Statement

of the

TMF – Technology, Methods, and Infrastructure for Networked Medical Research

on the

Guidelines 01/2025 on Pseudonymisation of the European Data Protection Board of 16 January 2025

Berlin, 14.03.2025

Summary

Pseudonymisation is a widely used and accepted measure in research to reduce the risks of processing personal data without restricting the necessary use cases or the benefits of processing. Due to the particular sensitivity of health data required in medical research, this measure is of particular relevance here. Pseudonymisation is also defined in the GDPR and therefore has a prominent place as a possible protective measure in the processing of personal data. In this respect, it is generally to be welcomed that the European Data Protection Board (EDPB) has published a guideline on what is meant by pseudonymisation within the meaning of the GDPR and what limits and boundaries exist in this regard.[1]

Pseudonymisation is defined in the GDPR in Art. 4 No. 5 on the one hand as a specific type of processing of personal data and on the other hand as processing with a specific purpose.

With regard to the **type of processing in the case of pseudonymisation**, the GDPR very clearly describes in the definition in Art. 4 No. 5 that this is about the separation of information and that some of the information that can obviously be used to identify data subjects is subject to certain technical and organisational measures. Although the EDPB Guideline analyses the requirements for the type of processing in the context of pseudonymisation, it fails to recognise that this is an essential core of pseudonymisation and therefore necessary characteristics of pseudonymisation. In this respect, measures for implementing pseudonymisation are proposed in many places in the guideline that do not even have the necessary components of pseudonymisation with regard to the type of processing. Some of the proposed measures are more in the nature of erasure or would be applicable in the context of anonymisation, but have nothing to do with the separation of certain information and its further processing in a protected form.

The **objective of pseudonymisation** is described in Art. 4 No. 5 GDPR in such a way that it must be assumed that the part of the data that remains after the separation of the data with special identification potential should be as good as anonymous only if this definition is considered. A detailed expert opinion obtained by the TMF on this a few years ago concluded that for a comprehensive understanding of the concept of pseudonymisation in the GDPR, it is not sufficient to use the definition in Art. 4 No. 5 alone, but rather all uses or references to the concept in the GDPR must be considered [1]. Accordingly, the result of pseudonymisation may well be that part of the data no longer has a personal reference after certain information has been separated. As a rule, however, the result is a partial data set that is still personal data, which is merely reduced in the sense of data minimisation so that this partial data set no longer contains any data that is not necessary for the intended purposes of the processing. Instead of distinguishing between two possible and clearly different objectives of pseudonymisation, the EDPB guideline tends to follow a fuzzy middle ground by suggesting that pseudonymisation should bring the part remaining after the separation of the obviously identifying data as close as possible to an anonymous data set, without necessarily achieving anonymity. In this respect, the EDPB guideline interprets the objective of pseudonymisation far too broadly and then proposes measures to achieve this objective that are no longer covered by the definition of pseudonymisation in terms of the type of processing. However, this specification is neither helpful for the users of pseudonymisation, who want to act in the

[1] see https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en

sense of data minimisation without restricting use cases or data quality, nor for the users who, as a result of pseudonymisation, may actually be accessing data that is no longer personal data in the legal sense. Here, the shift in the objective of pseudonymisation towards anonymisation threatens to make the measure unattractive, especially for research, where this measure is still widespread and widely accepted today. It cannot be ruled out that, as a result of such a reinterpretation, even the basic and in practice very helpful measure of separating directly identifying data will be used less frequently than before. Legal requirements for pseudonymisation at national or even European level could also rule out many applications in scientific research in the future.

Since pseudonymisation, as defined in the GDPR, is a process of splitting data with different characteristics with regard to identifiability and their separate further processing, the question of whether the **concept of personal data,** on which the GDPR is based, must be understood as being based on a subjective or objective criterion, which has long been the subject of controversy, arises here in particular. As the ECJ has already taken a position on this question and concluded that the concept of personal data in the GDPR is based on a subjective criterion [2; 3], it is regrettable that the EDPB Guideline does not take a clear position on this. On the contrary, there are many formulations in the guideline that suggest an absolute understanding in the sense of being based on an objective criterion and thus unnecessarily restrict the scope for pseudonymisation with regard to different objectives.

Furthermore, the guideline contains many legally and technically correct statements that may well be helpful for different users. However, as they are embedded in a vague understanding of pseudonymisation with regard to both the type and the objectives of processing, the usefulness of the guideline is severely limited. Particularly in research, the importance of pseudonymisation would decrease drastically if the understanding set out in the guideline were to prevail, as experience has shown that many of the measures described would be accompanied by a significant loss of data quality. In this respect, a fundamental revision of the guideline in line with the analysis presented here is required.

The type of processing of personal data in the context of pseudonymisation

Pseudonymisation is defined in the GDPR in Art. 4 No. 5:

> *'pseudonymisation' means 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;*

In this definition, references to a specific type of processing and the purpose of the processing are closely intertwined. If one considers only the part of the definition that deals with the type of processing, then it refers to information held separately and to the application of technical and organisational protection measures to which this separately held information is subject. These references to a specific type of processing are obviously incomplete in the brevity chosen here. In order for certain information to be kept separately, information must first be divided up and then separated.

It should no longer be possible to assign the data to the data subjects after the separation of certain information without the use of this separated information. However, this leaves open the possibility that an association may be possible with the use of the separated information. In these cases, the processing must be carried out in such a way that a link between the separated information and the data actually to be used must remain or be established. This link is typically created through the use of pseudonyms, which are also not mentioned in the definition. Separated information is then replaced by a pseudonym.

The definition also leaves open the nature of the technical and organisational protective measures to be applied to the separated information. In practice, for example, there is the use case on the one hand where the separated information is stored and processed by a trusted third party and on the other hand where a cryptographic transformation is applied to the separated information. This cryptographic transformation can be designed as reversible encryption or as one-way encryption. In all cases, reversibility of the pseudonymisation is technically possible. In the case of one-way encryption, the key for this is held in part by the pseudonymising controller or processor, who knows the procedure and the encryption parameters used, and in part by the data subject, who knows the data included in the encryption.

In Section 2 "Definitions and legal analysis", the EDPB Guideline only contains an incomplete and vague analysis of the type of processing of personal data in the context of pseudonymisation (see paragraphs 18–20). According to this, the generation and use of additional information should be an inherent part of pseudonymisation (para. 19), which would leave open the possibility that completely different processing operations could also be part of pseudonymisation. This vague description fails to recognise that the separation of information and the application of certain protective measures to this information is a necessary characteristic of pseudonymisation. A processing of data that does not in any way relate to such a separation of information and the application of protective measures to this separated information is therefore not pseudonymisation. In this respect, the description of pseudonymisation in the guideline is vague and, as a result, the characteristics of

pseudonymisation described here cannot be used to differentiate the term from other types of processing.

In section 3.1, the guideline then contains a more detailed description of pseudonymised transformations, which deals in detail with the separation, replacement and protection of the separated data. Here, too, it is not recognised that pseudonymisation only covers certain types of processing. Paragraph 84 contains the statement:

> *[…] Insofar as necessary for pseudonymisation to have the intended effect, it also modifies other attributes, e.g. by removal, generalisation and noise addition.*

As part of pseudonymisation, measures such as the removal, generalisation or noise reduction of information are therefore proposed here, which are in no way related to the measures of separation and replacement of certain identifying information and its protection. The guideline then devotes the entire section 3.1.3 "Modification of original data necessary for the objectives of pseudonymisation" to measures that go beyond pseudonymisation. It becomes clear that, on the one hand, the guideline overlooks the fact that pseudonymisation is also defined by the type of processing and, in this respect, inappropriate processing cannot be part of pseudonymisation. On the other hand, this excessive part is presented as necessary to achieve the legally prescribed objective of pseudonymisation, so that this objective also requires a precise analysis.

The aim of pseudonymisation

The objective of pseudonymisation is described in Art. 4 No. 5 GDPR:

> *[…] 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 […]*

In addition, the "additional information" should be protected with the help of technical and organisational procedures to ensure that

> *[…] that the personal data are not attributed to an identified or identifiable natural person;*

If only this definition were taken into account to determine the objective of pseudonymisation, it would have to be assumed that the part of the data that remains after the separation of the data with special identification potential should no longer be personally identifiable without the addition of the additional information. The concept of personal data is only formulated positively in the GDPR in Art. 4 No. 1 and not in differentiation from data that can no longer be related to a person. There is also no definition of anonymous data. This is consistent to a certain extent, as the GDPR is only applicable to personal or personally identifiable data. Recital 26 of the GDPR, however, contains the decisive information on the distinction between personal data and data that is no longer personal. The latter are also referred to here as anonymous data. A precise comparison of the wording on the objective of pseudonymisation in Art. 4 No. 5 GDPR with the wording on the limit of personal data in Recital 26 GDPR shows that the wording on the objective of pseudonymisation with regard to the data remaining after separation is even more absolute and stricter than the distinction between data that can still be personal data and data that can no longer be related to a person formulated in Recital 26

would actually require. In this respect, the findings in Recital 26 GDPR must be considered when interpreting the criteria in the definition of pseudonymisation in accordance with Art. 4 No. 5 GDPR.

The pseudonymisation of data in research was already widely used in the German legal area before the GDPR became applicable. For example, a guideline on data protection in medical research with detailed specifications on the pseudonymisation of health data and biosamples was developed by the TMF and extensively coordinated with the data protection authorities in Germany [4]. The basis for this was, among other things, a definition in the former Federal Data Protection Act. The definition did not stipulate that the data remaining after the separation of the directly identifying information must be as good as anonymous without adding the separated data, nor was this regularly implemented in the application.

Against this background, the question of the interpretation of the objective of pseudonymisation in Art. 4 No. 5 GDPR with the introduction of the GDPR was obvious. The TMF obtained a comprehensive expert opinion on this, which came to the conclusion that for a comprehensive understanding of the concept of pseudonymisation in the GDPR, the use of the definition in Art. 4 No. 5 alone is not sufficient, but rather all uses or references to the concept in the GDPR must be taken into account [1, p. 174ff]. The expert opinion commissioned by the TMF comes to the conclusion that the result of pseudonymisation may well be that part of the data no longer has a personal reference after the separation of certain information according to the criteria of Recital 26, but that the result is generally a partial data set that continues to be personal data, which is merely reduced in the sense of data minimisation in such a way that this partial data set no longer contains any data that is not necessary for the intended purpose of the processing. The expert opinion resolves the tension between the need for data that is as good as anonymous on the one hand and data that can be linked to individuals on the other, to the effect that two types of pseudonymous data are considered possible. In order to understand the reasoning of the expert opinion, it is necessary to look at the position of pseudonymised data between personal data and anonymous data.

On the position of pseudonymous data between personal data and anonymous data

The guideline largely misses the opportunity to clarify the concept of pseudonymisation and thus eliminate existing uncertainties among legal users. In view of the importance of pseudonymisation in the GDPR, a clear positioning would be desirable. In particular, the TMF misses a discussion of the consequences of a subjective criterion when determining identifiability in multi-person constellations.

Based on the definition of Art. 4 No. 5 GDPR, one interpretation of pseudonymisation is possible, according to which, after the separation of the additional information from the remaining data, this remaining and then pseudonymised data should no longer be personally identifiable. The guideline uses the term "additional information" indiscriminately for both information that was separated during the pseudonymisation process and information that exists elsewhere regardless of this (for example, the guideline mentions social media posts or posts in forums, para. 21). The question of the assignability of the latter information is not a special feature of pseudonymisation, but a general criterion for identifiability. And so, the definition of pseudonymisation in Art. 4 No. 5 GDPR also clearly refers only to the information

separated in the context of pseudonymisation with regard to the information relevant for the identifiability of the data remaining after separation, but not to any information elsewhere.

In fact, the term pseudonymisation is used in many places and in different contexts in the GDPR. Recital 26 of the GDPR states:

> *[…] Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. […]*

In this respect, a certain tension arises between the description of pseudonymous data without access to the separated information as virtually anonymous in the definition in Art. 4 No. 5 GDPR and the statement that pseudonymous data should always be considered to be information on an identifiable natural person in Recital No. 26 GDPR. The decisive factor here is obviously whether the controller who processes the pseudonymous data remaining after the separation can be held accountable for access to the separated information. Constellations are conceivable and in certain cases also implemented in which the controller who processes the pseudonymised residual data has no access to the separated information. For such constellations, it must be assumed according to the case law of the ECJ that this processing no longer constitutes processing of personal data [2; 3].

Against this background, the tension between the demand for data that is no longer personally identifiable on the one hand and the statement that it should always be assumed that data is personally identifiable on the other can only be resolved by considering two different possible outcomes of pseudonymisation. In the first case, the processing of the pseudonymised residual data can actually be assessed as the processing of anonymous data; in the second case, the processing of personal data can still be assumed. In the second case, the pseudonymisation is only to be understood as a data-minimising measure, as suggested by the reference in Art. 89 para. 1 GDPR:

> *[…] Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. […]*

The question of whether the processing of the pseudonymous data remaining after the separation of directly identifying information can be categorised as processing of anonymous data must take into account the accessibility of the separated information on the one hand and the accessibility of further information enabling identification on the other. Pseudonymisation itself only refers to the first possibility, but a comprehensive assessment is always necessary to determine anonymity, as is rightly stated in the guideline in paragraph 22:

> *[…] 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.*

The next step is to determine the position of the guideline in the context of this analysis and to critically examine it in detail.

The position of pseudonymous data according to the guideline

How does the guideline deal with these two possible outcomes of pseudonymisation? It is not apparent that the expert opinion [1, p. 174ff] itself has been considered in the guideline, nor that the guideline has followed a comparably thorough approach in interpreting the concept of pseudonymisation. Instead of distinguishing between two possible and clearly different objectives of pseudonymisation, the guideline rather follows a fuzzy middle course by suggesting that pseudonymisation should bring the part remaining after the separation of the obviously identifying data as close as possible to an anonymous data set, without necessarily achieving anonymity. As a criterion for the effectiveness of pseudonymisation, reference is only made to sufficient risk minimisation, which on the one hand depends on the requirements of the legislator or the criteria for the applicability of a specific legal basis for processing (para. 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.*

It is noticeable here that the guideline would also leave it to the national legislator to specify not only when pseudonymisation should be applied, but also when pseudonymisation achieves sufficient risk reduction. However, if the degree of risk reduction is a criterion for determining whether pseudonymisation is effective or not, the latter would mean that the national legislator could specify criteria for achieving pseudonymisation and thus also influence the definition of pseudonymisation. This obviously cannot be the intention of the European legislator.

On the other hand, the guideline leaves sufficient risk reduction to the controller (para. 25):

> *When such specific mandates for pseudonymisation are absent, controllers themselves may define the objectives 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.*

This focus on a criterion of sufficient risk reduction as the objective of pseudonymisation overlooks the fact that a risk analysis typically always considers the result of an entire bundle of technical and organisational measures in combination and does not look at one measure alone. This is also formulated very clearly in Art. 35 para. 7 lit. d on the necessary content of a data protection impact assessment:

> *[The assessment shall contain at least:] the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.*

In the absence of a clear definition of the different outcomes of pseudonymisation, from complete anonymity on the one hand to pure data minimisation in accordance with the criteria of necessity on the other, the guideline establishes a risk criterion that overburdens pseudonymisation as a technical and organisational measure, the effect of which must

generally be assessed in conjunction with other measures, thus making it more difficult to apply and restricting its use, e.g. for research purposes.

This becomes very clear when dealing with identification risks in a pseudonymised data set. Variables in such a dataset whose characteristics in combination could potentially allow the identification of individual data subjects – here called quasi-identifiers – should be analysed and require post-processing. It is precisely the existence of such identification risks that characterises personal data. If such risks did not exist, one would have anonymous data – assuming no access to the separated directly identifying data. Possible quasi-identifiers are mentioned (para. 101):

> *[…] 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. […]*

And for the post-treatment is then proposed in para. 101:

> *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.*

However, such anonymisation measures, which are accordingly irreversible, lead to a considerable impairment of the data quality of a dataset. The TMF has been working intensively on the anonymisation of research data for over 10 years and has developed a further training programme for this purpose, which has been used extensively for many years. However, it can also be deduced from this intensive involvement that such measures are only applicable – and even then, only in very specific cases – if the research question to be answered with the data is fully known.

However, as recital 33 of the GDPR correctly reflects, it is often not possible to fully predict at the time the data is collected for which important questions it may be used at a later date, especially in research. This necessary consideration of future questions means that no variable that could become important at a later point in time should be irreversibly coarsened or even deleted at this time. If, for example, the age of data subjects is to be used as a grouping criterion (independent variable) for a later analysis, random noise in the values would have a direct negative impact on the accuracy of future analyses. However, a generalisation in the direction of the later grouping limits would perhaps be harmless. If, on the other hand, the age of two patient groups were to be regarded as the variable to be evaluated (dependent variable), a small amount of noise would possibly be less critical than a rough generalisation.

Here, the shift in the objective of pseudonymisation towards anonymisation threatens to make the measure unattractive, especially for research, in which this measure is still widespread and widely accepted today. It cannot be ruled out that, as a result of such a reinterpretation, even the basic and in practice very helpful measure of separating directly identifying data will be used less frequently than before.

Where pseudonymisation is already prescribed in national or European law today or will be prescribed in the future, for example in the context of the European Health Data Space (EHDS),

many use cases from scientific research could be excluded by such an understanding of pseudonymisation.

With regard to the goal of a certain risk reduction, which the guideline places solely on pseudonymisation, it unfortunately also obscures the fact that pseudonymisation can of course be supplemented by other measures in order to achieve certain goals. In particular, measures named in section 3.1.3 of the guideline can represent such useful additions. Although such a possibility is hinted at in para. 96, it is ultimately not used to clearly distinguish pseudonymisation from other measures:

> *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.*

It is positive to emphasise that the guideline explicitly points out that pseudonymisation can support compliance with the requirements of data protection-compliant processing (para. 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 […]*

However, it must be made clear that this task of ensuring adequate protection of the rights and freedoms of data subjects is generally not the responsibility of pseudonymisation alone, but of the interaction of various technical and organisational measures.

This requirement of a risk reduction, however measured, is neither helpful for the users of pseudonymisation, who want to act in the sense of data minimisation without restricting use cases or data quality, nor for the users who, as a result of pseudonymisation, may actually be accessing data that is no longer identifiable in the legal sense.

Lack of recognition of a subjective criterion for determining identifiability

Since pseudonymisation, as defined in the GDPR, is a process of splitting data with different characteristics in terms of identifiability and their separate further processing, the question of whether the concept of identifiability, on which the GDPR is based, should be understood as being based on a subjective or objective criterion, which has long been the subject of controversy, arises here in particular. The subjective approach would mean that the processing of data from which all personally identifiable information has been separated in such a way that there is no longer any access to the separated data could in principle also be understood as the processing of anonymous data, which could be carried out independently of the requirements of the GDPR. The objective approach, on the other hand, does not look at what possibilities the controller has to identify a person on the basis of the data accessible to them, but rather at whether there is still data at any point that would enable identification. Pseudonymous data, for which, by definition, there is supplementary data that enables identification, would therefore always be personal data. As the ECJ has already taken a position on this issue and concluded that the GDPR is based on a subjective criterion to determine identifiability [2; 3], it is regrettable that the EDPB Guideline does not take a clear position on

this.[2] On the contrary, there are many formulations in the guideline that suggest an objective or absolute understanding of identifiability and thus unnecessarily restrict the scope for pseudonymisation (e.g. para. 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, and is therefore personal. This statement also holds true if pseudonymised data and additional information are not in the hands of the same person. […]*

Recognising the concept of identifiability as being determined using a subjective criterion in accordance with the case law of the ECJ would be desirable and helpful for many users.

[2] The Advocate General's Opinion in Case C-413/23 P also points in the same direction (https://curia.europa.eu/juris/document/document.jsf?pageIndex=0&docid=295078&doclang=EN&text=&cid=5888244).

Literature

1. Dierks, C., Roßnagel, A., *Sekundärnutzung von Sozial- und Gesundheitsdaten – Rechtliche Rahmenbedingungen.* 2019, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, https://mwv-open.de/site/books/10.32745/9783954665181/ (accessed: 2025-03-10).

2. *Judgment of the Court of Justice (Second Chamber) of 19 October 2016. Patrick Breyer v. Federal Republic of Germany. Case C-582/14*. 2016. European Court of Justice, https://curia.europa.eu/juris/document/document.jsf?docid=184668&doclang=EN (accessed: 2025-03-10).

3. *Judgment of the European Court* of *Justice (Third Chamber) of 9 November 2023 Gesamtverband Autoteile-Handel e. V. v. Scania CV AB. Case C-319/22*. 2023. European Court of Justice, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62022CJ0319 (accessed: 2025-03-10).

4. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, https://www.mwv-open.de/site/books/10.32745/9783954662951/ (accessed: 2025-03-10).

Abbreviations

GDPR	Regulation of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC – General Data Protection Regulation (Regulation 2016/679)
EDSA	European Data Protection Board (https://edpb.europa.eu)
EC	European Community
EHDS	European Health Data Space, regulation of the European Commission on a European Health Data Space (
ECJ	Court of Justice of the European Communities (http://curia.europa.eu)
lit	Littera / Letter
TMF	TMF – Technology, Methods, and Infrastructure for Networked Medical Research (www.tmf-ev.de)