Against Biometric Identifiability:

Part I: The Law of What Is Reasonably Likely

*Cesar Augusto Fontanillo Lopez**

The paper critiques the EU's bifurcated regime for biometric data, which treats this data category as either ordinary or sensitive personal data depending on identifiability considerations. It argues that this regime is unjustified because biometric data, like other personal data, is ultimately governed by the same general identifiability criteria, which already capture the full spectrum of identification relations regardless of data type. The paper offers a systematic analysis of the biometric data category, following a logical progression from broad to specific, and is divided into two parts. The first part introduces the notion of 'personal data' and traces its legal and jurisprudential development, focusing on the overarching concept of identifiability. In so doing, it seeks to build a foundation for the understanding of the main contentions of the second part. The second part then examines biometric data as 'personal data' through the lens of identifiability and then probes into biometric data as 'sensitive data' using the same analytical approach. The paper's main submissions are developed in the second part.

Keywords: biometric data; identifiability; relative approach; Breyer; Nowak; SRB

I. Introduction

Over the past decade, the processing of biometric data has become central to numerous aspects of everyday life, from accessing physical environments such as research facilities, offices, or border crossing points to daily activities like unlocking mobile devices or confirming online transactions.[1] As biometric data processing has become more prevalent, data protection law has focused on regulating its most paradigmatic uses, especially those concerning the identification and authentication of individuals.[2] When biometric data is used for these purposes, the stringent legal standards associated with the special categories of personal data (also known as 'sensitive' personal data, or simply as 'sensitive data') come into play to offer enhanced protection to individuals.[3] Conversely, in circumstances where these identification and authentication purposes are not in place, as is the case in the mere exchange of photos, audio, or video recordings, the European Union (EU) legislator does not deem it necessary to impose such stringent legal standards.[4]

DOI: 10.21552/edpl/2025/4/10

* Cesar Augusto Fontanillo Lopez, KU Leuven Centre for IT and IP Law. For helpful comments and insights on this piece, I want to thank, in alphabetical order, Bilgesu Summer, Catherine Jasserand, David Sella-Villa, Els Kindt, Gianclaudio Malgieri, Jan Czarnocki, Laura Drechsler, Orla Lynskey, and Pierre Dewitte. I would also like to thank my anonymous reviewers.

1 Els J. Kindt and Cesar Augusto Fontanillo Lopez, 'Legal Aspects of Image Morphing and Manipulation Detection Technology' in S Marcel et al (eds), *Handbook of Biometric Anti-Spoofing* (Springer Nature 2023) 547, 555 ff.

2 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 (GDPR) [2016] OJ L 119, art 4(14), recital 51; 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, PE/31/2018/REV/1, (EUDPR) [2018] OJ L 295, art 3(18), recital 29; 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/ (LED) [2016] OJ L 119, art 3(13).

3 GDPR, art 9(1); EUDPR, art 10(1); LED, art 10.

4 GDPR, recital 51; EUDPR, recital 29.

These other uses are hence governed by the 'standard' personal data processing regime,[5] so to speak, and the data processed qualifies as 'standard' personal data.[6]

The reason why the sensitive data processing regime only applies to these two 'biometric modalities'[7] is something open to debate.[8] However, it seems that a key principle guiding this regulatory choice is to prevent or minimise the risk of exploiting the identification capacity of biometric characteristics, which could lead to, albeit is not limited to, identity theft.[9] Such a risk is assumed to be particularly high when biometric characteristics are processed through identification and authentication technologies, and less so when they are processed for other purposes. This concern also underpins the legal concept of biometric data, which emphasises this precise identification interest by defining it as

> personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person.[10]

In recent years, however, technological advancements in biometrics, affective computing, and other algorithmic applications have significantly progressed, and with them, new possibilities to exploit biometric data in ways other than the traditional biometric modalities have also emerged. Consider, for instance, technologies that assign individuals to specific categories on the basis of their biometric characteristics or features, such as algorithms that use their hair colour, eye colour, or facial features to assign them to specific groups or categories, including a specific race, ethnicity, sex, philosophical or religious group, or even certain disabilities: the so-called biometric categorisation technologies.[11] Similarly, consider other algorithmic technologies that measure individuals' biometric characteristics or features, such as their facial expressions, body movements, or voice tone, and use this information to assess or predict what those individuals are, feel, or do: the so-called emotion recognition technologies.[12]

Because the evolving technological landscape increasingly allows individuals to be impacted in unforeseen ways, scholars have advocated, more and more, for updated laws that extend the protection of biometric data beyond the standard biometric modalities.[13] Such calls appear to have been acknowledged by the EU legislator, who has recently introduced important overhauls in biometric data regulation. As of 1 August 2024, the Artificial Intelligence Act (AIA)

5 GDPR, art 6(1); EUDPR, art 5(1).

6 GDPR, art 4(1); EUDPR, art 3(1). This applies as long as such processing does not involve other special categories of data. However, it should be noted that biometric recognition systems may process additional data beyond biometric data to improve accuracy, such as 'soft biometrics', which may encompass attributes like gender or height, but also sensitive data like data revealing racial or ethnic origin. Current research is therefore focused on preventing the extraction of soft biometrics to enhance individual privacy. See Blaž Bortolato et al, 'Learning Privacy-Enhancing Face Representations Through Feature Disentanglement' (2020) *15 IEEE International Conference on Automatic Face and Gesture Recognition* (FG 2020), Buenos Aires, Argentina, 495; Phillipp Terhörst et al, 'PE-MIU: A Training-Free Privacy-Enhancing Face Recognition Approach Based on Minimum Information Units' (2020) 8 IEEE Access 93635.

7 I shall refer to biometric identification and biometric authentication (also referred to as biometric verification) jointly as biometric modalities from time to time. I have adopted this term from Jasserand. Cf Catherine Jasserand, 'Biometric Data, Within and Beyond Data Protection' in Bart van Der Sloot and Sascha van Schendel (eds), *The Boundaries of Data* (Amsterdam University Press 2024) 296, 298.

8 Catherine Jasserand, 'Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data' (2016) 2(3) EDPL 297, 307 ff.

9 Hatef Otroshi and Sebastien Marcel, 'Comprehensive Vulnerability Evaluation of Face Recognition Systems to Template Inversion Attacks via 3D Face Reconstruction' (2023) 45(12) IEEE Transactions on Pattern Analysis and Machine Intelligence 14248; Hatef Otroshi and Sebastien Marcel, 'Face Reconstruction from Partially Leaked Facial Embeddings' (2024) *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, South Korea, 4930. Additional grounds for asserting the sensitivity of biometric data may include considerations of integrity, autonomy, dignity, discrimination, or procedural justice. Cf Nancy Yue Liu, *Bio-Privacy: Privacy Regulations and the Challenge of Biometrics* (Routledge 2012), 14 and 71-78.

10 GDPR, art 4(14); EUDPR, art 3(18); LED, art 3(13).

11 Vassiliki Andronikou, Angelos Yannopoulos and Theodora Varvarigou, 'Biometric Profiling: Opportunities and Risks' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen: Cross-Disciplinary Perspectives* (Springer 2008); Irena Barkane, 'Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance' (2022) 27(2) Information Polity, 147.

12 Miguel A Becerra et al, 'EEG-Based Biometric Identification and Emotion Recognition: An Overview' (2025) 14(8) Computers 299; Joseph Bullington, 'Affective Computing and Emotion Recognition Systems: The Future of Biometric Surveillance?' in *Proceedings of the 2nd Annual Conference on Information Security Curriculum Development (InfoSecCD '05)* (Association for Computing Machinery 2005) 95.

13 Jasserand (n 8), 297-311; Jan Czarnocki, 'Will new definitions of emotion recognition and biometric data hamper the objectives of the proposed AI Act?' (2021) *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany 1; Elisabeth Steindl, 'Does the European Data Protection Framework Adequately Protect our Emotions? Emotion Tech in light of the Draft AI Act and its Interplay with the GDPR, Practitioners' Corner (2022) 8(2) EDPL, 311; Bilgesu Sumer, 'The AI Act's Exclusion of Biometric Verification: Minimal Risk by Design and Default?' (2024) 10(2) EDPL, 150.

www.lexxion.eu

introduces bans on certain biometric AI systems or classifies them as high-risk depending on the risks they pose. In the riskiest cases, AI systems performing biometric categorisation and emotion recognition are prohibited.[14] In other cases, AI-powered biometric systems leading to discriminatory outcomes and adverse effects on individuals' rights are classified as high-risk.[15]

The AIA also introduces an expanded definition of biometric data. Unlike its definition under EU data protection instruments, the AIA's definition of biometric data is not based on specific identifiability considerations. In other words, while in the rhetoric of EU data protection law, the processing of information about the physical, physiological or behavioural characteristics of individuals for discriminatory purposes does not qualify as biometric data processing (as such practices do not trigger an identification interest in its strict sense), this information does qualify as biometric data under the AIA. This is so because the AIA seeks to protect wider interests and goals,[16] something that is reflected in the somewhat more expansive definition of biometric data as 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person'.[17] Yet for this same reason, the AIA definition remains misaligned with the EU data protection framework.[18]

Against this backdrop, a relevant question arises: Is it tenable to defend a biometric data regime based solely on identifiability considerations? Some may be tempted to answer this question in the negative, while others may do so otherwise. Those who support the former view, as I am inclined to do, might quickly rush to justify their stance in ways similar to those already hinted, claiming that the scope of biometric data should be enhanced to afford better protection to individuals, considering the evolving technological realities and possibilities to adversely affect them. Some other proponents might even claim that discussing this issue is unproductive or ill-advised, especially as the AIA already offers a way forward. The solution might be simple: extend the AIA's definition of biometric data to the GDPR to resolve the legal misalignment. While these arguments carry some weight, they remain highly contestable absent clear and compelling reasoning about why such an approach is justified in the first place.

This work seeks to lay such a groundwork for challenging the existing regulation of biometric data. It does so by asserting that if certain assumptions about the concept of identifiability governing this framework are accepted, their logical consequences would lead one to reject the current biometric data regime, irrespective of any party's agendas. The paper is thus not so much concerned about delving into what makes biometric data sensitive to craft a sound legal definition. Such a task would involve identifying a wide range of relevant values, including identity theft, non-discrimination, human dignity, and many others,[19] and elucidating a way for the biometric data category to address them. Rather, the purpose of this paper is less ambitious, as it merely aims to explore what the definition of biometric data would look like within a system of coherent laws that takes the concept of identifiability seriously. Yet still, my intuition is that the outcomes drawn by this bipartite piece may be a good fit to address such values.

II. Identifiability in Personal Data Processing

A logical starting point for this inquiry is to examine the concept of personal data. This is an essential step since any piece of information that qualifies as biometric data must first qualify as personal data. The definition of personal data is found in Article 4(1) GDPR and is as follows: 'any information relating to an identified or identifiable natural person.'[20]

14 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (AIA), [2024] OJ L 2024/1689, art 5(1)(f).

15 AIA, art 6(2) i.r.w. Annex III, art 1(c).

16 AIA, recital 1.

17 AIA, art 3(34).

18 The conceptual mismatch between biometric data definitions across EU instruments can lead to a weakening of effective protection for rights and freedoms. These undesirable consequences are not the main focus of this work. However, where relevant, they will be addressed in passing in the second piece.

19 See n 9.

20 I will predominantly focus my analysis on the definition of biometric data in the GDPR. However, since data protection instruments share the same definition, the findings of the following sections could be extended mutatis mutandis to the other data protection instruments, in particular, to the EUDPR and the LED.

Four main elements of analysis can be distinguished from this definition:
- a material element: 'any information';
- a referential element: 'relating to';
- a modal element: 'an identified or identifiable'; and
- a subjective element: 'natural person'.

While all of these elements must be present for information to qualify as personal data, it is undeniable that none of them has garnered more academic attention than the modal element itself, often referred to by scholars as 'identifiability' or the 'identifiability test'.[21] This test seeks to determine the 'specific quality' that data must possess for it to fall within the purview of the GDPR and instruments alike.[22] Such a specific quality, as Article 4(1) GDPR defines, is the condition that the data pertains to a natural person who is 'identified or identifiable'.

This section examines the concept of identifiability to set the stage for a deeper exploration of biometric data, which will be addressed in the second part of the paper. Although a wide range of legal sources will be covered, the analysis will primarily centre on the GDPR (and other data protection instruments, where applicable), the authoritative opinions of Article 29 Working Party (A29WP), and the case law of the Court of Justice of the European Union (CJEU).

1. Data Protection Instruments

Article 4(1) GDPR is pivotal in articulating what the 'identified or identifiable' natural person means in EU data protection law. As the provision suggests, the concept of identifiability encompasses an identification threshold comprised of two distinct limits. The upper limit represents the natural person who has been 'identified', while the lower limit represents the natural person who has not been identified yet, but who remains 'identifiable'.

A natural person can be deemed 'identified' within the upper limit, either directly or indirectly, through reference to 'identifiers' or specific 'factors' related to that individual. Similarly, a natural person can be deemed 'identifiable' within the lower limit, either directly or indirectly, through reference to such 'identifiers' or 'factors'. If the processing of information relating to a natural person falls within the boundaries defined by these two limits, the GDPR applies.[23]

Article 4(1) GDPR offers additional clarification on the terms 'identifiers' and 'factors'. On the one hand, 'identifiers' encompass pieces of information that identify the natural person, such as a name, an identification number, location data, or an online identifier. On the other hand, 'factors' include facts or situations contributing to the identification of a natural person, like information related to the physical, physiological, genetic, mental, economic, cultural or social identity of individuals. The use of identifiers, factors, or their combination delineates the range of possibilities through which a natural person can be considered identified or identifiable.

An important analytical consideration transpiring from Article 4(1) GDPR is that both identifiers and factors possess the capacity to 'directly' or 'indirectly' identify the individual. This suggests that identifiers and factors, either on their own or through their unique combination, can lead to the identification of the natural person. For instance, an identifier like a full name may 'directly' identify the individual in certain situations, as is potentially the case with the (relatively uncommon) name César Augusto Fontanillo López. Yet, in other situations, the full name alone may prove insufficient, as may be the case with more common names like James Smith or Muhammad Muhammad. In such cases, individuals might still be identifiable 'indirectly', or through a unique combination of identifiers and/or factors.

When conducting the identifiability test, it is the lower limit of the test, or the 'identifiable' natural person, that presents the most interpretative challenges. Assessing the identifiability of an individual can become a cumbersome and highly uncertain task,

21 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, 42; Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 1(1) International Data Privacy Law 11, 14; Gerard Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European General Data Protection Regulation' (2016) *7* JIPITEC 163, 165.

22 Article 29 Working Party, *Opinion 4/2007 on the Concept of Personal Data* (2007) WP 136 1, 3 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 2 May 2024. See also David Erdos, 'Identification in Personal Data: Authenticating the Meaning and Reach of Another Broad Concept in EU Data Protection Law' (2022) 46(1) Computer Law & Security Review 1, 1.

23 GDPR, art 1(1).

and processing entities often struggle to fully grasp or internalise this concept. To determine whether a natural person is 'identifiable', the EU legislator has introduced a substantive standard in Recital 26 GDPR to guide the assessment. Under this standard, account must be taken of 'all the means reasonably likely to be used by either the controller or another person'. This standard establishes the benchmark against which identifiability ought to be assessed,[24] specifying the relevant 'actors' and 'means' that any processing entity must be expected to co-opt in its legal assessment.

The term 'actors', Recital 26 GDPR suggests, refers to the 'controller or another person.' The use of the alternative conjunction 'or' (already present in Recital 26 of the former Directive 95/46)[25] has sparked academic debate on the personal scope of identifiability.[26] There is an interpretative approach suggesting that all actors who could 'possibly' identify the individual should be considered in the identifiability assessment. According to this view, if one actor qualifies information as personal data, such a qualification could also influence how other actors treat the same information. This approach, which suggests identifiability to be a transitive notion, is better known as the 'absolute approach', and taken to its extreme, it implies that information qualified as personal data by one actor will also qualify as such by any other actor.

Alternatively, there is an interpretative approach suggesting that only those actors who could 'probably' identify the individual should be considered in the identifiability assessment, and that the qualification of certain information as personal data by one actor may not necessarily affect how other actors qualify the same information. This approach, which suggests identifiability to be a notion with limited transitivity, is better known as the 'relative approach', and taken to its extreme, it implies that the qualification of information as personal data by one actor has no bearing whatsoever on how any other actor should classify the same information.

The term 'means', on the other hand, encompasses 'all' those that are reasonably likely to be used either by the controller or by another person. As Recital 26 GDPR clarifies, the assessment of such means shall consider 'all objective factors'. Objective factors include the costs and time required for identification, the available technology at the time of the processing, and technological developments. The CJEU's case law has further expanded this list to encompass legal means.[27] Importantly, unlike the ambiguity in the assessment of actors, Recital 26 GDPR provides clearer guidance on how the means should be assessed. The recital indicates that from the means that are 'reasonably likely to be used', the controller or another person must consider 'all' of them, and not only just 'some' of them, to effectively discharge their burden of proof in identifiability assessments and for fulfilling their accountability obligations under the GDPR.

As is apparent, the substantive law standard offers a much clearer framework for determining how means ought to be assimilated (through the criterion of 'reasonably likely use') than it does for the relevant actors. The latter remains loosely framed as 'either by the controller or by another person', leaving somewhat open the criterion by which those actors are to be selected.[28] Despite this, the term 'reasonably likely' remains open-ended and may be subject to interpretation. In my understanding of this standard, the term 'reasonably likely' may be taken as more affirmative than the term 'likely'. Something that is 'reasonably likely' is, in my view, 'more likely than not', and hence, it may be closer to the 'probability' than the 'possibility' spectrum.[29] Consequently, one might argue that the standard of substantive law governing the assimilation of means leans toward something akin to the relative approach, if such an approach could be somehow implemented in the assessment of means, so to speak.

Yet, as mentioned earlier, when considering the means that are 'reasonably likely' to be used, the con-

24 EUDPR, recital 16; LED, recital 21.

25 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 L 281.

26 Purtova (n 21), 46; Finck and Pallas (n 21), 17; Spindler and Schmechel (n 21), 165-168.

27 C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779, para 49.

28 The main question is whether the disjunctive operator should be read as inclusive or exclusive. An inclusive 'OR' would mean that the identifiability assessment is made from the perspective of 'the controller, another person, or both'. An exclusive 'OR' would mean that such assessment is made either 'by controller or by another person, but not both'. The traditional reading has been inclusive; however, recent CJEU's case law may point towards the opposite reading, as Section II.3 will show.

29 Consider, in this regard, what could be a potential scale of likelihood for assessing identifiability: certain > probably > likely > possibly > impossible.

troller or another person must take 'all of them'. Accordingly, the substantive law standard seems to suggest a two-step process in the assimilation of means. In the first step, the controller or another person will follow a 'probability approach', identifying which means are 'reasonably likely' to be used. In the second step, they will consider 'all of them', without being able to discard any. If this interpretative line is upheld, it may imply that one cannot regard EU data protection law as the law of everything, as has been suggested elsewhere,[30] but rather as a law that commits to cover only those questions of fact that are deemed 'reasonably likely', or within the bounds of what remains probable.

Yet still, the concept of reasonable likelihood can sometimes lead to interpretative problems, especially when some means appear to be at the borderline between what is 'reasonably likely' and what is not. In such cases, it is argued that the controller or another person should err on the side of inclusion and take those contested means into account, so as to secure 'effective', 'complete', and 'high' protection of fundamental rights and freedoms.[31] However, this submission should not be taken as collapsing the standard into a mere 'possibility approach'. As far as my opinion is concerned, that would run contrary to the law. Rather, it simply means that, in case of doubt, the disputed means ought to be included in the identifiability assessment.

2. Authoritative Opinions

Beyond what the corpus of law and its normative statements hold about it, the concept of identifiability has also been clarified by A29WP (the main EU body tasked with interpreting issues relating to privacy and data protection under the former Directive 95/46). In its 2007 opinion on the concept of personal data, A29WP asserted that an individual is considered 'identified' when, within a group of individuals, they are 'distinguished' or 'singled out' from all other members of the group.[32] Similarly, it established that an individual is deemed 'identifiable' if it is possible to identify them, even if their identification has not yet occurred.

A29WP also noted that an individual may be identified directly 'by name' or indirectly by 'unique combinations' of pieces of information (including a telephone number, a car registration number, a social security number, or by a combination of criteria allowing it to be narrowed down from the group to which it belongs).[33] By contrast, a mere hypothetical or negligible possibility to single out an individual is not enough for them to be regarded as identifiable.[34] These statements, which emphasise the importance of not overstretching the standard of identifiability to the realm of mere possibility or speculation, seem to support the idea that EU data protection law is best understood as 'the law of what is reasonably likely'.

Although A29WP's 2007 opinion on personal data represented a major advance in the legal discourse on identifiability, it was not until the release of its subsequent 2014 opinion on anonymisation techniques that a well-established test for identifiability was developed.[35] In this updated opinion, A29WP introduced three main identification criteria, the satisfaction of which would render information personal data. These comprise the ability to (i) single out a natural person, (ii) link records related to them, or (iii) infer information about them.[36] Singling out implies the 'possibility to isolate some or all records which identify an individual in the dataset';[37] linkability refers to the 'ability to link, at least, two records concerning the same data subject or a group of data subjects';[38] and inference refers to the possibility to 'deduce, with significant probability, the value of an attribute from the values of a set of other attributes.'[39]

30 Purtova (n 21). Along similar lines to this work, see Opinion AG Bobek in Case C-245/20 *X, V v Autoriteit Persoonsgegevens* [2021] ECLI:EU:C:2021:822, paras 58 ff.

31 Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* [2014] ECLI:EU:C:2014:317, para 53; C-362/14 *Maximillian Schrems v Data Protection Commissioner* [2015] ECLI:EU:C:2015:650, para 39.

32 A29WP, *Opinion 4/2007* (n 22), 12.

33 Ibid, 13.

34 Ibid, 15.

35 A29WP, *Opinion 5/2014 on Anonymisation Techniques* (2014) WP 216, 1 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed 3 May 2024.

36 Ibid, 8. In my view, the A29WP seems to suggest that these criteria are not cumulative as they aim to ensure that identification becomes 'reasonably impossible.'

37 Ibid, 11.

38 Ibid.

39 Ibid, 12.

With the inclusion of the linkability and inference criteria (which were not covered in the 2007 opinion), A29WP significantly expanded the scope of the identifiability test, resulting in what is currently understood as its standard articulation. Since the release of the 2014 opinion, these identification criteria are trusted to cover the whole identifiability spectrum and have become the cornerstone of identifiability assessments. As A29WP notes: 'Whenever a proposal does not meet one of the criteria, a thorough evaluation of the identification risks should be performed.'[40] These identification criteria are also widely employed by technical experts,[41] legal scholars,[42] data protection authorities,[43] EU agencies,[44] standardisation bodies,[45] and EU courts[46] to substantiate or validate whether a given information constitutes personal data.

3. Case Law

Over the years, the case law of the CJEU has also refined the concept of identifiability and addressed existing gaps in its interpretation, albeit not without some convulsion. During what could be considered an initial period between the 2000s and early 2010s, many judgments featured rather categorical determinations on the personal status of information. Perhaps, due to the nascent nature of the data protection framework and the limited understanding of the technical intricacies involved in identifiability assessments, the Court refrained from providing a comprehensive account of identifiability and from offering extensive justification for its findings, which often took the form of qualifying information as personal data without further elaboration. Notable examples of such rulings include the *Lindqvist*[47] and *Scarlet*[48] cases.

In the *Lindqvist* case, the CJEU examined whether the action of referencing various individuals on a webpage and identifying them by name or other means, such as their telephone numbers or information about their work and hobbies, qualifies as personal data processing.[49] While the Court did not outline the criteria necessary for holding individuals identifiable, it acknowledged that 'the name of a person in conjunction with his telephone coordinates or information about his working conditions or hobbies'[50] would undoubtedly render them so. This judgment could be seen as anticipating the criterion of linkability[51] at a time when the legal framework contained no reference to identification criteria.[52] As such, it can be regarded as a seminal case on identifiability, yet the absence of a framework for explaining how information might interact to render individuals identifiable makes the judgment remain somewhat prototypical.

Scarlet provided an early opportunity to develop such a framework, clarifying whether static IP ad-

40 Ibid, 24. See also European Data Protection Board, *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models* (2024) 1, 15, para 40 <https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf> accessed 25 July 2025.

41 Micah Altman et al, 'What a Hybrid Legal-Technical Analysis Teaches Us About Privacy Regulation: The Case of Singling Out' (2021) 2(1) Journal of Science & Technology Law 1, 19; Khaled El Emam and Cecilia Álvarez, 'A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques,' (2015) 5(1) International Data Privacy Law, 73.

42 Finck and Pallas (n 21), 16; Purtova (n 21), 44-47.

43 CNIL, *Sheet n°1: Identify personal data* (2020) <https://www.cnil.fr/en/sheet-ndeg1-identify-personal-data> accessed 3 May 2024; Irish Data Protection Commission, *Guidance Note: Guidance on Anonymisation and Pseudonymisation* (2019) 1, 5-7.

44 European Union Agency for Network and Information Security (ENISA), *Privacy by Design in Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics* (Version 1.0, December 2015); ENISA, *Pseudonymisation Techniques and Best Practices: Recommendations on Shaping Technology According to Data Protection and Privacy Provisions* (November 2019); ENISA, *Data Pseudonymisation: Advanced Techniques & Use Cases: Technical Analysis of Cybersecurity Measures in Data Protection and Privacy* (January 2021).

45 ISO/IEC 29134:2023, *Information Technology — Security Techniques — Guidelines for Privacy Impact Assessment* (2nd edn, May 2023), s 6.4.4.1 Privacy risk identification.

46 To the best of my knowledge, the case law of the CJEU has not yet issued a judgment where it explicitly identifies the three identification criteria all at once. However, these criteria have been separately recognised in various judgements. For instance, the criterion of singling out is acknowledged in *Pankki.* See Case C-579/21 *Pankki S* [2023] ECLI:EU:C:2023:501, para 44. Further, the criterion of information linkage is referenced in *Breyer*. See *Breyer* (n 27), para 20. Lastly, the criterion of inference is implied in relation with data revealing sexual orientiation in the *OT* case. See Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2022] ECLI:EU:C:2022:601, para 128.

47 Case C-101/01 *Criminal proceedings against Bodil Lindqvist* [2003] ECLI:EU:C:2003:596.

48 Case C-70/10 *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM)* [2011] ECLI:EU:C:2011:771.

49 *Lindqvist* (n 47), para 19.

50 Ibid, para 24.

51 Ibid, para 27.

52 Recital 26 Directive 95/46/EC did not explicitly refer to the concept of singling out at this time.

dresses are personal data,[53] yet its conclusions remain opaque. The case concerned the downloading and transmission of content infringing third-party intellectual property rights. The organisation managing these rights sought an injunction requiring Internet Service Providers to introduce a filtering system enabling the identification of IP addresses linked to the unlawful transmissions.[54] Here again, without engaging in a detailed justification of the identification mechanism, the Court merely stated that such IP addresses 'allow those users to be precisely identified.'[55] While the judgement also stands as a pioneering ruling on identifiability, its rather laconic reasoning left unresolved a crucial question that had been percolating alongside the advancement of the Internet and the growing capacity to target individuals online: whether an individual's identification should be understood broadly as their record singularisation, as A29WP had suggested in its 2007 opinion,[56] or more narrowly as the disclosure of their identity.

In the late 2010s, the CJEU appears to have shifted its approach to legal reasoning, moving toward more meticulous and analytically grounded judgements. This change coincided with a broader and more sophisticated discourse on the technicalities of identifiability, shaped in part by the growing academic research and the influential A29WP's 2007 and 2014 opinions. During this period, the Court's treatment of identifiability became increasingly nuanced, reflecting a deeper engagement with identification mechanisms, such as direct and indirect forms of identification. This shift was particularly evident in cases like *Breyer*[57] and *Nowak*,[58] where the Court established a structured framework for assessing identifiability across datasets, considerably extending the notion of personal data.

In the *Breyer* case, the CJEU expanded its analysis from *Scarlet* to dynamic IP addresses, which change each time there is a new internet connection.[59] The Court clarified that the notion of 'identification' concerns the revelation of an individual's identity, thereby addressing the ambiguity left unresolved in the earlier judgment.[60] It ruled that dynamic IP addresses, on their own, do not permit such identification because they do not 'directly' reveal the identity of individuals accessing websites.[61] This represented a departure from A29WP's more protective stance,[62] emphasising that the personal nature of data is determined not by the possibility of isolating a user's records, but by the ability to reveal their 'identity'. Nonetheless, the Court stressed that IP addresses could still be considered personal data if combined with additional information held by third parties.[63] In practice, this meant that 'incomplete' datasets could still acquire personal data status where a pathway exists for identifying individuals across different parties.[64] As several scholars have observed,[65] *Breyer* expanded the scope of data protection and implied an endorsement of the absolute approach, since information capable of identifying an individual no longer needed to be in the hands of a single entity to qualify as personal data.

But was it possible to construct a theoretical framework capable of delineating the conditions under which information becomes personal data? The answer to this question came one year later, in the *Nowak* case. The central question of this judgment concerned whether the written answers submitted by a candidate during a professional examination and the examiner's comments were identifiable data. The Court endorsed A29WP's 2007 opinion,[66] establishing two cumulative conditions for the data to be considered personal, namely that (i) it must relate to the individual in purpose, content, or effect,[67] and (ii) that the individual must be identified or identifiable.[68] The judgment also expanded the concept of

53 *Scarlet* (n 48), para 51.

54 Ibid.

55 Ibid.

56 A29WP, *Opinion 4/2007* (n 22), 14.

57 *Breyer* (n 27).

58 Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECLI:EU:C:2017:994.

59 *Breyer* (n 27), para 16.

60 Ibid, paras 19-25, and 38.

61 Ibid, para 38.

62 A29WP, *Opinion 4/2007* (n 22), 14; A29WP, *Opinion 5/2014* (n 35), 11.

63 *Breyer* (n 27), para 43.

64 Ibid, paras 47 and 49; Opinion AG Sanchez-Bordona in Case C-582/14 *Breyer* [2016] ECLI:EU:C:2016:339, para 68.

65 Finck and Pallas (n 21), 18; Purtova (n 21), 62 and 69-72; Frederik Zuiderveen Borgesius, 'Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition' (2017) 3(1) EDPL 130.

66 A29WP, *Opinion 4/2007* (n 22), 10, which says that for information 'to relate to an individual, a 'content' element OR a 'purpose' element OR a 'result' element should be present'.

67 *Nowak* (n 58), para 35.

68 Ibid, para 45.

personal data, confirming that this notion encompasses not only objective but also subjective information, such as opinions and assessments.[69] In the case at stake, both the candidate's answers and the examiner's assessments about those answers were information 'related' to the candidate that could 'identify' them directly by name or indirectly via an identification number or specific factors related to their identity.[70] This is why the Court concluded that both conditions were satisfied for the information to constitute personal data. Interestingly, *Nowak* maintained the connection between identification and the revelation of an individual's identity,[71] reinforcing *Breyer.*

Finally, entering the 2020s, the concept of personal data faced numerous attacks, which were aimed at testing its boundaries. While many judgments in this decade continue to reflect the broader interpretations on identification criteria developed in the 2010s,[72] three recent judgements stand out for the new insights they provide into the notion of identifiability. *OT* is particularly noteworthy for engaging with inferences, which appear far less frequently in case law than the more established criteria of singling out or linkability.[73] Equally significant are *SRB*[74] and *OC*[75] for their engagement with the theoretical framework set out in *Nowak,* revealing how divergent judicial interpretations may arise around the 'reasonable likelihood' standard. This tension, centred on how strictly or permissively the standard should be read, appears to have been recently resolved by the CJEU, which, by overruling some aspects of its prior case law in favour of the General Court's (GC) view, has reinforced the relative approach to personal data.

In *OT,* the Court had the opportunity to consider, albeit only in passing, the inference criterion. The case concerned a Lithuanian law that required recipients of public funds to declare their private interests, a measure to ensure impartial decision-making and prevent corruption. Such declarations included personal data like names, financial data, and information about spouses or partners. The key question was whether publishing this personal data could allow observers to infer sensitive information about the sexual orientation of individuals and hence constitute sensitive personal data processing. The Court confirmed that the potential to draw such sensitive inferences from personal data could indeed bring that data within the scope of the sensitive personal data processing regime.[76] It defined inference narrowly, as 'an intellectual operation involving comparison or deduction'.[77] Yet, the ruling left important questions open, namely, whether inductive and abductive reasoning fall within the remit of inferences (and if so, which standard of proof applies to them), and whether inferences can bring non-personal data within the scope of personal data just as they bring personal data within the scope of sensitive data.[78]

SRB, on its side, can be seen as a corrective to the broad interpretation of personal data in *Breyer* and *Nowak*. The case concerned whether individuals'

69 Ibid, para 34.

70 Ibid, paras 28–62.

71 Ibid, para 31.

72 Case C-319/22, *Gesamtverband Autoteile-Handel e.V. v Scania CV AB* [2023] ECLI:EU:C:2023:837, paras 45-49; Opinion of AG Campos Sanchez-Bordona in Case C-319/22 *Scania* [2023] ECLI:EU:C:2023:385, paras 34 and 39; Case C-683/21 *Nacionalinis visuomenės sveikatos centras prie Sveikatos apsaugos ministerijos v Valstybinė duomenų apsaugos inspekcija* [2023] ECLI:EU:C:2023:949, paras 54-56; Opinion AG Emiliou in Case C-683/21 *Nacionalinis* [2023] ECLI:EU:C:2023:376, para 52; Case C-487/21 *Österreichische Datenschutzbehörde and CRIF* [2023] ECLI:EU:C:2023:369, paras 22-26; Case C-604/22 *IAB Europe v Gegevensbeschermingsautoriteit* [2024] ECLI:EU:C:2024:214, paras 36-51. Importantly, although the singling out criterion has been acknowledged in several judgements, the Court has not provided a precise definition of what constitutes singling out. See *Pankki* (n 46), paras 44-45. However, *IAB* offers a detailed discussion on the underlying identification mechanism used to determine when information qualifies as personal data. See *IAB*, paras 41-47.

73 *OT* (n 46).

74 Case T-557/20 *SRB v EDPS* [2023] ECLI:EU:T:2023:219.

75 Case T-384/20 *OC v European Commission* [2022] ECLI:EU:T:2022:273.

76 *OT* (n 46), para 128.

77 Ibid, para 120.

78 If answered in the affirmative, this could suggest that the inference criterion developed by the A29WP may hold an independent and mutually reinforcing status in EU data protection law, standing alongside the notion of singling out, particularly if the latter is interpreted amply, as A29WP seems to suggest in its 2007 and 2014 opinions. Cf A29WP, *Opinion 4/2007* (n 22), 14-15; A29WP, *Opinion 5/2014* (n 35), 11. However, it could also be argued that such an interpretation may risk committing a fallacy of composition, as what is assumed to be true of a subset is vindicated as being equally true of the set, which may not necessarily be the case. In other words, what might be true from sensitive data as a subcategory of personal data (ie, the possibility of inferring sensitive data from personal data) might not be true for the category of personal data as a subcategory of data more generally (eg, the possibility of inferring personal data from non-personal data under similar conditions). This distinction becomes especially significant when comparing the notions of 'revealing' and 'concerning' that underpin specific types of sensitive data, with the notion of 'identified or identifiable' that underpins the definition of personal data. These notions might not be symmetrical and might be subject to different standards of proof. Although a full engagement with this problem lies beyond the scope of this paper, it remains an issue worthy of further exploration or judicial clarification. Cf Opinion of AG Pikamäe in Case C-184/20 *OT v Vyriausioji tarnybinės etikos komisija* [2021] ECLI:EU:C:2021:991, para 85.

comments analysed by a data processor, linked to alphanumeric codes, qualify as personal data, even though the data processor lacked access to the decoding database, which was held exclusively by the data controller. Drawing on *Nowak*,[79] the European Data Protection Supervisor (EDPS) argued that the comments 'related' to individuals based on their effects and thus were personal data,[80] while the data controller rejected this understanding.[81] The EDPS also argued that the data remained 'identifiable' for the processor because they were pseudonymised,[82] whereas the data controller maintained that the data were effectively anonymous to the processor.[83] In the end, the GC sided with the data controller, noting that the EDPS had neither assessed the 'relating to' element to the satisfaction of *Nowak*'s test nor 'identifiability' from the data processor's perspective.[84] *SRB* highlights the importance of satisfying the evidentiary standard for personal data to the court's conviction and implies a strong affirmation of the relative approach to personal data, underlining that the determination of personal data status is a context-dependent and actor-specific exercise.

On appeal, however, the CJEU overturned the GC's judgment. Following *Nowak*, it introduced a factual presumption regarding the connection of personal opinions or views to individuals.[85] This meant that the EDPS did not need to assess the 'relatedness' of such opinions and views to determine that they were indeed 'related' to individuals, thereby rejecting the GC's earlier finding. Nevertheless, the CJEU still upheld what remains arguably the GC's most significant contribution to the relative approach: it confirmed that the GC was indeed correct in concluding that pseudonymised data cannot under all circumstances be considered personal data, regardless of the entity involved.[86] It clarified that personal data should be assessed from the 'relevant perspective', which depends on the specific processing context.[87] Since, in the case at hand, the processing of the individual's comments formed part of a hearing procedure aimed at determining compensation and concerned the relationship between data subjects and the controller,[88] the 'relevant perspective' was that of the controller rather than the processor. While this finding led the Court to set aside the GC's judgment, its endorsement of personal data as a relative notion settled the long-standing debate between the absolute and relative approaches. Moreover, by stressing the importance of processing purposes, the ruling embeds this relativity within a broader teleological framework where purposiveness determines personal data status.

Lastly, the *O.C.* case is noteworthy for it prompted the GC to introduce a specific articulation of the identifiability test in the context of online information: the so-called 'average reader' test. This test imagines whether a regular internet user, using reasonable means (such as online searches), could identify an individual within a reasonable amount of time.[89] The test arose in a case concerning a press release that, although not identifying a researcher involved in an anti-money laundering procedure, allowed a journalist to do so. Like *SRB*, the significance of *O.C.* lies in clarifying the evidentiary threshold that a party must meet to qualify a dataset as personal data to the conviction of the court. This threshold, the GC ruled, demanded the researcher be identified by an average reader 'easily or quickly'[90] and with a 'simple objective reading'[91] of the press release. In the case at stake, however, a minor effort could not lead to the researcher's reidentification, so the Court concluded that the information did not constitute personal data. Importantly, the GC limited the evaluation to the content of the press release itself, excluding external sources,[92] and ruling out investigative journalists from the average reader test.[93] This approach nar-

79 *SRB* (n 74), paras 61-63.

80 Yet the EDPS itself had not conducted such an examination. Cf *SRB* (n 74), paras 73-74.

81 Ibid, para 60.

82 Ibid, para 83. Yet the EDPS failed to assess identifiability from the processor's perspective. Ibid, paras 103-105.

83 Ibid, para 77.

84 *SRB* (n 74), paras 73-74 and 100-106.

85 *Nowak* (n 58), paras 42-44; Case C-413/23 P *EDPS v SRB* [2025] ECLI:EU:C:2025:645, paras 57-61.

86 *SRB* [Appeal] (n 85), para 86.

87 Ibid, paras 99-100.

88 Ibid, paras 106-116.

89 *OC* (n 75), paras 61 and 76.

90 *Ibid, para* 67. This finding could be seen, in a way, as a legacy of the Explanatory Report to the Convention 108 of the Council of Europe, which stipulated that an '"identifiable person" means a person who can be easily identified: it does not cover identification of persons by means of very sophisticated methods'. Cf Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, *European Treaty Series* - No. 108, Strasbourg, 28.I.1981, art 2, litt a, para 28.

91 *OC* (n 75), para 68.

92 Ibid, para 49.

93 Ibid, paras 81-82, and 87.

rowed the scope of personal data and, to some extent, can be seen as a reaction to the broad interpretation of personal data in *Breyer*, implying that identifiability is assessed with respect to specific information and the perspective and capabilities of an average audience.

On appeal,[94] however, the CJEU reversed some of the GC's findings, confirming the broad scope of personal data again. It held that personal data cannot be restricted to specific pieces of information, as indirect identifications also fall within this concept,[95] and that professional journalists must be considered within the average reader test.[96] In reviewing the assessment of the evidence, the CJEU adopted a more permissive approach to the reasonable likelihood standard. Unlike the GC, it acknowledged that the identifiers in the press release could indeed allow an average reader to reidentify the researcher,[97] even though the researcher's identity had not been disclosed.[98] At the same time, the CJEU emphasised that a successful reidentification alone does not automatically render the information personal data; instead, the reidentification must remain reasonably likely.[99] In other words, the Court made clear that personal data status does not arise from any conceivable instance of reidentification; contrarily, cases involving disproportionate effort would escape legal scrutiny.[100] But more crucially, it noted that an individual need not adduce evidence that they were actually 'identified', since identifiability only requires them to be 'identifiable'.[101] This effectively lowered the evidentiary threshold for qualifying information as personal data, aligning with the preventive rationale of data protection law, which seeks to avert potential harm beyond addressing harm that has already occurred.

III. Interim Conclusions

The story of identifiability in EU data protection law is one of evolution, tension, and gradual clarification. Assessing whether information can identify an individual is both a complex and disputed task. Even with guidance from Recital 26 of the GDPR and the opinions of A29WP, uncertainty persists regarding how this test should be applied. At the heart of this uncertainty lies the longstanding debate between absolute and relative approaches, a debate that has left EU Courts oscillating for years without taking a clear stance until recently. While early cases, such as *Lindqvist* and *Scarlet*, did not clarify much, later decisions like *Breyer* and *Nowak* have been interpreted as significantly broadening the concept of personal data. In recent years, however, a series of cases have emerged which have severely tested the CJEU's willingness to sustain this expansive interpretation, prompting the Court to move toward a more flexible position.

Amid this progression, the activism of the GC has been crucial in keeping the law grounded. In *SRB,* the GC pushed back against the broad interpretation of personal data set out by the CJEU in *Breyer*, recognising that information may not necessarily carry the same qualification for all parties.[102] This created an

94 Case C-479/22 P *OC v European Commission* [2024] ECLI:ECLI:EU:C:2024:215.

95 Ibid, paras 52-55 and 64.

96 Ibid, para 57.

97 Ibid, paras 60-66. Importantly, the CJEU, following *Breyer*, reaffirmed what cannot be considered a reasonably likely identification, including that which may not happen due to a legal prohibition or practical impossibility to reidentify an individual (on account of disproportionate effort in terms of time, cost, and labour). In such cases, the risk of reidentification would be 'trivial or insignificant' in reality. Ibid, para 51; *Breyer* (n 27), para 46.

98 *OC* [Appeal] (n 94), para 75.

99 Ibid, para 58.

100 *OC* [Appeal] (n 94), paras 60-66. Importantly, the CJEU, following *Breyer*, reaffirmed what cannot be considered a reasonably likely identification, including that which may not happen due to a legal prohibition or practical impossibility to reidentify an individual (on account of disproportionate effort in terms of time, cost, and labour). In such cases, the risk of reidentification would be 'trivial or insignificant' in reality. *Ibid, para* 51; *Breyer (n* 27, para. 46.
In addition, it is worth mentioning that the Commission argued in the appeal that a person cannot be considered reasonably likely identifiable under the 'hypothetical possibility [they] may be distinguished,' a contention which aligns with A29WP's 2007 opinion, as discussed earlier, and which does not seem to be challenged by the CJEU. Thus, this suggests that a 'possibility' approach towards identifiability is not in line with the concept of personal data. Ibid, para 35; see also A29WP, *Opinion 4/2007* (n 22), 18.

101 Ibid, paras 35-36 and 60-61. Another key finding is that the identifiability test is an independent test that must be evaluated vis-à-vis data protection law and cannot be conflated with non-contractual liability provisions of Union institutions and bodies (ibid, paras 52-54). This last point is crucial as it highlights an interpretational divergency between the GC and the CJEU in relation to fundamental considerations of the right to the protection of personal data. While *OC* revealed a pragmatic GC, willing to strike 'a just equilibrium' between data protection objectives and EU interests to ensure that data protection law remains effective but practical, the CJEU was more reluctant about conflating or balancing them, maintaining a strict separation between legal regimes to avoid weighing the identifiability test against EU non-contractual liability provisions. Cf *OC* (n 75), para 46.

102 *SRB* (n 74), para 97.

interpretative tension between the courts, which the CJEU ultimately resolved by affirming the relative nature of personal data in *SRB* and *OC*. These later pronouncements reflect a gradual balancing between the teleological aims of data protection law and the need for a context and actor-sensitive understanding of personal data. The opinion of AG Spielmann in *SRB*'s appeal is particularly noteworthy, as it underscores the idea that adopting a relative approach to personal data is not inconsistent with the EU's overarching goal of ensuring a high level of rights protection,[103] a principle the CJEU has consistently upheld.[104]

The jurisprudential evolution of identifiability has been shaped not only by the court's conceptual framework but also by how evidence is evaluated. A pivotal factor has been the interpretation of the substantive law standard of 'reasonable likelihood'. The GC, applying stricter scrutiny, linked this standard to the notion of trivial or insignificant reidentification efforts.[105] This effectively lowered the amount of evidence that the Court required for holding data identifiable, since it was sufficient to show that identification could occur easily and quickly. This finding also narrowed the scope of what counts as personal data, as more onerous re-identification efforts would escape legal scrutiny, even if technically possible. The practical consequence was that less data could be classified as personal. In contrast, the CJEU applied a more lenient standard of review, reassessing the classification of facts and holding that the reidentification efforts regarded as non-trivial by the GC were, in fact, reasonably likely.[106] Coupled with the possibility of using external sources of information to reidentify the researcher and the inclusion of professional journalists in the average reader test (who had been excluded by the GC), the CJEU broadened the scope of admissible evidence a party could adduce to establish identifiability. As a result, the scope of personal data expanded once again, as a wider array of reidentification means can now satisfy the legal standard beyond mere triviality.

Alongside this jurisprudential evolution, A29WP's clarification of identification criteria represents a major step forward in the conceptualisation of identifiability. By clarifying how information can make individuals identifiable through information isolations, linkages, inferences, or a combination thereof, A29WP has helped create a comprehensive framework that has significantly broadened the scope of personal data. This framework, which has been gradually developed in its 2007 and 2014 opinions, has effectively become the gold standard for identifiability assessments, offering a benchmark for actors to validate their qualifications.

The singling out criterion, introduced by A29WP in its 2007 opinion,[107] did not feature in early CJEU judgments (such as *Lindqvist*, *Scarlett Extended*, *Breyer,* and *Nowak*), which often associated the term 'identification' with the revelation of an individual's 'identity'.[108] Yet at the time, Directive 95/46/EC did not explicitly mention this identification criterion either.[109] In the post-Directive era, however, as data-driven technologies advanced and the ability to target individuals grew, it became clear that the concept of identification needed refinement. This prompted A29WP to introduce the singling out criterion, noting that computers 'no longer necessarily require the disclosure of [an individual's] identity in the narrow sense' to attribute certain decisions to them.[110] This position, further developed in its 2014 opinion,[111] ultimately influenced the inclusion of this criterion in the GDPR and instruments alike. [112] Since then, the CJEU case law has reflected it,[113] and although some rulings continue equating the terms 'identification' and 'identity',[114] recent decisions such as *OC* draw a distinction between these terms, showing that an individual can be 'identified' without their 'identity' being disclosed.[115] This approach also brings the CJEU closer to A29WP's understanding of identifiability.

---

103 Opinion of AG Spielmann in Case C-413/23 P *SRB* [2025] ECLI:EU:C:2025:59, paras 51-60.

104 *Lindqvist* (n 47), paras 96-96; *Nacionalinis* (n 72), paras 29 and 77; *IAB* (n 72), para 53; *Pankki* (n 46), para 55; *OT* (n 46), paras 61 and 125; *Österreichische Datenschutzbehörde* (n 72), para 40. This is, however, not without limits, as this provision requires the data subject be identified or identifiable. See *SRB* (n 74), paras 66 and 88.

105 *OC* (n 75), paras 65-68;*OC* [Appeal] (n 94), paras 37, 61-87. This is an outcome that follows from *Breyer*. See *Breyer* (n 27), para 46.

106 *OC* [Appeal] (n 94), paras 60-66.

107 A29WP, *Opinion 4/2007* (n 22), 13-15.

108 *Breyer* (n 27), paras 19-21, 24-25; *Nowak* (n 58), para 31; but see also later judgements, eg, *OC* (n 75), para 58.

109 Directive 95/46/EC, recital 26 merely stated: 'to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by other person to identify the said person'.

110 A29WP, *Opinion 4/2007* (n 22), 14.

111 A29WP, *Opinion 5/2014* (n 35), 11.

112 GDPR, recital 26; EUDPR, recital 16; LED, recital 21.

113 *IAB* (n 72), para 40; *Pankki* (n 46), para 44; *OT* (n 46), para 17.

114 *SRB* (n 74), paras 18, 60, 82, 99; *SRB* [Appeal] (n 85), para 72.

115 *OC* [Appeal] (n 94), para 75.

The linking criterion has also played a central role in shaping identifiability. Tracing back *Lindqvist*,[116] linkability remains one of the most frequently cited identification criteria in EU case law.[117] Importantly, *Breyer* clarified that personal data need not be in the hands of a single processing entity,[118] broadening the law's scope to account for any reasonably likely linkage of information by any actor. This principle was reinforced by *OC*, which held that external sources beyond the information contained in a press release could be used to reidentify a data subject.[119] While this approach drew a wide-ranging, almost adversarial, landscape for data processing entities seeking to classify information as non-personal data, other landmark judgments, such as *SRB*, have sought to refine it, introducing the concept of 'relevant perspectives' that embed the legal nature of personal data within a purposive, context-sensitive framework.[120]

Finally, the inference criterion has begun to feature in EU case law, though less prominently than singling out or linkability. Prior to *OT*, A29WP had made clear that inferences imply 'the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes'.[121] Building on this guidance, *OT* emphasises that inferential processes rely on deductive reasoning and comparisons, yet it does not impose strict requirements regarding the accuracy of such inferences. Despite this, the Court's reference to 'deduction'[122] suggests a connection to truthfulness,[123] while it remains unclear whether the mention of 'comparisons'[124] allows room for inductive and abductive reasoning. Although data protection law does not assume a conception of information that is bound to truth or falsehood, recognising that even misinformation may impact individuals' rights,[125] A29WP and *OT* suggest that accuracy may still play a role. Crucially, while *OT* does not treat inferences as a standalone identification criterion capable of bringing information within the scope of personal data, it shows that inferences can nonetheless bring personal data within the scope of sensitive data.

Taken together, these developments suggest that the concept of identifiability is both central and increasingly nuanced in EU data protection law. The evolution of the identifiability test, shaped by legal instruments, institutional opinions, and the CJEU case law, indicates that personal data should be understood as a relative notion, assessed in context and from the relevant perspective, rather than as an absolute category. The interplay between probability and possibility and the distinctions between the identifiability criteria of singling out, linkability, and inference collectively establish a structured but evolving framework for determining when information qualifies as personal data. This implies that the law of personal data does not function as a law of everything, but as a calibrated system that targets risks that are 'reasonably likely' to occur, balancing the protection of rights and freedoms with practical constraints.

The second piece of this article, to be published in the next issue, will apply these insights to the regulation of biometric data. In that piece, the identification criteria elaborated by A29WP will serve as the analytical foundation to probe into the categories of biometric data, both as personal data and as sensitive personal data. By treating the concept of identifiability, as developed here, as a constant, the analysis will interrogate the so-called 'biometric identifiability test', assessing whether the dual-layer identifiability framework built around biometric data can still withstand scrutiny under the general principles and guidelines governing personal data. This allows for a refined evaluation of biometric data or, at least, one grounded in established identifiability standards rather than assumptions specific to biometric modalities.

116 *Lindqvist* (n 47), paras 18-19, 24.

117 *Breyer* (n 27), para 21; *Pankki* (n 46), para 43; *Scania* (n 72), paras 45 and 49; *Nacionalinis* (n 72), para 58; *Österreichische Datenschutzbehörde* (n 72), para 24; *IAB* (n 72), paras 25, 29, 37; *SRB* (n 74), paras 24, 69, 94; *OC* (n 75), paras 45, 56; *SRB* [Appeal] (n 85), paras 28, 49, 55-58.

118 *Breyer* (n 27), para 43.

119 *OC* [Appeal] (n 94), para 64.

120 *SRB* [Appeal] (n 85), paras 99-100; *OC* (n 73), paras 61 and 76; *OC* [Appeal] (n 94), para 57.

121 A29WP, *Opinion 5/2014* (n 22), 12.

122 *OT* (n 46), para 120.

123 Indeed, valid and sound deductive arguments guarantee true conclusions, yet it should be noted that deductively valid conclusions may still be false if the premises are false.

124 *OT* (n 46), para 120.

125 See the relevant EU case law affirming that personal data can be also subjective and comprise personal opinions, eg, *Nowak* (n 58), paras 34, 53-56; *Österreichische Datenschutzbehörde* (n 72), paras 23-24; *SRB* (n 74), paras 68 and 73; *SRB* [Appeal] (n 85), paras 49; *OC* [Appeal] (n 94), para 45; Case C-200/23 *Agentsia po vpisvaniyata* [2024] ECLI:EU:C:2024:827, para 130; see, to the same extent, A29WP's pronouncement on the truth value of personal data in A29WP, *Opinion 4/2007* (n 22), 6; see also the principle of accuracy and the rights to rectification, restriction or objection, which reject the veridicality thesis. Cf GDPR, arts 5(1)(d), 16, 18, and 21.