseudonymisation for controllers and processors.
2. The GDPR defines the term ‘pseudonymisation’ for the first time in EU law and refers to it several times as a safeguard that may be appropriate and effective for the fulfilment of data protection obligations. EU and Member State law is relying on that definition when requiring or recommending the use of pseudonymisation, see, e.g., Art. 17(1)(g) of Regulation (EU) 2023/2854 or Art. 44(3) of the European Commission’s Proposal for a Regulation on the European Health Data Space<sup>4</sup>.
3. Art. 4(5) GDPR defines pseudonymisation as a manner of processing with prescribed effects and calls for certain measures by which those effects are to be achieved.
4. The desired effect of pseudonymisation is to control the attribution of personal data to specific data subjects by denying this ability to some persons or parties. The GDPR does not specify who those persons or parties are to be, leaving it – absent specific requirements by other EU or Member State law – to the controller’s decision. Recital 29 makes clear that, when the pseudonymisation is carried out within the same controller, the effects might be confined to specific parts of the controller’s organisation.
5. There are three actions controllers should take to achieve the desired effect. First, they need to modify or transform<sup>5</sup> the data. Second, they need to keep additional information for attributing the personal data to a specific data subject separately, i.e. separate from those who are to be prevented from achieving such an attribution. Last, they need to apply technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. In particular, they need to prevent the unauthorised use of the

---

<sup>3</sup> References to “Member States” made throughout this document should be understood as references to “EEA Member States”.

<sup>4</sup> See <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197>.

<sup>5</sup> The guidelines use the terms “transform” and “transformation” to refer to a modification of the data for pseudonymisation and fitness for subsequent processing in pseudonymised form.

additional information they control and control the flow of pseudonymised data to the extent possible.

6. Pseudonymisation as a technical measure for the protection of the privacy of individuals has been around for a long time. The common understanding of pseudonymisation involves the replacement of identifiers of individuals by pseudonyms. In this process, the pseudonyms are to be chosen in a way that they do not reveal the identity of the individual they are assigned to. The legal definition presented by the GDPR differs from that understanding in three significant ways.
7. First, the legal definition takes a more comprehensive view of the effect of pseudonymisation. It shall no longer be possible to attribute the personal data to a specific data subject without the use of additional information. This requires a look at all parts of the personal data, not only the pseudonyms.
8. Second, it does not even explicitly require the replacement of direct identifiers<sup>6</sup> by pseudonyms. It is clear that direct identifiers need to be removed from data if those data are not to be attributed to individuals. Moreover, Art. 4(5) GDPR provides for the retention of additional information that allows attribution of the data to individuals. During attribution, a link will be made between the data or parts thereof to identifiers of the individuals. This link will usually, but not necessarily, start with pseudonyms inserted into the data, precisely with the aim of allowing for attribution in authorised circumstances.
9. Third, it requires more than just the transformation of data. It requires additional technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. Typically such measures limit access to the retained additional information (e.g. keys or tables of pseudonyms), and control the flow of pseudonymised data.
10. These guidelines will first have a closer look at the legal definition of pseudonymisation and the terms used therein. What is attribution? What is to be considered additional information? A key aspect evolving from this analysis are the many options for controllers to tailor their pseudonymisation processes to the objectives they intend to achieve. The guidelines introduce a new concept, called pseudonymisation domain, to capture one aspect of that freedom: to determine who should be precluded from attributing the pseudonymised data to individuals.
11. In a second step, the guidelines show how controllers and processors can use pseudonymisation to meet data-protection requirements. While pseudonymisation is a powerful and relevant measure, the document shows that it will always need to be complemented by further measures. The Guidelines highlight the benefits of pseudonymisation. They show in particular how pseudonymisation serves as a measure for data protection by design and by default, and as a measure contributing to ensuring a level of security appropriate to the risk of processing. At least in the latter case, the effect of pseudonymisation will have to be measured against the capabilities of persons or parties acting without authorisation.
12. In a third part, the guidelines will look at the implementation of pseudonymisation. How should personal data be transformed to pseudonymise it? How should unauthorised attribution be prevented? How should different pseudonymised data sets be linked, and how could such linkage be controlled?

---

<sup>6</sup> See the definition of this term in the glossary.

13. Often it is important to look beyond the confines of the organisation of a single controller pseudonymising the data. Personal data is frequently pseudonymised before it is shared with other controllers or to processors to limit the risks involved in that sharing. Pseudonymised data coming from different controllers might need to be brought together and linked. Or, in contrast, different data sets need to be pseudonymised in a way that assures that they cannot be linked.
14. The guidelines close with a summary of procedures for pseudonymisation, which is presented not as a prescription, but as guidance for the steps controllers and processors could take to ensure that the pseudonymisation they implement is effective.
15. Annexed to the guidelines, the readers will find several examples showing the use of pseudonymisation to limit risks for data subjects in real life scenarios.

## 2 DEFINITIONS AND LEGAL ANALYSIS

### 2.1 Legal definition of pseudonymisation

16. Pseudonymisation is defined in Art. 4(5) GDPR 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 is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”*
17. To **attribute** data to a specific (identified) person means to establish that the data relate to that person. To attribute data to an identifiable person means to link the data to other information with reference to which the natural person could be identified. Such a link could be established on the basis of one or several identifiers or identifying attributes.
18. Pseudonymisation generally requires the application of a **pseudonymising transformation**. This is a procedure that modifies original data in a way that the result—the **pseudonymised data**—cannot be attributed to a specific data subject without additional information. The pseudonymising transformation may and regularly does replace part of the original data with one or several **pseudonyms**—new identifiers that can be attributed to data subjects only using additional information. For details, see section 3.1.1. These guidelines will call controllers that use pseudonymisation as a safeguard and modify original data according to Art. 4(5) GDPR **pseudonymising controllers**. Similar terminology is used for processors.
19. **Additional information** is information whose use enables the attribution of pseudonymised data to identified or identifiable persons. The generation, or use of additional information is an inherent part of the pseudonymising transformation.
20. It includes information that is retained as part of the pseudonymisation process for consistent pseudonymisation of different items of personal data relating to the same data subject and information that is kept to be used for later reversal of pseudonymisation. Such additional information may consist of tables matching pseudonyms with the identifying attributes they replace. It may also consist of cryptographic keys. Additional information kept by a pseudonymising controller or processor must be subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. In particular, the additional information is not to be disclosed to or used by persons processing

the pseudonymised data. Such additional information may itself be personal data and so also subject to the GDPR.

21. Additional information may also exist beyond the immediate control of the pseudonymising controller or processor. The pseudonymising controller or processor should take such information into account in the assessment of the effectiveness of pseudonymisation to the extent such information can reasonably be expected to be available. For example, information from publicly accessible sources, such as posts in a social media or an online forum, may contribute to the attribution of pseudonymised data to data subjects. This assessment will help determine if any further measures need to be implemented to avoid attribution.
22. Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person,<sup>7</sup> and is therefore personal. This statement also holds true if pseudonymised data and additional information are not in the hands of the same person. If pseudonymised data and additional information could be combined having regard to the means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal. Even if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data becomes anonymous only if the conditions for anonymity are met.
23. Pseudonymisation is a technical and organisational measure that allows controllers and processors to reduce the risks to data subjects and meet their data-protection obligations, for example under Art. 25 or 32 GDPR. Therefore, if a controller processes personal data and applies pseudonymisation in the process, then the legal basis for the processing of the personal data extends to all processing operations needed to apply the pseudonymising transformation.
24. Union or Member State law may require pseudonymisation of personal data for the processing of personal data in specific situations, e.g. when providing for a legal basis under Art. 6(1)(c) or (e) GDPR in accordance with Art. 6(3) GDPR, or as a further condition in accordance with Art. 9(4) GDPR. In such cases, the law may also lay down specific requirements the pseudonymisation process or output has to meet, or the objectives it should achieve.
25. When such specific mandates for pseudonymisation are absent, controllers themselves may define the objectives<sup>8</sup> that pseudonymisation should achieve. Those objectives may be connected with the processing they intend to perform themselves or with any subsequent processing of the pseudonymised data by recipients of those data.

## 2.2 Objectives and advantages of pseudonymisation

26. In accordance with Rec. 28 GDPR, pseudonymising data reduces risks for data subjects while allowing general analysis.

### 2.2.1 Risk reduction

27. Pseudonymisation reduces *confidentiality risks* when done *effectively*, which presumes that the additional information referred to in paragraph 20 are subject to the measures provided in Art.

---

<sup>7</sup> Rec. 26 GDPR.

<sup>8</sup> These guidelines distinguish between the *purpose* of the processing of personal data according to Art. 5(1)(b) GDPR, and the *objective* of a safeguard like pseudonymisation employed during that processing, which consists in a certain aspect of the fulfilment of data protection obligations.

4(5) GDPR. It does so in two ways. First, it prevents the disclosure of direct identifiers of data subjects to some or all legitimate recipients of the pseudonymised data. Second, in the event of unauthorized disclosure or access to data that has been effectively pseudonymised, pseudonymisation can reduce the severity of the resulting confidentiality risk and the risk of negative consequences of such disclosure or access to the data subjects, provided that the persons to whom the data is disclosed are prevented from accessing additional data.

28. Pseudonymisation can reduce *risks of function creep*, i.e. the risk that personal data is further processed in a manner that is incompatible with purposes for which it was collected. This is because processors or persons acting under the authority of the controller or of the processor, who have access to the pseudonymised data, are not able to use those data for purposes whose fulfilment requires attribution to the data subjects. In particular, this concerns purposes whose fulfilment requires any direct interaction with the data subjects.
29. Finally, depending on the techniques used, assigning widely differing pseudonyms to persons with very similar identifying attributes, may not only enhance confidentiality, but also reduce risks to *accuracy* of the data by reducing the risk of incorrectly attributing data or objects to the wrong data subjects.<sup>9</sup>
30. The effectiveness of the implementation of pseudonymisation determines the extent of the reduction of risks for the data subjects and the benefits the controllers may derive from it, including the fulfilment of data-protection obligations according to Art. 24, 25 and 32 GDPR, see sections 2.4.1 and 2.4.2 below.

## 2.2.2 Analysis of pseudonymised data and planned attribution

31. Pseudonymised data can often be usefully analysed since, in large part, the information content of the original data can still be evaluated. Moreover, the insertion of pseudonyms enables the linkage of various records of pseudonymised data relating to the same person without the need to use additional information.<sup>10</sup>
32. After the analysis has been performed, pseudonymisation may be partially or completely reversed by
  - a. identifying the data subject,
  - b. linking pseudonymised to original data, or
  - c. reconstituting original data from pseudonymised data

using additional information kept by the controller for that purpose (planned attribution). This reversal should be performed by persons specifically authorised for this purpose, as per Rec. 29 GDPR. Under the same conditions, pseudonymisation may also be reversed in individual cases due to singular circumstances applying to them, while continuing to process the bulk of the data by default in a pseudonymised manner. See Example 3 in the annex.

---

<sup>9</sup> See Example 4 in the annex

<sup>10</sup> Such linkage might be required and lawful only under certain conditions. However, controllers can shape the pseudonymisation transformation in a way that limits the ability to link various items of pseudonymised data accordingly, see section 3.3.1.

33. Moreover, it may also be possible to use additional information in order to link different sets of pseudonymised data whose linkage has not been planned at the outset, i.e. at the time the purposes and means for processing have been determined by the controller or controllers involved. Processing implementing such linkage should likewise be performed only by persons specifically authorised for this purpose.
34. It needs to be noted that all processing operations mentioned in this section (including data set linkage) will need to be executed in compliance with the GDPR, in particular observing all data protection principles according to Art. 5 GDPR, and, especially, need to rely on a legal basis according to Art. 6 GDPR.

### 2.3 Pseudonymisation domain and available means for attribution

35. Controllers may define the context in which pseudonymisation is to preclude attribution of data to specific data subjects, generally on the basis of a risk analysis. They subject the additional information to technical and organisational measures to ensure that the pseudonymised data cannot be attributed to data subjects by persons operating within that context. This means in particular that additional information that would enable attribution is kept separate from it. These guidelines call this context (with the people operating in it and its attending physical and organisational aspects, including the IT assets available) the **pseudonymisation domain**.
36. The pseudonymisation domain may – by choice of the pseudonymising controller – coincide with a set of foreseen legitimate recipients of the pseudonymised data.
37. Additionally, the pseudonymising controller when defining the pseudonymisation domain may choose to include persons who are not legitimate recipients of the pseudonymised data, but may attempt to gain access to it anyway. The controller would do so in order to mitigate adverse effects of unauthorised access by those persons.
38. In sum, depending on the objective of pseudonymisation and their risk assessment the controller may define the pseudonymisation domain to encompass, e.g., only a single organisational unit of the controller, a single external recipient, all authorised or foreseen legitimate recipients, or a range of or all external entities that may attempt to gain access to the data without authorisation.
39. For effective pseudonymisation within a single organisational unit or a set of legitimate recipients, all involved controllers and processors should choose appropriate technical and organisational means—possibly including legal safeguards (e.g., contracts) if these can be effectively enforced—guaranteeing that pseudonymised data does not leave the pseudonymisation domain, which could lead to the circumvention of the protection afforded by pseudonymisation, see section 3.2.2.
40. Controllers that process pseudonymised data should also put in place such measures to ensure that actors within the pseudonymisation domain are not able to reverse the pseudonymisation. To that end, the controllers may for example choose to limit the resources available for processing the pseudonymised data and ensure that additional information allowing the attribution to data subjects does not enter the pseudonymisation domain.
41. If the pseudonymisation domain consists of a defined set of recipients and within the domain the measures mentioned in the previous paragraph are effectively enforced and maintained, then only those means need to be considered for attribution of the pseudonymised data to the data subject that can be used in the planned context of processing. In particular, if data is pseudonymised and then processed within the same controller, the pseudonymisation domain

does not encompass the controller as a whole, but only the persons processing the pseudonymised data under its authority (with the exception of those authorised to use additional information in order to attribute the pseudonymised data to individuals), the information they have at their disposal, and the systems and services they employ.

42. If a controller or processor wants to use pseudonymisation to reduce confidentiality risks from some or all unauthorised third parties, they will include those third parties in the pseudonymisation domain and assess the means they are reasonably likely to use for attribution. Relevant third parties include not only cyber-crime actors, but also employees or maintenance service providers acting in their own interests rather than on instructions from the controller. Taking due account of contextual elements and the circumstances at hand, it is recommended to consider both actions in good faith, and those executed with criminal intent.
43. For instance, pseudonymisation may be performed prior to transmission of the data to a processor or third party that ensures only a level of security that would not be appropriate for the processing of the original data, but is appropriate for the risk connected with the processing of data that cannot be attributed to data subjects. In this case, all means available to unauthorised parties that might access the pseudonymised data while the (authorised) recipient of that data processes them need to be considered.

## 2.4 Meeting data-protection requirements using pseudonymisation

44. Pseudonymisation can be used effectively by controllers and processors to meet certain data-protection requirements. Note, however, that while pseudonymisation is a valuable tool, it is often most effective when complemented by additional measures. Controllers need to assess the appropriateness of all its measures taken together in order to establish whether they suffice to meet the relevant data-protection requirements. Determining the effectiveness of pseudonymisation in preventing the attribution of pseudonymised data to data subjects is a building block of this assessment.

### 2.4.1 Pseudonymisation as an effective measure for data protection by design and by default

45. Pseudonymisation may be employed by controllers and processors as one of several technical and organisational measures in order to implement data-protection principles according to Art. 25(1) GDPR, in particular data minimisation and confidentiality. It may also contribute to safeguarding the lawfulness, fairness, purpose limitation and accuracy principles. The following paragraphs detail the application of pseudonymisation for those objectives, both in circumstances where pseudonymised data is processed internally, and where it is transmitted to other parties.

#### 2.4.1.1 Data minimisation, confidentiality, and purpose limitation in internal processing

46. For a controller's own processing, pseudonymisation may be an appropriate measure when the data being processed do not need to be attributed to data subjects for a given purpose. In those circumstances, pseudonymisation allows the linking to data subjects where this is required for the treatment of exceptional cases or for subsequent processing for another purpose. For an illustration of this use of pseudonymisation, see Examples 1 and 2 in the Annex.
47. In the case of internal processing, pseudonymisation can *effectively* contribute to the implementation of the abovementioned principles (see para. 56) provided that the following conditions hold for the persons handling the pseudonymised data:

- They are not able to reconstitute the original value of the attributes that have been omitted or transformed in the process of pseudonymisation.
- They cannot link the pseudonymised data to other data relating to the same person (unless it was pseudonymised “consistently”<sup>11</sup> with the first one).
- They are not able to single out the data subjects in other contexts on the basis of what they learned from handling the pseudonymised data.

#### *2.4.1.2 Data minimisation, confidentiality, and purpose limitation for a pre-defined set of recipients*

48. Pseudonymisation can also be used as an appropriate measure for the implementation of the data minimisation, confidentiality, and possibly also purpose limitation principles if data are to be transmitted to and processed by an **external recipient**, be it a processor or a controller. A typical objective is to prevent the recipient and the persons acting under its authority from learning identifying information they do not need for the data processing at hand. Additionally, the aim may be to ensure that the data subjects are not treated differently outside the context of the planned processing on the basis of the data received. Another objective might be to prevent that data is transmitted and then processed by the recipient for some incompatible purposes (like personalised advertisement) that would involve a data linkage in the recipient’s hands that is forestalled by pseudonymisation. For this, the pseudonymising controller sets up the pseudonymisation domain to include all intended recipients of the pseudonymised data.
49. Note that pseudonymisation by the original controller also aids controllers who are recipients of pseudonymised data in fulfilling their data protection obligations, in particular with regard to the data minimisation principle, data protection by default and the maintenance of an appropriate level of security.
50. Provided that safeguards (including contracts or legal acts) limit the disclosure of the pseudonymised data to a defined set of recipients, pseudonymisation can effectively contribute to the implementation of the three principles mentioned above if the conditions in paragraph 47 hold for all recipients. See Example 3 given in the Annex.
51. For external processing, i.e. processing under instruction by a processor or transmission to an independent controller, more extensive measures and risk assessment may be necessary to prevent attribution to data subjects. In particular, all intended recipients of the pseudonymised data need to demonstrably assure that the pseudonymised data are not disclosed to unauthorised recipients beyond the defined domain. For processors, additional tools under Art. 28 GDPR such as audits are available to support this assurance.
52. Transmissions of pseudonymised data might also occur within a group of collaborating controllers. These controllers might seek to prevent members of some organisational units with access to the pseudonymised data from being able to attribute the data to data subjects, even though other persons acting under the controllers’ authority might have the ability to attribute the data. The pseudonymisation domain in this case consists of those organisational units rather than the controllers themselves. In such a setup, the participating controllers need to demonstrably ensure that any relevant additional information they might have access to are not disclosed to unauthorised recipients beyond the defined domain and the capability to reverse

---

<sup>11</sup> Two sets of data are considered to be pseudonymised consistently if data contained in those sets and relating to the same person can be linked on the basis of the pseudonyms they contain, see section 3.3.1.

pseudonymisation is reserved to authorised persons. Technical and organisational measures and legal safeguards may be required for this purpose.

53. In the same way, pseudonymisation can also constitute an appropriate measure to be taken when processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, in particular in order to ensure respect for the principle of data minimisation, Art. 89(1) GDPR. This use is illustrated in Example 5 in the Annex.

#### 2.4.1.3 *Lawfulness, fairness and accuracy principles*

54. Specific EU or Member State law may require certain data to be pseudonymised as a condition for the lawfulness of its processing, thus making pseudonymisation an obligatory measure to meet the **lawfulness** principle.<sup>12</sup>
55. In the case of processing based on the legitimate interest provision in Art. 6(1)(f) GDPR, controllers other than public authorities in the performance of their tasks may consider the reduction of the risks to the rights and freedoms of the data subjects achieved by pseudonymisation (as by any other effective safeguard).<sup>13</sup> This may be the case when assessing whether their legitimate interests are overridden by the interests of the fundamental rights and freedoms of data subjects. The use of pseudonymisation for this purpose is illustrated in Example 7 in the Annex.
56. Pseudonymisation may also be an appropriate safeguard to be taken into account when considering “compatible purposes” in respect of further processing, since it may limit the possible consequences of the intended further processing for data subjects, in accordance with Art. 6(4)(d) and (e) GDPR.<sup>14</sup> Illustration for this use of pseudonymisation can be found in Examples 7 and 8 in the Annex.
57. The use of pseudonymisation within the implementation of the **fairness** principle is illustrated in Example 10 in the Annex.
58. Finally, considering the risk reduction described in paragraph 29, an appropriate pseudonymisation procedure can also contribute towards the **accuracy** principle as is illustrated in Example 4 in the Annex.

#### 2.4.2 *Ensuring a level of security appropriate to the risk*

59. Pseudonymisation may be employed as one of several measures contributing to a level of **security** appropriate to the risk of the data processing activity, in accordance with Art. 32(1) GDPR. Pseudonymisation may lower the severity of the consequences of unauthorised access to data. No one in the pseudonymisation domain, who accesses the pseudonymised data without authorisation, should be able to easily use the data to the disadvantage of the data subject, unless they also manage to (illegitimately) access the relevant additional information needed for attribution. Controllers and processors still have to provide a level of security appropriate to the remaining risks involved in the processing of the pseudonymised data. For processors this includes, as per Art. 28(1) GDPR, providing sufficient guarantees that appropriate technical and

---

<sup>12</sup> For example, Italian law mandates pseudonymisation in the course of the processing of genetic and judiciary data.

<sup>13</sup> Cf. Article 29 Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, p42-43.

<sup>14</sup> See also Recital 50 GDPR for further context.

organisational measures to ensure this level of security are implemented. The use of pseudonymisation for reducing security risks is illustrated in Example 6 in the Annex.

60. For pseudonymisation to be an effective security measure, additional information sufficient to attribute the pseudonymised data to identifiable natural persons should only be available outside the pseudonymisation domain. Therefore, the controller or processor needs to determine the actors that the pseudonymised data is to be protected from and assess whether there exists additional information relating to the data subjects accessible to those actors with reasonable means. Based on this assessment, for it to be effective, the controller needs to design the pseudonymisation procedure in such a way that additional information is required for attribution that goes beyond what the selected actors possess or could obtain with reasonable effort.
61. The controller will then have to take technical and organisational measures to prevent the use of that additional information by the actors in the pseudonymisation domain. This concerns not only the information needed to apply, or undo the pseudonymising transformation, but also the original personal data, if kept, or other data derived from it that continues to be stored. The security level reached with the help of pseudonymisation depends on the security level achieved for both pseudonymised and the relevant additional information. If it is easy for an unauthorised actor to obtain the relevant additional information, then the security benefit of pseudonymisation is small, and might well be negligible or lost.
62. Since effective pseudonymisation can mitigate adverse effects of data breaches, it may also be considered when assessing the obligations a controller has under Art. 33 and 34 GDPR. In particular, it may be regarded as an appropriate technical and organisational measure that limits the impact of a personal data breach in the sense of Art. 34(3)(a) GDPR. However, the content of data that was accessed without authorisation can still be analysed by the actor who accessed it. Careful analysis is required in this case to establish whether the pseudonymisation has reduced the risks resulting from the data breach sufficiently to render communication of the breach to the affected data subjects unnecessary, Art. 34(1) and (3) GDPR.<sup>15</sup>

#### 2.4.3 Pseudonymisation as a supplementary measure for third country data transfers

63. Pseudonymisation may constitute a so-called “supplementary measure” to ensure compliance with Art. 44 and 46(1) GDPR. In the absence of a decision pursuant to Art. 45(3) GDPR, a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available. The applied appropriate safeguards (e.g. binding corporate rules, standard contractual clauses) may, however, due to the legislation or practice of the third country, not be effective. Access to transferred data by third country public authorities may not be excluded. In this situation, pseudonymisation may constitute an effective measure to protect personal data transferred to a third country from disproportionate government access by public authorities of that country if the conditions enumerated in paragraph 85 of Annex 2 to the EDPB Recommendations 01/2020 are fulfilled.<sup>16</sup> See Example 9 in the Annex.

---

<sup>15</sup> See section IV.B in the Guidelines 9/2022 on personal data breach notification under GDPR..

<sup>16</sup> In paragraph 85 et seq. of the Recommendations 01/2020 under the heading “Use Case 2: Transfer of pseudonymised Data”, the EDPB identifies when pseudonymisation could be an effective measure that supplements transfer tools to ensure compliance with the EU level of protection of personal data.

64. The conditions include that

- the attribution of pseudonymised data to a specific data subject requires the use of additional information that public authorities of the recipient country neither possess, nor are able to obtain with reasonable effort,
- additional information is held exclusively by the data exporter and kept separately in a Member State or in a third country, by an entity trusted by the exporter in the EEA or under a jurisdiction offering an essentially equivalent level of protection to that guaranteed within the EEA,
- the authorities are not able to single out a data subject in the course of an interaction with members of a group of persons based on the pseudonymised data and information they are able to obtain with reasonable effort.

This implies that the public authorities, who would be able to have access to the pseudonymised data based on foreign law or practice, need to be framed within the pseudonymisation domain.

65. Thus, any design of a pseudonymisation procedure needs to start from an assessment of which information the public authorities of the recipient country can be expected to possess or to be able to obtain with reasonable means, even if those means may infringe the legal norms in the third country. This information must then be assumed to be available in the pseudonymisation domain.
66. As an additional supplementary measure, all entities holding additional information should provide sufficient guarantees to the exporter, and be bound by contract or legal act (e.g. by obligations of professional secrecy) not to disclose the additional information. Furthermore, where the importer has access to technical infrastructure of the exporter that is used to store additional information, the exporter needs to retain exclusive legal and administrative control over that infrastructure and ensure that access to additional information is effectively limited to its own employees.
67. Lastly, the data exporter or any other entity holding (part of) the additional information must prevent disclosure or unauthorised use of that additional information by appropriate technical and organisational safeguards.
68. Taken together, those measures can ensure that the data exporter retains control of any attribution of the pseudonymised data to specific data subjects.
69. Similarly, pseudonymisation may also be a suitable safeguard with regard to Art. 49(1) second sentence GDPR. In this case, if the application of pseudonymisation leads to such a reduction of the risks for the data subjects that those risks no longer override the legitimate interests of the controller, then a transfer may take place provided the other requirements of that provision are met.

## 2.5 Transmission of pseudonymised data to third parties

70. Pseudonymising controllers may need to consider whether the risk reduction achieved by pseudonymisation for internal processing still holds when data are transmitted to a third party. In this case, as a minimum the means available to the recipient for attribution of the data need to be identified and taken into account. This is particularly important if the transmission would only be lawful if the transmitted data remain pseudonymised on the recipient's side.

71. Prior to a transmission of pseudonymised data, controllers should also assess, in accordance with the principle of data minimisation, whether it is required for lawful purposes to transmit the full pseudonymised data including the pseudonyms. Pseudonyms might be needed to collate data records transmitted at different times when they relate to the same data subjects, or for the establishment of a back channel where transmitted personal data or the result of the processing of this data needs to be returned to the sender. In the absence of such reasons, pseudonyms should not be transmitted.
72. Moreover, controllers need to assess whether they should modify or replace the pseudonyms prior to transmission in order to minimise risks, including risks stemming from data breaches, that could arise if the data they continue to hold and data they have transmitted are brought together unlawfully or by unauthorised third parties. Controllers should treat this as a new pseudonymisation process requiring the same type of analysis and the same steps as the original pseudonymisation. In particular, they should define a suitable pseudonymisation domain for the newly transformed pseudonymised data and establish safeguards that those data do not leave it. If the recipient will be a controller in its own right, then it is good practice that the receiving controller informs the sending controller about the risks inherent in its own processing and aids in determining the way the sending controller performs the pseudonymising transformation.
73. The recipient themselves may intend to rely on the risk reduction achieved by pseudonymisation. Union or Member State law may also prescribe that they process personal data only in pseudonymised form. In such cases, they should ensure that the sender (or any other holder of the additional information) applies to the additional information technical and organisational measures to prevent its use for the attribution of the received data to identified or identifiable natural persons. The pseudonymisation domain should include the recipient, its processors and all persons acting under the authority of the recipient or one of its processors, at least to the extent they have access to the pseudonymised data. For this, they may take into account any legal obligations extending to the sender, e.g., rules of professional secrecy applying to that additional information. Insofar as it is needed to ensure appropriate treatment of additional information by the sender or any other holder of additional information that would enable the attribution of the received pseudonymised data, the recipient should enter into a legally binding agreement with those parties that allows for the enforcement of such treatment.
74. A particular case of transmission occurs when several controllers seek to combine different sets of pseudonymised data, alone or acting cooperatively. Clearly, they must have a legal basis for this operation including any transmissions involved, and for any further processing of its result. Apart from that, the pseudonymisation objectives of each individual party should be maintained and account should be taken of the possibilities for attribution of the data arising from linking or combination. As a result, the pseudonymisation domain designed by each party may need to be re-assessed and updated, especially if this processing was not initially envisaged. See section 3.3 for technical approaches to privacy preserving linkage.
75. Finally, controllers might also consider the possible transmission of the additional information they hold, which allows for the attribution of pseudonymised data to specific data subjects. Insofar as the additional information is held in the form of personal data, of course, all obligations for processing personal data apply. This concerns the original data in the state they were prior to the application of the pseudonymising transformation, but also, inter alia, tables matching pseudonyms and identifiers of data subjects, see paragraph 92. Independently of that, any transmission of additional information might have consequences for the effectiveness of pseudonymisation the controller needs to assess. In particular, any such transmission should

prevent the additional information from becoming available within the pseudonymisation domain. If necessary, the controller might need to enter into a binding agreement with the receiver ensuring that the received information are treated accordingly.

## 2.6 Implications for the rights of the data subjects

76. Since pseudonymised data, which could be attributed to a natural person by the use of additional information, is personal data, the rights of the data subject according to Chapter 3 GDPR apply.<sup>17</sup>
77. Art. 11 GDPR recognises that the controller may be able to demonstrate that it is not in a position to identify the data subject, including in pseudonymised data it holds. This may be the case if the controller does not have (or no longer has) access to additional information allowing attribution, is demonstrably unable to lawfully obtain such information and is demonstrably unable to reverse the pseudonymisation with the assistance of another controller. Consequently, except where the data subject (for exercising his or her rights) provides additional information enabling his or her identification, the rights of the data subjects enumerated in Art. 11(2) or 12(2) GDPR, respectively, shall not apply in this case. In compliance with Art. 11(2) GDPR, the controller has to inform the data subject accordingly, if possible.
78. For instance, if the data subject can provide the pseudonym or pseudonyms under which data relating to them is stored, and proof that those pseudonyms pertain to them, the controller should be able to identify the data subjects. In consequence, the data subject rights should apply in this case.
79. Therefore, in order to give full effect to the rights of the data subjects, the controller should indicate in the information provided to data subjects according to Art. 11(2) GDPR how they can obtain the pseudonyms relating to them, and how they can be used to demonstrate their identity. In this case, the controller may need to provide the identity and the contact details of the source<sup>18</sup> of the pseudonymised data or of the pseudonymising controller.

## 2.7 Unauthorised reversal of pseudonymisation

80. Any breach of security leading to the unauthorised reversal of pseudonymisation constitutes a personal data breach<sup>19</sup>, and may, in consequence, require the controller to notify the supervisory authority unless it is unlikely to result in a risk to the rights and freedoms of natural persons.<sup>20</sup>
81. If the unauthorised reversal of pseudonymisation is likely to result in high risks to the data subjects, the controller will need to communicate the nature of the data breach, and the further information stipulated in Art. 34(2) GDPR to the data subjects. If the controller is not in the position to communicate with the data subjects due to a lack of sufficient directly identifying information (even though pseudonymisation has been reversed), and other forms of communication (e.g., employing the services of a controller that possesses information sufficient for that purpose) would involve disproportionate effort, then the controller needs to notify by way of a public communication or similar equally effective measure.

---

<sup>17</sup> See also the EDPB Guidelines 1/2022 on Data Subject Rights - Right of access, No. 45, 66 and the WP29 Guidelines on the right to data portability - endorsed by the EDPB, p. 11.

<sup>18</sup> See Art. 14(2)(f) GDPR.

<sup>19</sup> See also EDPB Guidelines 01/2021 on Examples regarding Data Breach Notification.

<sup>20</sup> There is an obligation to document the breach pursuant to Art. 33(5) GDPR independent of the risk.

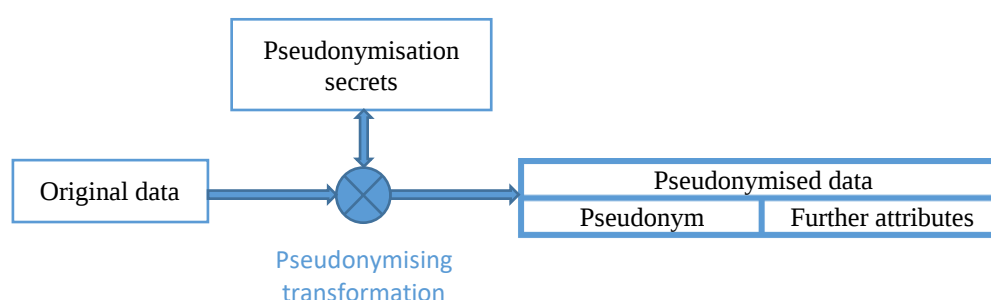
82. As per Art. 29 GDPR, a processor or a person acting under the authority of the controller or of the processor acts unlawfully if it reverses pseudonymisation in contravention to the instruction of the controller. This holds in particular if they do so in order to pursue a purpose other than the one they were instructed to carry out on behalf of the controller.

### 3 TECHNICAL MEASURES AND SAFEGUARDS FOR PSEUDONYMISATION

#### 3.1 Pseudonymising transformation

##### 3.1.1 Structure of the pseudonymising transformation

83. In order for pseudonymisation to be effective, pseudonymised data must not contain direct identifiers (e.g. national id numbers) whenever those direct identifiers could be used in the pseudonymisation domain to easily attribute the data to the data subjects. To this end, those identifiers are removed in the course of the pseudonymising transformation. Direct identifiers may, however, be replaced by new identifiers that can be attributed to data subjects only with the use of additional information. Such identifiers are called pseudonyms.
84. The *pseudonymising transformation* implements this replacement. Insofar as necessary for pseudonymisation to have the intended effect, it also modifies other attributes, e.g. by removal, generalisation and noise addition.



85. In order to prevent unauthorised attribution of pseudonymised data, the pseudonymising transformation regularly involves secret data. The controller may choose these data prior to the execution of the transformation. They may also choose or generate it in the course of performing the transformation. These data are often either cryptographic keys (for encryption or one-way functions) or tables matching pseudonyms with the personal data they replace. Hereafter, they will be called “**pseudonymisation secrets**”.
86. Since the pseudonymised secrets allow attribution of the pseudonymised data, they form part of the additional information in the sense of Art. 4(5) GDPR. Hence, controllers need to keep them separately and subject them to technical and organisational measures that ensure their confidentiality and prevent their unauthorised use.<sup>21</sup>

---

<sup>21</sup> If the controller keeps the original data in the form they had prior to pseudonymisation, those original data also constitute part of the additional information that have to be kept separately.

### 3.1.2 Types of pseudonymising transformations

87. Two classes of replacement procedures are commonly applied as pseudonymising transformations: cryptographic algorithms and lookup tables.<sup>22</sup>
88. It should not be possible within the pseudonymisation domain to attribute pseudonymised data relating to a data subject whose identifiers are known. This could be done by applying the pseudonymisation transformation to those identifiers, obtaining the pseudonym, and locating the pseudonymised data attached to that pseudonym.<sup>23</sup> Hence, the transformation needs to involve information that the pseudonymising controller keeps secret and an unauthorised person cannot use. Only possession of the secret information should enable the computation of the pseudonym given the identifier. In order to limit the likelihood of a successful guess or brute-force searching, the secrets should have sufficient entropy.<sup>24</sup> For the first class, cryptographic algorithms, this information takes the form of secret parameters or keys. For the second class, lookup tables, the controllers keep the tables themselves secret.
89. The first class of transformations consists of cryptographic algorithms. Examples of suitable algorithms are cryptographic one-way functions like Message Authentication Codes (MACs) or encryption algorithms. Preference should generally be given to one-way functions due to the difficulty of their reversal even when the secret parameters are known.<sup>25</sup> However, special demands of the use case – in particular the need to easily reverse pseudonymisation in authorised scenarios – might call for the use of encryption. If hash functions are used as building blocks for the cryptographic one-way functions used in the pseudonymising transformation, then it is advisable to use specialized hash functions designed for secure password authentication.<sup>26</sup>
90. The secret parameter(s) or key(s) involved in the cryptographic algorithms will be the pseudonymisation secrets. A suitable choice of the algorithm and the technical and organisational measures applied to the pseudonymisation secrets can make it hard to calculate the pseudonyms and locate any data relating to a specific data subject within the pseudonymised dataset given only the value of the original identifier for that subject. It also makes it hard for anyone without access to the pseudonymisation secrets to determine the original identifiers given the pseudonym by brute-force computation of all possible pseudonyms.
91. Note that procedures from this class, and the choice of the parameters might become vulnerable to breach, including due to cryptanalytic and technological advances. Hence, the controllers need to draw up a plan for replacement of weak algorithms should it become necessary. This plan needs

---

<sup>22</sup> For a more detailed treatment of various pseudonymisation procedures, see ENISA: Pseudonymisation techniques and best practices, Chapter 5.

<sup>23</sup> For instance, if you know the transformation is merely a SHA256 hash of a name, you could apply this to all names you have elsewhere and then see which hashes match in the pseudonymised dataset.

<sup>24</sup> Entropy refers here to the randomness of the secret parameter. For example: If the controller selects the date when the pseudonymising transformation was applied as parameter, the entropy of this parameter will be very low. However, if the controller chooses a random generated string of 20 alphanumerical characters as secret parameter, the entropy is high.

<sup>25</sup> Several privacy enhancing techniques that controllers may apply to ensure data protection by design involve pseudonymisation in the course of the application of sophisticated cryptographic techniques. In this case, the pseudonymising transformations are tailored to the respective cryptographic protocol. See, e.g., ENISA, “Data Pseudonymisation: Advanced Techniques & Use Cases”, 2021 <https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases>

<sup>26</sup> EU or Member State agencies (ENISA or, e. g., the German BSI) provide advice regarding such hash functions that make it much harder to brute force. At the time of writing of these guidelines, argon2 is a typical example.

to foresee a procedure to replace the pseudonyms *already* generated, if possible by working within the pseudonymisation domain. It is best practice to foresee a procedure that allows making the change without having to reconstitute the original personal data. In many contexts the determination of the new pseudonyms can be achieved by applying a second function (a keyed one-way function or encryption algorithm) that is still secure to the old pseudonyms.

92. The second class consists of procedures that create lookup tables matching identifiers with the pseudonyms used to replace them. Whenever the procedure encounters a new identifier value<sup>27</sup>, a pseudonym is generated as a uniquely chosen value, and a row is added to the table with the replaced identifier and the pseudonym. If an unauthorised prediction of the generated pseudonyms from a few observed values can lead to the attribution of the pseudonymised data to data subjects, then the controller will need to consider a more secure way of generating pseudonyms, e.g., by use of an effective hardware random number generator or a cryptographically secure pseudo-random number generator.
93. Note that procedures involving lookup tables require the storage of at least one record for each data subject present in the original dataset. Lookup tables are personal data since they allow the identification of data subjects. Since they are parts of the pseudonymisation secrets, they need to be protected from unauthorised access and use. Thus, controllers need to weigh the disadvantage of securely storing this possibly large set of personal data against the reduced or avoided susceptibility to cryptanalytic attacks in comparison to the first class of procedures, which is particularly important wherever long-term guarantees for irreversibility of the pseudonymising transformation are needed.

### 3.1.3 Modification of original data necessary for the objectives of pseudonymisation

94. In order to decide which attributes need to be replaced or modified by the pseudonymising transformation, controllers should refer to the objectives they seek to achieve with pseudonymisation, establish the pseudonymisation domain, choose the technical and organisational measures to be applied in it, and also determine the means that could be applied in the domain for the attribution of data to data subjects, see paragraph 41.
95. In the course of this process, they should consider that those means may be applied not only to individual pseudonymised records, but also to the result of a linkage of records relating to the same data subject. Such linkage may happen within the same dataset and with other data that were pseudonymised in the same or a similar way. Linked data might allow for attribution to a specified person while individual data records do not, because it contains more attributes to match with other data. The extent to which linkage has to be considered depends first on the design of the pseudonymisation transformation, see section 3.3.1, and second, less importantly, on the technical and organisational measures implemented to effectively separate data sets that are not to be linked.
96. Controllers can benefit from a potential trade-off: the smaller the pseudonymisation domain and the more restrictive the access to pseudonymised data and other relevant information sources within the pseudonymisation domain, the less need there is in general, considering the remaining circumstances, to modify the original data.

---

<sup>27</sup> The generation of a new uniquely chosen pseudonym and creation of a new entry in the table might also be triggered by other factors, see section 3.3.1.

#### 3.1.3.1 *Determination, substitution and removal of directly identifying attributes*

97. For effective pseudonymisation, directly identifying attributes need to be replaced or be discarded. The choice between replacement or deletion of those attributes depends on the purpose of the processing and the objectives of the pseudonymisation (see section 2.2).
98. For purposes that do not require linkage of records, data protection by design calls for the removal of individuals' "long term" identifiers (e.g. a "health service ID") while replacing transactional or "short term" identifiers (e.g. a "case number") by pseudonyms.
99. If, however, linkage of records is required—e.g., when collating records referring to events over a long period of time for longitudinal analysis—then it might be necessary to replace long-term identifiers by pseudonyms, while other identifiers are discarded. Such long term pseudonyms should, however, only be used if they are required for the purposes of the processing.
100. Purposes that involve linking across pseudonymised data sets need to use identifiers contained in both sets as a basis for the pseudonymisation transformation.

#### 3.1.3.2 *Determining and treating quasi-identifiers*

101. One way to attribute data to a natural person is by looking at several attributes contained in the data that reveal information about the physical, physiological, genetic, mental, economic, cultural or social identity of the data subject. If a combination of those attributes are sufficient to attribute at least part of the pseudonymised data to data subjects, then they are called *quasi-identifiers*. Demographic data are prime examples of such attributes: age, gender, languages spoken, marital or family status, profession, income. If the data concern employees, then other relevant data may be structural role, number of working hours, length of service. Persons handling pseudonymised data may well know the values of those attributes of some of the individuals to whom the pseudonymised data relate. This would enable them to attribute the data to those persons without the use of the pseudonymisation secrets, i.e. without the need to reverse the pseudonymising transformation.
102. The most direct way to prevent attribution based on quasi-identifiers is their removal. A second approach lies in their modification by generalisation and randomisation.
103. A third approach, which is particularly applicable when pseudonymised data are processed internally by the pseudonymising controller, is to reduce the number of attributes that need to be considered quasi-identifiers in the pseudonymisation domain by minimising the information available there. This can be achieved by limiting the pseudonymisation domain to few employees and applying technical controls to restrict the information they can access. If an assessment shows that there is an insignificant risk that the pseudonymised data are linked to other information, then quasi-identifiers may be kept.
104. It needs to be noted, however, that the latter methods are not available when pseudonymisation is intended to protect data against security risks posed by unauthorised third parties external to the organisation of the controller.

#### 3.1.4 *Pseudonymisation in the course of data collection*

105. There are two approaches to introduce pseudonymisation into the *collection* process:
  - a. "Pseudonymisation proxy": All relevant incoming data is first processed by a dedicated, separate team. The persons authorised to reverse pseudonymisation (Rec. 29 GDPR, second

sentence) work in that unit. Whenever the special circumstances of a case require it, pseudonymisation is reversed, and the original collected data turned over for processing.

- b. “Pseudonymisation at the source”: Pseudonymisation is already performed by the controller that is the source of the information, prior to transmission to the entity processing the pseudonymised data.

In the case that there is any doubt regarding the trustworthiness of the source when the latter approach is used, then cryptographic methods can be employed to allow the verification of attributes omitted or transformed during pseudonymisation and provided later upon request.<sup>28</sup>

### 3.2 Technical and organisational measures preventing unauthorised attribution of pseudonymised data to individuals

106. In order to prevent the unauthorised attribution of pseudonymised data, measures should be taken in three directions: First, the pseudonymising transformation should be protected against reversal by choosing a suitable design, and ensuring an appropriate level of security for the pseudonymisation secrets. Second, quasi-identifiers should be appropriately handled, see section 3.1.3.2. Third, data controllers should ensure that their assumptions about the scope of the pseudonymisation domain, about the use of pseudonymised data and about the accessibility of relevant information sources within it are met. In the following, these points are addressed in more detail.

#### 3.2.1 Preventing reversal of the pseudonymising transformation

107. For pseudonymisation to be effective, it must not be possible with reasonable effort to reverse the chosen pseudonymising transformation based on its output alone. When using lookup tables for the pseudonymising transformation, it suffices to choose randomly generated pseudonyms. When using cryptographic algorithms, suitable building blocks include (keyed) pre-image resistant<sup>29</sup> cryptographic one-way functions (like HMACs) or encryption schemes guaranteeing cipher text indistinguishability<sup>30</sup> (like symmetric block ciphers used in a suitable mode).

108. If lookup tables or reversible cryptographic algorithms are used, then it is clear that the pseudonymisation secrets need to be kept confidentially. (For added security, they may also be divided up, e.g. by secret sharing, and stored by different entities.) However, this confidentiality requirement holds for all types of pseudonymisation secrets, and needs to be extended to measures that protect against unauthorised use of those secrets, since such use may permit the construction of lookup tables that will allow the reversal of pseudonymisation.

109. In consequence, access to the systems performing the pseudonymising transformation and their interfaces needs to be strictly controlled. Integrity and confidentiality of the processing systems and services themselves must be ensured. Appropriate technical measures may include network

---

<sup>28</sup> For this purpose, the pseudonym needs to contain what is called a cryptographic commitment to those attributes. Provided a state-of-the-art protocol is used for the commitment, the source of the data is unable to produce a second set of attributes yielding the same pseudonym. For an example of this approach, see Example 2 in the Appendix.

<sup>29</sup> Pre-image resistance in a one-way function guarantees that finding the input corresponding to a given output is a hard task.

<sup>30</sup> The ciphertext indistinguishability property ensures that ciphertexts do not reveal anything about the corresponding plaintexts, making it hard to tell which plaintext corresponds to a given ciphertext.

segmentation, secret key storage in hardware security modules, secure authentication for Application Programming Interface (API) access, and rate limiting and logging of the execution of both the pseudonymising transformation and, in particular, its reverse application whenever that is available. Appropriate organisational measures include the employment of vetted, specifically authorised personnel for the operation of the systems used for the execution of the pseudonymising transformation and the storage of the pseudonymisation secrets. Controllers need to ensure that those employees, and all employees that are tasked with both interacting with data subjects and accessing pseudonymised data (e.g., in order to grant data subject rights) are properly trained.

110. Insofar as pseudonymised data or additional information are stored on devices used by data subjects—e.g. in order to enable data subjects to claim their rights, enhance transparency, or minimise centralised storage of data—controllers should take technical measures that maintain the validity of the assumptions they make about the accessibility and flow of the data. Since controllers usually do not have control over the devices, those measures might in particular involve the application of cryptographic techniques or leverage of secure elements present in those devices. An assessment of device features for effectiveness should also be conducted, as device producers can take different approaches in design and scope.

### 3.2.2 Securing the pseudonymisation domain

111. The pseudonymisation domain needs to be properly secured and separated from additional information for pseudonymisation to be effective. Appropriate measures should be in place to ensure additional information does not enter the pseudonymisation domain. Likewise appropriate measure should also be in place to ensure pseudonymised data does not leave it whenever this is possible, i.e., whenever it is restricted to the original controller or a well-defined set of recipients.

112. As is true for any personal data, the flow of pseudonymised data should be tightly controlled. Controllers holding pseudonymised data should define who the data should be disclosed to, and to which extent. Access control systems should be in place, and APIs should be secured against unauthorised use. Copies of data should be deleted as soon as they are no longer needed. Transmission of pseudonymised data to other entities should proceed only upon authorisation, guaranteeing that it is never transmitted out of the established pseudonymisation domain.

113. For any measure to be effective against unauthorised actors, controllers need to ensure the ongoing confidentiality, integrity, and resilience of the processing systems and services that are used to process additional information or the pseudonymised data.

114. Whenever the pseudonymisation domain is to consist of a defined set of recipients, the responsibilities of all parties involved should be defined by an arrangement, preferably in contractual form. Those arrangements should reflect the need to keep the pseudonymised data within the pseudonymisation domain, and to limit the inflow of or access to information that might allow attribution of pseudonymised data to data subjects, including among the recipients. Moreover, whenever relevant, the arrangements should regulate the process to be followed when assumptions about the pseudonymisation domain need to be adapted. Note, however, that such arrangements on their own are not sufficient to ensure a proper separation of the pseudonymisation domain from additional data without corresponding effective enforcement.

## 3.3 Linking pseudonymised data

### 3.3.1 Controlling the scope for the linkage of pseudonymised data

115. In order to allow the linkage of several pieces of pseudonymised data referring to the same data subject with the same pseudonym, the pseudonymising transformation is regularly performed deterministically.<sup>31</sup> Based on the objectives of pseudonymisation, controllers need to define which sets of personal data will be pseudonymised *consistently*. For example, they may decide to pseudonymise all data they collect on the same day consistently allowing for the linkage of two data records pertaining to the same data subject and collected on the same day, but preventing linkage of records of data collected on different days.<sup>32</sup> More generally, two pieces of pseudonymised data can be linked if the original data came from the same set and they both relate to the same data subject. In particular, three ways to arrange for controlled linkage of pseudonymised data are widely used: person, relationship, and transaction pseudonyms. Note, however, that other ways to segment the pseudonymised data are available and may be appropriate for the respective use case.
116. One or several controllers may choose to pseudonymise all data they process relating to the same data subjects consistently. The corresponding pseudonyms are usually called *person pseudonyms*. Their use requires long-term storage of the pseudonymisation secrets. The use of such pseudonymisation is admissible if and only if the linking of different pieces of pseudonymised data relating to the same person may become necessary and will be lawful in this case. The risk of unauthorised attribution is comparatively high. Correspondingly, this type of pseudonymisation may not significantly reduce the severity of risks associated with unlawful or unauthorised disclosure of the pseudonymised data.
117. A controller may also choose to pseudonymise all data consistently that it intends to process for one or several particular purposes defining a certain type of relationship of data subjects with that controller. For instance, a data subject may be assigned different pseudonyms depending on whether the data concern their relationship with controllers as employees or customers. In this case, pseudonymisation secrets (or parts thereof) are maintained only for the time the relationship with the data subject lasts. The resulting pseudonyms are called *relationship pseudonyms*.<sup>33</sup> The use of such pseudonymisation is only admissible if linking of different pieces of pseudonymised data relating to the same person in the same relationship to the controller may become necessary and will be lawful in this case. This condition is often fulfilled if there is only one common purpose, or the various purposes are compatible.
118. In order to generate relationship pseudonyms by a cryptographic algorithm, the secret values or keys need to be chosen dependant on the relationship, i.e. its type or the partners involved. If relationship pseudonyms are kept in lookup tables, they need to be generated and stored in separate tables according to the relationship.

---

<sup>31</sup> Deterministic pseudonymisation replaces the same identifying attributes with always the same pseudonym. For a comparison between deterministic and randomised pseudonymisation see ENISA Data Pseudonymisation: Advanced Techniques & Use Cases, p. 12.

<sup>32</sup> This can be achieved by choosing independently and randomly chosen pseudonymisation secrets once for each day and using those secrets for determining all pseudonyms on that day.

<sup>33</sup> This type of pseudonym is also called role-relationship pseudonym, see Pfitzmann/Hansen: *Anonymity, Unobservability, Pseudonymity, and Identity Management – A Proposal for Terminology*, and distinguished from more general role and relationship pseudonyms which are consistent across types of relationships, or partners of relationship, respectively.

119. Finally, the controller may choose to pseudonymise each individual transaction of a data subject with the controller differently. For example, the controller might pseudonymise each record capturing an interaction of a vehicle with an intelligent transport system service using time-dependent keys resulting in different pseudonyms for the same vehicle for each interaction. The resulting pseudonyms are called *transaction pseudonyms*. Pseudonymisation of this type, when applicable, contributes most effectively to data minimisation and data protection by default, since it prevents unlawful or unauthorised linkage of pseudonymised data across transactions by the pseudonyms they contain. Moreover, this form of pseudonymisation is well suited to mitigate risks connected with unlawful or unauthorised disclosure of pseudonymised data.

120. In order to generate transaction pseudonyms by a cryptographic algorithm, the pseudonyms should be computed on the basis of identifiers that are unique for each transaction. If transaction pseudonyms are stored in lookup tables, they should be randomly generated and stored for each transaction anew.

121. In order to comply with the data minimisation principle and data protection by default, the controller should define the sets of data that are to be pseudonymised consistently as small as possible. In particular, where consistent with the nature, scope, context and purposes of processing, the controller should prefer transaction pseudonyms to other types of pseudonyms.

### 3.3.2 Linking data pseudonymised by different controllers

122. In certain circumstances, two or more controllers may lawfully link different sets of pseudonymised data they hold. The goal is to process the linked data sets in pseudonymised form within a newly defined pseudonymisation domain.

123. Several approaches to implementing pseudonymising transformations allow for the controlled linking of pseudonymised data held by different controllers: Sharing of pseudonymisation secrets among controllers, jointly using a trusted service provider for carrying out the pseudonymisation, or a combination of the two. In such a combination, some pseudonymisation secrets are split between controllers and a trusted service provider and the service provider does not learn the identities of the data subjects. Finally, cryptography also allows for the computation of common pseudonyms without revealing direct identifiers or long-term pseudonyms for natural persons who are data subjects of pseudonymised data held by one party, but not the other (private set intersection).

124. Note that, for all these cases, it is a prerequisite that: a) the resulting pseudonymising transformation is the same for all data controllers, and b) the pseudonyms derived by each controller are based on original identifiers of the data subjects common in the different data sets.

125. The first and simplest approach consists of using cryptographic algorithms for the pseudonymising transformation and sharing the pseudonymisation secrets needed for consistent pseudonymisation among all controllers involved.<sup>34</sup> However, controllers that intend to use pseudonymisation to protect data against unauthorised attribution to specified data subjects need to consider the disadvantages of this approach in the course of their risk assessment according to Art. 25, 32, 35 and 36 GDPR: a) The pseudonymisation secrets are stored in multiple locations which increases the chance of unauthorised access or use. b) All controllers are enabled

---

<sup>34</sup> For example in Germany, this approach has been used to link data from different regional cancer registries in order to create a national pool of epidemiological data. Each of the contributing regional registries used their own pseudonym for storage, and a common national pseudonym for transmission.

to attribute not only those data records they have pseudonymised themselves to specific data subjects, but also data records pseudonymised by other controllers. c) The complexity of renewing the secrets is increased, which becomes relevant when pseudonymised data are used for long periods and in particular if the pseudonymisation secrets are compromised. In consequence, this approach is generally not recommended.

126. The second approach requires that the controllers jointly enter into a contract among themselves, and individually into contracts with the trusted service provider.<sup>35</sup> This service provider may perform its task as a processor for each contributing controller. It may also act as a controller if it has been given the independent power to decide whether to pseudonymise an individual data record or to reverse its pseudonymisation on the basis of legal or ethical considerations. The processor or trusted third party needs to know only the identifiers of the data subjects on the basis of which it will compute the pseudonyms, and no other data. Hence, the controllers should transmit only those identifiers combined with ephemeral numbers assigned to the records that contain them. The service provider applies a pseudonymising transformation to the identifiers that is uniform for all controllers, and obtains the pseudonyms. It returns those pseudonyms together with the respective record numbers. Subsequently, the pseudonyms can be joined with the data records using the record numbers, which are then deleted.

127. As a result, all pseudonymised data relating to the same data subject at any of the controllers contain the same pseudonym allowing for the desired linkage. If at a later time a controller wants to attribute pseudonymised data to a specified data subject (provided the controller could lawfully do so), then the procedure is executed in reverse. The advantage of this approach is that each controller, if given access to the linked data set, would at most be able to reattribute the linked records to which it contributed back to the corresponding data subjects.

128. The third approach is a variation of the second, and requires the same contractual guarantees as the previous one. It evades the necessity of disclosure of identifiers to the trusted service provider in cases where this constitutes a significant risk to the data subjects or where Member State law (e.g. regarding the maintenance of professional secrecy) precludes such disclosure. The procedure consists of several steps: During the set-up phase, the controllers agree on a common pseudonymisation secret. They use this common secret to compute a first-level pseudonym as in paragraph 125. Afterwards they transmit it to the trusted service provider who in turn uses a pseudonymisation secret of its own to compute a second-level pseudonym, which is the one to be used for linkage and use of the linked pseudonymised data. The advantage of this approach is that the trusted service provider does not (and cannot) learn the identifiers. Moreover, using two pseudonymisation secrets and storing them at different entities makes it harder to reverse the pseudonymising transformation without authorisation.

129. It is possible, and preferable to compute the common pseudonyms from data that is already pseudonymised without reconstituting identifying attributes using the pseudonymisation secrets. In this case, controllers hold pseudonymised data, which they process for their own purposes. If the need for linkage arises, then their private pseudonyms are transformed directly into a common pseudonym. They may do this by themselves as in paragraph 125, or have it done by a service provider as in paragraph 126. They could also perform a transformation into a common first-level pseudonym as in paragraph 128. This process should not affect the property that a pseudonym

---

<sup>35</sup> See Example 3.

occurring in pseudonymised data held by a contributing controller cannot be attributed to a specific data subject without the use of additional information held by that controller.

### 3.4 Summary of procedures for pseudonymisation

130. Controllers<sup>36</sup> who intend to implement pseudonymisation should determine the objectives they intend to achieve with this measure in order to define the domain of the pseudonymisation and decide which sets of data are to be processed consistently, see sections 2.3 and 3.3.1, respectively. Then the controllers perform the following steps:

131. At the time of the determination of the means for processing, they should analyse the data, and establish:

- which attributes contained in the personal data that is to be pseudonymised can be used alone or in combination to identify the data subjects directly (identifiers);
- which attributes should be used to determine (using cryptographic algorithms) or (using lookup a table) linked with the pseudonyms, applying the criteria set out in section 3.3.1;
- which method is to be used to replace those attributes with pseudonyms, and, in particular,
- which parameters (like size of group or key length for the cryptographic algorithms employed) are to be applied in the course of the pseudonymising transformation;
- which information is to be retained as additional information that can be used to attribute the pseudonymised data to a specific data subject;
- whether and which attributes contained in the personal data can be used alone or in combination to attribute some of the data to data subjects, directly or indirectly, within the pseudonymisation domain, considering information that can be accessed with reasonable effort from within it;
- which method is to be used to modify or remove those attributes in order to guarantee that the personal data are not attributed to an identified or identifiable natural person without use of the additional information while retaining the ability to perform general analysis on the resulting pseudonymised data. Available methods are, among others, omission, generalisation, and randomisation;
- which party or parties— controllers, processors, or specialised third parties entrusted with safeguarding the transformation<sup>37</sup>—are to execute the pseudonymisation transformation (individually or jointly), and
- who will store which pseudonymisation secrets or other additional information, and which technical and organisational measures will be applied to ensure that they cannot be used from within the pseudonymisation domain, that their integrity and confidentiality is

---

<sup>36</sup> The procedure also applies *mutatis mutandis* to processors using pseudonymisation as a safeguard.

<sup>37</sup> The involvement of a single trusted third party in the pseudonymisation of several controllers provides the additional benefit that data records from those differing sources can be pseudonymised in a way that allows for *accurate linkage* of pseudonymised records relating to the same data subject, if there is a legal basis for doing so. See section 3.3.2.

maintained, and that they are only used to attribute pseudonymised data to data subjects when authorised.<sup>38</sup>

Importantly, after the pseudonymising transformation is defined, the controller also needs to assess the risk of attribution in the pseudonymisation domain, and ascertain that it is insignificant.

132. When applying the pseudonymisation transformation, the controllers:

- (optionally) establish which data records pertain to the same data subjects, and assign unique identifiers of the respective data subjects to those data records,
- replace the chosen attributes that identify the data subjects and the unique identifier added before (if any was inserted) with pseudonyms by applying the method established previously, removes all other identifiers and stores separately from the pseudonymised data any pseudonymisation secrets generated in or derived from this process,
- modify or remove the quasi-identifiers by applying the method defined for this end.

133. All involved controllers apply the planned technical and organisational measures to additional information that they keep to attribute pseudonymised data to data subjects when a legitimate need for this arises, or that they otherwise retain and that might enable such an attribution. In particular, they restrict access to and use of the pseudonymisation secrets.

134. All recipients apply appropriate technical and organisational measures to safeguard that pseudonymised data does not leave the pseudonymisation domain, and also ensure that no information known to allow attribution enters it.

135. Finally, the controllers restrict the handling of the pseudonymised data to the extent this is necessary to mitigate any remaining risk of reversal of pseudonymisation.

---

<sup>38</sup> This includes the measures taken to secure any pseudonymised data, pseudonymisation secrets or other additional information stored on devices used by data subjects, see paragraph 110.

## ANNEX – EXAMPLES OF THE APPLICATION OF PSEUDONYMISATION

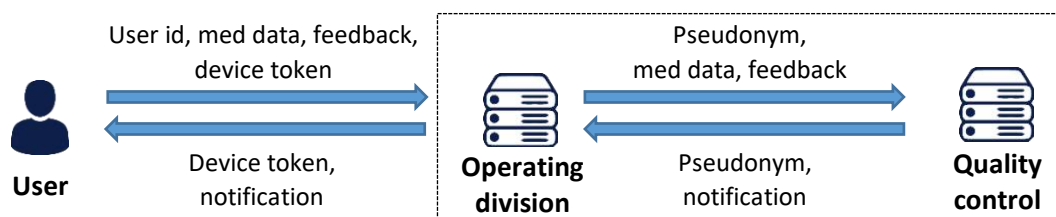
Section 2.3 highlighted the benefits of pseudonymisation in light of some of the relevant GDPR principles and GDPR provisions (data protection by design and default, processing for research and statistical purposes, security of processing, processing for the purpose of the legitimate interests of the controller, further processing and transfer of pseudonymised data).

The ten following sections intend to illustrate by way of real-world scenarios the use and benefits of pseudonymisation. These examples are listed in Table 1 in function of the GDPR provisions that pseudonymisation helps implement. Note that Member State law may require modifications to the setups described here in some cases.

| GDPR Article | GDPR Provisions                                                                                                                            | Example numbers |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| Art. 5(1)(c) | Data minimisation                                                                                                                          | 1, 2 and 3      |
| Art. 5(1)(b) | Purpose limitation                                                                                                                         |                 |
| Art. 5(1)(f) | Confidentiality                                                                                                                            |                 |
| Art. 5(1)(d) | Accuracy                                                                                                                                   | 4               |
| Art. 89(1)   | Safeguard for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes | 5               |
| Art. 32(1)   | Security of processing                                                                                                                     | 6               |
| Art. 6(1)(f) | Lawfulness of processing for the purposes of legitimate interests                                                                          | 7               |
| Art. 6(4)    | Processing for a purpose other than that for which the personal data have been collected (further processing)                              | 7 and 8         |
| Art. 46      | Transfers subject to appropriate safeguards                                                                                                | 9               |
| Art. 5(1)(a) | Fairness                                                                                                                                   | 10              |

Table 1: Examples of use and benefits of pseudonymisation

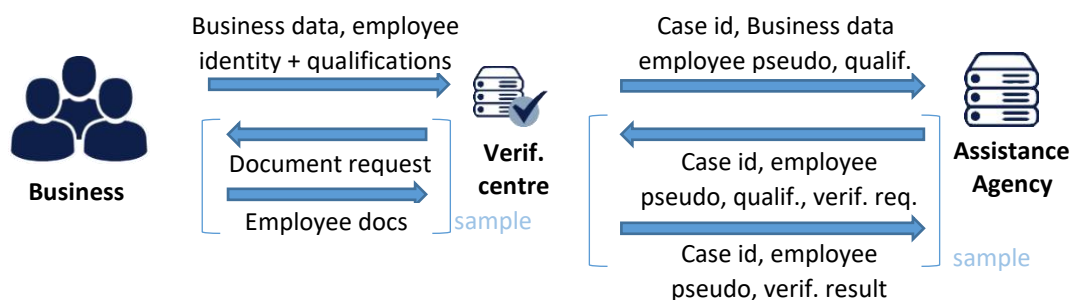
### Example 1: Data minimisation and confidentiality in internal analysis



|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | A Company provides an app that dispenses medical advice based on the description of symptoms entered into the app by users. It has tasked one of its divisions to perform quality control. In the course of quality control, it is established (using data collected with explicit consent of the app users) whether the dispensed advice conforms with established medical knowledge, and it is established whether and which patients need to be notified in critical cases of inappropriate advice given by the app. |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

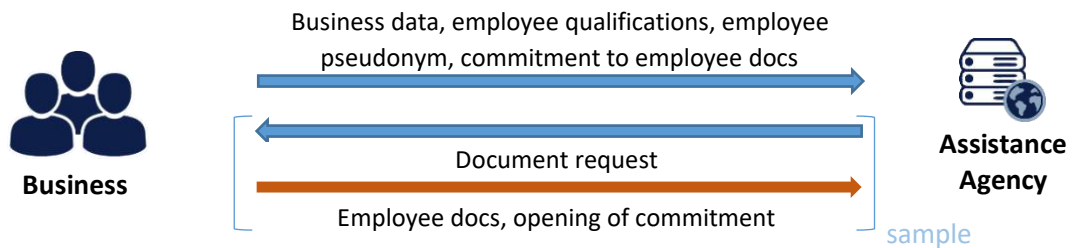
|                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | In order to meet the last purpose, the analysed data need to retain a link to the data subjects. Notifications to patients are not directly issued by the quality control division, but by regular operation staff.                                                                                                                                                                                                                                                                                                                                                                            |
| Problem to be solved             | Preserve the link between data and data subjects while ensuring compliance with the data minimisation principle, Art. 5(1)(c) GDPR, and data protection by default, Art. 25(2) GDPR, in particular with regard to access to data allowing attribution of data to data subjects, as well as reducing confidentiality risks thereby contributing to compliance with Art. 5(1)(f) GDPR.                                                                                                                                                                                                           |
| Original Data                    | Records containing:<br>the user id,<br>the device token,<br>the symptoms recorded by the app,<br>the advice dispensed,<br>the user feedback (optional).                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Pseudonymisation domain          | Quality control division.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Pseudonymised Data               | Records containing:<br>a pseudonym based on the user-id,<br>(categorized) symptoms recorded by the app,<br>the advice dispensed,<br>the user feedback.                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Additional information           | Table linking user-id and pseudonym<br>Lookup table linking user id and device token.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Processing of pseudonymised data | The quality control division is provided with the (pseudonymised) extract of data received by the backend of the app. The members of the division are not involved in, and have no further access to data stemming from service provision.<br><br>It performs the analysis. If the need arises to inform a data subject, it conveys the pseudonym, and the message to the operative division. The operative division has access to the additional information. Hence, it is able to identify the users, and convey the messages to them employing a notification service and the device token. |

### Example 2: Separation of functions allowing for data minimisation, purpose limitation, and confidentiality



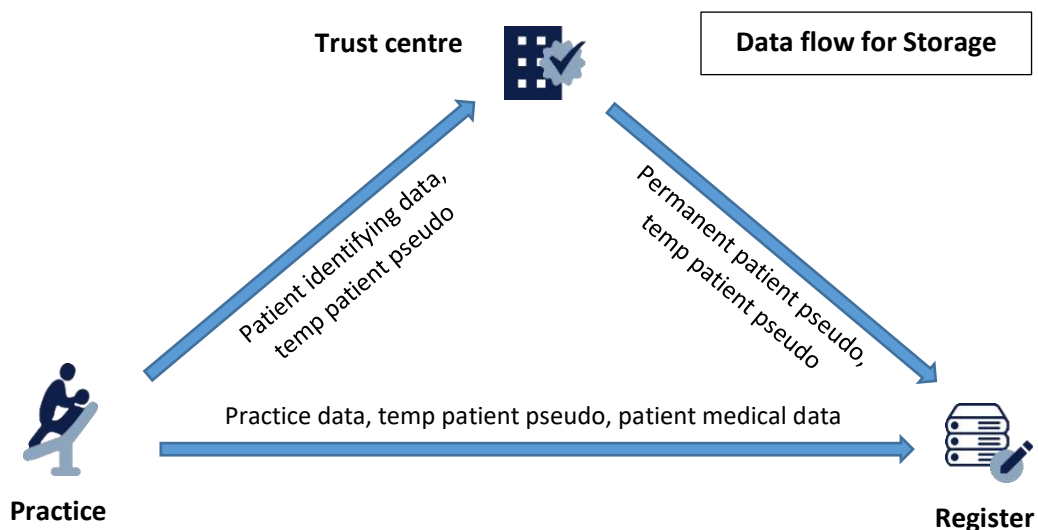
|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | <p>An agency is tasked with paying out subsidies to businesses according to criteria applying to both the businesses themselves, and their employees.</p> <p>Interested businesses submit applications containing data that prove that they meet those criteria, e.g. data on turnover and employee qualification.</p> <p>The agency verifies that the criteria are met. For a random sample of applicants, it requests further documents proving identity and qualifications of employees.</p>                                                                                                                                                                                                                                                                                      |
| Problem to be solved              | <p>Minimise access to employee data while retaining the ability to check the identity of the employees—in randomly chosen cases or in cases of special concern (suspicion of fraud)—in order to comply with the data minimisation principle, Art. 5(1)(c) GDPR, and data protection by default, Art. 25(2), with regard to access to data allowing attribution of data to employees, as well as reducing confidentiality risks and the risk that the received data about an employee is used outside the context of the application it is contained in.</p>                                                                                                                                                                                                                          |
| Original data                     | <p>The bulk of the data processed is non-personal business data. Data concerning the employees contain information about their identity (name, demographics), and professional qualifications.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Pseudonymisation domain           | <p>The Assistance agency.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Pseudonymisation                  | <p>The Agency sets up a separate organisational unit, which serves as a verification centre with the sole task of safeguarding the identity of the employees by handling the pseudonymisation and, if called upon in individual cases, the verification of the integrity of the applications. The verification centre receives the applications, stores all attributes describing the civil identity of the employees (names, date of birth, etc.) in a lookup table connecting this data with the application registration number and a pseudonym, replaces all those attributes by the pseudonym and turns over the result to the Agency.</p> <p>For each application, new pseudonyms are randomly chosen in order to prevent the linkage of data records across applications.</p> |
| Pseudonymised data                | <ul style="list-style-type: none"> <li>– Data concerning the qualification of the employees</li> <li>– Pseudonyms for employees</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Additional information            | <p>Lookup table linking employee pseudonyms with identifying data.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Processing of pseudonymised data  | <p>The agency assesses the applications.</p> <p>In randomly chosen cases or in cases of special concern, it turns over the pseudonym of an employee and the data concerning their qualification to the Verification centre for verification.</p> <p>The Verification centre uses the lookup table to establish the civil identity of the employees and directs an inquiry to the business requesting additional documents for verification of identity and qualifications claimed.</p>                                                                                                                                                                                                                                                                                               |

*Variant (using commitments)*

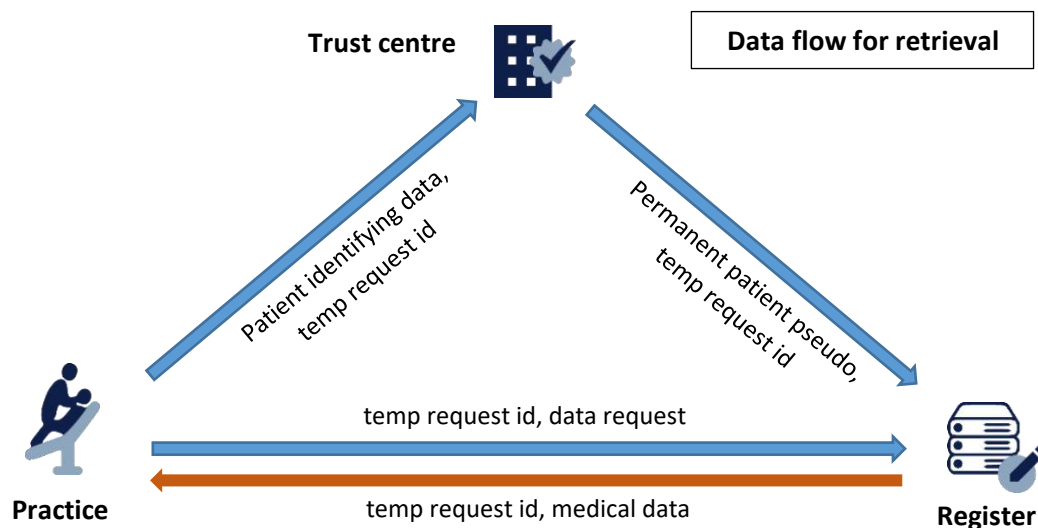


It is possible to avoid the establishment of the additional Verification centre by using cryptographic means. The Agency provides the businesses with a web app that allows the businesses to perform the pseudonymisation themselves. The pseudonyms include cryptographic commitments<sup>39</sup> of the employee documentation. In randomly chosen cases or in cases of special concern, the agency (or possibly a dedicated organisational unit thereof) requests the original documents proving identity and qualification of certain employees chosen by the Agency from the respective business. The binding property of the commitment assures that the Agency has full control over which documents to request, and the business is not able to substitute data of one employee for that of another.

**Example 3: Data minimisation and purpose limitation in the course of external analysis**



<sup>39</sup> Roughly speaking, a cryptographic commitment is a cryptographic protocol that allows one party (called the prover) to commit to holding some data by sending a message  $m$  which is derived from the data to another party (called the verifier) while hiding its content from the verifier. The verifier may ask the prover to disclose the original data, and is able to ascertain whether the message  $m$  has actually been computed starting from the original data as presented. One says that the prover is bound to the data. For a simple (but not verifiably strong) commitment, it suffices to compute  $m$  as the cryptographic hash of the input data extended by a secret random nonce of sufficient entropy.



|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | A Register collects data about dental implants for purposes of quality control. The Register uses the data to analyse the quality of the implants, and provide a summary of the results to the companies providing the material for making them. It also provides feedback to the practices on the quality of the care provided. Moreover, the stored data may be retrieved by subsequent caregivers upon consent by the data subjects. (Persons working for the register have no access to medical data beyond what is stored in the register.)   |
| Problem to be solved              | Retain the link between data and data subjects while ensuring compliance with the data minimisation principle, Art. 5(1)(c) GDPR, and data protection by default, Art. 25(2)—in particular with regard to access to data allowing attribution of data to data subjects—as well as strengthening purpose limitation, and reducing confidentiality risks. Nobody accessing data in the register should be able to attribute it to the data subjects and use them for incompatible purposes, e.g., to address data subjects for advertising purposes. |
| Original Data                     | <ul style="list-style-type: none"> <li>– Data identifying patient</li> <li>– Information about the implant, the operation, and other medical data about the patient</li> <li>– Data about the dental practice</li> </ul>                                                                                                                                                                                                                                                                                                                           |
| Pseudonymisation domain           | Register                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Pseudonymised data                | <ul style="list-style-type: none"> <li>– Patient pseudonym</li> <li>– Information about the implant, the operation, and other medical data about the patient</li> <li>– Data about the dental practice</li> </ul>                                                                                                                                                                                                                                                                                                                                  |
| Additional information            | <ul style="list-style-type: none"> <li>– Lookup table collating patient identifying data with pseudonyms</li> <li>– Original medical data together with patient identifying data held by the dentists' practices</li> </ul>                                                                                                                                                                                                                                                                                                                        |
| Pseudonymisation process          | Dentists transmit medical data and data relating to their practice accompanied with a temporary patient pseudonym to the Register. They also transmit those temporary pseudonyms together with patient                                                                                                                                                                                                                                                                                                                                             |

|                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                               | <p>identifying data to a designated Trust Centre<sup>40</sup> for safeguarding. The Trust Centre assigns a permanent patient pseudonym (either an existing pseudonym if the Trust Centre has a record for the patient on file, or a newly generated one), stores the new entry (if any) in the lookup table and transmits the permanent pseudonym along with the temporary pseudonym to the Register. The Register stores the data it received from the dentist together with the patient pseudonym it received from the Trust Centre. All parties delete the temporary patient pseudonym.</p> <p>When patients opt to allow dentists and other medical practitioners treating them subsequently to retrieve data relating to them, those practitioners send the data request by the same procedure to the Registry. The registry is able to lookup all data relating to the patient, and transmit the retrieved data back to the requesting practice.</p> |
| Processing of pseudonymised data                              | <p>The Register is able to link all cases relating to a given patient, or a given practice. Data from a given practice is analysed to provide aggregated data on the quality of care provided by this practice. Data relating to a given practice can be conveyed using the procedure described above to any subsequent practitioner treating that patient. All medical data, including data on the implants used, are analysed to obtain findings regarding the quality of those implants.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Additional safeguards particularly pertinent in this scenario | <ol style="list-style-type: none"> <li>1. The original data are kept confidential by the controllers who collected them, under obligation of professional secrecy.</li> <li>2. The Trust Centre safeguards the lookup table connecting the civil identity of the data subjects with pseudonyms used for long-term storage.</li> <li>3. All participating entities are bound by contract or another legal act to execute the protocols for the exchange of data faithfully.</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

#### Example 4: Safeguarding identity – confidentiality and accuracy

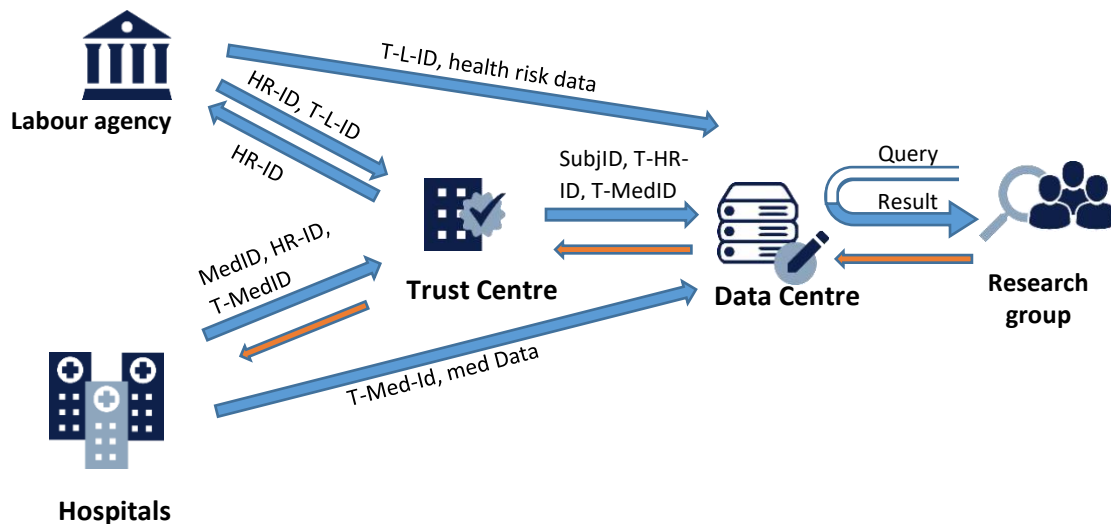
A medical laboratory wants to notify test results to its users via a mobile message. For this purpose, it enrolls users' mobile phone numbers (applying the necessary confirmation procedure). Before medical analysis is carried out, the laboratory transforms the identity and contact data of the patients and those relating to the date, time and scope of the test into a pseudonym. Those pseudonyms are coded as barcode or a QR codes, which is attached to test tubes containing the patients' samples. The pseudonymisation procedure assures that even samples pertaining to data subjects with very similar identity and contact data carry widely differing pseudonyms. Personal and contact data in intelligible format are kept separately by the laboratory. The analysis is carried out using only the pseudonyms to label the case. Afterwards, the procedure for notifying the results of the examination to the customer can be automated with lower risk of human errors and potential identity exchanges (for example in case of homonymy or in presence of similarities in contact data) and the accuracy of the data is reinforced. The richer the number of attributes that

---

<sup>40</sup> Here, a trust centre is an entity that performs security critical processing operations under contract with the relying parties.

are transformed into a pseudonym, the less likely it is that test results are inappropriately assigned to data subjects, and the lower is the likelihood of negative impacts on data subjects.

#### Example 5: Secondary use for research

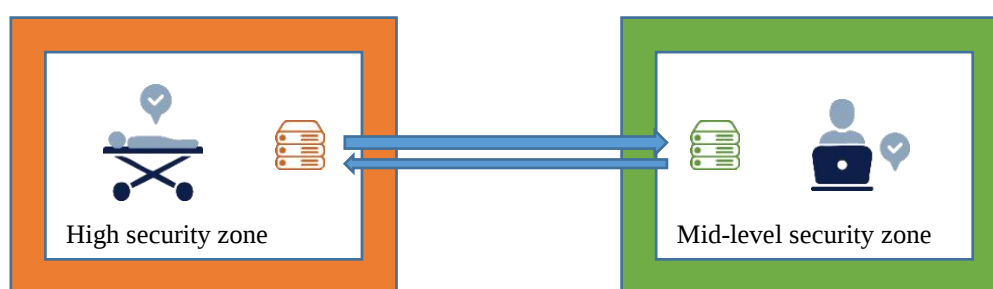


|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | <p>A Data Centre (established by a consortium of universities as a separate organisational unit at one of its members) collects data about the health and medical treatment of participants of a large longitudinal research project as well as data about occupational exposure to health hazards.</p> <p>The Data Centre receives health data from participating university hospitals, collects the data about occupational exposure to health hazards from a Labour agency that this agency has previously collected from employers. The centre provides the results of queries on the data to individual studies upon approval of the request by the data access board. It also co-ordinates access to original medical records for quality control purposes and informs patients of any significant unanticipated risks that studies may have identified.</p> |
| Problem to be solved              | Collect and link data from independent sources, maintain the link to records at the contributing institutions and to data subjects, while preventing attribution of the data to data subjects by the employees of the data centre and the research groups in compliance with Art. 89(1) GDPR.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Original Data                     | <ul style="list-style-type: none"> <li>– Data directly identifying the patient / employee</li> <li>– Medical data</li> <li>– Data about occupational exposure to health hazards</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Pseudonymisation domain           | <p>Data centre</p> <p>Research groups at participating universities.</p> <p>Members of these groups have no access to health records relating to the care of patients at their respective university's hospital.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Pseudonymised data                | <ul style="list-style-type: none"> <li>– Different pseudonyms at various stages of processing</li> <li>– Medical data</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

|                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | <ul style="list-style-type: none"> <li>– Data about occupational exposure to health hazards</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Additional information           | <ul style="list-style-type: none"> <li>– Original data maintained at the source institutions (hospitals, employers, labour agency)</li> <li>– Similar data about data subjects held by other medical service providers or by institutions with insight into the employment situation of the data subject provided it is linkable to the above mentioned original data without using directly identifying data</li> <li>– Pseudonym lookup tables held by the Trust Centre</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Pseudonymisation process         | <p>For performing the main pseudonymisation processes the consortium employs a Trust Centre.</p> <p>When data subjects sign-up for participation in the project at one of the members of the consortium, they are assigned a medical data ID (MedID), which is computed from data in the Electronic Health Record all members of the consortium that have treated the patient have access to. The hospital collects the human resources ID (HR-ID) used by the Labour agency, transmits it to the Trust Centre together with the MedID and then erases it.</p> <p>Hospitals transmit medical data together with a temporary pseudonym T-Med-ID to the Data centre, and the MedID together with the same temporary pseudonym to the Trust Centre. The Trust Centre requests data about occupational exposure to health hazards from the Labour agency using the HR-ID, which the Labour Agency subsequently transmits to the Data Centre. Again temporary pseudonyms (T-L-ID) are used for this transmission, which are also transmitted to the Trust Centre together with the HR-ID included in the request.</p> <p>The Trust Centre generates a data subject ID (SubjID) for each data subject and maintains a lookup table connecting MedID, HR-ID and SubjID. The SubjID is then combined with the temporary pseudonyms and send on to the Data Centre, which replaces the temporary pseudonyms with the data subject ID (SubjID) in all incoming data and links all data it obtains that contain the same SubjID.</p> |
| Processing of pseudonymised data | <p>The Data Centre provides the collected data to research groups upon approval of the request by the data access board. As part of the decision about access to data, the data access board seeks contractual guarantees from the receiving institution that all members of the research group are prevented by technical and organisational safeguards from access to any additional information that would allow attribution of the pseudonymised data to data subjects. Moreover, the institution commits to proceed with any further processing of the data it receives only upon approval by the data access board. Study groups do not receive the raw data stored in the Data Centre, but only the result of queries on the data executed within a secure processing environment.</p> <p>If access to the original data is requested to assure the quality and integrity of research, or if the data subject needs to be informed about hitherto unknown and significant individual risks, then the pseudonymisation process is reversed at the Trust Centre in order to</p>                                                                                                                                                                                                                                                                                                                                                                                                                                      |

|                                                               |                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                               | enable the necessary processing for those purposes. (The corresponding data flow is depicted with red arrows in the graphic above.) The hospital which submitted the last set of medical data relating to the data subject is responsible for contact with this individual.                                                       |
| Additional safeguards particularly pertinent in this scenario | Employees working in the Data Centre have no access to medical data from treatment at their institution, which is assured by separating it using organisational and technical means.<br>The Trust Centre is an independent service provider working under contract with and taking instructions only from the consortium's board. |

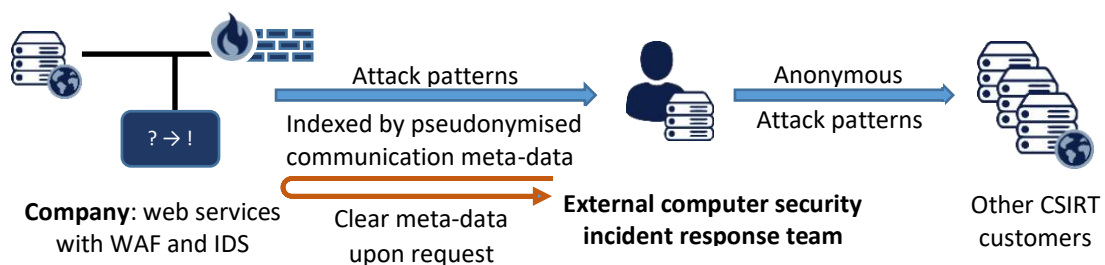
### Example 6: Reduction of confidentiality risks



|                                   |                                                                                                                                                                                                                                                                                                         |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | A large university hospital seeks to optimise its service portfolio and billing procedures by analysing treatment data.                                                                                                                                                                                 |
| Problem to be solved              | Allow the analysis of highly sensitive medical data by non-medical administrative staff operating in a mid-level security environment. The ability to provide feedback to care managers on a case specific basis needs to be retained in case irregularities are found in the data.                     |
| Original Data                     | Per case: <ul style="list-style-type: none"> <li>– diagnoses,</li> <li>– length of stay,</li> <li>– resources spent on the care of the patient,</li> <li>– diagnostic procedures and therapeutic interventions applied,</li> <li>– patient and case ID,</li> <li>– patient identifying data.</li> </ul> |
| Pseudonymisation domain           | All entities not having legal access to original treatment data identifying the patients                                                                                                                                                                                                                |
| Pseudonymised Data                | Per case: <ul style="list-style-type: none"> <li>– diagnoses,</li> <li>– length of stay,</li> <li>– resources spent on the care of the patient,</li> <li>– diagnostic procedures and therapeutic interventions applied,</li> <li>– encrypted patient and case ID.</li> </ul>                            |

|                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Additional information           | <ul style="list-style-type: none"> <li>– Encryption key</li> <li>– Original hospital records</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Pseudonymisation process         | For transmission to a database which operates outside the medical network zone, attributes relevant to the analysis are selected omitting highly individual documents (like discharge letters) and attributes which allow employees outside the medical departments to identify the patients directly. The selected attributes are transmitted together with the encrypted patient and case id for all records not presenting particular confidentiality risks e.g. due to the notoriety of a case, public interest in the patient, or affiliation of the patient with the hospital.                                                   |
| Processing of pseudonymised data | <p>The analysis of the pseudonymised data is performed relying exclusively on the data in the dedicated database. Only non-medical personnel that has no access to the hospital information system is allowed to work with the database.</p> <p>Pseudonymisation contributes to the security of the data: A person accessing the database without authorisation and without prior knowledge of the health status of the selected patients will not be able to draw conclusions about the health status of any individual. Accordingly, a placement of the database in a mid-level security environment can be considered adequate.</p> |

#### Example 7: Risk reduction as a factor in the balancing of interests, and ascertainment of compatibility of purposes

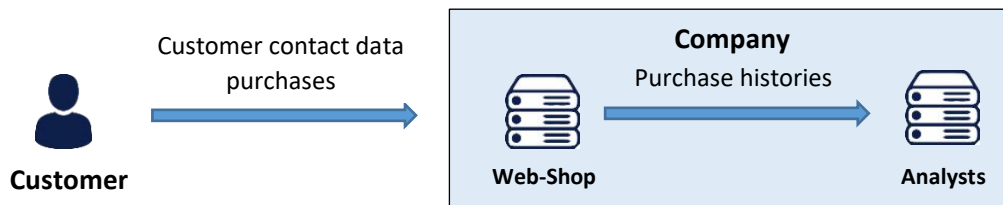


|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | <p>A Company provides various services to the public, which are provisioned by web services, and service interfaces placed in a demilitarized zone of the company network. Those services have varying sensitivity, and include online counselling in the course of which information might be revealed that indicates behaviour that, if known, could lead to the data subject being ostracised and severely disadvantaged in public, including very serious discrimination and severe bodily harm (Example: paedophilia).</p> <p>The company employs a web application firewall (WAF), an intrusion detection system and various system logs for detecting attacks against the security of its systems and services.</p> <p>In the case of a security incident, the Company intends to grant access to logged data to an external Computer Security Incident Response Team (CSIRT) for forensic analysis. For purposes acc. to Rec. 49, the</p> |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | CSIRT will also use the data for security services it extends to other customers.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Problem to be solved             | Generally, the grant of access to the data by the Company to the CSIRT, and its subsequent processing by the CSIRT can be considered to be based on legitimate interests, Rec. 49. Due to the sensitivity of some of the services, and the data processed therein, however, those interests may in turn be overridden by the interests of the data subject provided the processed data can be attributed to the data subjects. Likewise, in view of possible consequences of the intended further processing for data subjects, the purposes pursued by transmission to the CSIRT may not be compatible to the purposes of the original processing (online counselling).                                                                |
| Original Data                    | Traffic and content data (e.g. queries that triggered the WAF).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Pseudonymisation domain          | CSIRT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Pseudonymised Data               | Filtered traffic and content data with identifying information removed or transformed (in particular IP addresses, access tokens, and login credentials).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Pseudonymisation process         | After real time analysis, data is filtered, identifying information (IP addresses, access tokens, login credentials) transformed by a keyed cryptographic one-way function (provided the information contains sufficient entropy) or removed (otherwise), and the resulting data sets collected in a centralised log repository from which they are extracted for transmission to the CSIRT. Moreover, during the extraction process any content data still contained in the repository is reduced to fragments that do not permit the derivation of any information concerning data subjects beyond the fact that a query they have issued via their browser in the course of the use of Company's services contained those fragments. |
| Additional information           | <ul style="list-style-type: none"> <li>– Original log data.</li> <li>– Cryptographic key.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Processing of pseudonymised data | <p>The CSIRT analyses the data describing the security incident. In this process, it is able to link various log entries by the filtered and transformed traffic data (e.g., by time, source, and destination), including likewise transformed access tokens or other credentials.</p> <p>The CSIRT may request to obtain those IP addresses in the clear that are clearly linked to the attack and not to legitimate users of the services.</p> <p>The CSIRT anonymises the data to produce information about attack methodology or source, and transmits this information to other customers.</p>                                                                                                                                     |
| Effect                           | Under these conditions, the pseudonymised data transmitted to the CSIRT do no longer permit attribution of the data to specific data subjects by the CSIRT (with the possible exception of persons involved in the attack). Considering those measures, the Company and the CSIRT may consider the risk reduction achieved by pseudonymisation in their assessment whether Art. 6(1)(f) GDPR is a suitable legal basis for their data processing (insofar as it is not already covered by the legal basis                                                                                                                                                                                                                               |

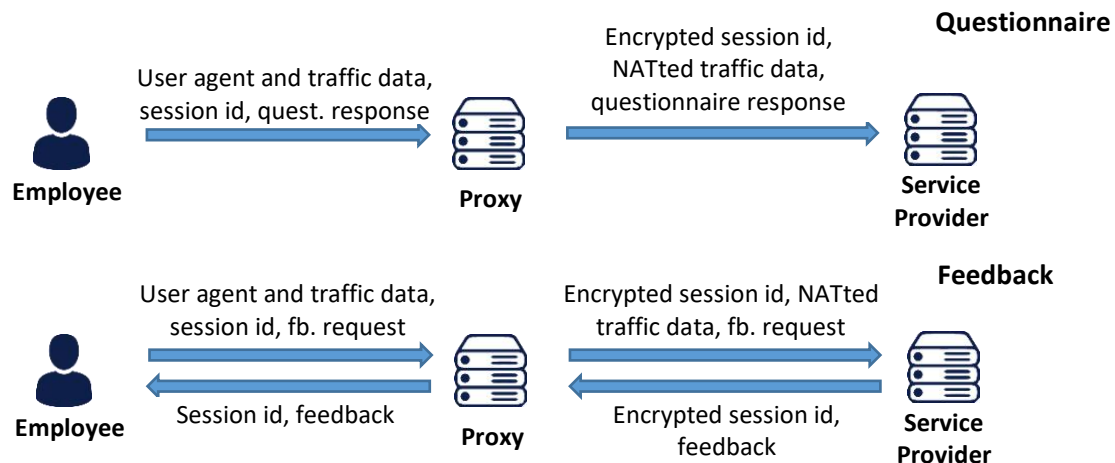
|  |                                                                                                                                                        |
|--|--------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | that allowed the data collection). Moreover, the Company can do likewise in its assessment of compatibility of purposes in light of Art. 6(4)(e) GDPR. |
|--|--------------------------------------------------------------------------------------------------------------------------------------------------------|

#### Example 8: Risk reduction justifying further processing



|                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of processing | A Company operates a large web-shop for a variety of products. Data about customers' purchases is stored and presented in customer accounts. The Company intends to extract data from the underlying database to find correlations between the products or services purchased.                                                                                                                                                                                                                                                                                                                                                             |
| Problem to be solved              | Due to the wide spectrum of products and services offered by the Web-Shop, purchase records may allow significant conclusions to be drawn regarding the data subjects, and may allow an evaluation of personal aspects relating to the economic situation, health, personal preferences, interests, or behaviour of data subjects. In order to be considered compatible to the purpose for which the personal data were initially collected, and avoid profiling of the customers acc. to the criteria in Art. 4(4) GDPR, the data has to be processed in a manner that the analysts can no longer attribute it to specific data subjects. |
| Original data                     | <ul style="list-style-type: none"> <li>– User profile</li> <li>– purchase history</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Pseudonymisation domain           | Team of Analysts                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Pseudonymised data                | Purchase history with all individualised entries removed (e.g., clothing with lettering chosen by the customer)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Additional information            | Original customer account.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Pseudonymisation procedure        | The Company extracts the purchase history omitting all individualised entries and directly identifying attributes, and assigns the analysis to an Organisational Unit of Analysts with no access to further customer data.                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Processing of pseudonymised data  | The Analysts perform the desired analysis, and summarise the results in aggregate form. Afterwards, the Organisational Unit erases all personal data it holds.                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Effect                            | The processing performed in this way is unlikely to affect data subjects. The controller can use this effect of pseudonymisation in its assessment of compatibility of purposes according to Art. 6(4) GDPR. Taking also into account the other factors mentioned in Art. 6(4) GDPR and depending on the particularities of the concrete case, the assessment may arrive at the conclusion that the purpose of the analysis can be considered compatible with the purpose for which the personal data were initially collected.                                                                                                            |

## Example 9: Supplementary measure



|                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Context and purpose of the processing | <p>A Company that belongs to a group of undertakings controlled by another company outside the EEA would like to use a personnel survey to improve work conditions and talent retention. The company has performed a careful assessment of the rules and requirements placed on it by Member State law according to Art. 88 GDPR, and put in place all necessary safeguards to guarantee lawfulness of the processing, including voluntariness of participation.</p> <p>Like all members of the group, the Company avails itself of the services of a Service Provider, which is located in a third country outside the EEA.</p>                                                                                                                                                  |
| Problem to be solved                  | <p>The export of personal data has to conform to the requirements of Chapter V of GDPR. Even though the Company and its Service Provider have concluded a contract containing standard data protection clauses acc. to Art. 46(2)(c) GDPR, their transfer impact assessment identified that the Service Provider would not be able to comply with certain provisions of the clauses because of conflicting requirements in its domestic legal system that go beyond what is necessary and proportionate in a democratic society.</p>                                                                                                                                                                                                                                              |
| Original Data                         | <ul style="list-style-type: none"> <li>– Traffic data stemming from the interaction with the online questionnaire.</li> <li>– Questionnaire response with closed-ended answers mostly regarding personal outlook, attitudes and assessments of the work environment, but also including a very small number of coarse demographic attributes about gender, age group, time spent in the employment of the Company, and current role.</li> </ul> <p>The selection of those attributes is carefully calibrated to ensure that there are at least 5 (or no) employees in each category formed by any combination of them. No other attributes describing characteristics of the data subjects that can be observed by a third party are contained in the questionnaire response.</p> |
| Pseudonymisation domain               | Service Provider, and any other non-EEA entity                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pseudonymisation procedure      | The Service Provider operates a server that provides an online questionnaire, which the Company offers to a section of its personnel (not including middle and upper management) through a proxy operated by itself. Employees use dedicated and company supplied disposable browser instances to interact with the online questionnaire. All interactions of an employee with the questionnaire form one session, which is assigned a unique session identifier chosen from a sufficiently large pool and displayed to the employee. The proxy replaces all data describing the user agent with dummy data, substitutes client IP and port by NAT, encrypts the session id in http requests <sup>41</sup> , and decrypts them in http responses. <sup>42</sup>                                                                                                                                         |
| Pseudonymised data              | <ul style="list-style-type: none"> <li>– Encrypted session id substituting all client traffic data with the exception of the client network address, which is transformed by Network Address Translation.</li> <li>– Questionnaire response.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Additional information          | <ul style="list-style-type: none"> <li>– Encryption key</li> <li>– Original client traffic data at the time of processing</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Processing of pseudonymous data | The Service Provider collates all questionnaire responses by pseudonymous session id. Upon receipt of all questionnaire responses, the Service Provider performs the requested analysis. It submits recommendations to the Company and provides the aggregated survey results derived from the responses it received to demonstrate the basis for its recommendations. Moreover, it provides individual feedback to all employees who have indicated that they wished to receive it, and consented to the processing involved. In order to receive it, employees have to note down the session id assigned to them when they fill out the questionnaire, and enter it into the feedback form. The form is provided through the proxy in the same way as the questionnaire encrypting and decrypting the session id as needed. After the performance of the task, all personal data received is deleted. |
| Effect                          | The carefully controlled environment in which the questionnaire is filled out assures that the interaction with the questionnaire cannot be attributed to any other online activity by the respective employee. Moreover, the questionnaire responses by itself do not allow attribution to specific natural persons either. Hence, even if authorities of the third country obtain the data records held by the Service Provider, they will not be able to attribute the data to the corresponding data subjects. Hence, upon careful analysis, including that the pseudonymisation measures have been effective to achieve their stated purpose, the preconditions of Art. 46 GDPR can be considered to be fulfilled in this example.                                                                                                                                                                 |

---

<sup>41</sup> The transformation of the session id can also be effected by creating random substitutes and storing them in lookup tables.

<sup>42</sup> In order to protect the identity of employees with unusual working hours either batch processing can be employed (in which case the granularity of the submission time is reduced) or the data collection is limited to usual working hours (e.g., by shutting down the service providing the dedicated browser used for submission).

## Example 10: Granting access rights to pseudonymised data

A Company is using the services of an Identity Provider for identification and authentication of customers. The Company does not keep information about the legal identity of their customers, but stores all data labelled with the pseudonym assigned to the customer by the Identity Provider. When a customer asserts her rights to access or data portability, the Company does not attempt to ascertain the legal identity of the customer<sup>43</sup>, but—after due information of its customers about this process—uses the communication channel that already exists between the data subject and the controller via the Identity Provider. Upon authentication of the data subject, the latter can deliver the authentic pseudonym to the Company, which in turn provides the customer with a copy of her or his data.

Note that Art. 11(2) GDPR applies, and Art. 15 and 20 do not apply if data subjects are not in the position to provide the pseudonym that relates to them, and substantiate this relationship, e.g., in the case that they deregistered from the Identity Provider's service.

Note further that this example also shows the use of pseudonymisation as part of the implementation of the fairness principle.

## GLOSSARY

### **Additional information**

Additional information is information whose use enables the attribution of → pseudonymised data to identified or identifiable persons.

### **Attribution of pseudonymised data to data subjects**

Process that establishes that → pseudonymised data relate to an already identified person, or links the data to other information with reference to which the data subjects could be identified.

### **Consistent pseudonymisation**

Two sets of data are considered to be pseudonymised *consistently* if data contained in those sets and relating to the same person can be linked on the basis of the pseudonyms they contain.

### **Direct identifier**

A direct identifier is a data element (or set thereof) that has been assigned or is being used to distinguish the data subject it refers to from all others in the given context without requiring the use of → additional information. Examples are passport or social security numbers, or the set consisting of first and last name as well as date of birth.

### **Pseudonym**

Identifier that is added to data in the course of the → pseudonymising transformation and set in such a way that it can be attributed to data subjects only using → additional information.

### **Pseudonymised data**

Result of applying the → pseudonymising transformation to some personal data. Cannot be attributed to a specific data subject without → additional information.

---

<sup>43</sup> The WP29 Guidelines on the right to data portability, endorsed by the EDPB, state: "The ability for the data controller to request additional information to assess one's identity cannot lead to excessive demands and to the collection of personal data which are not relevant or necessary to strengthen the link between the individual and the personal data requested."

**Pseudonymisation domain**

Environment in which the controller or processor wishes to preclude → attribution of data to specific data subjects. May incorporate persons acting under the authority of the controller or processor, respectively, other natural or legal persons, public authorities, agencies or other bodies, and their respective technological and informational resources. Does not include persons authorised to process additional data allowing the attribution of the → pseudonymised data to data subjects.

**Pseudonymisation secrets**

Data that is used in the application of the → pseudonymising transformation or is created during that process. Usually tables matching → pseudonyms with identifiers of data subjects or cryptographic keys. Allows the computation of pseudonyms from certain identifying attributes. Part of → additional information.

**Pseudonymising controller or processor**

Controller or processor that uses pseudonymisation as a safeguard and modifies original data according to Art. 4(5) GDPR.

**Pseudonymising transformation**

Procedure that modifies original data in a way that the result cannot be attributed to a specific data subject without → additional information.