Douwe Korff

THE DATA PROTECTION OFFICER'S HANDBOOK:
A Practical Guide to the EU GDPR

[DRAFT]

EXCERPT: Chapter 7 (Definitions), section 2:[1]

2. Data subject, personal data, pseudonymous and anonymous data

2.1 "Personal data" and "data subject"

Article 4(1) GDPR:

'**personal data**' means any information **relating** to an identified or identifiable natural person ('**data subject**'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Recital 26:

"To determine whether a natural person is identifiable, account should be taken of **all the means reasonably likely to be used**, such as **singling out**, **either by the controller or by another person** to identify the natural person directly or indirectly."

(Emphases added; the emphasised elements are discussed below.)

In 2007, the Article 29 Working Party issued an opinion on the concept of personal data.[2] In it, it clarified first of all that:

> in order to consider that the data "relate" to an individual, a "**content**" element OR a "**purpose**" element OR a "**result**" element should be present.
>
> …
>
> These three elements (content, purpose, result) must be considered as alternative conditions, and not as cumulative ones. In particular, where the content element is present, there is no need for the other elements to be present to consider that the information relates to the individual. A corollary of this is that the same piece of information may relate to different individuals at the same time, depending on what element is present with regard to each one. The same information may relate to individual Titius because of the "content" element (the data is clearly about Titius), AND to Gaius because of the "purpose" element (it will be used in order to treat Gaius in a certain way) AND to Sempronius because of the "result" element (it is likely to

---

1 With minor edits, in particular to cross-references to other sections in the book and footnotes.

2 Article 29 Working Party, Opinion 4/2007 on the concept of personal data (WP136), adopted on 20 June 2007, p. 13, available at:
https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf
Although this opinion was not endorsed (as various other WP29 opinions were) by the EDPB at its first meeting on 25 May 2018 in the form of Endorsement 1/2018 it is, for now, still an authoritative document on the issue. However, in October 2025, the European Data Protection Board announced a "stakeholder event" on anonymisation and pseudonymisation following the clarification on the scope of the concept of personal data provided by the Court of Justice of the European Union (CJEU) in its judgement in *EDPS v Single Resolution Board (SRB)*, discussed below and in the next two sub-sections. See:
https://www.edpb.europa.eu/news/news/2025/anonymisation-and-pseudonymisation-take-part-stakeholder-event_en
This event "will inform and support the EDPB's ongoing work on these topics as per its work programme 2024-2025". This is likely to result in updated guidance. In the meantime, I have submitted the present section, with my take on that judgment and the broader issues, to the EDPB for information.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

> have an impact on the rights and interests of Sempronius). This means also that it is not necessary that the data "focuses" on someone in order to consider that it relates to him. Resulting from the previous analysis, the question of whether data relate to a certain person is something that has to be answered for each specific data item on its own merits.
>
> (pp. 10 and 11 – 12, emphases and underlining original)

This approach was endorsed by the Court of Justice in its 2017 *Nowak* judgment.[3] However, in its judgment in *EDPS v. SRB* (further discussed in the next sub-section), the Court held that it was not necessary to always examine all three of these *nexus* in detail: since in that case it was clear from previous case-law[4] that the data in question (opinions and assessments expressed by a reviewer) constituted personal data by virtue of their obvious link to the data subjects (i.e., those reviewers), there was no need to further examine whether they might be personal data by reference to the three specific tests.[5]

In relation to the "identifying" or "singling out" of individuals, the Article 29 Working Party Opinion clarified that:

> [On the one hand, a] very common family name will not be sufficient to identify someone - i.e. to single someone out - from the whole of a country's population, while it is likely to achieve identification of a pupil in a classroom. [On the other hand,] ancillary information, such as "the man wearing a black suit" may identify someone out of the passers-by standing at a traffic light. So, the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case. (p. 13)

More recently, the European Data Protection Board, too, stressed that if a person can be "singled out" from a dataset, even without being named, and different data ("records") relating to that singled-out person can be linked, so that conclusions can be inferred from the links, those data are personal data:[6]

> Evaluating the robustness of anonymisation relies on three criteria: (i) **singling-out** (isolating an individual in a larger group based on the data); (ii) **linkability** (linking together two records concerning the same individual); and (iii) **inference** (deducing, with significant probability, unknown information about an individual).

However, as I will note in the next sub-section, **the rules on pseudonymisation, and the definition of "identification" as used in that context, appear to differ and divert from this.**

In relation to the need, spelled out in Recital 26 (quoted above) to take account of "all the means reasonably likely to be used", the Court of Justice in its *Breyer* judgment agreed with

---

[3] Case C 434/16, EU:C:2017:994, para. 35, referenced (with subsequent cases in which the same view was adopted) in the Court judgment of 4 September 2025 in Case C-413/23 P, *European Data Protection Supervisor (EDPS) v. Single Resolutions Board (SRB)*, ECLI:EU:C:2025:645, at para. 55.

[4] Specifically, the Court's judgment in *Nowak*, referenced in the previous footnote.

[5] CJEU, *EDPS v. SRB* (footnote 3, above), paras. 58 – 60.

[6] EDPB, Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, adopted on 21 April 2020, para. 16, emphases added, available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-042020-use-location-data-and-contact-tracing_en

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

the Advocate General's Opinion in the case that there would be no "means likely reasonably to be used to identify the data subject" where:[7]

> the identification of the data subject was **prohibited by law or practically impossible** on account of the fact that it requires **a disproportionate effort** in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.

This issue, too, is of particular importance in relation to pseudonymised or (supposedly) fully-anonymised data, as I will also discuss (with some critical remarks) in the next sub-section.

2.2 "Pseudonymous" and "anonymous" data[8]

> **Article 4(5) GDPR:**
>
> '**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.
>
> * The additional data needed to directly identify the data subject is usually referred to as the [re-identification] "key" – DK
>
> **Recital 26:**
>
> ['**pseudonymous data**'] "should be considered to be information on an identifiable natural person".
>
> ['**anonymous data**' are] personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable"; the GDPR "does not … concern the processing of such anonymous information, including for statistical or research purposes."

The first point to be noted concerns the concept of "attribut[ing] [personal data] to a specific data subject". In ordinary language this is generally referred to as "identifying" the data subject. However, in this context, it cannot have the same meaning as "identifying" data subjects more generally, in line with recital 26, in relation to which the Article 29 Working Party and the EDPB rightly noted that it included "singling out" an individual from a larger data set, even if not by name (see sub-section 2.1, above).

In fact, the whole point of creating data sets with pseudonymised data is usually specifically to allow cross-referencing of data on the same individual, without that individual being "identified" in the more specific sense of being identified by name etc. Thus, in longitudinal medical research, patient data that is (often rather flimsily) pseudonymised is analysed to see if a specific patient (only "identified" in the data set with a patient number) who underwent certain treatment later developed medical issues or complications that could be attributed to the earlier treatment. If the data on the earlier treatment and on the later issues are correlated (which is the whole point), that patient is undoubtedly "singled out" from the other patients in the data set – and should therefore, under the general approach indicated in

---

[7] CJEU, second chamber judgment in Case C-582/14 of 19 October 2016, *Patrick Breyer v. Bundesrepublik Deutschland*, ECLI:EU:C:2016:779, para. 46, emphasis added.

[8] On 16 January 2025, the EDPB issued Guidelines 01/2025 on Pseudonymisation (v.1 for consultation), to which I will briefly refer several times. However, it does not address some of the issues covered by me here.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

recital 26, be regarded as "identified". But the wording of Article 4(5) GDPR suggests that the data are not "attributed to" the patient unless and until someone links the patient number to the name etc. of the patient. I believe this issue has not been sufficiently widely noted or addressed.[9]

Secondly, there have been conflicting views on whether the concept of "identifiable" data is relative or absolute. Those holding the former view argue that what is and what is not "personal data" is relative and depends on who processes the data and what knowledge and means they have to identify the data subjects (in the Article 4(5) sense):[10]

> Data can be 'just' pseudonymised [data] for one [entity] but anonymised [data] for another [entity].

According to the EDPS, in the *SRB* case, the General Court also took this view:[11]

> [A]ccording to the interpretation adopted by the General Court, personal data change in nature when they are transmitted to an entity outside the controller, which does not have the additional information enabling the data subject to be identified.

In other words, in this view, pseudonymised data are personal data for the entity that holds "the additional data" that are needed to re-identify the data subjects (in the Article 4(5) sense), but they are not personal data for any recipient who does not have access to this additional data.[12]

The EDPS rejected this approach because, in his opinion:[13]

> That interpretation would allow such a controller unduly to remove personal data from the scope of EU law on the protection of such data, even where the processing by the external entity would expose the data subjects to significant risks.

Moreover, the EDPS argued that:[14]

> [B]y introducing the concept of pseudonymisation, the EU legislature clarified that, in order to exclude personal data from the scope of EU law on the protection of such data,

---

9 The Data Governance Act, briefly discussed in Part One, chapter 5, section 3 of the handbook, similarly somewhat misleadingly (intentional or otherwise) defines "**anonymisation**" as: "*the process of changing documents into anonymous documents which do not relate to an identified or identifiable natural person, or the process of rendering personal data anonymous in such a manner that the data subject is not or no longer identifiable*" (Article 2(7); see also recital 52), without mentioning the reference to "singling out" of individuals, set out in recital 26 to the GDPR. This clearly suggests that the drafters of that Act intended to allow processing of pseudonymised data (such as the patient data referred to in the above example, when held by public sector entities) with minimal restraint – in fact, as noted in that earlier chapter, that is the main aim of the Act. But that ignores the constraints I discuss here that flow in particular from the fact that whatever the Act says, the act of anonymising data is by definition performed on still-identified data and is therefore a form of processing of personal data that must conform to the GDPR in all respects.

10 "*Daten [können] für die einen 'nur' pseudonyme und für andere anonyme sein*": Gola/Heckmann, DS-GVO & BDSG Kommentar, Commentary on Article 4 DS-GVO, Margin Number 50, with reference to Rossnagel and Gierschmann.

11 CJEU, *EDPS v. SRB* (footnote 3, above), para. 64.

12 I discuss the reference to "the" additional data in my critical comments later in this section.

13 *Idem*.

14 *Idem*, para. 65. Note again the reference to "the" additional data that allow re-identification.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

> it is not sufficient to separate those data from the additional information enabling the data subject to be identified.

In other words, in this view, pseudonymised data should continue to be treated as personal data even in relation to the processing by the entity that receives only the pseudonymised data without the additional information required for (re-)identification of – or rather: the re-attribution of the data to – the data subjects in the Article 4(5) sense by means of the re-identification "key" (in the medical research example: by re-linking the patient data and patient number to the named patient).

In its judgment in *EDPS v. SRB*, the Court, after analysing these different views, appears to have taken a more nuanced approach (but this should be seen as related to the rather special nature of the data that were disclosed in that case, as noted below). Briefly, the case concerned the collection, by SRB (an EU agency),[15] of comments from shareholders in a bank subject to a preliminary decision by SRB on the effect of that decision on their interests, and the transfer of those comments in aggregated form to the consultancy firm Deloitte for assessment:[16]

> The comments transferred to Deloitte were filtered, categorised and aggregated. Where the comments were duplicates of earlier comments, only one version was transmitted to Deloitte [with an alphanumerical code that related to the comment, rather than to the individual or individuals who submitted the comment]. This meant that individual comments that had been duplicated could not be distinguished within a single theme, and Deloitte was unaware whether a comment had been made by one or more participants in the right to be heard process.
>
> [SRB could, via the code and its own database, still link individual comments (including duplicate ones) to individual shareholders who had submitted them. But Deloitte did not have access to that database.]

In other words, the data passed on to Deloitte were not the usual, typical kind of pseudonymised data, i.e., data from which the most immediate identifying information (typically the data subjects' names) has been stripped and replaced with a code or pseudonym that is still uniquely linked to the individual. In the above medical research example: the patient number. In the SRB case, as I understand it, Deloitte could not in any way link specific comments to specific shareholders: it could not "single [them] out". The case must be read in the light of that special feature, although, perhaps regrettably, the Court did not really dwell on it.

Below, I paraphrase, first, as much as possible in the Court's own words, the Court's general reasoning and its reasoning in relation to the relative/absolute issue, and to the further issue

---

15 See: https://www.srb.europa.eu/en/about On the implications, see the next footnote.

16 For details, see Part II of the judgment, *Background to the dispute*, paras. 9 – 28. The quote in the text is from para. 27, with the text in square brackets added; that information is taken from para. 28. Since SRB is an EU agency, the case was addressed under Regulation 2018/1725 containing the rules relating to processing of personal data by EU bodies and institutions. However, as the Court rightly noted, Recital 5 to that regulation makes clear that: "*Whenever the provisions of [that] Regulation follow the same principles as the provisions of [the GDPR], those two sets of provisions should, under the case-law of the [Court of Justice], be interpreted homogeneously*". (para. 2). The reasoning of the Court in this case therefore applies equally to the corresponding (often identical) provisions in the GDPR.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

(addressed in an earlier case)[17] of non-personal data in some cases still having to be treated as personal data. I then return to the issue of "means likely reasonably to be used to identify the data subject", again as much as possible in the Court's own words.

In the opinion of the Court:[18]

> **General points:**
>
> "*[P]seudonymisation … refers to the establishment of technical and organisational measures to reduce the risk of a data set being* ***correlated with the identity***** of data subjects.*"
>
> The main objective of pseudonymisation in relation to the points at issue is "*to* ***prevent*** *the data subject from being identified* solely by means of pseudonymised data.*" (para. 74).
>
> ***Note** that the Court here appears to use the word "identified" in the Article 4(5) GDPR sense: see also the Court's references in the quotes above and below to "correlation" and "attribution" of the data [read: with or to the names of the individuals concerned]. If it had followed the recital 26 approach, all data in a pseudonymised data set that allows cross-referencing between the same data subjects (without knowing their names) should still be regarded as "identifying" those data subjects because they could be – and usually would be – "singled out" from the other individuals in the set. See the medical research example given above.
>
> However, "*[a]ccording to [recital 28 of the GDPR], pseudonymisation 'can [only]* ***reduce*** *the risks' of such correlation for those data subjects and, in doing so, '****help*** *controllers and processors to meet their data protection obligations'*" (para. 72).
>
> More specifically, in relation to pseudonymised data, [Article 4(5) GDPR] requires that "*the identifying information[19] be kept separately and that it be subject to technical and organisational measures [including legal measures – DK] '****to ensure that the personal data are not attributed*** *to an identified or identifiable natural person*'" (para. 74).
>
> **On the relative/absolute/non-personal data issues:**
>
> "*[P]rovided that such technical and organisational measures are actually put in place and are [actually – DK] such as to prevent the data in question from being attributed to the data subject, in such a way that the data subject is not or is no longer identifiable,*

---

[17] CJEU judgment in Case C-319/22 of 9 November 2023, *Gesamtverband Autoteilehandel e.V. v Scania CV AB*, ECLI:EU:C:2023:837. The case concerned the processing of vehicle identity numbers uniquely assigned to cars by car manufacturers, that could be linked to the car owners (who are mostly natural persons) by car repair shops etc. The Court held that consequently the numbers constituted "indirect personal data", i.e., data that indirectly identify (or can be indirectly linked to) a specific individual; and that the car manufacturers should therefore also treat the numbers as personal data. See para. 49 in particular.

[18] The text paraphrases in particular the essential points set out in paras. 68 – 90 of the judgment, relating to "the first complaint of the second part of the first ground of appeal" submitted by the EDPS, with direct quotes in italics; emphases in bold and underlinings added; case references in the text abbreviated or removed to footnotes; references to the (essentially concurring) views of the Advocate General omitted; and references to recitals and articles in Regulation 2018/1725 replaced by references in square brackets to the corresponding (identical) recitals and articles in the GDPR. On that latter point, see footnote 11, above.

[19] As already mentioned in footnote 11, above, I discuss the reference to "the" additional – here: "the" identifying – data in my critical comments later in this section.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

pseudonymisation may have an impact on whether or not those data are personal within the meaning of [Article 4(1) GDPR]." (para. 75).

More specifically, in the case at hand (as is usually the case for controllers who have pseudonymised data), "*the SRB does have additional information enabling the comments transmitted to Deloitte to be attributed to the data subject*". Therefore, in the Court's view, "*those comments are, in spite of pseudonymisation, still personal in nature [for SRB].*" (para. 76).

But in relation to Deloitte (and any others who may have access to pseudonymised data), the issue is more complicated. It all depends on the effectiveness of the technical and organisational (and legal) measures put in place in relation to the pseudonymised data:

"[C]ontrary to what the EDPS maintains, pseudonymised data must not be regarded as constituting, in all cases and for every person, personal data for the purposes of the application of [the GDPR], in so far as pseudonymisation may, depending on the circumstances of the case, effectively prevent persons other than the controller from identifying the data subject in such a way that, for them, the data subject is not or is no longer identifiable."(para. 86).

Here, it is not clear whether the Court here refers to "identification" in the recital 26 ("singling out") or Article 4(5) ("identify" by name) sense. Even so:

In the Court's view, if an entity is given (or otherwise obtains, perhaps even unlawfully: see the second critical comment under the next heading) access to pseudonymous data subject to measures that really, actually and effectively prevent that entity from identifying the data subjects (in either sense?), then those data do not constitute personal data for that entity.

Crucially, in the special circumstances of the SRB case, Deloitte was actually, really prevented from "identifying" the data subjects (the shareholders who submitted specific comments to SRB) even in the recital 26 ("singling out") sense, because SRB provided Deloitte only with the comments in aggregate. In most cases of disclosure of pseudonymised data to third parties that is not the case.

Conversely, if, in a specific context, the measures taken do not really, actually and effectively "prevent" re-identification (in either sense?), then the relevant pseudonymous data must be regarded as still constituting personal data (at least for any entity that can overcome those measures).

Indeed, as the Court held in its *Gesamtverband Autoteilehandel* judgment,[20] **data which are in themselves impersonal may become 'personal' in nature where the controller puts them at the disposal of other persons who have means reasonably likely to enable the data subject to be identified:** "*It is apparent, in particular, from [that] judgment that – where those data are put at their disposal – those data are personal data both for those persons and, indirectly, for the controller.*" (para. 84).

This appears to confirm the "relative" view of the concept of personal data. However, it is important to note the **significant qualifications** that the Court imposed in this respect, as

[20] Footnote 17, above, paras. 46 and 49.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

noted in the paragraphs quoted next. I sum up the main thrust after that in a text box, followed by some critical comments and a discussion of the implications.

> The EDPS and the EDPB argued that a "relative" view of the concept of personal data would "*undermine the high level of protection pursued by the EU legislature and required by the Charter*" and would "*entail[ ] the risk that pseudonymised data could be processed without restrictions under the GDPR and Regulation 2018/1725, including the sharing, publication and transfer to third countries of those data*." (para. 66).
>
> The Court dismisses this by noting that it follows from its judgment in *Gesamtverband* that "*the fact [that pseudonymised data are not, as the case may be, personal in nature for persons to whom the controller transfers the pseudonymised data] has no bearing on the assessment of the personal nature of those data in the context, inter alia, of a potential subsequent transfer of those data to third parties.*" (para. 85).

In other words, pseudonymised data may not constitute personal data for an entity to which they are disclosed but that cannot (re-)identify the data (presumably in the Article 4(5) sense), but, if they are transferred (or further transferred) to a third party that would have available to it the means to re-identify the data (in that sense), then (as with the kind of non-personal data addressed in the *Gesamtverband* judgment) those data would again become personal data "*both for [the entity to which they were disclosed] and, indirectly, for the controller [or other transferring entity]*".

This makes the issue of what can be regarded as "**all the means reasonably likely to be used**" by any recipient of pseudonymised data crucial. As already noted, the Court held in *Breyer* that there would be no "means likely reasonably to be used to identify the data subject" where:

> the identification of the data subject was **prohibited by law or practically impossible** on account of the fact that it requires **a disproportionate effort** in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.

The Court repeated this *verbatim*, first in its judgment in *OC v. the European Commission*,[21] and then again in *SRB*.[22]

The above requires five critical comments.

First of all, there is a problem with the reference in Article 4(5) – repeated in almost every discussion of the issue including in the *SRB* judgment – to certain specific data, "*the* additional data" or "*the* identifying data" – i.e., to a specific "re-identification key" – that can be used to re-link pseudonymised data to a specific data subject. In the medical research example given earlier, this is the patient number: medical staff can use this to link medical information on patients to the relevant specific patients, but researchers who are given the data without access to the database in which the names and the numbers are linked are supposed to not be able to link the data to "identified" patients (in the Article 4(5) sense). As I already noticed, this makes no sense if a data subject can be said to be "identified" if she or he can be "singled

[21] CJEU, judgment of 7 March 2024 in the case of *OC v. the European Commission* (case C-479/22 P, ECLI:EU:C:2024:215), para. 51.
[22] At para. 82.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

out" from the other data subjects in the data set, even if not by name (i.e., if the person can be "identified" in the recital 26 sense). In the recital 26 sense, much if not all of the data in a typical pseudonymised data set can be "identified" (linked to a "singled out" individual) and should therefore, at least to that extent, be regarded as – and treated as – "personal data" in the Article 4(1)/recital 26 sense (but that was not the case in relation to the data in the a-typical data set provided to Deloitte in the SRB case).

Second, the suggestion that the "means reasonably likely to be used" by someone obtaining access to pseudonymised data do not cover *unlawful means* must be seen as specific to the *Breyer* case, which concerned the transmission of identifying data by an internet service provider to online media services – which was allowed only under special legal rules. In my opinion, the uncritical and unqualified repeating of that phrase in *OC* and *SRB* is wrong: as Chris Pounder pointed out in relation to a UK case, *DSG Retail*, in which credit card details were stolen, but without the names and details of the card holders, it would be absurd to hold that the hacker would not be "reasonably likely" to also use unlawful means to obtain the card holders' details (typically, from the "dark web") and thus re-identify the data (and carry out fraud).[23]

The uncritical re-use of the phrase by the Court in *OC* and *SRB* suggests that controllers need not guard against unlawful attacks on pseudonymised data– which is patently absurd, and cannot be the way in which the EU GDPR should be applied.

On the other hand, "legal means" such as strong contractual clauses can be part of the "technical and organisational measures" that can be adopted to protect pseudonymous data against re-identification (in the Article 4(5) sense).[24] Of course, the propriety of the entity to which such data are disclosed and the professionalism of its staff (and internal rules within the entity) should be assessed and taken into account. I will return to that below, in relation to the implications of the *SRB* judgment.

Third, (non-)identifiability can change over time and with a change in context. As the WP29 put it in its 2007 opinion:[25]

> [The test of identifiability] is a dynamic one and should consider the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed. Identification may not be possible today with all the means likely reasonably to be used today. If the data are intended to be stored for one month, identification may not be anticipated to be possible during the "lifetime" of the information, and they should not be considered as personal data. However, it they are intended to be kept for 10 years, the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment. The system should be able

[23] See Chris Pounder's blog at https://amberhawk.typepad.com/ for 23/10/2024, "*Upper Tribunal undermines data breach reporting under the UK_GDPR?*"

[24] The EDPB, in its guidelines on pseudonymisation (01/2025) (footnote 8, above), discusses in useful detail, with informative examples, the technical and organisational measures (including contractual stipulations) that can be used to protect pseudonymous data against re-identification, including in contexts in which various parties share pseudonymised data, etc.

[25] WP136 (footnote 2, above), p. 15, under the heading "*Means to identify*".

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

> to adapt to these developments as they happen, and to incorporate then the appropriate technical and organisational measures in due course.

The WP29 re-emphasised this view in its later opinion on anonymisation technologies:[26]

> The identification risk may increase over time and depends also on the development of information and communication technology. ...

Or as the EDPB put it in its guidelines on the use of location data in the context of the COVID-19 pandemic:[27]

> Anonymisation refers to the use of a set of techniques in order to remove the ability to link the data with an identified or identifiable natural person against any "reasonable" effort. This "reasonability test" must take into account both objective aspects (time, technical means) and contextual elements that may vary case by case (rarity of a phenomenon including population density, nature and volume of data). If the data fails to pass this test, then it has not been anonymised and therefore remains in the scope of the GDPR.

Fourth, recital 26 says that

> [The GDPR] does not … concern the processing of [personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable].

However, crucially:

The very act of pseudonymising or (attempted) anonymising of personal data is a form of processing of personal data – a core GDPR concept that covers "any operation or set of operations which is performed on personal data or on sets of personal data" including "adaptation or alteration"[28] of the data. And this act of processing – which is performed on data which at that time are still identifiable (in any sense) – is subject to the GDPR.

Thus, the GDPR *is* concerned with:

(a) the acts of pseudonymising or (supposedly) anonymising personal data (because those are forms of processing);

(b) the possibility of re-identification of pseudonymised data (which in a typical pseudonymised data set almost always remain identifiably in the recital 26 ["singling out"] sense) even if not in the Article 4(5) ("identification" by name) sense; and

(c) the possibility of re-identification of supposedly fully anonymised data (in either the recital 26 ("singling out") or the Article 4(5) ("identification" by name) sense.

The increasing use of "Big Data" datasets and the use of massive amounts of data collected in ever more sophisticated, especially AI-supported analytical processing of such data makes it correspondingly more difficult to prevent re-identification, even of supposedly fully-

---

[26] Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP216), adopted on 10 April 2014, p. 9, emphases added.

[27] EDPB, Guidelines 04/2020 (footnote 6, above), para. 16.

[28] Or "transformation" or "modification" – which are the terms used by the EDPB in its 01/2025 guidelines (footnote 7, above): see in particular para. 18.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

anonymised data. I am using the adjectives "attempted" and "supposedly" in relation to anonymised data precisely because of this:

It is becoming extremely difficult to "render[ ] [personal data] anonymous" in such a manner that the data subject is really, actually and effectively not or no longer "identifiable" (in either the recital 26 or the Article 4(5) GDPR sense).

As the European Data Protection Board noted in relation to the use of location data in the context of the COVID-19 pandemic, but with clearly wider implications:[29]

> The concept of anonymisation is prone to being misunderstood and is often mistaken for pseudonymisation.[30] While anonymisation allows using the data without any restriction, pseudonymised data are still in the scope of the GDPR.
>
> Many options for effective anonymisation exist[31], but with a caveat. Data cannot be anonymised on their own, meaning that only datasets as a whole may or may not be made anonymous. In this sense, any intervention on a single data pattern (by means of encryption, or any other mathematical transformations) can at best be considered a pseudonymisation.
>
> Anonymisation processes and re-identification attacks are active fields of research. It is crucial for any controller implementing anonymisation solutions to monitor recent developments in this field, especially concerning location data (originating from telecom operators and/or information society services) which are known to be notoriously difficult to anonymise.
>
> Indeed, a large body of research has shown[32]10 that location data thought to be anonymised may in fact not be. Mobility traces of individuals are inherently highly correlated and unique. Therefore, they can be vulnerable to re-identification attempts under certain circumstances.
>
> A single data pattern tracing the location of an individual over a significant period of time cannot be fully anonymised. This assessment may still hold true if the precision of the recorded geographical coordinates is not sufficiently lowered, or if details of the track are removed and even if only the location of places where the data subject stays for substantial amounts of time are retained. This also holds for location data that is poorly aggregated.
>
> To achieve anonymisation, location data must be carefully processed in order to meet the reasonability test. In this sense, such a processing includes considering location

---

29 EDPB, Guidelines 04/2020 (footnote 6, above), paras. 17 – 22.

30 Or deliberately misrepresented, as when controllers often refer to "de-identified data", which suggests the data are anonymised, when in fact the data in question are only pseudonymised. That kind of obfuscation should be avoided.

31 (de Montjoye et al., 2018) "*On the privacy-conscientious use of mobile phone data*" (original footnote that links to: https://pmc.ncbi.nlm.nih.gov/articles/PMC6289108/)

32 (de Montjoye et al., 2013) "*Unique in the Crowd: The privacy bounds of human mobility*" and (Pyrgelis et al., 2017) "*Knock Knock, Who's There? Membership Inference on Aggregate Location Data*" (original footnote that links to:
https://www.nature.com/articles/srep01376 (de Montjoye et al., 2018)
https://arxiv.org/abs/1708.06145 (Pyrgelis et al., 2017))

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

> datasets as a whole, as well as processing data from a reasonably large set of individuals using available robust anonymisation techniques, provided that they are adequately and effectively implemented.

Fifth, there is the issue of the use of anonymous data to create profiles that are then applied to individuals. In that regard, the WP29 considered the following:[33]

> [It would be negligent to] **not consider[ ] the impact on individuals**, under certain circumstances, [of processing of even] properly anonymised data, **especially in the case of profiling**. The sphere of an individual's private life is protected by Article 8 of the ECHR and Article 7 of the EU Charter of Fundamental Rights; as such, even though data protection laws may no longer apply to this type of data, the use made of datasets anonymised and released for use by third parties may give rise to a loss of privacy. **Special caution is required in handling anonymised information especially whenever such information is used (often in combination with other data) for taking decisions that produce effects (albeit indirectly) on individuals.** As already pointed out in this Opinion and clarified by the Working Party in particular in the Opinion on the concept of "purpose limitation" (Opinion 03/2013),[34] **the data subjects' legitimate expectations concerning further processing of their data should be assessed in the light of the relevant context-related factors** – such as the nature of the relationship between data subjects and data controllers, applicable legal obligations, transparency of processing operations.

In sum:

Pseudonymous data do not "escape" the GDPR: (i) because the GDPR always applies (in full) to the very act of pseudonymising personal data; (ii) because in a typical pseudonymised data set most if not all of the data will remain "identifiable" in the recital 26 ("singling out") sense; and (iii) because, even if in some special cases (such as in the SRB case) the recipient really actually cannot re-identify the data (even in the recital 26 "singling out" sense), if that were to change (e.g., because of new data being added, or because of new technologies), or if the data were to be passed on to yet another entity that could re-identify the data, then the GDPR would also again fully apply.

Moreover, contrary to what many controllers believe, the assertion in recital 26 that "*[the GDPR] does not concern the processing of [fully anonymised] data*" does not give them a *carte blanche* to (a) anonymise any personal data they hold and (b) provide the supposedly anonymised data to any other party, for any other purpose. In particular, the very act of anonymising the data (or at least attempting that) is a form of processing, which means that in that regard they must fully comply with the GDPR (as discussed in the next section). Moreover, such controllers must still at all times still guard against the anonymised (no longer personal/non-personal) data again becoming identifiable – in which case those data (like the non-personal car identity data in *Gesamtverband*), and the supposedly still anonymised data all again become personal data.

---

[33] WP216 (footnote 25, above), p. 11, emphases added.

[34] Opinion 03/2013 on purpose limitation (WP203), adopted on 2 April 2013. [original footnote with title added] (footnote 28 above)

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

2.3 Implications and conclusions

2.3.1 Implications

As noted above, the very act of pseudonymising or (attempted) anonymising of personal data is a form of processing of personal data. This means that that act must in all respects comply with the requirements of the GDPR. In the *SRB* judgment, several, but not all, implications of this are noted.

- **Informing of the data subject:**

The judgment rightly stresses that the obligations relating to the informing of data subjects of details of the intended processing, as set out in Article 13 and 14 GDPR, with further requirements in Article 12, apply to pseudonymisation. In that respect, the Court held that:

> [I]t must be held that, for the purposes of applying the obligation to provide information laid down in [Article 13(1)(e) of the GDPR], the identifiable nature of the data subject must be assessed at the time of collection of the data and from the point of view of the controller.
>
> It follows that [in the case at hand] the SRB's obligation to provide information was applicable in the present case prior to the transfer of the data at issue and irrespective of whether or not those data were personal data, from Deloitte's point of view, after any potential pseudonymisation.
>
> (Paras. 111 – 112, reference to Article 15(1)(d) of Regulation 2018/1725 replaced by reference in square brackets to the corresponding provision in the GDPR. See also paras. 102 – 105)

To put that in more general terms:

If an entity collects personal data directly from a data subject and, at that time, already intends to pseudonymise the data and pass the pseudonymised data on to another entity, then the collecting entity must inform the data subjects at the time of collection of that intention, and of the identity of the intended recipient of the pseudonymised data, or if the collecting entity already at that time intends to disseminate the pseudonymised data more widely, of the "categories of recipients of the [pseudonymised] data": see Article 13(1), introductory sentence and point (e) GDPR.

The Court did not address the issue in relation to a situation in which a controller collected personal data other than from the data subject directly, because that did not arise in the case. However, under the same reasoning, such a controller, too, must inform the data subject of any intended pseudonymising of their data and of the intended recipients or categories of recipients of the pseudonymised data; and since this relates to a situation in which "disclosure to another recipient is envisaged", this informing must take place "at the latest when the [pseudonymised] data are first disclosed" (Cf. Article 14(1)(e) and (3)(c) GDPR).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

Although the Court did not mention this, it is also clear that in both these cases, the controller (the data collecting entity) should also inform the data subjects of the purpose and legal basis of the pseudonymising and of the disclosure(s) of the pseudonymised data (see Article 13(1)(c) and 14(1)(c) GDPR). More specifically, as is stipulated in recital 61 GDPR (and in identical terms in recital 36 of Regulation 2018/1725):

> Where the controller intends to process the personal data for a purpose other than that for which they were collected, the controller should provide the data subject prior to that further processing with information on that other purpose and other necessary information.

This also applies if the disclosure is for a "compatible" ("not incompatible") purpose such as scientific or historical research or the creation of statistics. Thus:

If a controller intends to pseudonymise personal data it collects and disclose the pseudonymised data to another entity for a different purpose than the one for which the data were collected – even if that different purpose is a "compatible" purpose – then that controller must inform the data subjects of that intended pseudonymising, and of the recipients or categories of recipients of the pseudonymised data, and of the purpose(s) for which they pseudonymised data are going to be used, and (if the legal basis for the disclosure of the pseudonymised data is different from the legal basis under which the personal data were obtained) of the (new) legal basis for that. See note

Note: in my opinion, the pseudonymising of the data in the *SRB* case and the disclosure of the pseudonymised data to Deloitte in that case was *not* for a different purpose: the evaluation of the comments of the shareholders was part and parcel of SRB's official task. I note an important implication in note (1) under the first text block under the next sub-heading, *Consent, fair processing and the right to object*.

The Court also did not address this issue in the situation in which a controller only decides at some later stage to pseudonymise the data and pass the pseudonymised data on to other entities (because again that issue did not arise in the case). But logically, to such a case the same should apply as noted in relation to collecting of personal data from another entity than the data subject: the controller should inform the data subjects of this (new) intention to pseudonymised the data and to disclose the pseudonymised data, and of the purpose and legal basis for this; and this informing, too, should take place "at the latest when the [pseudonymised] data are first disclosed" (but I would say preferably, and arguably obligatorily, before that, especially if the "right to object" applies, as noted below).

- **Consent, fair processing and the right to object:**

In its *SRB* judgment, the Court rightly stressed the link between informing, consent, fair processing and the right to object:

> … [W]here the collection of [personal] data from the data subject is … based on that data subject's consent, the validity of the consent given by that data subject depends, inter alia, on whether that data subject has previously obtained the information in the light of all the circumstances surrounding the processing of the data in question to which he or she was entitled, under [Article 13 GDPR], and which allow him or her to give consent in full knowledge of the facts ( ).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

> Moreover, as regards a situation in which there is an obligation incumbent on the data subject to provide personal data to the controller, [recital 60 GDPR] specifies, in the fourth sentence, that the data subject should also be informed whether he or she is obliged to provide the personal data and of the consequences, where he or she does not provide such data; that bears out the importance of the information required under [Article 13 GDPR], specifically at the moment that those data are collected from the data subject.
>
> … [O]ne of the purposes of the obligation to provide the data subject – at the time of collection of the personal data linked to him or her – with information relating to the potential recipients of those data is to enable that data subject to decide, in full knowledge of the facts, whether to provide or, on the contrary, refuse to provide the personal data being collected from him or her.
>
> [This] is … also essential in order for the data subject to be able to defend his or her rights against those recipients subsequently. However, the obligation to provide that information at the time of the collection of personal data ensures, inter alia, that those data are not collected by the controller against the will of the data subject, or even transferred to third parties against his or her will.
>
> (Paras. 106 – 109, references to Article 15 and recital 35 of Regulation 2018/1725 replaced by references in square brackets to Article 13 and recital 60 GDPR; case-law reference omitted and replaced by ( ))

Recital 32 and 33 GDPR and recital 19 of Regulation 2018/1725 stipulate in identical terms that:

> Consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given [read: separately – DK] for all of them. …
>
> It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.

In my opinion, it follows that if a controller processes personal data on the basis of the consent of the data subject, and wants to pseudonymise or anonymise those data and disclose the pseudonymised or anonymised data to another entity for a different purpose[See note 1] – even if that is a "compatible" purpose such as research – the controller must obtain the separate consent of the data subject for that pseudonymising or anonymising and disclosure for the different purpose; and the data subjects should be able to consent to the processing for the primary purpose but to not consent to the pseudonymising or anonymising and disclosure of their data for the secondary purpose. Moreover, if the data subjects did initially consent also to the pseudonymising or anonymising and disclosure, they must still be able to withdraw that consent at any time, in which case any further pseudonymising or anonymising and disclosure of their data must cease (see Article 7(3) GDPR).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

Notes:

(1) As noted above, in my opinion, the pseudonymising of the data in the *SRB* case and the disclosure of the pseudonymised data to Deloitte in that case was *not* for a different purpose. In such a case, the data subjects can of course not decide to consent to part of the processing -*in casu*, the collecting of their comments – and not to consent to another integral part of the same processing operation – *in casu*, the evaluation of the comments by Deloitte. But in many cases pseudonymised data are disclosed for purposes that are not the same – that are not intrinsically part of – the overall operation.

(2) As further discussed in chapter 8, section 1.3.2, sub-section 1.3.3.2 of the handbook, in my opinion, pseudonymising or anonymising data generated by users of software underpinned by AI models (such as educational software ["EdTech"] or gaming software) and using the pseudonymised data to **train or improve the AI model** in question may well be a compatible ("not incompatible") use of those data, but it is still a clearly distinct purpose from the primary purpose: to allow learning or gaming etc. In such cases, in my opinion, the data subjects should therefore still be asked for their separate consent for this pseudonymising and use of their pseudonymised data.

The right to withdraw consent is distinct from **the right to object** as enshrined in Article 21 GDPR. This latter right applies to processing that is "necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller" or that is done "for the purposes of the legitimate interests pursued by the controller or by a third party" (but that are not "overridden by the interests or fundamental rights and freedoms of the data subject") (Article 6(1)(e) and (f) GDPR).[35] But it extends to profiling based on those provisions (*idem*). Where this applies, and where the data subject exercises this right:

> The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.
>
> (Article 21(1), last sentence, GDPR)

Such "compelling, overriding legitimate grounds" for the continued pseudonymising or anonymising of personal data and for the disclosure of the pseudonymised or anonymised data could exist, for instance, in relation to pseudonymised or anonymised reporting of serious transmissible diseases. However:

In many – I would say most – cases, and especially in contexts in which the pseudonymising or anonymising and disclosure – and any associated profiling – merely serves a commercial "legitimate" interest, a controller will have to cease pseudonymising or anonymising personal data and disclosing of the pseudonymised or anonymised data if the data subject objects to this.

- **Risk- and data protection impact assessments and consulting of supervisory authority:**

  Risks assessments generally:

Under the "accountability" principle set out in Article 5(2) GDPR that underpins the GDPR, any controller intending to carry out any (form of) processing of personal data – including

[35] Under Regulation 2018/1725, the right to object only applies to the former (Article 5(1)(a)), because public authorities cannot rely on the "legitimate interest" legal basis in any case: see the last sentence in Article 6(1) GDPR.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

pseudonymising or attempted anonymising of such data – must assess the risks involved in or possibly resulting from that processing, including any risk of re-identification at some future stage, and the effect that would have on the data subjects. As the EDPB puts it:[36]

> Controllers may [I would read that as should – DK] 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 [should] 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.

In-depth data protection impact assessments:

Under Article 35 GDPR, if there is a serious risk that pseudonymised or supposedly anonymised data may become re-identifiable at some later stage because of new, more powerful technical capabilities such as AI or if merged with and matched against other datasets (or typically both), and if this can have a serious effect on individuals – i.e., if re-identification would pose "a high risk to the rights and freedoms of natural persons" (here: to the data subjects whose identity was to be shielded by the pseudonymisation or anonymisation) – then the controller will be required to carry out a **data protection impact assessment** prior to the pseudonymisation or attempted anonymisation and to adopt compensating measures against such possible future re-identification such as, e.g., strong contractual obligations on the part of the recipient of the supposedly fully anonymised data to not try to re-identify the data and to not merge them with or match them against other data sets.

The UK Information Commissioner's Office (ICO) has issued useful guidance in this respect. It summarises this "at a glance", as follows:[37]

> **At a glance**
>
> - Anonymisation ensures that the risk of identification is sufficiently remote to minimise the risks to people arising from the use of their information.
> - Identifiability is a wide concept. A person can be identifiable from many factors that can distinguish them from someone else, not just a name.
> - Identifiability exists on a spectrum. When assessing whether someone is identifiable, you **should** take account of the "means reasonably likely to be used to enable identification". There are likely to be many borderline cases where you **should** use careful judgement based on the specific circumstances of the case.
> - You do not need to take into account any purely hypothetical or theoretical chance of identifiability, rather, what is reasonably likely relative to the circumstances.

---

[36] EDPB, Guidelines 01/2025 (footnote 7, above), para. 35, with useful further discussion of the different contexts.

[37] ICO, *How do we ensure anonymisation is effective?*, original emphases, available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/#dataprotectionlaw
As the header says, strictly speaking this guidance only relates to anonymisation. However, the same considerations arise in relation to pseudonymisation. The ICO notes that "Due to the [UK] Data (Use and Access) Act coming into law on 19 June 2025, this guidance is under review and may be subject to change."

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

- You **should** consider both the information itself, and who may get (or want to get) access to it.
- Before you release data to the world at large, you **should** consider using robust techniques to reduce the higher risk of unauthorised personal data disclosure compared to intentional and controlled data release to known recipients.
- You **should** also consider potential unauthorised access by people (eg hacking or the actions of rogue employee).
- Applying a "motivated intruder" test is a good starting point to consider identifiability risk.
- You **should** review your risk assessments and decision-making processes regularly.

The ICO expands on each of these issues and especially on the "motivated intruder" test, in detail in the guidance on the website.

In simple terms: whenever a controller is thinking about releasing data they hold in pseudonymised or supposedly fully-anonymised form to another entity – or more problematically, to a range of other entities or to the public at large – it is crucial that the controller assesses (a) the effect that any unintended re-identification of the data by anyone (not just the legitimate recipient(s) of the data) would have on the rights and interests of the data subjects, and (b) whether any of the technical and organisational measures that are under consideration to protect against this are really, actually effective in that regard, *vis-à-vis* anyone including a "motivated [even illegal] intruder". And of course, the more sensitive the original data are, either because they fall within the special categories of data listed in Article 9(1) GDPR,[38] or more generally (e.g., financial or credit card data), the more serious these tests should be applied. Moreover, under the GDPR a record must be kept of the relevant assessments and of the measures adopted, with the reasons for believing they are sufficient in the relevant context.

Failure to carry out a DPIA in cases in which one should be carried out constitutes a breach of the GDPR – which could make the controller liable if re-identification occurs because the tests were not or insufficiently rigorously done and/or the measures inadequate.

Moreover, the last point made by the ICO, that controllers should keep their risk assessments under review, is especially important in the light of what I wrote earlier, in my second "critical comment" in section 2.2, above, about (non-)identifiability being able to change over time depending on the context and technical developments.

Consulting the competent data protection supervisory authority/ies:

If a data protection impact assessment "indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk", and no effective measures of the relevant kind are available, the controller must consult the relevant (competent) supervisory authority – or in cross-border cases possibly several such authorities

[38] I.e., "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, … genetic data, biometric data [when used] for the purpose of uniquely identifying a natural person, data concerning health [and] data concerning a natural person's sex life or sexual orientation".

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

– prior to processing (Article 36 GDPR). The authority/ies can issue advice, and may even order the processing – here: the pseudonymisation or attempted anonymisation and the disclosure of the pseudonymised or supposedly anonymised data – to not happen (Article 36(2) read together with Article 58(2)(f)(GDPR).

Failure to consult the authority/ies in cases in which they should be consulted also constitutes a breach of the GDPR – which again could make the controller liable if re-identification occurs because the above tests were not or insufficiently rigorously done and/or the measures inadequate. Controllers would of course also ignore any "advice" issued by those authorities at their peril, and would breach the GDPR if they ignored orders to cease the processing (i.e., the pseudonymisation or (supposed) anonymisation.

2.3.2 Conclusions

The GDPR rules on pseudonymisation, and the definition of "identification" as used in that context (Article 4(5) in particular), differs from the broader use of that term in Article 4(1) if read with recital 26. Rather than including "singling out" of a person, the rules only apply it to re-linking pseudonymised or supposedly anonymised data to a named individual.

In its *SRB* judgment, the Court of Justice appears to have endorsed the view that pseudonymised or anonymised data can be "personal data" in relation to one entity – typically, the entity that has retained the additional information that can be used to re-identify the individuals concerned – while not constituting "personal data" for another entity that does not have this additional data – or, I should add, that does not otherwise have the means to re-identify the data subjects, including means to "single [them] out" from the data set.

This will probably be seized on by various entities to argue that they can now much more freely pseudonymise or (supposedly) anonymise personal data they hold, and can then freely use the pseudonymised or (supposedly) anonymised data.

They would be wrong in this. As I believe to have shown in this section:

Even after the *SRB* judgment, pseudonymising personal data or attempting to fully anonymise personal data, and the subsequent disclosure of the pseudonymised or supposedly fully-anonymised data remains subject to crucial constraints.

This follows in particular from the fact that the very act of pseudonymising or (attempted) anonymising of personal data is a form of processing of personal data – a core GDPR concept that covers "any operation or set of operations which is performed on personal data or on sets of personal data" including "adaptation or alteration" of the data. And this act of processing – which is performed on data which at that time are still identifiable (in any sense) – is subject to the GDPR.

This has a range of **important implications**. Specifically:

- controllers who wish to make pseudonymised or indeed (supposedly) anonymised data available to others must at all times still guard against the data again becoming identifiable – in which case those data (like the non-personal car identity data in *Gesamtverband*), and the pseudonymised or supposedly still anonymised data all again become personal data;

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

- if a controller intends to pseudonymise or anonymise personal data it collects and disclose the pseudonymised or (supposedly) anonymised data to another entity for a different purpose than the one for which the data were collected – even if that different purpose is a "compatible" purpose – then that controller must inform the data subjects of that intended pseudonymising or anonymising, and of the recipients or categories of recipients of the pseudonymised or anonymised data, and of the purpose(s) for which they pseudonymised or anonymised data are going to be used, and (if the legal basis for the disclosure of the pseudonymised or anonymised data is different from the legal basis under which the personal data were obtained) of the (new) legal basis for that;
- if a controller only decides at some later stage to pseudonymise or anonymise the data and pass the pseudonymised or anonymised data on to other entities, the same applies: the controller should inform the data subjects of this (new) intention to pseudonymised or anonymise the data and to disclose the pseudonymised or anonymised data, and of the purpose and legal basis for this; and this informing, too, should take place "at the latest when the [pseudonymised or anonymised] data are first disclosed" (but I would say preferably, and arguably obligatorily, before that, especially if the "right to object" applies, as noted below);
- if a controller processes personal data on the basis of the consent of the data subject, and wants to pseudonymise or anonymise those data and disclose the pseudonymised or anonymised data to another entity for a different purpose – even if that is a "compatible" purpose such as research – the controller must obtain the separate consent of the data subject for that pseudonymising or anonymising and disclosure for the different purpose; and the data subjects should be able to consent to the primary processing but to not consent to the pseudonymising or anonymising and disclosure of their data for the secondary purpose. Moreover, if the data subjects did initially consent also to the pseudonymising or anonymising and disclosure, they must still be able to withdraw that consent at any time, in which case any further pseudonymising or anonymising and disclosure of their data must cease;
- in many – I would say most – cases, and especially in contexts in which the pseudonymising or anonymising and disclosure – and any associated profiling – merely serves a commercial "legitimate" interest, a controller will have to cease pseudonymising or anonymising personal data and disclosing of the pseudonymised or anonymised data if the data subject objects to this;
- whenever a controller is thinking about releasing data they hold in pseudonymised or supposedly fully-anonymised form to another entity – or more problematically, to a range of other entities or to the public at large – it is crucial that the controller assesses (a) the effect that any unintended re-identification of the data by anyone (not just the legitimate recipient(s) of the data) would have on the rights and interests of the data subjects, and (b) whether any of the technical and organisational measures that are under consideration to protect against this are really, actually effective in that regard, *vis-à-vis* anyone including a "motivated [illegal] intruder". And of course, the more sensitive the original data are, either because they fall within the special categories of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864

data listed in Article 9(1) GDPR, or more generally (e.g., financial or credit card data), the more serious these tests should be applied. Moreover, under the GDPR a record must be kept of the relevant assessments and of the measures adopted, with the reasons for believing they are sufficient in the relevant context;

- failure to carry out a DPIA in cases in which one should be carried out constitutes a breach of the GDPR – which could make the controller liable if re-identification occurs because the tests were not or insufficiently rigorously done and/or the measures inadequate;
- failure to consult the authority/ies in cases in which they should be consulted also constitutes a breach of the GDPR – which again could make the controller liable if re-identification occurs because the above tests were not or insufficiently rigorously done and/or the measures inadequate;
- controllers would of course also ignore any "advice" issued by those authorities at their peril, and would breach the GDPR if they ignored orders to cease the processing (i.e., the pseudonymisation or (supposed) anonymisation.

- o – O – o -

Douwe Korff (Prof.), Cambridge, 24 November 2025

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5734864