European Journal of Risk Regulation (2025), 1–19
doi:10.1017/err.2025.16

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Identifiability, as a Data Risk: Is a Uniform Approach to Anonymisation About to Emerge in the EU?

Sophie Stalla-Bourdillon

VUB, Brussels, Belgium
Email: sophie.stalla-bourdillon@vub.be

Abstract

The concept of identifiability remains a foundational yet contentious criterion in European Union (EU) data protection law. Similarly, anonymisation has sparked intense debate.

This paper examines recent developments that have shaped the EU's approaches to identifiability and anonymisation, including trends in the Court of Justice of the European Union (CJEU) case law, national supervisory authority (SA) assessments of anonymisation processes, and the recent European Data Protection Board (EDPB) Opinion 28/2024 addressing the anonymity of artificial intelligence models and EDPB Guidelines 01/2025 on pseudonymisation.

The paper explores how the balance between over-inclusiveness and under-inclusiveness is being calibrated, suggesting the emergence of a functional definition of personal data in CJEU case law. It underscores the importance of the burden of proof in evaluating anonymisation processes, as confirmed by national SA assessments. Finally, it highlights how to ensure consistency between the GDPR and data sharing mandates stemming from the new generation of EU data regulations.

Keywords: anonymisation; data protection; data sharing; GDPR; Identifiability

I. Introduction

The concept of *identifiability* as a foundational criterion in European Union (EU) data protection law has attracted considerable criticism. Some scholars argue that the scope of identifiability is, on the one hand, too broad, making data protection law the "law of everything."[1] In this view, data protection law risks becoming an overreaching framework that applies to virtually all forms of data, diluting its effectiveness and focus. On the other hand, some argue that the criterion is too narrow, leaving certain data practices, such as some forms of profiling, beyond the law's reach.[2] Furthermore, some critique identifiability as a poor proxy for harm, suggesting that a focus on harm rather than

[1] See, eg, N Purtova, "The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law" (2018) 10 Law, Innovation and Technology 40. Purtova's claims should be read in the light of a more recent contribution N Purtova, "From Knowing by Name to Targeting: The Meaning of Identification under the GDPR" (2022) 12 International Data Privacy Law 163.

[2] See eg, S Wachter, "Data Protection in the Age of Big Data" (2019) 2 Nature Electronics 6; W Schreurs and Others, "Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector" in M Hildebrandt and S Gutwirth (eds), *Profiling the European Citizen: Cross-Disciplinary Perspectives* (Springer Netherlands 2008) <https://doi.org/10.1007/978-1-4020-6914-7_13>.

identifiability would better address the underlying concerns privacy or data protection laws attempt to address.[3]

With its third generation of data laws,[4] the EU confirms that identifiability remains a key risk factor for data sharing and reuse and, in an attempt to increase the level of data sharing and reuse within the EU data economy, promote anonymisation processes in different ways.[5] While data risks take various forms, identifiability is often one of the most talked about in practice.

Anonymisation, as a related EU data protection law concept, has thus been debated for many years, both under the Data Protection Directive (DPD)[6] and the General Data Protection Regulation (GDPR).[7] While national supervisory authorities (SAs) appear to converge in rejecting inadequate anonymisation practices,[8] determining when and if the anonymisation threshold is met remains a challenging task, which led a fair amount of consultants and practitioners taking the view that anonymisation under EU law can only be achieved when at least two conditions are met: the data is aggregated and the raw data, ie, the individual-level data, has been destroyed.

It is therefore not surprising that the European Data Protection Board (EDPB) has been delaying the release of its guidelines on anonymisation – intended to update its 2014 opinion – for several years.[9]

Despite this delay, several recent trends are worth investigating, as they have direct implications for the EU approach to anonymisation. Recent case law of the Court of Justice of the European Union (CJEU) suggests that the concept of identifiability is gradually maturing, much like the assessment of anonymisation processes by national SAs. EDPB Opinion 28/2024, issued on 17 December 2024 and addressing implications of personal data processing in the context of Artificial Intelligence (AI) model training and deployment, also has direct consequences for the anonymisation of personal data,[10] which now seems to include AI model information when certain conditions are met.[11] In addition, EDPB

[3] See eg, DJ Solove, "Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data" (2024) 118 Northwestern University Law Review 1081.

[4] Assuming the first generation started with the Data Protection Directive.

[5] See Section IV for references to various regulations.

[6] 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, OJ L 281, 23.11.1995, pp 31–50.

[7] See eg, S Stalla-Bourdillon and A Knight, "Anonymous Data v. Personal Data – A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data" (2017) 34 Wisconsin International Law Journal 284; M Finck and F Pallas, "They Who Must Not Be Identified – Distinguishing Personal from Non-Personal Data under the GDPR" (2020) 10 International Data Privacy Law 11; Purtova (n 1); K El Emam and C Álvarez, "A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques" (2015) 5 International Data Privacy Law 73.

[8] See eg, *Taxa 4x35* [2019] Datatilsynet n°2018-41-0016; *Doctissimo* [2023] CNIL n°SAN-2023-006; *Finanstilsynet* [2022] Datatilsynet n°2020-442-8099; *Sean Serios SL* [2021] AEPD n°PS-00520-2021.

[9] See EDPB, "EDPB Work Programme 2024/2025" (2024) <https://www.edpb.europa.eu/system/files/2024-10/edpb_work_programme_2024-2025_en.pdf> accessed 6 February 2025; EDPB, "EDPB Work Programme 2021/2022" (2021) <https://www.edpb.europa.eu/system/files/2021-03/edpb_workprogramme_2021-2022_en.pdf> accessed 6 February 2025; EDPB, "EDPB Work Programme 2023/2024" (2023) <https://www.edpb.europa.eu/system/files/2023-02/edpb_work_programme_2023-2024_en.pdf> accessed 6 February 2025.

[10] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (2024) Opinion of the Board (Art 64) <https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf> accessed 6 February 2025.

[11] Ibid 29, 37.

Guidelines 01/2025 on pseudonymisation, while not directly tackling anonymisation practices, include important considerations related to risk assessment.[12]

By exploring these trends, this paper aims to highlight how the EU approach to anonymisation has been evolving in the last few years. For this purpose, it unpacks the balance between over-inclusiveness and under-inclusiveness in CJEU case law and suggests that a functional definition of personal data is emerging, while making recommendations for refining the legal analysis and preserving the preventive goal pursued by EU data protection law.[13] This paper then stresses the importance of the burden of proof for assessing anonymisation processes and identifies ways to ensure consistency between the GDPR and newer data regulations.

This paper is structured as follows. Section II gives an overview of the identifiability test derived from the CJEU case law. Section III draws some lessons from national SA assessments of anonymisation processes and EDPB guidance. Section IV assesses the implications of the identifiability test for data sharing obligations stemming from recent data regulations. Section V concludes.

II. The CJEU's approach to identifiability

The CJEU's approach to identifiability is characterised by a focus upon anticipated uses of the data and does not expressly stress the importance of considering a wide range of stakeholders, including unintended recipients.

1. A focus upon anticipated uses

a. Breyer

For quite some time, the Breyer[14] case has been the main CJEU case touching upon the concept of identifiability. Let's recall the facts, as they are important to shed light upon the effects of the CJEU's holding. In Breyer, publicly accessible websites operated by German Federal institutions stored traffic data "on all access operations in logfiles."[15] More specifically,

> "[t]he information retained in the logfiles after those sites have been accessed include the name of the web page or file to which access was sought, the terms entered in the search fields, the time of access, the quantity of data transferred, an indication of whether access was successful, and the IP address of the computer from which access was sought."[16]

This browsing information was retained for cybersecurity purposes, ie, for "preventing attacks and making it possible to prosecute 'pirates.'"[17]

The first question posed to the CJEU in this case was whether "an internet protocol address (IP address) which an [online media] service provider stores when his website is accessed already constitutes personal data for the service provider if a third party (an access provider) has the additional knowledge required in order to identify the data

[12] EDPB, "Guidelines 01/2025 on Pseudonymisation" (2025) <https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en> accessed 6 February 2025.

[13] Data protection law is designed to avert risks before harm occurs, recognising that compensation for harm caused by data misuse is often impractical or inadequate. Once data is misused, the harm can be difficult, if not impossible, to remedy fully, although the right to receive compensation remains a foundational remedy.

[14] *Patrick Breyer v Bundesrepublik Deutschland* (2016) ECLI:EU:C:2016:779.

[15] Ibid, 14.

[16] Ibid.

[17] Ibid.

subject"[18] under Article 2(a) of the Data Protection Directive. Said otherwise, the German Federal Court of Justice was essentially asking whether a (dynamic) IP address[19] should be considered personal data when it is within the hands of an online service provider that is not an internet access service provider, ie, a provider that is able to combine IP addresses with subscriber information.

In Breyer, the CJEU easily admits that a dynamic IP address does not constitute data relating to an identified natural person. It states that

> "a dynamic IP address does not constitute information relating to an 'identified natural person', since such an address does not directly reveal the identity of the natural person who owns the computer from which a website was accessed, or that of another person who might use that computer."[20]

The more difficult question to answer is whether a dynamic IP address constitutes data relating to an identifiable natural person.

The CJEU refers to Recital 26 of the Data Protection Directive and rightly states that "it is not necessary that that information alone allows the data subject to be identified"[21] to characterise it as personal data.

According to the CJEU, the applicable test in the case at hand to be able to determine whether a dynamic IP address amounts to personal data is "whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject."[22]

The CJEU refers back to the opinion of its Advocate General and notes that "legal channels exist so that the online media services provider is able to contact the competent authority, so that the latter can take the steps necessary to obtain that information from the internet service provider and to bring criminal proceedings."[23] Said otherwise, the CJEU refers to the purpose of the anticipated and subsequent use (ie, to bring criminal proceedings against attackers) to determine the relevant means likely reasonably to be used under Recital 26.

From a technical perspective, the CJEU's solution can be explained through the concepts of distinguishability and availability, two core attributes of personal identifiers.[24] Distinguishability is the ability of a data point or a set of data points to distinguish or single out a data subject or closed group of data subjects within a broader open group. Availability is the ability of a data point or set of data points to be accessed by a situationally relevant entity who is said to be in a position to combine the data point or set of data points with additional personally identifying information, eg, official identification

[18] Ibid, 30.

[19] The CJEU defines dynamic IP addresses as "provisional addresses which are assigned for each internet connection and replaced when subsequent connections are made, and not 'static' IP addresses, which are invariable and allow continuous identification of the device connected to the network." Ibid, 36.

[20] Ibid, 38.

[21] Ibid, 44.

[22] Ibid, 45.

[23] Ibid, 47.

[24] The concepts of distinguishability and availability are borrowed from de-identification methods developed for the purposes of de-identifying health data. See eg, B Malin, "Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule" (2012) <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf> accessed 6 February 2025; Committee on Strategies for Responsible Sharing of Clinical Trial Data; Board on Health Sciences Policy; Institute of Medicine, "Appendix B, Concepts and Methods for De-Identifying Clinical Trial Data" (2015) Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk <https://www.ncbi.nlm.nih.gov/books/NBK285994/> accessed 6 February 2025. A third attribute, ie, replicability, could also be added to further reduce the list of personal identifiers although it is preferable to consider replicability through the lenses of availability.

numbers of natural persons, full person names, home addresses . . . [25] A situationally relevant entity is an entity of which means should be taken into account for the purposes of the GDPR identifiability test. As a reminder, Recital 26 GDPR refers to the data controller as well as other persons.

In Breyer, the CJEU thus considers dynamic IP addresses as distinguishing references – data points capable of differentiating a natural person or a closed group of natural persons within a broader open group. To determine whether such a data point constitutes personal data, the CJEU also considers its availability, meaning whether additional personally identifying information exists and whether it is reasonably accessible to a situationally relevant entity, ie, the online service provider in this case, allowing the data point to be linked to the natural person it pertains to. As mentioned above, the anticipated and subsequent processing purpose is a key consideration for determining whether the data point is available.

b. Scania

Scania[26] is a 2023 CJEU judgment. It involved a vehicle manufacturer sharing vehicle, repair and maintenance information with independent repairers. Other independent operators brought a suit against a vehicle manufacturer and claimed access to vehicle, repair and maintenance information on the basis of Regulation 2018/858.[27] The Regional Court of Cologne decided to stay the proceedings and to refer three main questions to the CJEU for a preliminary ruling. The last question asked whether Article 61(1) of Regulation 2018/858 constitutes, for vehicle manufacturers, a legal obligation within the meaning of Article 6(1)(c) GDPR, which justifies the disclosure of vehicle identification numbers (VINs) or information linked to VINs to independent operators acting as controllers. To answer this question, the CJEU had to determine whether VINs, when within the hands of the vehicle manufacturer, amount to personal data.

Following its Advocate General, the CJEU holds in Scania 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 is for the referring court to determine, that VIN constitutes personal data for them, 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."[28]

This sentence is relatively ambiguous and could mean two things: either the VIN is personal data within the hands of the vehicle manufacturer when the vehicle belongs to a natural person; or the VIN is personal data within the hands of the vehicle manufacturer the moment the independent operator submits a request to access the data. In both cases however, because the vehicle manufacturer does not know whether the vehicle has an owner and whether an access request will be made by independent operators, due diligence should imply treating all VINs as if they were personal data.

Just like what the CJEU did in Breyer, the court in Scania takes into consideration the purpose of the anticipated and subsequent use, ie, the offering of services to vehicle

[25] Personally identifying information is usually considered to be made of direct identifiers.

[26] *Gesamtverband Autoteile-Handel eV v Scania CV AB* [2023] CJEU C-319/22, ECLI:EU:C:2023:837.

[27] 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, OJ L 151, 14.6.2018, pp 1–218.

[28] *Gesamtverband Autoteile-Handel eV v Scania CV AB* (n 26) para 49.

holders. To pursue such a purpose, independent operators will combine the data point at stake (the VIN) with additional information that is linked to a natural person, ie, the owner of the vehicle. The CJEU however goes further than in Breyer in that it considers two types of situationally relevant entities in which hands the data is indirectly personal or personal, ie, the initial data holder and the subsequent data recipients, that is to say the vehicle manufacturer and independent operators.

From a technical perspective, the CJEU's solution in Scania can again be explained through the concepts of distinguishability and availability. The CJEU considers VINs as distinguishable references – data points capable of differentiating a natural person or a closed group of natural persons within a broader open group. To determine whether such a data point constitutes personal data, the CJEU assesses its availability, meaning whether additional personally identifying information exists and whether it is reasonably accessible to a situationally relevant entity, ie, the independent operator in this case, allowing the data point to be linked to the natural person it pertains to. This time the anticipated and subsequent processing purpose is a key consideration for determining whether the data point is available within the hands of both the initial data holder and the subsequent data recipient.

As a result, in both Breyer and Scania, the CJEU is dealing with a distinguishing reference, which is considered to be a personal identifier when it is considered available, ie, when it is reasonable to assume that additional personally identifying information will be associated with the reference. For the purposes of these two cases, a personal identifier is thus understood as a distinguishing and available reference to a natural person or a closed group of natural persons within a broader open group.

c. IAB Europe

IAB Europe,[29] a 2024 CJEU judgment, is even more interesting, as it marks a significant evolution. In IAB Europe, the data points at stake are not unique or closed group references. They are non-identifying user preferences, which can be shared across the whole population of users. When a user consults a website or application for the first time, a Consent Management Platform ('CMP') appears in a pop-up window to enable the user to consent or object to the subsequent processing of her personal data for pre-determined purposes, eg, for marketing or advertising. User preferences related to the subsequent processing of personal data are described as being "encoded and stored in a string composed of a combination of letters and characters."[30] This Transparency and Consent String ('the TC String') is shared with data brokers and advertising platforms participating in the OpenRTB protocol, so that these stakeholders can adapt their processing on the basis of the TC String. When operating, the CMP places a cookie (euconsent-v2) on the user's device. When combining the TC String with the euconsent-v2 cookie, the TC string can be linked to the user's IP address.

The CJEU is asked two main questions in IAB Europe. The first one touches upon the test how to determine whether a string of preferences amounts to personal data. It is worth noting that the question is posed in the light of the roles played by two different stakeholders: IAB Europe, which is described as a "a sectoral organisation which makes available to its members a standard whereby it prescribes to them how that string should be generated, stored and/or distributed practically and technically,"[31] and participants to the OpenRTB protocol, ie, the parties that have implemented that standard on their websites or in their apps and thus have access to that string.

The CJEU initiates its reasoning by repeating a principle found in Breyer: "in order to treat information as personal data, it is not necessary that that information alone allows

[29] *C-604/22 IAB Europe* [2024] CJEU C-604/22, ECLI:EU:C:2024:214.
[30] Ibid, 25.
[31] Ibid, 31.

the data subject to be identified."[32] The CJEU then adds that "it is not required that all the information enabling the identification of the data subject must be in the hands of one person."[33]

In IAB Europe, the CJEU expressly characterises an IP address as an identifier and states that this information "may make it possible to create a profile of that user and actually identify the person specifically concerned by such information."[34] The CJEU is thus viewing IP addresses as distinguishing and thereby as tantamount to personal identifiers.

In IAB Europe, and contrary to Breyer although the CJEU refers to Breyer without distinguishing it from IAB Europe, the CJEU does not consider the availability of IP addresses, ie, whether it is reasonable to assume that IP addresses will be combined with additional personally identifying information. For the CJEU it is enough to consider whether it is reasonable to assume that the TC string could be combined with an IP address or other identifying information.[35] The CJEU finds that

> it is apparent from the documents before the Court, and in particular from the decision of 2 February 2022, that the members of IAB Europe are required to provide that organisation, at its request, with all the information allowing it to identify the users whose data are the subject of a TC string.[36]

The CJEU thus concludes that Recital 26 GDPR's test is met and that IAB Europe has "reasonable means allowing it to identify a particular natural person from a TC String, on the basis of the information which its members and other organisations participating in the TCF are required to provide to it."[37]

Is the IAB Europe judgment departing from the Breyer judgment? It is not entirely clear, but the CJEU's position is certainly evolving and maturing. The CJEU is now presuming or more radically mooting considerations related to the availability of the distinguishing reference, in particular for IP addresses, when the anticipated processing implies or enables the profiling of natural persons. The CJEU expressly refers to Recital 30 GDPR to justify its approach.[38]

This solution should be welcome as profiling, even when there is no identity inference – the deducing of an individual's identity based on available data – potentially jeopardises the autonomy of the natural person. Yet, autonomy is a fundamental value protected by the rights to privacy and data protection.[39] As a result, the anticipated processing purpose in IAB Europe is leveraged to set aside the criterion of availability.

[32] Ibid, 39.

[33] *Patrick Breyer v Bundesrepublik Deutschland* (n 14) para 40.

[34] *C-604/22 IAB Europe* (n 29) para 44.

[35] Access to IP addresses should not be a necessary condition, as the TC string is combined with a set of data points (age, user's location, age and search and recent purchase history) that is distinguishing. Ibid, 24.

[36] Ibid, 48.

[37] Ibid, 49.

[38] Ibid, 45.

[39] This is clearly stated in various guidelines. See eg, EDPB, "Guidelines 03/2022 on Deceptive Design Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them – Version 2.0" (2023) <https://www.edpb.europa.eu/system/files/2023-02/edpb_03-2022_guidelines_on_deceptive_design_patterns_in_social_media_platform_interfaces_v2_en_0.pdf> accessed 6 February 2025; EDPB, "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0" (2020) <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> accessed 6 February 2025. See in relation to consent, *Orange România SA v Autoritatea Națională de Supraveghere a Prelucrării Datelor cu Caracter Personal (ANSPDCP)* (2020) ECLI:EU:C:2020:158 (Advocate General Szpunar) [44]. For the EDPB autonomy is a fairness consideration as confirmed in *Binding Decision 3/2022* (EDPPB) [222].

d. Summary and implications for SRB

Table 1 compares the solutions adopted in Breyer, Scania and IAB Europe.

From Table 1, it appears that the data should be characterised personal in the following scenarios:

1. When the situationally relevant entity is in a position to process distinguishing and available information (Breyer).
2. When the situationally relevant entity is expected to share the data points at stake with other situationally relevant entities who are in a position to process distinguishing and available information (Scania).
3. When the situationally relevant entity is in a position to access distinguishing information if the anticipated processing performed by other situationally relevant entities implies or enables user profiling (IAB Europe).

How will the CJEU then manage upcoming cases? In SRB,[40] which will soon be decided by the CJEU in appeal, one key question to solve for the General Court was whether the data recipient (ie, Deloitte), with whom the data controller (ie, the Single Resolution Board) had shared key-coded individual-level data, held personal data, as Deloitte had alleged that it was not in a position to combine the data points at stake with additional personally identifiable information.[41]

Applying the CJEU reasoning emerging from Breyer, Scania and IAB Europe, it is worth noting two things. First, the data recipient held data that is distinguishing. Second, the data provider was expected to learn from the data held by the data recipient to assess the situation of specific identified natural persons. To put it another way, there was a feedback loop.[42]

Assuming the CJEU gives importance to this feedback loop, it should thus find that the data within the hands of the data recipient is personal data. It is with the feedback loop in mind that one could and should try to make sense of paragraph 105 of the General Court's judgment.[43] Although the EDPS should have investigated whether Deloitte had legal means available to it to achieve re-identification, there is a reasonable argument that because the anticipated use implies re-identification the data held in the hands of Deloitte should remain personal data, or at the very least indirect personal data.

In his opinion delivered on 6 February 2025, the Advocate General Spielmann (AG) acknowledges that there is no reason to exclude that pseudonymisation could pave the way towards anonymisation.[44] Although he argues that there is no need to determine the status of the data held in Deloitte's hand to characterise a transparency violation on the part of the SRB,[45] he opines that to determine the status of the data once it is in the hands of Deloitte, the EDPS should have demonstrated "for what reason, legal or technical, the pseudonymisation process (. . .) was not sufficient and should have led to the conclusion

[40] *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)* [2023] General Court Case T-557/20.

[41] More specifically, in SRB, Deloitte had been transferred shareholder and creditor comments, which had been generalised and bore an alphanumeric code. Ibid, 24.

[42] A feedback loop should not necessarily imply that the natural persons are identified. If natural persons are being profiled, distinguishability should suffice.

[43] "Therefore, since the EDPS did not investigate whether Deloitte had legal means available to it which could in practice enable it to access the additional information necessary to re-identify the authors of the comments, the EDPS could not conclude that the information transmitted to Deloitte constituted information relating to an "identifiable natural person" within the meaning of Article 3(1) of Regulation 2018/1725." *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)* (n 40) para 105.

[44] *EDPS v SRB* [2025] Advocate General Spielmann Case C-413/23 P, ECLI:EU:C:2025:59 [51–52].

[45] Ibid, 74.

https://doi.org/10.1017/err.2025.16 Published online by Cambridge University Press

Table 1. Comparison between Breyer, Scania and IAB Europe

Case	Data Point(s)	Additional Information	Data Controller(s)	Anticipated data recipient/data user	Anticipated processing purpose	Reasonable means
Breyer	Dynamic IP addresses	Subscriber information	Online service provider	Internet access provider	To bring criminal proceeding against suspected data subjects	The online service provider has legal means to require access to additional information
Scania	VINs	Vehicle owner information	Vehicle manufacturer and independent operators	Independent operators	To offer services to data subjects	Independent operators have access to both VINs and vehicle owner information
IAB Europe	User preferences	IP address and other device information	Standardization organisation and participants to the OpenRTB protocol	Participants to the OpenRTB protocol	To supply targeted advertising to data subjects	IAB Europe has legal means to require access to additional information

that [the data recipient] was processing personal data."[46] The AG, however, does not seem to be viewing Deloitte's processing as being ancillary to SRB's processing and does not expressly point to the existence of a feedback loop. The AG's interpretation is problematic for at least two other reasons, as explained below: unintended attackers seem to be ignored and it is on the EDPS to demonstrate that pseudonymisation is insufficient to achieve anonymisation.

2. *The missing piece: unintended attackers*

Although it is possible to present a relatively neat description of the CJEU case law by introducing the concepts of distinguishability and availability, the CJEU's test remains too elliptic, as it seems to be only focusing (at least explicitly) upon anticipated uses to determine what the means reasonably likely to be used are in a particular case. In other words, in its case law the CJEU seems to focus either upon the situation of the initial data controller, ie, the online service provider in Breyer, the vehicle manufacturer in Scania, IAB Europe in IAB Europe, or the situation of the potential anticipated recipient of the data or anticipated provider of additional information, ie, the internet access providers in Breyer, the independent operators in Scania, the participants to the OpenRTB protocol in IAB Europe, to assess the means reasonably likely to be used to re-identify individuals.

Yet, if one adopts a robust approach to re-identification risks, the assessment should also take into account unintended attackers that are situationally relevant, which would need to be distinguished from the data controllers or anticipated recipients of the data or providers of additional information.

A situationally relevant attacker should be defined as an adversary whose capabilities, motivations and resources align with the specific context of the data and its environment, including its accessibility, sensitivity, and potential value. A situationally relevant attacker is thus an attacker that is plausibly positioned to exploit specific vulnerabilities of the data environment to access the data.

The Information Commissioner's Office (ICO) conceptualises such an attacker through the prism of the test of the motivated intruder, which is a particular type of situationally relevant attacker with the following characteristics: the motivated intruder does not have prior knowledge but wishes to identify an individual within the data source and is relatively competent.[47] Depending upon the type of data at stake and the traits of its environment, it may make sense to use another model such, as the prosecutor model, and assume that the adversary knows that a target individual is in the data source.[48]

As a consequence, to determine whether the identifiability test is met, the CJEU should ascertain two things:

1. Whether it is reasonable to expect that situationally relevant entities could combine the data points at stake and additional personally identifying information.
2. Whether data holders, both the initial data holder and the recipient of the data, have put in place appropriate measures to prevent situationally relevant attacks.

[46] Ibid, 96.

[47] See earlier and more recent versions of the guidelines produced by the ICO: Information Commissioner's Office, "Anonymisation: Managing Data Protection Risk Code of Practice" (2012) <https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 6 February 2025; Information Commissioner's Office, "Chapter 2: How Do We Ensure Anonymisation Is Effective?" (2021) Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance <https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf> accessed 6 February 2025.

[48] Information and Privacy Commissioner of Ontario, "De-Identification Guidelines for Structured Data" (2016) <https://www.ipc.on.ca/sites/default/files/legacy/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf> accessed 6 February 2025.

What is more, it is important to have a broad understanding of the concept of additional information and consider not only the additional information held or governed by a situationally relevant entity but also the additional information that is publicly available. What this means is that to determine whether the anticipated data user is in a position to combine the data points at stake with additional personally identifying information and whether data holders have put in place appropriate measures to prevent situationally relevant attacks, it is not sufficient to consider the data held by the entities sharing or governing the data.

For example, if it is reasonably likely that the data source contains demographic data points, which may be publicly available through public registries for example, a data recipient could be deemed to be in a position to re-identify even if it does not have access to the personally identifying information held by the data provider.

III. National supervisory authorities' and the EDPB's approaches to anonymisation

Just like the CJEU's interpretation of the concept of personal data has been maturing over time, SA assessments of anonymisation processes have progressively become quite detailed, as illustrated by two recent SA decisions. EDPB Opinion 28/2024 is also rich in implications for the assessment of anonymisation processes, while EDPB 01/2025 Guidelines on pseudonymisation confirms the importance of a couple of key steps for risk assessment.

1. Doctissimo

Doctissimo is a 2023 decision from the French SA. Doctissimo operates a website that provides articles, tests, quizzes, and discussion forums related to health and well-being.

Privacy International had complained to the French SA regarding all personal data processing carried out by Doctissimo on its website, particularly the placement of cookies on users' devices, the legal basis for processing personal data collected during health-related tests, transparency obligations, information provided to website users, and data security. The complaint led to a series of investigation missions performed by the French SA, acting as the lead SA in this case.

Since Doctissimo claimed that some of the retained data had been anonymised, the French SA needed to evaluate the robustness of the anonymisation process implemented. As regards personal data generated in the context of tests and quizzes, the SA finds that the anonymisation process is not good enough to meet the GDPR standard.[49] In particular, it notes that the anonymisation process performed by Doctissimo's processor consists in transforming only two data elements: IP addresses, which are hashed without a hashing key,[50] and pseudonyms which are said to be replaced by a string of random numbers and letters.

The French SA therefore anticipates situationally relevant attacks by both the data controller and third parties and observes (1) Doctissimo was aware of the hashing parameters and, given the finite and known number of IP addresses, could use brute force to reasonably determine the IP addresses of individuals who completed the tests,[51] and

[49] *Doctissimo* (n 8) para 46.

[50] When a hash function like SHA256 is used on its own, without a hashing key, it takes an input (eg, an IP address) and produces a fixed-length output (the hash value). This process is deterministic, meaning the same input will always generate the same hash. Deterministic hashing makes it vulnerable to brute-force attacks or dictionary attacks, particularly when the input (eg, IP addresses) comes from a limited or predictable space.

[51] *Doctissimo* (n 8) para 37.

(2) retained forum posts could include identifying information on their own as they had not been treated in any specific way.[52] To use the terminology developed in Section II, the information held by Doctissimo is therefore both distinguishing (each user is given a user ID) and available (forum posts may contain additional identifying information). Notably, Doctissimo's practice amounts to profiling: profiles are organised by user IDs.

One key lesson from this case is that it is not enough for the data holder to claim that the possibility and risk of re-identifying individuals is not demonstrated to make a case for anonymisation.[53] Although this is not made explicit in the decision, as the data holder, ie, Doctissimo, had not established the implementation of a thorough evaluation of the identification risks,[54] the French SA had no choice but to limit its analysis to the re-identification risks mentioned in Opinion 05/2014 on Anonymisation Techniques issued by the EDPB's predecessor.[55] This meant starting with an analysis of the risk of singling out, which is interpreted as requiring a determination as to whether the information at stake is distinguishable.[56]

2. *The THIN database*

The Italian SA's 2023 decision in the THIN database case is another interesting decision on the topic of anonymisation. The case was brought by a general practitioner (GP) who had reported an alleged violation of the GDPR by the company Thin S.r.l. in the context of the THIN project (the Project).

The Project aimed to improve patient care and clinical outcomes by analysing real-life anonymised data from GPs. This data was intended to help in understanding patient care pathways and driving advancements in healthcare. GPs using the "Medico 2000" software (developed by Mediatec Informatica S.r.l.) to manage patient data could leverage an add-on module that had been provided to automatically transform patient data from the software and transmit it to the company conducting the Project (Thin S.r.l.). Participating GPs received, in addition to monetary compensation to cover the expenses related to the installation of the software, additional services, including access to complete and updated information on their patients' health conditions and prescriptions, a support tool to assist in their professional activities, and resources to help them become "researcher physicians" and contribute to medical knowledge.[57]

The scientific value of the Project was said to be confirmed by the fact that the European Medicines Agency, following an open tender[58] had selected, for the Italian territory, the THIN database for the development over the next six years of analyses on effectiveness and safety based on the population receiving primary care, presumably having assessed, both in terms of regulatory compliance and scientific adequacy, the creation methods, the level of anonymisation, and the data quality.

Thin S.r.l. thus alleged that the data transferred by the GPs was anonymised and stated that the anonymisation process was under the responsibility of GPSs acting as data controllers as per the contract concluded with GPs.

In regard to the anonymisation process, Thin S.r.l. stressed three points:

[52] Ibid, 45.

[53] Ibid, 43.

[54] Article 29 Data Protection Working Party, "Opinion 05/2014 on Anonymisation Techniques" (2014) WP216 24 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed 6 February 2025.

[55] Article 29 Data Protection Working Party (n 54).

[56] *Doctissimo* (n 8) para 41.

[57] *The THIN database* [2023] GPDP [9913795], Registro dei provvedimenti n. 226 del 1° giugno 2023 1.

[58] EMA/2021/01/TDA 'Real World Data Subscription' Lot 1: Primary health care or claims database from a Southern European country.

1) an independent and accredited third party, Edgewhere, acting as a processor for the GPs, operates at a centralised level to measure and ensure the effectiveness of the anonymisation process before the data is transmitted to Thin S.r.l., with the aim of eliminating any possible residual risk of singling out from the transformed dataset.[59]
2) Records with a frequency lower than ten are systematically blocked, and therefore Thin S.r.l. will not have access to such records.[60]
3) Edgewhere officially stated that it is impossible to proceed with the irreversible re-identification of the data contained in the Thin database.[61]

Notably, Thin S.r.l. had produced a document titled "Results of the Anonymization Tests Conducted on the THIN Dataset" (prepared with the support of Bl4ckswan S.r.l.), containing the results of analyses conducted on three datasets created using the THIN Database as a source, with the goal of assessing the risk of re-identification following the application of a diversified set of de-identification techniques.[62]

Despite the implementation of a compound set of organisational and technical measures and an allegation that the external guidance produced by the European Medicines Agency had been followed,[63] the Italian SA refused to consider that the data was anonymised to the GDPR standard once in the hands of Thin S.r.l. Referring to Opinion 05/2014 on Anonymisation Techniques, the SA considered that the risk of singling out, which requires a determination as to whether the information at stake is distinguishable, had not been properly mitigated.[64]

Using the terminology developed in Section II, one reading of the SA's decision would be to say that to determine whether the data at stake is personal, it is enough to find that the information is distinguishable.

What is peculiar about the setting of the THIN database case is that a feedback loop seems to be maintained between the data sent to Thin S.r.l. and the data held by GPs,[65] which would make it hard to argue that no profiling is happening. In fact, profiling would be an anticipated use if THIN data is intended to enrich patient data held by GPs.

3. EDPB Opinion 28/2024 on AI models[66]

The EDPB adopted Opinion 28/2024 on 17 December 2024 on the basis of Article 64(2) GDPR. This opinion addresses data protection concerns related to AI models. Requested by the Irish SA in September 2024, the opinion offers non-exhaustive guidance on interpreting GDPR provisions when training and deploying AI models. It follows the Report of the work undertaken by the ChatGPT Taskforce adopted on 23 May 2024.[67]

The Irish SA's request raised four main questions: (1) When is an AI Model considered to be anonymous? (2) How can controllers demonstrate that they meet the test for the

[59] *The THIN database* (n 57) s 3.2. The goal of eliminating any potential residual risk of singling out is questionable, as achieving a zero-risk anonymisation process is impossible under real-world conditions.
[60] Ibid.
[61] Ibid.
[62] Ibid, 5.
[63] Ibid.
[64] Ibid, 7.3.
[65] Ibid, 1.
[66] This section is based upon the following blog post: S Stalla-Bourdillon, "EDPB Opinion 28/2024 on Personal Data Processing in the Context of AI Models: A Step Toward Long-Awaited Guidelines on Anonymisation?" (*European Law Blog*, 12 January 2025) <https://www.europeanlawblog.eu/pub/zh4uxsfq/release/1> accessed 6 February 2025.
[67] EDPB, "Report of the Work Undertaken by the ChatGPT Taskforce" (2024) <https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf>.

legitimate interest legal basis when they create, update and/or develop an AI Model? (3) How can controllers demonstrate that they meet the test for the legitimate interest legal basis when they deploy an AI Model? (4) What are the consequences of an unlawful processing of personal data during the development phase of an AI model upon subsequent phases, such as deployment?

In answering question 1, the EDPB draws two conclusions. First, adopting what seems to be a subjective approach and considering AI model designers' intention, it states that for AI models that have been "specifically designed to provide personal data regarding individuals whose personal data were used to train the model, or in some way to make such data available,"[68] the models include personal data.

Second, and for other AI models, the EDPB stresses the possibility of an AI model memorising training data including personal data. Personal data in this case is said to be "absorbed in the parameters of the model, namely through mathematical objects."[69] The EDPB's conclusion is pretty clear: whenever memorisation may happen, the model may not be anonymous and therefore its storage may amount to the processing of personal data. The use of the auxiliary *may* is important. What this means is that ultimately the answer to question 1 is case specific.

What is this EDPB's response telling us about its interpretation of the test for personal data anonymisation under the GDPR?

It is possible to draw five important observations from Opinion 28/2024.

First, the EDPB seems to confirm that Opinion 05/2014 on Anonymisation Techniques embeds a two-prong test and the prongs are actually alternative.[70] In other words, to determine whether anonymisation is achieved controllers have two options:

1) Option 1: to demonstrate that that three re-identification risks (singling out, linkability, and inference) are all mitigated.
2) Option 2: "whenever a proposal does not meet one of the [3] criteria, a thorough evaluation of the identification risks should be performed."[71]

To paraphrase one more time the EDPB, para. 40 seems to be saying that as a matter of principle it is not because singling out, linkability or inference risks are not all mitigated that anonymisation cannot be achieved under the GDPR.[72]

Second, the EDPB stresses the importance of not focussing only upon the means available to the intended recipient of the information (or the model) to evaluate the risks of re-identification but of widening the net to consider unintended third parties as well.[73] This is particularly important in the light of the EU case law discussed in Section II.

Third, the EDPB draws an important distinction between information, including AI model information, that is publicly available and information that is not publicly available. The EDPB is thus saying that depending upon the release setting of the information at stake, (ie, open or

[68] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (n 10) para 29.

[69] Ibid, 31. The EDPB thus appears to depart from the position of the Hamburg Commissioner for Data Protection and Freedom of Information as it states that information that is not organised in a way that makes the relationship with an individual apparent can amount to personal data. Ibid, 37.

[70] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (n 10) para 40.

[71] Ibid.

[72] This aligns with the external guidance on the implementation of the European Medicines Agency policy 0070 on the publication of clinical data for medicinal products for human use, which is expected to be updated in early 2025.

[73] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (n 10) para 42.

closed settings) SAs will have to consider "different levels of testing and resistance to attacks."[74] The EDPB thus indirectly confirms the relevance of context controls, ie, controls that do not transform the data as such, but impact its environment so that the likelihood of attacks is reduced. Notably, there is no mention in EDPB Opinion 28/2024 that the training data should be destroyed to be able to claim that the model is anonymised, which may suggest that data segmentation controls within the hands of a single controller could make it possible to effectively shield identifying data sources from non-identifying data sources.

Fourth, the EDPB makes it very clear that anonymisation processes are governed by data protection law and that the principle of accountability[75] is particularly relevant in this context. Appropriate documentation should therefore be produced to evidence the technical and organisational measures taken to reduce the likelihood of identification and demonstrate their effectiveness, which may include quantification of risks through the use of relevant metrics, as well as to describe the roles of the stakeholders involved in the data flows.

Notably, in the THIN database case such documentation had been produced but did not manage to convince the Italian SA. The Italian SA adopts a rather restrictive approach as it seems to reject option 2 when a k-anonymisation method is applied on the data and either progressive or randomized codes are used. The severity of the Italian SA could however be explained by the behaviour of THIN S.r.l., which had refused to bear any responsibility for the contentious anonymisation process.

Fifth, and this is probably the most contentious part, the EDPB seems to opine that if personal data is processed unlawfully and then successfully anonymised, the GDPR does not apply to the anonymised data.[76]

4. EDPB Guidelines 01/2025 on pseudonymisation

EDPB Opinion 28/2024 has been followed by EDPB 01/2025 Guidelines on pseudonymisation, which have been adopted on 16 January 2025 and are expected to be revised after the consultation.[77]

While these guidelines do not discuss CJEU case law, they attempt to unpack the main tenets of the GDPR definition of pseudonymisation and its implications for performing pseudonymisation on the ground. The EDPB does so by stressing that pseudonymisation under Article 4(5) GDPR does not simply mean removing or transforming direct identifiers into pseudonyms. It also entails developing and documenting a systematic approach to re-identification risks by clearly delineating a pseudonymisation domain. Such a stance should be welcome.[78]

[74] Ibid, 46.

[75] Art 5(2) GDPR.

[76] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (n 10) para 134. The EDPB is clearly concerned about AI models being developed in breach of applicable data protection rules and seems to be suggesting controllers a way to regularise the situation. However, as anonymisation should be conceived as being fundamentally contextual, meaning the status of the data, including model information, could evolve depending upon who is holding and shielding it, such a solution should be limited in scope. This solution could also raise some fairness issues. It would probably have been enough to mention that SAs, in case of infringement, have a range of corrective measures at their disposal, which should be selected upon consideration of the circumstances of each case, which is what the EDPB does at para. 114.

[77] EDPB, "Guidelines 01/2025 on Pseudonymisation" (n 12).

[78] Ibid, 26. See also S Stalla-Bourdillon, "The EDPB 01/2025 Guidelines on Pseudonymisation: A Step in the Right Direction?" (*European Law Blog*, 4 February 2025) <https://www.europeanlawblog.eu/pub/tfef074h/release/1> accessed 6 February 2025; R Hu and Others, "Bridging Policy, Regulation and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR" in R Leenes and Others (eds), *Data Protection and Privacy: The Age of Intelligent Machines* (Hart Publishing Ltd 2017) pp 115–142.

More specifically, the EDPB recalls that "information from publicly accessible sources, such as posts in a social media or an online forum" is relevant for assessing the strength of the data transformation process.[79]

In addition, the EDPB confirms that the range of potentially relevant stakeholders for assessing the strength of the data transformation process includes both anticipated and unanticipated recipients.[80]

The terminology used by the EDPB remains however confusing in that it does not distinguish between distinguishability and availability: pseudonyms seem to be categorised as direct identifiers although it is stated that "it is clear that direct identifiers need to be removed from data if those data are not to be attributed to individuals."[81]

Moreover, although a strict interpretation of these guidelines does not rule out the possibility that pseudonymised data could be considered anonymised under certain conditions, it is not clear what those conditions might be.[82]

IV. Implications for new data sharing mandates

This section explores the implications of the trends discussed in Sections II and III for the new data-sharing mandates, following a deeper examination of the concept of profiling.

1. Profiling as a deal breaker

From Sections II and III, it should be possible to posit that if the anticipated use does not entail nor the profiling of data subjects nor the combination of the data points at stake with additional personally identifying information, these data points could be considered anonymised as long as appropriate controls have been in put in place to mitigate against situationally relevant attacks.

As a result, when data reuse does not involve data subject profiling, anonymisation through a risk-based approach remains possible. As mentioned in Section III(3), a risk-based approach to anonymisation had not been mooted by Article 29 Working Party (WP29). This seems to have been confirmed by the EDPB in Opinion 28/2024.[83]

One key parameter to determine when anonymisation could be achieved is to properly delineate the concept of profiling. As mentioned in Section II(1), when profiling is anticipated, under IAB Europe, distinguishability appears to be the only relevant criterion for characterising personal data.

For the purposes of the GDPR, profiling covers "automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."[84] Profiling therefore has the potential to have drastic consequences for individuals. By ignoring availability when profiling is anticipated, the CJEU is thus adopting a functional definition of personal data, which is guided by considerations related to potential harm.

[79] EDPB, "Guidelines 01/2025 on Pseudonymisation" (n 12) para 21.

[80] Ibid, 37.

[81] Ibid, 8. See also Stalla-Bourdillon (n 78).

[82] Stalla-Bourdillon (n 78).

[83] EDPB, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (n 10) para 40.

[84] Art 4(4) GDPR.

When compared with other definitions of profiling such as Hildebrandt's working definition,[85] the GDPR definition appears to kick in as early as the discovery phase. With this said, WP29 adopts a relatively narrow definition of profiling in its 2018 guidelines and states that a simple classification of natural persons does not necessarily lead to profiling.[86]

The direct consequence of the CJEU approach in IAB Europe is that when profiling is anticipated, a risk-based approach to anonymisation should be mooted. What this implies is that the processing purpose ought to be a fundamental consideration for determining whether the anonymisation process is successful. The importance of purposes is confirmed by WP29 in its Opinion on Anonymisation Techniques.[87]

Is this approach compatible with the framing of new data sharing mandates introduced by recent data regulations? Let's look at two of them to better understand whether the overall jigsaw could make sense.

2. *The Data Act*

The Data Act (DA)[88] is intended to open up the pot of industrial data and stimulate its reuse.[89] The DA does not aim to derogate from the GDPR.[90] Building upon the Data Governance Act (DGA),[91] it acknowledges the relevance of anonymisation for sharing covered data, without suggesting that the raw data should be destroyed to achieve anonymisation.[92]

Following the CJEU's approach to identifiability, it seems reasonable to find that data access and sharing under Chapter II, which governs access to and sharing of IoT data, will often involve personal data or at the very least mixed data. This is because when product or related service data is granted access to natural persons or third parties permissioned by natural persons under Chapter II the data will often amount to a profile. Even if personally identifying information is not within the data source, the data is likely to contain very rich relational information.[93] In addition, the data user or third party is likely

[85] "The process of 'discovering' correlations between data in databases that can be used to identify and represent a human or nonhuman subject (individual or group) and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category." M Hildebrandt, "Defining Profiling: A New Type of Knowledge?" in M Hildebrandt and S Gutwirth (eds), *Profiling the European Citizen: Cross-Disciplinary Perspectives* (Springer Netherlands 2008) <https://doi.org/10.1007/978-1-4020-6914-7_2>.

[86] Article 29 Data Protection Working Party, "Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679" (2018) WP251rev.01 7 <https://ec.europa.eu/newsroom/article29/items/612053/en> accessed 6 February 2025.

[87] Article 29 Data Protection Working Party (n 54) 25.

[88] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), OJ L, 2023/2854, 22.12.2023.

[89] European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A European Strategy for Data" (2020) COM/2020/66 final <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066> accessed 6 February 2025.

[90] Article 1(5) DA; Recitals 7, 34 DA.

[91] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJ L 152, 3.6.2022, pp 1–44. Interestingly, Recital 7 DGA specifies that "re-identification of data subjects from anonymised datasets should be prohibited," as an attempt to extend context controls. The DGA also contains some data reuse and sharing provisions in Chapter II and provides in Article 7 that bodies competent to assist public sector bodies authorising the re-use of covered data shall provide support for both anonymisation and pseudonymisation.

[92] Recitals 7, 26 DA. Of note, the DA in its Chapter V includes a bespoke data reuse and sharing regime for public sector data when an exceptional need arises.

[93] "Relational information can be used to probabilistically identify an individual. General examples include sex, geographic indicators (such as postal codes, census geography, or information about proximity to known or

to have access to additional personally identifying information. In other words, the anticipated and subsequent processing purposes are likely to imply a combination of both non-personally identifying and personally identifying information.

Profiling by third parties is only partially prohibited by the DA: Article 6(2)(b) prohibits third parties from using the data received for profiling, "unless [profiling] is necessary to provide the service requested by the user."[94]

With this said, the drafters of the DA seem to have been well-aware of the functional definition of personal data and have therefore included within covered data both personal and non-personal data.[95]

3. The European Health Data Space Regulation

Under the Health Data Space Regulation (EHDS) adopted on 8 January 2024,[96] two types of uses are distinguished: primary use and secondary use. Article 2(2)(e) defines secondary use of electronic health data by excluding the initial purposes for which the data was collected or produced and referring to the list of purposes found in Article 53. The introduction to the compromise text had a generic description of these purposes: "research, innovation, regulatory and public policy purposes across the EU."[97]

It is clear that the EHDS and the GDPR should be read together, as the data subject rights granted by the GDPR continue to apply when the EHDS applies.[98] The GDPR identifiability test should therefore be relevant for applying the EHDS and arguing that destroying the raw data is necessary to achieve anonymisation does not make sense in this context. With this said, interpreting Article 2(2)(b) EHDS in the light of Article 4(1) GDPR will require some agility. Indeed, it is not clear whether the category of personal electronic health data exhausts the category of personal data pertaining to a patient, although Recitals 6 and 7 seem relatively open.[99] Said otherwise Article 2(2)(b) should not be read as meaning that data that is not personal data concerning health or genetic data will always be non-personal.[100]

unique landmarks), and event dates (such as birth, admission, discharge, procedure, death, specimen collection, or visit/encounter)." K El Emam, "Methods for the De-Identification of Electronic Health Records for Genomic Research" (2011) 3 Genome Medicine 25.

[94] See also Recital 39 DA. The DA's definition of profiling found in Art 2(20) refers to Art 4(4) GDPR.

[95] See for example Art 4(1) DA, which covers readily available data, be it personal or non-personal.

[96] Regulation of the European Parliament and the Council on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. 2022/0140(COD).

[97] See Council of the European Union, Proposal for a Regulation on the European Health Data Space – Analysis of the final compromise text with a view to agreement, e: 2022/0140(COD), 18 March 2024.

[98] Art 1(2)(a) EHDS, Recital 8 EHDS.

[99] Note that the GDPR adopts quite a broad definition of personal data concerning health, as Recital 35 makes it clear ("Personal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject. This includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council (9) to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test.")

[100] For more definitional challenges, see R Rak, "Anonymisation, Pseudonymisation and Secure Processing Environments Relating to the Secondary Use of Electronic Health Data in the European Health Data Space (EHDS)" (2024) 15 European Journal of Risk Regulation 928.

Article 66 makes it clear that the secondary use of electronic health data is based on pseudonymised or anonymised data.[101] More specifically, Recital 65 seems to suggest that anonymisation of microdata sets is possible.[102]

When considering the relationship between health data holders and health data users under the EHDS and within the context of secondary use, anticipated uses do not seem to cover the recombination of data held by data providers and data users.

Assuming there is no feedback loop and following the CJEU case law, SA assessment methods and the EDPB guidance, the data should be personal in two situations:

1. The data user is deemed profiling individuals.[103]
2. The data user is in a position to distinguish data subjects and situationally relevant attacks have not been mitigated.

One last point should be stressed. When health data is kept at the individual level and is rich of relational information, it is hard to claim that the data is effectively anonymised without giving effect to contextual controls as defined by WP29,[104] which should combine both technical and organisational measures.

V. Conclusion

This paper has unpacked the identifiability test that is being progressively shaped by the CJEU and has described a sample of SA assessment methods deployed when evaluating the robustness of anonymisation processes.

It shows that profiling without the processing of personally identifying information is governed by the GDPR and that therefore the definition of personal data is functional. It also shows that it is on the data controller to prove the robustness of the anonymisation process, and the application of the accountability principle implies producing a detailed and rigorous documentation of such a process.

This paper has then assessed the implications of these emerging approaches to identifiability and anonymisation for recent data sharing mandates included in the Data Act and the Health Data Space Regulation.

It highlights that, in a data sharing context, a risk-based approach to anonymisation should remain workable outside of profiling scenarios but its effectiveness is dependent upon the richness of the relational information contained within the data source to be shared and the set of both data and context controls.

Assuming these interpretation and assessment trends are acknowledged and confirmed, a uniform approach to anonymisation may be on the verge of emerging within the EU. This approach does not entail that the raw data should be destroyed to achieve anonymisation and does not exclude that pseudonymisation could lead to anonymisation when additional conditions are met. There are also good reasons to be cautious about mere aggregation and not make it a conclusive consideration.[105] However, much will depend upon whether it is possible to generate consensus about what a thorough evaluation of identification risks should look like in case it is not possible to mitigate singling out, linkability and inference all together.

Competing interests. The author declares no competing interests related to this publication.

[101] See also Recitals 53 and 72 EHDS. See also Recital 77 EHDS.

[102] "This includes rules for pseudonymisation and anonymisation of micro datasets." Recital 65 EHDS.

[103] Without a feedback loop and given WP29's definition, profiling seems unlikely.

[104] Article 29 Data Protection Working Party (n 54) 24–5.

[105] S Stalla-Bourdillon and A Rossi, "Aggregation, Synthesisation and Anonymisation: A Call for a Risk-Based Assessment of Anonymisation Approaches" in D Hallinan and others (eds), *Data Protection and Privacy: Data Protection and Artificial Intelligence* (Hart Publishing 2021) pp 111–144.