2.3.2.2 Data subject, personal data, pseudonymous data and anonymous data

i. "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. [144] 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 <u>about</u> Titius), AND to Gaius because of the "purpose" element (it will be used <u>in order</u> to treat Gaius in a certain way) AND to Sempronius because of the "result" element (it is likely to have an <u>impact</u> 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)

In relation to the "identifying" or "singling out" of individuals, it 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)

[144] Article 29 Working Party, <u>Opinion 4/2007 on the concept of personal data</u> (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 <u>Endorsement 1/2018</u>, it is still an authoritative document on the issue.

And in relation to the need to take account of "all the means reasonably likely to be used", the Court of Justice in its *Breyer* judgment agreed with 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:[145]

> the identification of the data subject was … **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.

ii. "pseudonymous" and "anonymous" data[146]

> **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" [read: by anyone (see text)]; the GDPR "does not … concern the processing of such anonymous information, including for statistical or research purposes."

It has been argued 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.[147] This is of particular importance in relation to pseudonymised data: if the "relative" view is maintained, pseudonymised data could be freely disclosed to other parties

[145] 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. The Court held that this also applied where reidentification was "**prohibited by law**" (the words omitted from the above quote). However, that must be seen as specific to the 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. 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. See Chris Pounder's blog at https://amberhawk.typepad.com/ for 23/10/2024, "*Upper Tribunal undermines data breach reporting under the UK_GDPR?*" Remarkably, the UK data protection supervisory authority, the ICO, in its Anonymisation Code of Practice nevertheless elevates this idea, that "means likely to be used" by any third party with access to pseudonymous data only covers lawful means, to a general principle. It 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 will be applied.

[146] 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.

[147] "*Daten [können] für die einen 'nur' pseudonyme und für andere anonyme sein*" (*"Data can be 'just' pseudonymised [data] for one [entity] but anonymised [data] for another [entity]*"): Gola/Heckmann, DS-GVO & BDSG Kommentar, 3rd ed., 2022, Commentary on Article 4, Margin Number 50, with reference to Rossnagel and Gierschmann. But this was written before the *Gesamtverband/Scania* judgment, discussed in the text. The UK Upper Tribunal, in the *DSG Retail* ruling (footnote 145, above), also took this "relative"Pounder view. If this is upheld on appeal it would constitute a sigificant divergence of UK data protection law from the EU GDPR.

that are "reasonably unlikely" to have the means to re-identify the data and in particular do not have access to the reidentification "key".

However, in its judgment in *Gesamtverband/Scania*, the Court of Justice has made clear that data that are not "in themselves" personal data must still be treated as personal data by *anyone* processing those data if *any* entity that is provided with the data can link the data to specific natural persons.[148] In my opinion, this debunks the "relative" view which, in relation to pseudonymised data, also squarely contradicts the clarification in recital 26 that pseudonymous data "should be considered to be information on an identifiable natural person" and should therefore still be treated as personal data (by all concerned).

Consequently, only data that are fully anonymised, in such a way that they can not (no longer) be linked to any specific individual, by anyone, fall outside the GDPR (as recital 26 also makes clear).[149] However, there are <u>three caveats</u> to this.

<u>First</u> of all, (non-)identifiability can change over time. As the WP29 put it in its 2007 opinion:[150]

> [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 to adapt to these developments as they happen, and to incorporate then the appropriate technical and organisational measures in due course.
>
> (Original italics, emphasis in bold added)

The WP29 re-emphasised this view in its later opinion on anonymisation technologies:[151]

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

<u>Secondly</u>, although 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] –

[148] CJEU, third chamber judgment in Case C-319/22 of 9 November 2023, *Gesamtverband Autoteile-Handel 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 or allow the singling out of) a specific individual; and that the car manufacturers should therefore also treat the numbers as personal data.

[149] Cf. also the EDPB <u>Guidelines 01/2025</u> (footnote 146, above), para. 22.

[150] WP136 (footnote 144, above), p. 15, under the heading "*Means to identify*".

[151] Article 29 Working Party, <u>Opinion 05/2014 on Anonymisation Techniques</u> (WP216), adopted on 10 April 2014, p. 9, emphases added. The EDPB, in its guidelines on pseudonymisation (01/20225, footnote 146, 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.

in fact the very act of pseudonymising or (attempted) anonymising of personal data is a form of processing – 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"[152] of the data (see sub-section 2.3.2.3, below). And this act of processing – which is performed on data which at that time are still identifiable – is subject to the GDPR. Thus, the GDPR *is* concerned with (a) the act of (supposedly) anonymising personal data (because that is a form of processing) and (b) the possibility of re-identification of supposedly fully anonymised data; and the regulation of course fully applies if and when any of the pseudonymised or previously deemed to be anonymised data are in fact re-identified.

Indeed, 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. I am using the adjectives "attempted" and "supposedly" in relation to anonymised data precisely because of this: it is extremely difficult to "render[ ] [personal data] anonymous in such a manner that the data subject is not or no longer identifiable".

(There is the associated issue that 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.)

<u>Thirdly</u>, 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:[153]

> [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),[154] **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.

I will come to the taking of automated decisions and the use of profiles in section 2.3.4, sub-section 2.3.4.7, below. Suffice it to note here that the European data protection authorities clearly do not feel that processing of supposedly anonymised data is of no concern.

<u>In sum</u>: 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.

[152] Or "transformation" or "modification" – which are the terms used by the EDPB in its guidelines: see in particular para. 18.

[153] WP216 (footnote 150, above), p. 11, emphases added.

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

On the contrary, under the "accountability" principle underpinning the GDPR (further discussed in section 2.3.3, sub-section 2.3.3.1, below), any controller intending to carry out any (form of) processing of personal data – including 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:[155]

> 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.

If there is a serious risk that the data may become re-identifiable at some later stage (because of new, cheaper 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, the controller will be required to carry out a **data protection impact assessment** prior to the attempted anoymisation 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 (including by singling out the individual from the data: see above, at i) and to not merge them with or match them against other data sets.[156]

This poses a delimma: in-depth analysis and data matching is often precisely what research and commercial entities understandably want to do with bought-in supposedly anonymised data. The point I am making here is that the suppliers of the supposedly anonymised data and the end-users, rather than trying to pretend that their activities are completely outside the scope of the GDPR, should carry them out in accordance with the regulation – which in many ways clearly seeks to facilitate legitimate research and data analysis, subject to "appropriate safeguards".[157] Those safeguards should include safeguards against re-identification/singling out of data subjects and/or against the use of the results of the processing in ways that unfairly affect the data subjects. If controllers who make supposedly anonymised data available to third parties do not adopt such safeguards, they may well end up becoming responsible – and liable – for breaches of the GDPR. DPOs should guard against this.

All that said, the EDPB still rightly notes that pseudonymisation can be a useful means to facilitate compliance with the GDPR "by design" by reducing *confidentiality risks*, *the risk of function creep*, and by assigning widely differing pseudonyms to persons with very similar identifying attributes, *risks to accuracy*.[158]

END OF EXTRACT

- O – o – O -

[155] EDPB, Guidelines 01/2025 (footnote 146, above), para. 35, with useful further discussion of the different contexts.

[156] I discuss the carrying out of risk assessments and DPIAs further in Part Three, Tasks 3 and 4.

[157] See in particular Article 89 GDPR on *Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes*.

[158] See EDPB, Guidelines 01/2025 (footnote 146, above), section 2.2.1, paras. 27 – 30.