EVOLUTION OF PERSONAL DATA IN THE COURT OF JUSTICE'S CASE LAW AND IMPLICATIONS FOR SCIENTIFIC RESEARCH

LINE LUNDSTRÖM & SANTA SLOKENBERGA*

This article provides an in-depth analysis of the concept of personal data under the General Data Protection Regulation (GDPR) in light of the Court of Justice's jurisprudence and explores its implications for such a data-intensive field as scientific research. It traces the evolution of the core definitional elements set out in the definition – 'any information', 'relating to', and 'an identified or identifiable natural person' – highlighting how their interpretation has developed through case law, with particular attention to such newest cases as IAB Europe and SRB. Through the analytical approach it takes, this article enhances clarity in data processing activities in scientific research, and illuminates questions that necessitate further inquiries. Moreover, by linking case-law developments to ongoing discussions on the notion of personal data, it enhances awareness of pseudonymisation and its legal nature, and nuances the boundaries between data processing that triggers the GDPR and that which does not.

1 INTRODUCTION

The General Data Protection Regulation (GDPR) establishes a comprehensive, far-reaching framework for the protection of personal data across the European Union and has been in effect for almost a decade.[1] Central to its scope is the definition of personal data, provided in Article 4(1) as 'any information relating to an identified or identifiable natural person'. This definition is not new in the EU legal culture; it can already be found in the previous EU personal data framework, the Data Protection Directive, adopted in 1995,[2] and in addition to the GDPR it features in other legal instruments, notably, the Law Enforcement

* Uppsala University. Line Lundström conducted the case law research under the supervision of Santa Slokenberga, contributed to early discussions on the structure of the manuscript, and had main responsibility for the initial draft. Santa Slokenberga conceived and conceptualized the study, contributed to early drafting and undertook revisions of the manuscript, including in response to peer review. Both authors reviewed and approved the final version for publication. The authors are grateful to the anonymous reviewer for insightful and constructive comments.

This work has been supported by the PROMOT project. It has received funding from – the Canadian Institutes of Health Research (CIHR), and in partnership with L'Agence Nationale de la Recherche (France), Health Research Board (Dublin, Ireland), Instituto de Salud Carloss III (Spain), IRSC – IG, Swedish Research Council (Sweden), and Swiss National Science Foundation (Switzerland) – partners of the EJP RD, under grant agreement No02424 – 000.
The EJP RD initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No825575.

[1] See Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ 2016 L119/1, which will be cited as the 'GDPR'.

[2] See Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31, which will be cited as the 'Directive' or the 'Data Protection Directive'.

Directive (LED)[3] and the Data Protection Regulation for EU institutions, bodies, offices, and agencies (EUDPR).[4] While deliberately broad to accommodate a wide range of data types and contexts,[5] the definition's interpretation – particularly in technologically complex and data-driven environments – has given rise to legal uncertainty and judicial scrutiny.[6]

Because the GDPR applies to the processing of personal data pursuant to Article 2(1) GDPR, the assessment of whether the data in question constitutes personal data is central to all activities involving the processing of information about individuals, including such data-intensive fields as biomedical research. It has implications not only for ascertaining whether the GDPR applies and, if so, what obligations must be complied with, but also for how the intended data-processing aims can be effectively met. In scientific research, if a dataset is treated as falling outside the scope of personal data, this may have implications for decisions regarding how data processing activities and data flows are organised, and even for the extent to which the intended study is legally feasible.[7] This is also particularly relevant for consortia in which datasets are located both within different EU Member States, and within and outside the EU, and in which data are shared with third countries. When data are located across different EU Member States, administrative and practical hurdles may make data sharing difficult, if not impossible. When data are located both within the EU and in third countries, and a data transfer is crucial to achieving the scientific aims, at times, a study may not appear feasible given the requirements set out in Chapter V of the GDPR for international data transfers, legal realities in respective third countries, and the risks that research partners in the EU are willing to assume.[8]

A correct assessment of whether a dataset in question constitutes personal data or not is also crucial for other considerations. For example, in Sweden, if data-driven scientific research involves processing sensitive data within the meaning of the GDPR, it requires ethical approval, whereas the processing of a health data dataset that is not personal under the GDPR does not.[9] Whilst ethical approval is an important safeguard, it is also resource-intensive. It must be used when required. However, if ethical approval is sought when it is not required, whilst it can be perceived as an additional safeguard, it is not serving

[3] See Article 2 of Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L119/89, which will be cited as 'LED'.

[4] See Article 2 of Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC [2018] OJ L295/39, which will be cited as 'EUDPR'.

[5] See, by analogy, Case C-434/16 *Nowak* EU:C:2017:994 paras 33–34 and Case C-604/22 *IAB Europe* EU:C:2024:214 para 36.

[6] Compare for example the referring courts' questions in Case C-582/14 *Breyer* EU:C:2016:779 and *IAB Europe* (n 5).

[7] For example, in organisational decisions regarding AI development, whether to train AI in a centralised data pool or to adopt a federated model. Whereas the data-minimisation principle may incline towards the latter, in cases where the intended processing objectives could be compromised by federated learning, a centralized data pool might be the only viable option.

[8] See Santa Slokenberga et al, 'EU Data Transfer Rules and African Legal Realities: Is Data Exchange for Biobank Research Realistic?' (2019) 9(1) International Data Privacy Law 30.

[9] Lag (2003:460) om etikprövning av forskning som avser människor (The Ethical Review of Research Involving Humans Act) Section 3.1.

the intended purpose, and it is adding not only to the costs but also is placing a burden on the respective ethics body, and hampering the effectiveness of the achievement of the intended research objectives.

Over the years, the Court of Justice of the European Union (CJEU), and in particular the Court of Justice, has played an important role in developing the concept of personal data and in understanding its constituent elements. As one of the central pillars of the right to data protection, the notion of personal data has appeared in a broad range of cases, from the era of the aforementioned Data Protection Directive to the present, under the GDPR, the LED, and the EUDPR. As the concept of personal data is essentially the same across all areas of EU data protection law,[10] the CJEU has assigned it a uniform interpretation across the different instruments,[11] allowing case law clarifying the notion under one instrument to be applied under another. Despite the notion of personal data having a prominent role in EU law, the concept has been substantiated in only a relatively small number of decisions,[12] including the well-known cases of *Nowak*[13] and *Breyer*,[14] adjudicated in the previous decade. In those cases, the CJEU has gradually provided guidance on how key notions included in the definition – any information, its relation to a natural person, and identifiability – should be interpreted and applied, allowing one to pursue a devil's advocate's position, highlighting that even information about weather can constitute personal data, and caution that the GDPR is becoming the law of everything.[15]

In scientific research, working with pseudonymised data is commonplace. It has been a prevailing argument that '[t]the GDPR explicitly defines data that have undergone pseudonymization as personal data, thus falling within the scope of the regulation'.[16] Interpretations like this have not only provoked considerable debate about the GDPR in scientific research but also given rise to practical consequences, in particular, the necessary arrangements to ensure compliance with the GDPR for pseudonymous data.

In 2024-2025, the Court of Justice delivered preliminary rulings in two tone-setting cases: *IAB Europe* (request for a preliminary ruling)[17] and *SRB* (appeal of the General Court judgment).[18] These cases add further nuance to the concept of personal data, its constitutive elements, and its interplay with anonymous data, and challenge the view that treats pseudonymised data as personal data in all cases. They also arguably open up the possibility of drawing a bit clearer boundaries between when data are personal and when they are not, within the meaning of the GDPR. In a scientific research context, this means a new opportunity, with potentially greater legal certainty, to make more nuanced assessments of

[10] Compare Article 2(a) of the Data Protection Directive (n 2), Article 3(1) LED (n 3), Article 3(1) EUDPR (n 4) and Article 4(1) GDPR (n 1).
[11] See Case C-413/23 P *SRB* EU:C:2025:645 para 52, *IAB Europe* (n 5) para 33 and compare Case C-180/21 *Inspektor v Inspektorata kam Visshia sadeben savet* EU:C:2022:967 para 12.
[12] Central cases are reviewed in Section 3.
[13] *Nowak* (n 5).
[14] *Breyer* (n 6).
[15] Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10(1) Law, Innovation and Technology 40.
[16] Mahsa Shabani and Luca Marelli, 'Re-identifiability of Genomic Data and the GDPR' (2019) 20 EMBO Rep EMBR201948316. For a broader insight into the scholarly debates, see Section 2.
[17] *IAB Europe* (n 5).
[18] *SRB* (n 11).

whether a dataset is personal, and, when it is not, to avoid the associated legal requirements.[19]

Taken together, the body of case law on the notion of personal data provides guidance on how the constituting elements of that notion should be interpreted and combined. Moreover, it also demonstrates how the concept of personal data functions within the broader structure of the GDPR and in light of its underlying purposes. These cases gradually clarify the boundaries of the GDPR by adding interpretive detail and methodological nuance on a case-by-case basis, and contributing to an increasingly substantive understanding of what constitutes personal data. At the same time, their contextual foundations allow for nuanced – and at times diverging – interpretations and methodologies to be applied, leaving the legal landscape with some ongoing uncertainty about the true meaning of the concept and how it ought to be assessed.

This article examines the meaning of personal data in light of recent CJEU rulings, situates it within the broader trajectory of the concept's evolution as shaped by prior case law, and explores its implications for such a data-intensive field as scientific research. It further considers the implications of this evolving conceptualisation for delineating the substantive scope of data protection under the GDPR, and illuminates a way forward for enhancing legal certainty for those acting in good faith in their attempts to comply with the GDPR, whilst fulfilling aims of public importance. Whilst the primary objective of the article is to offer a doctrinal analysis of the evolution of the understanding of personal data, it has two secondary implications. On a practical level, it enhances clarity in data processing activities in scientific research, including the design of data flows. Conceptually, by linking case-law developments to ongoing discussions on the notion of personal data, it aims to enhance awareness of pseudonymisation and its legal nature, and to nuance the boundaries between data processing that triggers the GDPR and that which does not.[20]

2 SOME REMARKS ON THE NOTION OF PERSONAL DATA

2.1 THE PROVISIONS OF THE GDPR

Following Article 4(1) GDPR, '"personal data" means any information relating to an identified or identifiable natural person ("data subject")'. It clarifies that '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'. Neighbouring to that definition is the definition of pseudonymised data set out in Article 4(5) GDPR. It states that '"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'.

These two definitions are primarily grounded in Recital 26, although subsequent

[19] Even though the data might not be personal data within the meaning of the GDPR, these data might, nonetheless, be subject to other legal or ethical requirements.

[20] For reasons elaborated in Section 4, anonymous data are not contrasted with personal data.

recitals have also appeared to be helpful interpretative tools. It states that '[t]he principles of data protection should apply to any information concerning an identified or identifiable natural person'. In regard to pseudonymised data, it affirms that '[p]ersonal 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'. Thereafter, it guides that

> [t]o 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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

The limits on the application of the GDPR are highlighted by a further affirmation in that recital, namely, that

> [t]he principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

2.2 THE SCHOLARLY DEBATE AND POLICY APPROACHES

What data fall within the scope of the personal data definition, and what fall outside it, have generated considerable scholarly debate and attracted policymakers' attention. To illustrate briefly, computer science has its roots in technology, including the development of privacy-preserving tools, and, against that background, understanding the requirements and challenges posed by the legal standards of anonymisation and personal data. Law, especially the EU law, has its roots in preserving the right to privacy and the right to data protection against evolving technology, whilst not placing undue obstacles to market objectives. Both converge somewhat on the understanding that anonymised information cannot be linked back to an individual and therefore cannot be used to cause harm to the individual, while also opening up discussion of the construct of anonymisation and its content. The data protection rules in the EU are anchored in the understanding that an individual shall be protected insofar as the scope of personal data stretches. Where data are rendered anonymous, even if they come from an individual, they no longer benefit from the legal protection, as the data subject has ceased to exist in that particular context.[21]

Anonymisation is not defined in the GDPR. In scholarship, it has been referred to as

[21] See Santa Slokenberga, 'You can't put the genie back in the bottle: on the legal and conceptual understanding of genetic privacy in the era of personal data protection in Europe' [2021] BioLaw Journal - Rivista di BioDiritto 223.

an approach to find a balance between sharing data and protecting an individual's privacy.[22] From a technical perspective, the standard of anonymity has been achieved as a requirement against privacy attacks. Consequently, anonymous data should be robust against attempts to (re)identify and learn about individuals.[23] A key question has been re-identification and the extent to which risks are acceptable for the dataset to remain regarded as anonymous. This bears on how privacy is generally approached and on how the notion of personal data is understood, whether as a static or context-dependent concept. Against this, various approaches to privacy are also located. That of pragmatists, who focus on the risk of re-identification, and that of formalists, who focus on mathematical rigour when defining privacy.[24] Put in different terms, whether privacy is relative or objective, understood also as absolute, a debate that has existed for a considerable time, and was already picked up by Advocate General in the case of *Breyer*, as a background to the questions raised in that case.[25]

A body that can be said to have pursued a zero-risk approach is the WP 29, established under the Data Protection Directive and now replaced by the EDPB. It stated that 'anonymisation is a technique applied to personal data in order to achieve irreversible deidentification'.[26] This approach can be contrasted with that of the GDPR, where the above-quoted Recital 26 takes a relativist approach by focusing on 'reasonably likely'.[27] While some actors, such as the Irish Data Protection Authority, have adopted a relativist approach,[28] a zero-risk approach is also evident. In 2020, Finck and Pallas reported that this has been the case with regard to the French Data Protection Authority (Commission nationale de l'informatique et des libertés). The finding holds even in February 2026, when the CJEU has taken the opportunity to clarify its case law on the fundamental concepts surrounding the notion of personal data, and notably the case of *SRB* in 2025. In particular, it still explicitly guides that 'L'anonymisation consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de toute personne par quelque moyen que ce soit'.[29] That is, at a time when awareness of perfect anonymisation as a myth, against the growing number of reported cases that identify the anonymised in

[22] Andrea Gadotti et al, 'Anonymization: The Imperfect Science of Using Data While Preserving Privacy' (2024) 10(29) Science Advances eadn7053.
[23] Gadotti et al (n 22).
[24] See in that regard Ira Rubinstein and Woodrow Hartzog, 'Anonymization and Risk' (2016) 91 Washington Law Review 703, 715–716.
[25] Opinion of Advocate General Campos Sánchez-Bordona in Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* EU:C:2016:339 para 52.
[26] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniqus', 0829/14/EN WP216, 7.
[27] See in that regard, e.g., Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10(1) International Data Privacy Law 11.
[28] Data Protection Commission, Guidance on Anonymisation and Pseudonymisation, 'Anonymisation and Pseudonymisation - Latest April 2022.Pdf', 2-3 <https://www.dataprotection.ie/sites/default/files/uploads/2022-04/Anonymisation%20and%20Pseudonymisation%20-%20latest%20April%202022.pdf> accessed 12 February 2026.
[29] "Anonymization involves using a set of techniques to make it practically impossible to identify any person by any means whatsoever," authors' translation. 'Comment prévenir les risques et organiser la sécurité de vos données ?' <https://www.cnil.fr/fr/comment-prevenir-les-risques-et-organiser-la-securite-de-vos-donnees> accessed 12 February 2026.

different datasets, has already existed for a while in the scholarship.[30]

The absolutist or relativist approach to understanding anonymisation has also had implications for the discussion on pseudonymisation. Article 4(5) GDPR approaches 'pseudonymisation' as a manner of processing of data.[31] However, it does not necessarily preclude treating pseudonymisation as an action in which personal data undergoes pseudonymisation. Indeed, the EDPB also highlights that pseudonymisation requires a pseudonymisation transformation of data.[32]

The scholarship is rich in approaches that treat pseudonymous data as personal data, either explicitly[33] or rather more subtly.[34] These conclusions are not difficult to achieve against such considerations as recital 28, stating that

> [t]he application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations. The explicit introduction of 'pseudonymisation' in this Regulation is not intended to preclude any other measures of data protection.

A rather radically different take, considering the time, has been made by Mourby et al, who, against the background of the *Breyer* case, have argued in 2018 that 'it should be possible for these data [data which have undergone pseudonymisation under the GDPR, author's note] to be rendered anonymous in some circumstances'.[35] Ideas along those lines can also be found elsewhere,[36] and potentially read between the lines in policy documents, such as those of the Irish Data Protection Authority.[37] Thus, the idea that pseudonymised data are not necessarily personal data has lingered for some time, though it has not dominated the scholarship. Alongside these discussions, pseudonymisation as means has been widely discussed, with much of the focus on datasets that undergo pseudonymisation.

3 PERSONAL DATA CLOSE UP IN THE CASE LAW

3.1 'ANY INFORMATION'

---

[30] See Rubinstein and Hartzog (n 24).
[31] See EDPB, 'Guidelines 01/2025 on Pseudonymisation' (16 January 2025) para 3 onwards.
[32] ibid para 18 onwards.
[33] 'As the definition makes clear, "pseudonymised data" remains "personal data" which are within the scope of the GDPR, and the data subject rights set out in Articles 15–20 still apply'. Claudia Irti, 'Personal Data, Non-Personal Data, Anonymised Data, Pseudonymised Data, De-Identified Data' in Roberto Senigaglia, Claudia Irti, and Alessandro Bernes (eds), *Privacy and Data Protection in Software Services* (Springer Singapore 2022) 54.
[34] See, 'The result of pseudonymisation is pseudonymised data which remain personal data but being protected through coding or encryption'. Gauthier Chassang, 'The Impact of the EU General Data Protection Regulation on Scientific Research' (2017) 11 Ecancermedicalscience 709.
[35] Miranda Mourby et al, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' (2018) 34(2) Computer Law & Security Review 222, section 1.3.
[36] See note 6, where the authors reflect on pseudonumisation and state that '[t]his raises the potential problem of pseudonymised data being used for identification purposes, which makes it in effect personal data' in Mark Elliot et al, 'Functional Anonymisation: Personal Data and the Data Environment' (2018) 34(2) Computer Law & Security Review 204.
[37] 'Where a pseudonym is used, it is often possible to identify the data subject by analysing the underlying or related data. The Authority uses often, as popsed to always, which opens up for our interpretation'. Data Protection Commission (n 28) 3.

Article 4(1) GDPR sets out 'any information' as a starting point in the personal data definition. A case that illustrates the application and interpretation of the first criterion in the definition of personal data, the criterion of 'any information', is *Nowak*. In *Nowak*, the CJEU examined whether written answers submitted by a candidate at a professional examination, and an examiner's comments with respect to those answers, constitute personal data within the meaning of Article 2(a) of the Data Protection Directive.[38] The CJEU famously held that

> [t]he use of the expression 'any information' in the definition of the concept of 'personal data', within Article 2(a) of Directive 95/46, reflects the aim of the EU legislature to assign a wide scope to that concept, which is not restricted to information that is sensitive or private, but potentially encompasses all kinds of information […] provided that it 'relates' to the data subject.[39]

This notion was driven by the recognition that the scope of the Directive was broad, encompassing diverse categories of personal data.[40] However, as reflected in the quote, the expression 'any information' is qualified by the requirement that the data must 'relate to' an individual. This first criterion ensures that the scope is not artificially limited by excluding certain types of information, thereby preserving the GDPR's protective function. Rather than narrowing the concept of personal data itself, potential concerns arising from an overly broad scope are addressed through additional criteria – such as identifiability and the 'relates to' test – which ensure that the application of the rules remains proportionate and coherent.[41]

As elaborated in case law, 'any information' potentially encompasses all kinds of information provided that it 'relates to' the data subject. It includes not only objective, but also subjective information.[42] Moreover, whilst the CJEU has long accepted that erroneous data can be protected as personal data, in *Nacionalinis visuomenės sveikatos centras,* it noted that fictitious data do not constitute personal data.[43] However, the reasoning appears not to be that such data fail to qualify as information, but rather that they are not linked to a real, existing natural person and therefore are not personal data.[44]

In the recent *SRB* judgment, the Court of Justice nuanced that the nature or type of information may have implications for the application of the subsequent criteria.[45] It derives from the case that comments reflecting the opinions and views of data subjects constitute information relating to those individuals by reason of their content alone.[46] This illustrates that the nature of the information can be significant in determining whether it constitutes personal data under the GDPR.[47] Accordingly, where the information in question consists of the personal views or opinions of its author, the 'relates to' criterion is, in effect, fulfilled. Such information inherently establishes a connection to the individual, not through external

---

[38] See *Nowak* (n 5).
[39] ibid para 34.
[40] Compare ibid at paras 33–34.
[41] Compare the use of the principle of proportionality to limit unwarranted obligations following a wide interpretation of the scope of the Directive in Case C-101/01 *Lindqvist* EU:C:2003:596 para 88.
[42] See e.g., Case C-487/21 *Österreichische Datenschutzbehörde and CRIF* EU:C:2023:369 para 23; Case C-479/22 *OC v Commission* EU:C:2024:215 para 45; *IAB Europe* (n 5) para 36; and *SRB* (n 11) para 54.
[43] Compare Case C-683/21 *Nacionalinis visuomenės sveikatos centras* EU:C:2023:949 para 55.
[44] ibid.
[45] *SRB* (n 11).
[46] See ibid paras 52–61.
[47] See ibid para 56.

effect or purpose, but by its content alone.[48]

While it is clear that the nature of the information can influence the application of the 'relates to' criterion, its connection to the identifiability criterion warrants further examination, particularly in light of the Court of Justice's judgment in *IAB Europe*.[49] This case concerned the so-called Transparency and Consent String (TC String), in which user preferences were encoded and stored as a string of letters and characters.[50] This string formed part of a broader transparency and consent framework designed to facilitate the communication of user choices within real-time bidding – an automated system for auctioning advertising space on the internet based on user profiles.[51] Through this system, advertising technology companies could bid, in real time, for the opportunity to display personalised advertisements tailored to individual user profiles. In addition to the TC String, a cookie was placed on the user's device once he or she had consented to or objected to the processing of their data. When combined, the TC String and the cookie could be linked to the user's IP address.[52] In assessing whether the TC String constitutes personal data, the Court of Justice found that, although it did not contain direct identifiers, it encoded user preferences that originated from a specific natural person and related to that individual.[53] On this basis, the Court concluded that the TC String fell within the scope of personal data under Article 4(1) GDPR.[54]

The reasoning of the *IAB Europe* can be argued to present a shift in emphasis: rather than requiring that identifiability be established from the outset, the Court treated the human origin of the data – and its potential linkability to other information – as sufficient for the 'relates to' criterion to be met. In reaching this conclusion, the Court drew on Recital 26, which frames identifiability in risk-based terms, as well as the concept of pseudonymisation, despite the TC String not involving the typical removal of identifiers.[55] The judgment thus appears to decouple 'any information' from immediate identifiability, raising broader questions about the threshold for classifying data as personal under the GDPR when the data stems from a human. It thus appears that the nature of the information, as assessed under the first criterion, may influence the application of the subsequent criteria – both in determining whether the information relates to a natural person and whether that person is identifiable.

3.2 'RELATING TO'

The second essential criterion for information to qualify as personal data is the 'relates to' criterion, which requires that the information must relate to a natural person – that is, it must be personal. In contrast to the broader 'any information' element, which received a defining interpretation in *Nowak*,[56] this criterion has prompted more diverse interpretations in case

[48] Compare *SRB* (n 11) para 60 and *Nowak* (n 5) as referenced.
[49] *IAB Europe* (n 5).
[50] See ibid para 25.
[51] See ibid para 26.
[52] ibid para 25.
[53] See ibid para 43.
[54] See ibid para 51.
[55] Compare ibid paras 39–40.
[56] See *Nowak* (n 5) para 34.

law – some adopting a more teleological approach, while others take a more systematic or strict stance.[57]

In a joined case dating back to 2014, the Court of Justice elaborated on the criterion 'relating to' enshrined in the definition of personal data in Article 2(a) of the Data Protection Directive. The case *YS and Others* illustrates that the criterion serves to limit the otherwise broad scope of the term 'any information' in Article 4(1) GDPR.[58] In line with the purpose of the Regulation, it is not sufficient for information to be merely linked to an identified or identifiable individual; rather, the information must concern the person in a manner that could justify the exercise of data subject rights. The core questions in the preliminary ruling concerned draft documents and attached 'minutes' drawn up by the case officer of the Immigration and Naturalisation Service regarding applications for residence permits. A minute may include information about the case officer and the applicant, as well as applicable legal provisions, and an assessment of the foregoing information in the light of the legal provisions.[59] The national courts sought the Court of Justice's advice on whether the contents of such a minute constitute personal data. The Court of Justice initially noted that there is no doubt that information relating to the applicants who are identified in the minutes, in particular, by their name must be considered personal data.[60] However, the same cannot be said for the legal analysis contained within a minute.[61] The Court of Justice arrived at that conclusion against an interpretation of the wording of Article 2(a) of the Data Protection and the purpose of the Directive. The rights of the data subject, such as the right of access, uphold the rights to privacy and respect for one's private life by allowing the data subject to be certain that the data concerning him or her are correct and lawfully processed.[62] However, as a legal analysis is not itself subject to scrutiny for accuracy or lawfulness by data subjects, treating a legal analysis as personal data would not serve the Directive's purpose.[63] While this preliminary ruling is important for various reasons, including demarcating the scope of the regulation of national administrative procedures, it is also important for understanding how far the criterion 'relating to' in the personal data definition extends. Nonetheless, this balance can be very fine and context-dependent. In case *YS and Others*, access to documents, a matter commonly falling under the national constitutional and administrative law traditions, was at stake. However, with respect to medical information, one can readily argue that medical assessments of an individual constitute personal data,[64] unless the health records hold very general information that has lack of relevance to an individual's particular situation.

A few years after the *YS and Others* ruling, in *Nowak* in 2017, the Court of Justice elaborated on the meaning of the term 'relating to' a data subject in light of the purpose of

---

[57] Compare Joined Cases C-141/12 and C-372/12 *YS and Others* EU:C:2014:2081 with *Nowak* (n 5).
[58] See *YS and Others* (n 57).
[59] See ibid paras 13–14.
[60] ibid para 38.
[61] ibid para 39. The Court explained that 'such a legal analysis is not information relating to the applicant for a residence permit, but at most, in so far as it is not limited to a purely abstract interpretation of the law, is information about the assessment and application by the competent authority of that law to the applicant's situation, that situation being established inter alia by means of the personal data relating to him which that authority has available to it'. *YS and Others* (n 57) para 40.
[62] See *YS and Others* (n 57) para 44.
[63] ibid para 46.
[64] See Case C-21/23 *Lindenapotheke, ND v DR* EU:C:2024:846.

the law.[65] Instead of adopting a broad interpretation of the 'relates to' element in the definition of personal data, the Court adopted a more structured approach, setting out criteria to assess whether information has a sufficiently close connection to an individual to justify the application of data protection rights. It explained that information relates to a data subject where the Court, by reason of its *content, purpose or effect*, is linked to a particular person.[66] The CJEU pointed out that in this case, the information in question, the candidate's written answers, and the comments made by the examiner relate to the candidate on account of all three mentioned bases.[67] Recently, in *SRB*, the Court reaffirmed that the three elements are alternative criteria, as signalled by the conjunction 'or'.[68] However, *SRB* introduces some uncertainty regarding how detailed the assessment must be. As noted in relation to the first criterion, the presence of personal opinions or views necessarily implies a close link to the natural person concerned.[69] It appears that for certain types of data, the 'content' element may be deemed satisfied without a detailed assessment of the data's connection to the individual concerned. That could be the case when data originates from a sample of human biological material. This is further supported by the Court's reasoning in *IAB Europe*, where the user preferences contained in a TC String were considered to relate to a natural person without requiring a more elaborate analysis, rendering the content of the information essentially decisive.[70] Thus, while *Nowak* involved a detailed assessment of each of the three elements relevant to the case, despite them being alternative, such an in-depth analysis does not appear to be a requirement in itself. It is sufficient that one element only is met.

In light of these cases, it appears that the essential function of the second criterion in the definition of personal data is to limit the scope of the GDPR in a manner consistent with its underlying purpose. However, one might question whether the CJEU's approach is in line with that ambition. In a recent case delivered after *IAB Europe*, the definition of personal data was once again called into question. In *Ministerstvo zdravotnictví*, the Court was asked to consider whether the personal contact information of natural persons acting on behalf of a legal person could be regarded as personal data of those individuals.[71] Of particular interest is the Court's finding that the name, surname, and signature of a natural person qualify as personal data where the individual is identifiable.[72] However, it left the assessment of other types of contact details – such as email addresses, telephone numbers, and websites – to the national courts, which were to make that determination by reference to factors such as those discussed in *IAB Europe*.[73] The Court indicated that the criterion of whether data 'relates to' a data subject may require a more detailed assessment in such cases. Information directly linked to the identity of a data subject appears to constitute personal data without much controversy. In contrast, information that pertains to an individual in a more indirect or functional manner may require more careful consideration to determine whether it genuinely

[65] See *Nowak* (n 5).
[66] ibid para 35.
[67] ibid paras 37-39.
[68] *SRB* (n 11) para 56.
[69] See ibid para 58.
[70] See *IAB Europe* (n 5) para 43. This is also paving pathway for pseudonymisation being affirmed as a state of data. See below 4.
[71] See Case C-710/23 *Ministerstvo zdravotnictví* EU:C:2025:231.
[72] ibid para 24.
[73] ibid para 25.

'relates to' a natural person under the GDPR. In such instances, the relates to criterion seems to overlap with the identifiability criterion, as the question of whether data relates to someone may depend, at least to some extent, on whether that person can be identified, directly or indirectly, through the information in question.

3.3 'AN IDENTIFIED OR IDENTIFIABLE NATURAL PERSON'

The third and final criterion – the identifiability criterion – focuses less on the inherent qualities of the information and more on the risk associated with its potential to identify an individual. For the GDPR to apply, the data must relate to someone who is either identified or identifiable, aligning the scope of the Regulation with its risk-based approach to protecting individuals from potential misuse of their personal data. This criterion has given rise to legal uncertainty regarding its precise meaning despite, and perhaps partly because of, the clarifications provided in Recital 26. The text of the GDPR provides that the identifiability of an unidentified data subject depends on the means reasonably likely to be used, by the controller or any other person, taking into account all objective factors.[74] However, this formulation leaves certain ambiguities unresolved. Is it sufficient that any person has the means to identify the data subject for the information to be considered personal data? Can the same data be personal in relation to one actor, while remaining non-personal in relation to another? And where, precisely, is the threshold drawn between 'unreasonable' means and those 'reasonably likely' to be used? These questions have gradually been addressed through the development of case law.

The most prominent case exploring the meaning of personal data prior to the GDPR's entry into force was *Breyer*.[75] In *Breyer*, the CJEU established a framework for assessing whether a natural person is identifiable, shaping our understanding of the risk-based identification criterion – namely, whether there are means reasonably likely to be used to identify an individual. The Court's approach permitted a broad interpretation of the concept of personal data, allowing information held by another person to be considered when assessing an individual's identifiability in relation to the controller, as reflected now in Recital 26 of the GDPR. At the same time, the judgment introduced uncertainty regarding whether the assessment of identifiability should be objective or subjective. While Recital 26 states that identifiability depends on the means reasonably likely to be used by the controller or by any other person, the Court appeared to base its analysis primarily on the means accessible to the controller – either directly or indirectly through another party. In its judgment, the Court appears to have adopted a relative understanding of the concept of personal data – assessing identifiability on a subjective basis with reference to a specific data holder. However, it remains unclear whether this was the Court's intended approach, or whether the means most relevant to assess in this case were simply those available to the controller with particular emphasis on the potential for indirect identification.

The Court held that it follows from the wording of the personal data definitions set out in the Data Protection Directive that it is not necessary that information alone allows the data subject to be identified for it to be considered personal data of an indirectly identifiable

[74] Recital 26 GDPR (n 1).
[75] *Breyer* (n 6).

person.[76] As guided by Recital 26 of that directive, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify a person to determine whether that person is identifiable.[77] Regarding 'any other person', the Court of Justice held that it is not required that all the information enabling the identification of the data subject must be in the hands of one person.[78] The question that subsequently emerges, however, is whether the Court thereby accepted an objective assessment of identifiability – based on the hypothetical availability of identifying means to any third party – or whether the assessment remains relative, grounded in what is reasonably accessible to the specific data controller in question.

In this case, the Court of Justice then proceeded with a subjective assessment, examining whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means reasonably likely to be used to identify the data subject and stated that it would not be the case if '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'.[79] It held that it appears that the service provider had the means reasonably likely to be used to identify the data subject with the assistance of other persons.[80] Thus, *Breyer* demonstrated that there is reason to conduct a subjective assessment of whether the means are reasonably likely to be used by a specific person. However, it cannot be conclusively said that the judgment rules out an objective approach. In accordance with the GDPR, information must be reasonably likely to be used to identify an individual for the information to qualify as personal data. Therefore, an assessment based on the capabilities of a specific party – often the controller – may be unavoidable.

The CJEU has continued down the path of a relative concept of personal data also in cases following *Breyer*. In a case concerning the market surveillance of motor vehicles,[81] *Gesamtverband Autoteile-Handel*, the Court of Justice was asked to clarify whether Vehicle Identification Numbers (VINs) qualify as personal data under EU law.[82] The case endorsed a contextual, or relative, approach to identifiability, highlighting that a natural person may be identifiable in relation to a particular actor, rather than in absolute terms. The Court noted that data which is not personal in itself may become personal data for a person who has the means reasonably likely to associate it with a specific individual.[83] It stated that

> where independent operators may reasonably have at their disposal the means enabling them to link a VIN to an identified or identifiable natural person, which it

[76] *Breyer* (n 6) paras 40–41.
[77] Which is now comparable to Recital 26 of the GDPR. See *Breyer* (n 6) para 42.
[78] ibid para 43.
[79] ibid para 46.
[80] ibid para 48. The Court referred to the fact that, in case of a cyber-attack, legal channels that allow an online media services provider to contact the competent authority which in turn can take the steps necessary to obtain that information from the internet service provider existed. *Breyer* (n 6) para 47.
[81] See Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC [2018] OJ L151/1.
[82] See Case C-319/22 *Gesamtverband Autoteile-Handel* EU:C:2023:837.
[83] ibid para 46.

> is for the referring court to determine, that VIN constitutes personal data *for them* [emphasis added], within the meaning of Article 4(1) of the GDPR, and, indirectly, for the vehicle manufacturers making it available, even if the VIN is not, in itself, personal data for them, and is not personal data for them in particular where the vehicle to which the VIN has been assigned does not belong to a natural person.[84]

The case appears to build on the Court's reasoning in *Breyer*, which has been interpreted as supporting a relative approach to identifiability. However, as in *Breyer*, the Court in this case did not reject an absolute approach outright but rather limited its assessment to whether the individual was identifiable based on the means available to a specific person. It also illustrates that data may become personal data once the identification of an individual becomes both reasonably possible and contextually relevant.

Building on earlier case law, the Court of Justice recently clarified the assessment of the identifiability criterion in the *SRB* judgment. The case helped resolve lingering uncertainty about the relative nature of the concept of personal data, with reference to the distinction between anonymous data and pseudonymised data. It arose in the context of a resolution process conducted by the European Single Resolution Board (SRB), the central authority within the Banking Union. As part of this process, shareholders and creditors were invited to express their interest in exercising the right to be heard regarding potential compensation. The SRB collected their comments and pseudonymised them before transferring the data to Deloitte for an independent valuation. The question was whether these comments should be considered personal data. While acknowledging that pseudonymisation is not part of the definition of personal data but rather a process aimed at reducing the risk of associating a dataset with the identity of a data subject, the Court of Justice held that pseudonymisation can, in certain circumstances, alter the nature of personal data to the extent that it no longer qualifies as personal data.[85] This conclusion rests on the distinction that pseudonymised data remain personal data from the perspective of the controller who performed the pseudonymisation, but may no longer qualify as personal data in relation to a third party who lacks the means to re-identify the data subjects.[86] The Court of Justice stated that the assessment of 'reasonably likely means' set out in Recital 26 would be deprived of any practical effect if pseudonymised data were considered personal data in all cases and for every person.[87] Instead, whether data qualifies as personal is not determined once and for all, but may vary depending on the context and the party processing the data. As shown in *Gesamtverband Autoteile-Handel,* data which are in themselves impersonal may become personal in the hands of others with the means reasonably likely to be used to identify the individuals behind the data.[88] Conversely, personal data that has undergone pseudonymisation may cease to be personal in relation to another party who lacks access to the additional information necessary for re-identification.[89] While affirming that pseudonymisation is not part of the definition of personal data, the Court nonetheless relies on its relevance to the identifiability criterion,

[84] *Gesamtverband Autoteile-Handel* (n 82) para 49
[85] See *SRB* (n 11) paras 72 and 75.
[86] Compare ibid paras 76–77.
[87] ibid para80.
[88] See ibid para 84.
[89] *SRB* (n 11) para 86.

using the concept to assess the nature of identifiability.

The question remains: where does the threshold for the identifiability criterion lie? In dealing with the appeal following a decision of the General Court, in *OC v Commission*, the Court of Justice further clarified identifiability, focusing on its relative – or subjective – dimension.[90] The case concerned a press release issued by the European Anti-Fraud Office about the fraudulent use of research funding. The key question was whether the researcher mentioned in the press release could be identified by reasonable means from the information in the press release, and thus whether the information in that press release constituted personal data. In the case at hand, a journalist had successfully identified the researcher based on the press release, using additional information. However, the CJEU held that this individual instance of identification was not, in itself, sufficient to conclude that the press release contained personal data.[91] Rather, it emphasised the need to assess the broader risk of identification. The Court of Justice found that the risk of identification could not be regarded as insignificant, as individuals working in the same scientific field would be able to identify the researcher, considering easily accessible other information on the internet, and thus doing that without disproportionate effort in terms of time, cost, or labour, thus meeting the threshold for identifiability under data protection law.[92] However, while initially the Court placed emphasis on a defined group, individuals working in the same scientific field, it also went further in its analysis and acknowledged that the issue as such can likely arouse interest among the public and induce readers to investigate who is the subject of the press release. In particular, it noted that the internet searches that could be conducted to identify the person in question did not render the effort disproportionate, and consequently, the identification risk could not be insignificant.[93] When pointing at the errors of the General Court, the Court of Justice went even further and not only emphasised the means reasonably likely to be used by the readers of the press release, but also that the simple, objective reading of the press release might be sufficient to identify the data subject in question.[94] Thus, while the analytical side focuses on the reasonable efforts – internet search and publicly available information – which do not require advanced knowledge of the field or computing skills, the Court also includes a consideration that, for some readers, even that might not be necessary. This raises the question of the spectrum of reasonably likely means. This case appears to fall at one end, requiring limited analytical inquiry or professional expertise. Towards the other end, however, are questions of identifiability from genetic data alone or coupled with other information, the efforts, skills, and resources required, and whether, in a research context, those could make an individual reasonably likely to be identified.

The case of *OC v Commission* underscores two key principles for interpreting identifiability under EU data protection law. First, it confirms that identifiability is not a black and white concept but lies on a spectrum that requires a case-by-case assessment of what constitutes 'reasonable means' of identification. Second, it highlights that the risk of identification is inherently context-dependent, particularly in relation to the characteristics of

[90] See *OC v Commission* (n 42).
[91] ibid para 58.
[92] ibid para 61.
[93] See ibid paras 62–63.
[94] ibid para 64.

the data recipient and their ability to access and combine additional information. While the judgment does not provide a definitive standard for how context and risk should be factored into the analysis, it offers important indications. In line with the risk-based approach adopted in EU data protection law, the Court appears to favour a more objective criterion – focused not on whether any individual could identify the data subject, but whether identification would be reasonably likely for an average recipient. In this context, the average reader of a press release may serve as a reference point for assessing whether, when combined with reasonably accessible information, the data constitutes personal data.

The identifiability, but from a different angle, was further elaborated in *IAB Europe*, a decision delivered on the same day as the decision in *OC v Commission*. In the *IAB Europe* case, the data in question[95] were still considered personal data in relation to IAB Europe, since the organisation's members were required to provide IAB Europe with the relevant identifiers upon request, and IAB Europe thus had means reasonably likely to be used to identify the data subjects.[96] Even without a pseudonymisation process, data that were pseudonymous with respect to IAB Europe, as they contained individuals' choices and were considered personal due to the means available to the organisation, were considered personal despite the fact that these choices, in themselves, did not contain direct identification details such as name. Following that argument, if a party has the means to obtain identifiers that enable inherently impersonal data to be linked to an individual – for example, in the form of legal access – such data must be regarded as personal data from that party's perspective.[97] In such a case, it seems that the party cannot avoid the application of the GDPR by relying on anonymisation, as the means to identify the data subjects remain reasonably likely. Treating data that is pseudonymous by nature as personal, depending on, for example, the legal restrictions or options available to a processor, may appear very broad. However, the Court of Justice introduced an additional nuance in the *SRB* judgment. Since the concept of identifiability is relative, the identifiable nature of the data subject must be assessed at the time of relevance for the specific provision in question and from the perspective of the relevant party.[98] Hence, it can be argued that data may be regarded as personal in some cases and as non-personal in others, thereby potentially affecting the scope and the application of a legal instrument and the rights and obligations set therein.

4 NOTIONS NEIGHBOURING PERSONAL DATA IN CASE LAW

4.1 PSEUDONYMISED DATA

The case of *IAB Europe* is transformative for understanding pseudonymisation. In this case, the Court of Justice treated a TC-string containing the individual preferences of a specific user regarding his or her consent to the processing of personal data concerning him or her as personal data.[99] It arrived at the conclusion by contrasting indirect identification, which does not require that the information alone allows the data subject to be identified, with

[95] That were later described as 'inherently impersonal' when referred to in *SRB,* see *SRB* (n 11) para 83.
[96] See *IAB Europe* (n 5) paras 46–48.
[97] Compare *Breyer* (n 6) para 46.
[98] See *SRB* (n 11) para 111.
[99] *IAB Europe* (n 5) paras 39–43.

the definition of pseudonymisation and Recital 26. It can be said that the CJEU has implied, by this, that some data may have been pseudonymised by virtue of their very existence, such as the TC string in this case.

The *SRB* case adds further nuance to the understanding of pseudonymisation, particularly regarding the link between pseudonymous and anonymous data. Having acknowledged that pseudonymous data 'are not mentioned in the legislative definition of the concept of "personal data"',[100] the Court of Justice followed Advocate General's opinion, and emphasised that pseudonymised data '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'.[101] In its view, 'the concept of "pseudonymisation" presupposes the existence of information enabling the data subject to be identified'.[102] In the Court's view, the 'the objective of pseudonymisation is, among other things, to prevent the data subject from being identified solely by means of pseudonymised data'.[103] Against that background, the Court then established that

> provided that such technical and organisational measures are actually put in place and are 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, pseudonymisation may have an impact on whether or not those data are personal.[104]

This reasoning allowed the Court of Justice to clarify that while for the controller, who has pseudonymised data, the data may be personal, it might not necessarily be the case for the recipient of the personal data. That is the case if the recipient 'is not in a position to lift those measures during any processing of the comments which is carried out under its control', as well as the pseudonymisation measures are in fact such as to prevent the recipient from attributing the pseudonymised data 'including by recourse to other means of identification such as cross-checking with other factors, in such a way that, for the company, the person concerned is not or is no longer identifiable'.[105] This approach was ultimately located against the wording of Recital 26 that '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'.[106] This is where the reasonable likelihood of identification comes into play,[107] and must be assessed in a particular case at hand.

Against the backdrop of *OC v Commission*, the Court of Justice took a chance to elaborate in detail on the limits and nature of pseudonymous data. It affirmed that, in *OC v Commission*, it did not confine itself to finding that the EU body that published the press

---

[100] *SRB* (n 11) para 71.
[101] ibid para 72.
[102] ibid para 73.
[103] ibid para 74. Finally, 'the requirement that the identifying information be kept separately and that it be subject to technical and organisational measures "to ensure that the personal data are not attributed to an identified or identifiable natural person", […], indicates that the objective of pseudonymisation is, among other things, to prevent the data subject from being identified solely by means of pseudonymised data'.
[104] ibid para 75.
[105] ibid para 77.
[106] *SRB* (n 11) para 78.
[107] ibid para 79.

release possessed all the information necessary to identify that person. Instead, it 'examined whether the statements contained in that press release reasonably enabled the public concerned to identify that person, in particular by combining those statements with information available on the internet'.[108] It endorsed the *OC v Commission*, stating that the Court held in that case

> that a means of identifying the data subject is not reasonably likely to be used where the risk of identification appears in reality to be insignificant, in that the identification of that data subject is prohibited by law or impossible in practice, for example because it would involve a disproportionate effort in terms of time, cost and labour.[109]

It went on to further emphasise that '[t]hat case-law bears out the interpretation that the existence of additional information enabling the data subject to be identified does not, in itself, mean that pseudonymised data must be regarded as constituting, in all cases and for every person, personal data'.[110] Consequently, the Court of Justice moved on to clarifying the logic in *IAB Europe* and *Breyer*, referring to data originating from humans as 'inherently personal' and being treated as personal data because the controller had the means to identify the data subject,[111] and pointing out the lesson of *Gesamtverband Autoteile-Handel* that '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'.[112]

These cases provide several important lessons. They affirm that some data may be pseudonymous by their nature, as it derives predominantly from *IAB Europe*. Data that are 'inherently personal' may become personal data, but they may also not be such, depending on whether the party in question has the means to identify the data subject with a reasonable likelihood. A risk that appears insignificant should not be such, following *OC v Commission*, that it renders the data personal. The same reasoning also holds for impersonal data. Whilst the language may suggest that they do not concern personal data, in the hands of those with a reasonable likelihood of identification, provided that this identification risk is not insignificant, they could become personal data.

4.2 ANONYMOUS DATA

In contrast to personal data, anonymous data does not relate to an identified or identifiable natural person.[113] The protection of personal data is also limited to personal data which has not been rendered anonymous in such a manner that the data subject is no longer identifiable. The concept of anonymous data has two dimensions: it encompasses both data that has never been related to an identified or identifiable natural person and data that has been effectively anonymised so that it no longer constitutes personal data.[114] Case law predating

[108] ibid para 81.
[109] ibid para 82.
[110] ibid para 82.
[111] ibid para 83.
[112] ibid para 84.
[113] Recital 26 GDPR (n 1).
[114] Compare Recital 26 GDPR (n 1).

SRB appears to hint at a distinction between anonymised and pseudonymised data while defining personal data under the GDPR,[115] albeit in a subtle way, keeping the actual relationship unelaborated.[116] This unelaborated relationship allowed the European Data Protection Supervisor to submit in the *SRB* case that pseudonymised data constitute personal data – in all cases and for every person.[117] However, following the Court of Justice's reasoning in *SRB*, this take was rejected at its roots, and pseudonymisation may appear to be sufficient, in certain contexts, for data to be regarded as no longer identifiable.[118]

As evidenced by more recent CJEU case law, the concept of pseudonymised data has become somewhat clearer, with implications for the understanding of anonymous data. Interestingly, though, the Court of Justice refrains in *SRB* from attaching the anonymous data label to such data that are pseudonymised by a controller, transferred to another party, and not identifiable by that party. One way to look at the question of pseudonymised data versus anonymous data is to consider pseudonymised data as data that can be regarded as rendered not personal and thereby anonymous in relation to a specific person, while data is truly anonymous if it has never been related to a natural person.[119] However, it does not seem appropriate to limit the concept of anonymous data to non-human data, and to require an active anonymisation process for human data relating to an unidentifiable natural person to fall outside the scope of the GDPR. Yet, recent developments in case law call for a reassessment of what anonymous data truly is. As demonstrated in *SRB*, the concept of personal data is inherently relative. If, as suggested in *Nacionalinis visuomenės sveikatos centras*,[120] the distinction between pseudonymous and anonymous data mirrors the boundary between personal and non-personal data, then it raises the question: Is the concept of anonymous data itself also relative?

A reading of Recital 26 may suggest that anonymous information is assessed based on the same principles as the identifiability criterion within the concept of personal data, as indicated by the use of the word 'therefore' to describe anonymous information in relation to that assessment. It would also be consistent with prior case law, such as *Nacionalinis visuomenės sveikatos centras*, if the relative nature of personal data similarly influenced the understanding of anonymous data. However, if anonymous information were to be judged on an objective basis, in relation to any potential individual, there would be a divide between pseudonymised data that can no longer be attributed to a data subject by reasonable means by a specific person, and anonymous data which could not be attributed to a data subject by reasonable means by any person. In that case, the CJEU's choice to refer to the de-identification of data in relation to a third party as 'pseudonymisation' rather than 'anonymisation' makes sense, even though the data is considered non-personal in relation to that third party.[121] However, because the wording of the *SRB* judgment is not entirely clear, it remains unclear whether anonymous information is defined as data not relating to an identifiable person in a relative or absolute sense.[122]

---

[115] Compare *Nacionalinis visuomenės sveikatos centras* (n 43) paras 57–58.
[116] ibid.
[117] *SRB* (n 11) para 86.
[118] See ibid para 87.
[119] Compare the reasoning in *Nacionalinis visuomenės sveikatos centras* (n 43) paras 55 and 57–58.
[120] Compare *Nacionalinis visuomenės sveikatos centras* (n 43) paras 57–58.
[121] Compare *SRB* (n 11) para 75.
[122] Compare the sentence in *SRB* (n 11) para 73: 'The very existence of such information precludes data that

5 SYNTHESIS OF THE FINDINGS: WHERE DO WE STAND NOW?

The concept of personal data has demonstrated both its significance and complexity in CJEU case law, serving as a crucial threshold for the scope and application of data protection laws. The seemingly straightforward wording of Article 4(1) GDPR – defining personal data as 'any information relating to an identified or identifiable natural person' – reveals a far greater significance than it may at first suggest, shaping the very boundaries of data protection and determining the scope of obligations for those who process personal information, and brings along implications for data-intensive fields, such as biomedical research. The Court's clarifications on its application are relevant to understanding when data is inherently protected, when it is not, and when its protection is contingent on individual actions. Recent cases not only clarify the meaning of the concept, but also shed light on pseudonymisation and anonymisation as activities or states in which data can exist. Building on these developments, the analysis now turns to the concept of personal data, examining how its defining criteria, clarifications elaborated in case law, and the Court's key legal reasoning have shaped its current interpretation under the GDPR. To recall, not all of the cases shaping the understanding have been addressed under the GDPR.

First, it has become increasingly clear that the definition of personal data is functional in nature; its interpretation has evolved through the gradual development of CJEU case law, guided by the underlying principles of data protection set out in the relevant secondary legislation at issue in those cases. The first element of the definition, 'any information', serves as a neutral entry point, setting the tone for a technologically neutral framework intended to protect natural persons.[123] As the well-known formulation in *Nowak* illustrates – '[t]he use of the expression "any information" in the definition of the concept of 'personal data' […] reflects the aim of the EU legislature to assign a wide scope to that concept, which is not restricted to information that is sensitive or private, but potentially encompasses all kinds of information […] provided that it "relates" to the data subject'[124] – what qualifies information for protection is not its sensitivity or private character, but the fact that it relates to a specific individual. It thus appears that the first criterion functions less as a limitation and more as a prerequisite for regulating any processable information that concerns a person. Rather than narrowing the scope, it affirms its breadth – emphasising that the concept of personal data is intended to be expansive. That said, in keeping with a framework guided by overarching principles rather than rigid methodology, this criterion acquires meaning through its relationship with the subsequent elements of the definition.

As highlighted in the citation from the *Nowak* judgment above, the 'relates to' criterion serves to qualify what may fall under the broad category of 'any information' within the scope of the GDPR. The 'relates to' criterion may be assessed not only in terms of the content of the data itself, but also with regard to the purpose of the processing or its potential effect on

have undergone pseudonymisation from being regarded, *in all cases* [emphasis added], as anonymous data, which is excluded from the scope of that regulation'.

[123] In line with the principle set out in Recital 15 GDPR (n 1).

[124] *Nowak* (n 5) para 34.

the individual.[125] Although the information need not be inherently personal to fall within the scope of personal data, its content may nevertheless affect the extent to which it is linked to an identifiable individual. The examples of information that relate to an individual by reason of its content, as found in the CJEU case law – such as personal opinions or views[126] – highlight that while 'any information' can constitute personal data, some types of information are more likely to qualify as such because they relate to an individual by their content. This is precisely where information originating from an individual, such as human genetic data, comes in. At the same time, although the Court has refrained from limiting the concept of personal data solely on the basis of the first criterion, its connection to the 'relates to' criterion and its principle-based interpretation mean that some information might not necessarily be considered personal data. As demonstrated in the Court's reasoning in *YS and Others*, information that is clearly linked to personal data but remains too abstract to reveal anything about the individual is unlikely to qualify as personal data.[127] It appears that the 'any information' criterion, when considered alongside the 'relates to' criterion, requires the information to reveal something about the individual that justifies granting them control over the rights protected by the GDPR.[128] Information originating from the individual – whether directly or indirectly, and especially if of a private nature – tends to be inherently revealing. Conversely, information obtained from other sources linked to an individual could qualify as personal data only if it reveals something about that individual that satisfies the effect or purpose criteria. As clarified in *YS and Others*, any information may qualify as personal data, but it should do so only when it concerns an individual in a way that justifies the protection of their rights under the GDPR.[129]

Second, the concept is also relative. This relative dimension of personal data is highlighted in the Court's interpretation of the identifiability criterion. As early as seen in *Breyer*, the identifiability of a natural person must be assessed from the perspective of a specific actor.[130] However, some uncertainty persisted as to whether identifiability – when assessed from the perspective of a relevant party – should be understood in relation to all parties or only specific ones. In the cases following *Breyer*, the Court moved towards a clearly relative understanding of personal data, interpreting it as a concept defined in relation to a particular actor. With the culmination of this approach in the recent *SRB* case, it is now evident that the identifiability criterion is relative, thereby rendering the concept of personal data itself relative. However, some questions remain regarding how identifiability should be assessed in relation to potential data processors or other persons with access to these data. While the assessments in *Breyer* and *SRB* focused on a specific individual in light of a particular provision of the legal instrument at issue, in *OC v Commission,* the Court of Justice departed from this approach. It disregarded the actual identification of the data subject by a specific individual and instead examined the risk of identification by a member of a defined group, and subsequently extended that approach to encompass the public and readers more broadly. A realised risk of identification is thus not sufficient to establish that identification

[125] As seen in *Nowak* (n 5) para 35 and further highlighted in *SRB* (n 11) para 56.
[126] See *SRB* (n 11) para 58.
[127] See *YS and Others* (n 57) paras 39–40.
[128] Compare ibid.
[129] Compare ibid paras 40–46.
[130] See *Breyer* (n 6).

was reasonably likely.[131] This ties into other elements of relative identifiability – namely, that identifiability is also relative in time, as determined by the specific underlying legal provision at issue. The appropriate time for assessing identifiability appears to depend on the context and purpose of the provision in question under data protection law.[132] In *SRB*, the relevant provision concerned the obligation to provide information to the data subject at the time of data collection. Accordingly, the assessment of identifiability had to be made at that specific point in time, rather than retrospectively.[133] However, in *OC v Commission*, the applicant sought compensation for non-material damage under Article 268 TFEU,[134] allegedly caused by a press release issued by an EU institution.[135] In that case, the Court of Justice explicitly held that the question of what constitutes personal data 'cannot be confused with the question relating to the conditions necessary for the European Union to incur non-contractual liability' and that it 'must be assessed exclusively in the light of the conditions laid down by that provision [Article 3(1) of Regulation 2018/1725 – comparable to Article 4(1) GDPR] and therefore, […] cannot depend on considerations relating to the imputability of an act to the European Union'.[136] It thus seemed that the Court of Justice viewed the identifiability criterion as an element which should be interpreted in isolation – disconnected from the provision relevant to the case – contrary to the approach portrayed in *SRB,* where the relevant provision, the obligation to provide information, guided the interpretation. This raises doubts about whether the identifiable nature of a data subject varies with the applicable legal provision, or whether the provision is relevant only in specifying the time at which identifiability should be assessed.[137]

Further, with the relative understanding of the identifiability criterion, it appears to have become more distinct from the 'relates to' criterion. While it is true that information that enables the identification of an individual typically also relates to that individual, and vice versa, the relative approach to identifiability allows information to relate to a natural person to a greater extent while still remaining non-personal data with respect to certain parties. In this line of reasoning, the identifiability criterion is separated from the 'relates to' criterion, and its relevance seems to come in at a later stage than the first two criteria. This line of reasoning aligns with the view that the first criterion – qualified by the second – determines what merits protection, while the third sets out the conditions under which such protection is justified. The definition of personal data, therefore, incorporates both a teleological (purpose-driven) dimension and an element of risk. However, the distinction between the 'relates to' and identifiability criteria becomes less clear when examined in the context of pseudonymisation and pseudonymised data. If data can be considered pseudonymous by nature – such that they are 'inherently impersonal' to the entity processing them – but become personal data when they convey information about an identifiable person to another party, then the differentiation between the two criteria is maintained. Yet, if data of human origin are understood to carry an inherent element of

131 Compare *OC v Commission* (n 42) paras 57–58.
132 Compare *SRB* (n 11) para 111.
133 Compare ibid para 106.
134 Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/47.
135 *OC v Commission* (n 42) para 23.
136 ibid para 54.
137 Compare also *SRB* (n 11) para 85.

identifiability, the 'content' aspect of the 'relates to' criterion begins to overlap with the identifiability criterion, thereby weakening the latter's distinct function. Such an interpretation of *IAB Europe* should be cautioned against, as it suggests that a risk of identifiability is presumed for certain categories of data by default.

Third, the concept of personal data can be argued to be subject to a proportionality assessment through the 'means reasonably likely to be used' test within the identifiability criterion, with a de-minimis criterion that the identification risk is not insignificant. This limits the broad scope of the concept of personal data by setting a threshold for identifiability.[138] The risk-based assessment described in *Breyer* as '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'[139] clearly incorporates a proportionality element. Coupled with the fact that actual identification does not necessarily mean the identifiability criterion is fulfilled,[140] it is not proportionate to accept an insignificant risk of identification for the criterion to be met. However, the wording in *Breyer*, describing acceptable risk as 'insignificant', appears to set a relatively low threshold for identifiability, which may seem at odds with the language of Recital 26, which refers to the 'means reasonably likely to be used'. Whether there is a discrepancy between the wording of the provision and the Court of Justice's interpretation remains open to debate, but it is clear that objective factors must be taken into account. While identifiability should be assessed from the perspective of a specific person or group, the means of identification should not be judged based on actual or subjective possibilities, but rather on objective factors such as lawfulness, or time and cost. However, there is a caveat to the proportionality of the concept of personal data: data must be presumed to be personal data in relation to a third party if it cannot be ruled out that the third party has means reasonably likely to be used to identify the data subject.[141]

6 CONCLUSIONS

The more recent cases clearly indicate that, contrary to what was commonly discussed in the scholarship and even maintained by prominent data protection authorities such as the European Data Protection Supervisor, pseudonymised data are not necessarily personal data. An assessment must be made, and where the person lacks reasonably likely means of identifying individuals, those data may be treated as non-personal in the particular case. Whereas one limb of the assessment focuses on a legal prohibition, the other on a case-by-case assessment mandating it being '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 guidance provided by the Court sheds light on how to reason about these assessments, but deeper inquiries, especially in the research context, are needed, with the understanding that the relevant data category is not homogeneous.

---

[138] *SRB* (n 11) para 88.
[139] *Breyer* (n 6) para 46.
[140] As seen in *OC v Commission* (n 42).
[141] *SRB* (n 11) para 85.

Overall, especially the more recent cases, notably *IAB Europe* and *SRB,* are leading in nuancing the understanding of essential concepts for the right to data protection; other cases, including *OC v Commission,* also play a considerable role. However, taken together, they raise questions that merit further inquiry, areas of uncertainty, and thereby mandate further analysis in a scientific research context. In August 2025, at the Nordic Biomedical Law Conference at Lund University, Professor Thérèse Murphy, in her keynote address, highlighted the scholarly trend toward cross-disciplinary writing, setting somewhat aside doctrinal inquiries, and recalled the value of legal scholarship. Arguably, cases like the more recent ones and the questions at their core underscore the need for deeper doctrinal inquiry, including an understanding of *de lege lata* and the actual areas of uncertainty. After all, the Court of Justice continues to signal that all the interpretations set forth in these cases are already contained in the text of the law and in prior case law, predominantly that of *Breyer*. However, this holds true only if one turns a blind eye to interpretative tweaks – analytical elements that may not be readily foreseeable in advance. This is not to argue that work across disciplines is unimportant. It is crucial for several reasons, including understanding the law in context. However, this situation suggests caution against adopting a light-touch approach to the law.

There are various, rather easily accessible repositories in the scientific research context. There is also otherwise easily available information that can facilitate identifiability, including that accessible through internet searches, as in *OC v Commission*. Assessments on whether the data in question are personal continue to be needed, now in some cases, with somewhat greater clarity, but not necessarily with all answers in hand. One way to alleviate the regulatory burden and enhance certainty for researchers is a legal measure prohibiting the re-identification of data disclosed to a particular actor in a research context. Whereas a national legal measure might be sufficient, in light of the growing critique of the GDPR and the need to revisit it,[142] for coherence and in line with the European Health Data Space Regulation, such a clause could also be accommodated within the GDPR. The underlying aim would be to preserve non-identifiability, as well as support certainty. In particular, in light of the unclear dimensions of the reasonable means to identify, a prohibition could support certainty for those researchers who work on data originating from individuals, but do not contain such details that allow them to be directly identified, and do not risk liability under the GDPR, whilst acting in good faith and in line with the established research traditions. Any steps in that direction from third countries could facilitate cross-border collaboration,[143] especially when a recipient country is not covered by an adequacy decision,[144] and help address long-standing research ethics issues, such as parachuting.

[142] See note 111 in Vera Lúcia Raposo, 'Can personal data be recycled? The reuse and repurposing of data under the EHDS' [2025] 33 International Journal of Law and Information Technology <https://academic.oup.com/ijlit/article/doi/10.1093/ijlit/eaae016/8157201> accessed 3 October 2025.
[143] Slokenberga et al (n 8).
[144] See PROMOT, 'Privacy and data protection internal guidance report' (version 1, 2024), available upon request.

LIST OF REFERENCES

Chassang G, 'The Impact of the EU General Data Protection Regulation on Scientific Research' (2017) 11 Ecancermedicalscience 709
DOI: https://doi.org/10.3332/ecancer.2017.709

Elliot M et al, 'Functional Anonymisation: Personal Data and the Data Environment' (2018) 34(2) Computer Law & Security Review 204
DOI: https://doi.org/10.1016/j.clsr.2018.02.001

Finck M and Pallas F, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10(1) International Data Privacy Law 11
DOI: https://doi.org/10.2139/ssrn.3462948

Gadotti A et al, 'Anonymization: The Imperfect Science of Using Data While Preserving Privacy' (2024) 10(29) Science Advances eadn7053
DOI: https://doi.org/10.1126/sciadv.adn7053

Irti C, 'Personal Data, Non-Personal Data, Anonymised Data, Pseudonymised Data, De-Identified Data' in Senigaglia R, Irti C, and Bernes A (eds), *Privacy and Data Protection in Software Services* (Springer Singapore 2022)
DOI: https://doi.org/10.1007/978-981-16-3049-1_5

Mourby M et al, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' (2018) 34(2) Computer Law & Security Review 222
DOI: https://doi.org/10.1016/j.clsr.2018.01.002

Purtova N, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10(1) Law, Innovation and Technology 40
DOI: https://doi.org/10.1080/17579961.2018.1452176

Raposo V L, 'Can personal data be recycled? The reuse and repurposing of data under the EHDS' [2025] 33 International Journal of Law and Information Technology <https://academic.oup.com/ijlit/article/doi/10.1093/ijlit/eaae016/8157201> accessed 3 October 2025
DOI: https://doi.org/10.1093/ijlit/eaae016

Rubinstein I and Hartzog W, 'Anonymization and Risk' (2016) 91 Washington Law Review 703

Shabani M and Marelli L, 'Re-identifiability of Genomic Data and the GDPR' (2019) 20 EMBO Rep EMBR201948316
DOI: https://doi.org/10.15252/embr.201948316

Slokenberga S et al, 'EU Data Transfer Rules and African Legal Realities: Is Data Exchange for Biobank Research Realistic?' (2019) 9(1) International Data Privacy Law 30
DOI: https://doi.org/10.1093/idpl/ipy010

Slokenberga S, 'You can't put the genie back in the bottle: on the legal and conceptual understanding of genetic privacy in the era of personal data protection in Europe' [2021] BioLaw Journal - Rivista di BioDiritto 223