A critical appraisal of the European Data Protection Board's Guidelines 02/2026 on Anonymisation

Cesar Augusto Fontanillo Lopez, Khaled El Emam, Lisa Pilgram

1 Introduction

On 7 July 2026, the European Data Protection Board (EDPB) adopted Version 1.0 of its Guidelines 02/2026 on Anonymisation (hereinafter, the Guidelines).[1] The Guidelines update Article 29 Working Party's (A29WP) influential 2014 Opinion on anonymisation techniques.[2] They respond to a substantially changed legal and technological environment, including recent judgments of the Court of Justice of the European Union,[3] the development of European data spaces,[4] and advances in artificial intelligence[5] (AI) and re-identification techniques.[6] The Guidelines' principal contribution is an integrated legal and technical framework for determining whether information has been successfully anonymised. At its core are two legal questions, two approaches to the assessment, a standard of substantive law, and three identifiability criteria. More fundamentally, the document confirms

[1] EDPB, 'Guidelines 02/2026 on Anonymisation' (Version 1.0, adopted 7 July 2026).
[2] A29WP, 'Opinion 05/2014 on Anonymisation Techniques' (WP 216, adopted 10 April 2014).
[3] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* EU:C:2016:779; Case C-434/16 *Peter Nowak v Data Protection Commissioner* EU:C:2017:994; Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* EU:C:2018:388; Case C-252/21 *Meta Platforms Inc and Others v Bundeskartellamt* EU:C:2023:537; Case C-319/22 *Gesamtverband Autoteile-Handel eV v Scania CV AB* EU:C:2023:837; Case C-807/21 *Deutsche Wohnen SE v Staatsanwaltschaft Berlin* EU:C:2023:950; Case C-604/22 *IAB Europe v Gegevensbeschermingsautoriteit* EU:C:2024:214; Case C-479/22 P *OC v European Commission* EU:C:2024:215; Case C-413/23 P *European Data Protection Supervisor v Single Resolution Board* EU:C:2025:645.
[4] European Commission, 'A European Strategy for Data' COM(2020) 66 final, 21–22; Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 [2025] OJ L 2025/327; 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) [2022] OJ L 152/1; 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) [2023] OJ L 2023/2854.
[5] Stanford Institute for Human-Centered Artificial Intelligence, The 2026 AI Index Report (Stanford University 2026) ch 2 ('Technical Performance'). On the regulation of AI in the European Union, see Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 2024/1689.
[6] EDPB (n 1) para 2; Luc Rocher, Julien M Hendrickx and Yves-Alexandre de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models' (2019) 10 *Nature Communications* 3069.

that anonymity is a contextual conclusion that may differ between entities and change over time.

This article presents a detailed analysis of the Guidelines, and argues that, although the Guidelines provide a more structured and context-sensitive account of anonymisation, they also introduce several contestable developments. In particular, they risk collapsing relatedness into identifiability, qualifying identifiability through an effects-based conception of differentiated treatment, and reconstructing it through technical understandings that may not fully capture the normative nuance of privacy and data protection law. Our analysis is partially informed by our work in health research and also uses the recent Ontario de-identification guidelines (IPC Guidelines)[7] and the ISO/IEC 27559[8] international standard as contemporary baselines reflecting current best practices in the area of anonymisation.

2 The two legal questions

Under the General Data Protection Regulation (GDPR), personal data is defined as *any information relating to an identified or identifiable natural person*.[9] Anonymous data might be understood as the complement of that category. The EDPB, accordingly, separates the anonymisation inquiry into two questions:

1. Does the information relate to a natural person? If so,
2. Is that person identified or identifiable?[10]

These questions are separate but cumulative. Information that *relates* to a natural person may concern either an *identified* or *identifiable* person or a *non-identified* or *non-identifiable* person. On the other hand, information that *does not relate* to the natural person *cannot* concern an *identified or identifiable* person. Seen the other way around, *non-identified or non-identifiable* information can be either *related* or *unrelated* to a natural person, while *identified or identifiable* information must always be information

[7] Information and Privacy Commissioner of Ontario, *De-identification Guidelines for Structured Data* (15 October 2025) https://www.ipc.on.ca/en/resources/de-identification-guidelines-structured-data accessed 30 October 2026.
[8] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), Information Security, Cybersecurity and Privacy Protection – Privacy Enhancing Data De-identification Framework (ISO/IEC 27559:2022, November 2022).
[9] 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 [2016] OJ L 119/1 (General Data Protection Regulation), art 4(1).
[10] EDPB (n 1) para 6.

related to the natural person. Identifiability, therefore, is nested within relatedness; it is its strict subset. Let U be the universe of all information, R the set of information relating to a natural person, and I the set of information relating to an identified or identifiable natural person; then, $U \supset R \supset I$.[11] Relatedness, hence, is a necessary but insufficient condition for personal data status, while identifiability is a sufficient condition for relatedness. Their interaction produces three mutually exclusive and collectively exhaustive categories.

Category	Related?	Identified or identifiable?	Legal status
I	Yes	Yes	Personal data
$R \setminus I$	Yes	No	Anonymous data[12]
R^c	No	Not applicable	Anonymous data

The order of the questions, therefore, remains important. If information does not relate to an individual, there is no need to proceed to the question of whether that person is identified or identifiable. The EDPB crystallises this sequential reasoning by ending the first question with the conditional phrase *if so*, indicating that the second question presupposes a positive answer to the first.[13]

2.1 Relatedness

To answer the first question concerning relatedness, the EDPB retains A29WP's tripartite framework of *content*, *purpose* and *effect*.[14] The Guidelines largely reproduce

[11] I am assuming that not all information is related to the natural person and not all information that is related to the natural person is information about an identified or identifiable natural person. Otherwise, we may say $U \supseteq R \supseteq I$.

[12] Some may distinguish between *anonymised* and *anonymous data*. See, for instance, A29WP, 'Opinion 4/2007 on the Concept of Personal Data' (WP 136, adopted 20 June 2007) 21. Anonymised data is narrower than anonymous data, as it continues to relate to individuals, but not to identified or identifiable individuals. Often, anonymised data is the result of a masking or data aggregation process. Anonymous data, on the other side, is data that does not relate to individuals whatsoever. This is why some authors prefer to talk about *apersonal data.* See, to this extent, Mark Elliot, Elaine Mackey, Kieron O'Hara and Caroline Tudor, The Anonymisation Decision-Making Framework (UK Anonymisation Network 2016) 9. I do not assume any of these differences for simplicity purposes and to remain closer to the terminology used in data protection law. Cf GDPR, rec 26.

[13] EDPB (n 1) para 6.

[14] A29WP (n 10) 10–12. A precursor of this framework might be A29WP, 'Working Document on Data Protection Issues Related to RFID Technology' (WP 105, adopted 19 January 2005) 8. Case law has echoed this framework in *Nowak* (n 3) para 35; Case C-487/21 F.F. v Österreichische

the earlier formulation without clarifying how the three criteria interact or what independent analytical function each is intended to perform. This is particularly problematic in relation to purpose and effect relations. Information used for the purpose of evaluating a person will often also affect how that person is treated, so the same processing operation may satisfy both criteria. For example, aggregate creditworthiness data used to determine whether somebody receives a loan relates in purpose and effect to the relevant person being assessed. The difficulty, hence, is that the Guidelines do not sufficiently explain how purpose and effect relations are to be distinguished, where their overlaps lie, or whether one adds anything to the other in specific cases. By the same token, no additional clarification has been provided about the justificatory basis of this tripartite framework, and why it is supposed to constitute a complete theory of relatedness. It is just taken for granted. More guidance would be welcome to clarify these issues.

Perhaps the most significant issue in the Guidelines' treatment of relatedness concerns the extent to which it is contextual. The Guidelines require relatedness to be assessed "in light of all the circumstances of the specific case", and recognise that a relationship may not be readily apparent but may emerge through additional processing, including techniques capable of revealing individual-level information from aggregate data.[15] More ambiguously, paragraph 6 states, after setting out both relatedness and identifiability as the two elements of the anonymity test, that "these answers may vary from one entity to another".[16] Read literally, this formulation could suggest that relatedness itself, like identifiability, may vary according to the relevant entity's perspective.

On the strongest reading, such entity-relativity might be taken as extending aspects of the *reasonable likelihood* inquiry to the question whether information relates to an individual.[17] If so interpreted, this reasoning may be itself novel, but it would sit uneasily within the tenor of the law. Indeed, Recital 26 GDPR confines the reasonable likelihood threshold to identifiability, not to relatedness. This is also reinforced by the wording of Article 4(1) GDPR, which defines personal data as any information *relating* to an identified or identifiable natural person; it does not refer to information that is *related or relatable* to them. Relatedness is presented grammatically

Datenschutzbehörde EU:C:2023:369, para 24; *IAB Europe* (n 3) para 37; *OC* (n 3) para 45; *SRB* (n 3), paras 55–56.

[15] EDPB (n 1) paras 16-17.

[16] Ibid, para 6.

[17] GDPR, rec 26.

as an existing relation that either obtains or does not, while identifiability is framed as contextual, potential, and prospective. The risk of that phrasing, thus, is to incur a category error by collapsing two conceptually distinct elements.

To be able to explain recent case law, such as *SRB* findings,[18] the legal framework requires a global domain of information that relates to natural persons irrespective of anyone's perspective, or, if a lower bound is advocated, one that is wider than the entity-relative category of identifiability. This is the only way to ensure that future identifiability remains capable of operating over the full domain of information that relates to natural persons.[19] If identifiability is not able to operate upon that wider, pre-existing relatedness domain, we risk collapsing the ontological hierarchy of personal data and preventing the 'any future possibility of identification' argument from working.[20] Otherwise, information that is not presently related to a person from the relevant entity's perspective would fall outside the domain before future identifying means could ever be considered, making the argument circular. In other words, identifiability could not later attach to information that had already been excluded for lack of relatedness. The Guidelines would therefore benefit from clarifying that contextuality at the stage of relatedness does not entail the same entity-relative and likelihood-based inquiry that governs identifiability.

2.2 Identifiability

On the other hand, the Guidelines clarify what it means to identify someone. They seem to abandon the concept of identification as the discovery of a person's civil identity, as case law has sometimes portrayed.[21] Instead, identifiability is framed now broadly, as *distinguishability* from others within a given context in a way that makes it possible to *treat them differently*.[22] Direct identifiers such as names and identification numbers can achieve this, but so can unique combinations of attributes or indirect identifiers.[23] This conceptual shift is important because, as some have

[18] *SRB* (n 3) paras 112-116.
[19] See, to this extent, the previous analyses of Section 2.
[20] Rec 26 GDPR allows for this argument when it talks about the objective factor of *the available technology at the time of the processing and technological developments*, making anonymisation a defeasible legal claim. This also finds support in EDPB (n 1) paras 27, 29, 35, 84, and 88-93. I will expand on this argument in a future contribution on a general theory of identifiability.
[21] *Breyer* (n 3), paras 19-21, 24-25; *Nowak* (n 3), para 31; but see also later judgments, eg, *OC* (n 3), para 58; César Augusto Fontanillo López, 'Against Biometric Identifiability: Part I: The Law of What Is Reasonably Likely' (2025) 11(4) *European Data Protection Law Review* 465.
[22] EDPB (n 1) Executive Summary.
[23] Ibid, para 19.

argued, many contemporary data practices do not need to attach a legal name to a person.[24] Tracking, targeting, and personalised treatment can all occur through either direct identifiers or attribute combinations. Furthermore, such differential treatment of individuals can occur without individual identifiability, for example, if a person is part of a group with certain characteristics, and this issue is addressed further in the section below on inferences. This is illustrated in quadrant (b) in the table below, which, under this classification of the Guidelines, would result in the data not being identifiable.[25]

At the same time, this formulation seems to introduce a functional dimension to identifiability, as it connects contextual distinguishability with the possibility of *differentiated treatment*.[26] Although the legal significance of that connection is not entirely clear, drawing such a connection may make the satisfaction of the identifiability threshold more onerous. This would be especially so if identifiability is interpreted as requiring that distinguishability be operationally capable of determining how the individual is treated instead of being merely explanatory. The Guidelines do not clearly resolve which of these interpretations should prevail. Indeed, a person may be distinguished in a dataset by a unique record or combination of attributes, while the entity processing the data may lack any means to influence how they are treated. For instance, a researcher might isolate a unique row in a dataset but possess no capabilities whatsoever at a given moment to target or affect the individual or treat them differently. This is illustrated by quadrant (c) in the distinctions shown below. In such circumstances, data protection law tells us that the record is personal data, yet that person may not be identifiable in the sense that the Guidelines seem to suggest if that functional reading is adopted.

---

[24] Frederik J Zuiderveen Borgesius, 'Singling Out People without Knowing Their Names: Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 *Computer Law & Security Review* 256.

[25] Consider the example where a hospital anonymizes its data and provides it to a data analytics company. The data has a unique token for each patient. The company has no way of attaching that token to a patient or re-identifying the records. The data analytics company matches the patient information with clinical trials that are running. When they find a match, they send that information back to the hospital (the token and the clinical trial name, as well as a justification for the match). Now the hospital can reverse the token and contact the patient to suggest that they participate in the clinical trial, which can be very beneficial and life saving to the patient. In this case the company did not have identifiable information but was able to perform analysis that would result in differential treatment of the patient.

[26] EDPB (n 1), Executive Summary and para 19.

Distinguishability		Differential Treatment	
		No	Yes
	No	(a) Non-identifiable	(b) Non-identifiable
	Yes	(c) Non-identifiable	(d) Identifiable

Of course, the Guidelines do not seem either to require proof of this, yet it remains an important addition to the understanding of identifiability. It is important because this approach may risk importing an effects-based inquiry into the concept of personal data, while Article 4(1) GDPR formulates identifiability only in epistemic terms. The practical consequence, hence, may be a narrowing effect on the material scope of personal data, and that without further justification of why such a functional approach follows from the GDPR wording or why we should be better off following it. If, by contrast, differentiated treatment merely explains why contextual distinguishability matters, the formulation would not materially alter the notion of identifiability.

One way that the concept of differential treatment could be operationalised would be through an ethics lens. An ethics review process can add the subjective judgement that is consistent with prevailing cultural norms about whether differential treatment is detrimental to the individual. However, such an assessment would be orthogonal to an anonymity evaluation and raises the question of whether the definition is overloading the concept of anonymity.

Further clarification is therefore needed.

3 The two approaches

Reflecting recent case law developments, the Guidelines recognise that the same information may be personal data for one entity and anonymous for another. In so doing, they reproduce the traditional doctrinal distinction between the *relative* and *absolute* approaches to personal data, familiar to all of us in the data protection law domain.[27] The EDPB rebrands the two stances, though. While the relative approach

[27] Fontanillo López (n 18) 475–77; Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10 *International Data Privacy Law* 11, 17–18; Gerald Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European General Data Protection Regulation' (2016) 7 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 163, 165–66.

is renamed as the *contextual* approach, the absolute approach becomes the *simplified* approach.

3.1 The contextual approach

The contextual approach makes the notion of personal data acquire an intransitive character in the sense that the legal qualification of an entity's data may not affect how other entities classify that same data.[28] For instance, a controller may be able to identify the individuals represented in a dataset because it holds relevant additional information, whereas another controller or processor receiving that same dataset may lack means reasonably likely to be used for identification. In such circumstances, the information may constitute personal data for the former entity while being anonymous for the latter. The contextual approach, thus, is structured around the notion of *relevant perspective*, as incorporated by CJEU's case law.[29] This perspective asks in relation to whom the data is intended to be anonymous,[30] and thus, it depends heavily on the *entities* that are relevant in the particular processing context.[31] Typically, the relevant perspective will be that of the processing entity itself, but not always, as sometimes another entity's relevant perspective will be used, and other times, many different perspectives will be relevant.[32] The relevant perspective, thus, is context-dependent and may vary from case to case.[33]

As a contrast, in other jurisdictions such as the HIPAA Privacy Rule Expert Determination method, the relevant perspective is defined as the anticipated recipient.[34] This is but only one possible relevant perspective that can be taken under the CJEU definition, however.

---

[28] Fontanillo Lopez (n 18) 469.
[29] *SRB* (n 3), paras 99–112.
[30] EDPB (n 1) para 12.
[31] Ibid, para 11.
[32] EDPB (n 1) para 11-12.
[33] Other factors for assessing the relevant perspective include the nature and context of the processing, such as who has access to or control over the data, whether data will be disclosed or transferred to other parties, the relationship between those parties, and whether a recipient of the data might process it on behalf of another party. The relevant perspective is also connected with GDPR obligations and the notion of *controllership* and whether an entity can determine by itself the means and purposes of the processing. Cf. EDPB (n 1) paras 13-15.
[34] 45 CFR § 164.514(b)(1); US Department of Health and Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (26 November 2012), section 2.1, https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html accessed 30 October 2026

In determining which relevant perspective to take, each relevant entity must examine the means available to that entity and beyond, and whether those means are reasonably likely to be used. This inquiry, the EDPB argues, is susceptible to error. An entity may overlook a relevant actor, an available dataset, or a chain of entities through which identification becomes possible,[35] thereby misclassifying personal data as anonymous. The EDPB describes this as a risk of *false positives*.[36] To reduce the risk of misclassification, it provides several interpretative anchors.[37] For instance, where an entity processes data on behalf of a controller, the data status of the processor should be assessed from the controller's perspective.[38] Otherwise, controllers could potentially escape GDPR obligations simply by outsourcing processing to an entity for which the data might not be personal.[39] Despite these risks, the EDPB acknowledges that the contextual approach reflects the full nuances of the legal standard of anonymisation.[40]

Other jurisdictions have taken a different approach where the concept of separations within the controller explicitly allows identifiable and non-identifiable information to co-exist within the controller.[41] Separations include controls such as different individuals working on the data, physically separate locations, different access controls, and different reporting lines. This allows organisations, such as academic medical centres, to both provide care and perform secondary analysis on data. Then, by extension, if data can be considered non-identifiable with the separations at the controller, they would still be non-identifiable with similar administrative and technical controls at the processor.

The other important factor that determines the contextual approach is the *capabilities* of the relevant entities to access or control the data, considered over time.[42] The notion of capabilities includes elements like the accessibility of the data, in particular whether it is released publicly or has access restrictions; organisational, legal or contractual, and technical measures, in particular the implementation of encryption schemes; the possibility of data leakages that may give entities access to data

[35] EDPB (n 1) para 47; César Augusto Fontanillo López and Abdullah Elbi, 'On the Legal Nature of Synthetic Data' (NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research, New Orleans, 2 December 2022), 5-8.
[36] EDPB (n 1) para 47.
[37] Ibid, paras 12-15, 20-21, 29-30, 31-33, 47-49.
[38] Ibid, para 15.
[39] Ibid.
[40] Ibid, Executive Summary, para 47.
[41] IPC Guidelines, 22-23, 58-60.
[42] Ibid, para 88.

previously unavailable; or the technical developments that may reduce the time and cost of re-identification.[43] Notably, the EDPB emphasises that AI and agentic systems are likely to reduce such costs in terms of time and resources.[44] For that reason, it argues that contextual assessments should include adequate *safety margins* that remain robust as capabilities evolve and, where appropriate, be periodically revisited.[45] What a safety margin shall mean, however, is not clarified in the Guideline.

With respect to the appropriate period for the contextual re-assessments, we can consider precedents for DPIAs: the A29WP and the Irish DPC considered that DPIAs should be revisited after 3 years as a good practice,[46] while the Belgian DPA at 'at least' 3 years.[47] The Ontario guidelines set the re-identification risk interval, in practice, between two or three years, yet this has to be determined on a case-by-case basis.[48] These can help make the planning for periodic assessments more concrete.

3.2 The simplified approach

The simplified approach, on the other hand, makes the notion of personal data acquire a transitive character in the sense that the legal qualification of an entity's data may affect how other entities classify that same data.[49] This approach disregards differences between entities' access to means reasonably likely to be used, additional information, or other resources that contribute to re-identification.[50] It asks instead whether relevant identification methods exist without first determining precisely who is able and reasonably likely to use them.

Importantly, the simplified approach is not an alternative standard for identifiability, meaning that third parties cannot use the outcomes of this approach themselves to

[43] Ibid, paras 88-93.
[44] EDPB (n 1) paras 92-93.
[45] Ibid, para 88.
[46] Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is 'Likely to Result in a High Risk' for the Purposes of Regulation 2016/679 (WP248 rev 01, 4 October 2017) 12; Data Protection Commission (Ireland), Data Protection Impact Assessments https://www.dataprotection.ie/en/organisations/know-your-obligations/data-protection-impact-assessments accessed 30 October 2026.
[47] Belgian Data Protection Authority, *Recommendation No 01/2018 of 28 February 2018 on Data Protection Impact Assessments and Prior Consultation* (28 February 2018) paras 96–100.
[48] IPC Guidelines, 31.
[49] Fontanillo Lopez (n 18) 469.
[50] EDPB (n 1) paras 45 and 48.

challenge the anonymisation qualification done by the relevant entity.[51] This approach, which has a voluntary character, is meant to manage risk in a more secure way, deliberately shifting the possibility of *false positives* of the contextual approach to *false negatives*.[52] That is, information that may legally be anonymous for a particular entity is conservatively treated as personal data. This results in overprotection beyond what the legal standard strictly requires, but gives the controller greater confidence that it has not wrongly placed personal data outside the GDPR.[53]

Such a conservative approach has been considered suitable for public data release where no additional administrative, technical, or contractual controls can be put in place.[54] For public releases of data, the risks are higher, and therefore, the shift towards a false negatives approach may be more justified. At the same time, this has an impact on data utility and means that data for public release will be of less value almost by definition. This has implications for open government and open science efforts, and the transparency agenda in general, that advocate public release of information. Therefore, a recognition that overprotection also has broader practical consequences that are not explicitly accounted for in the Guidelines may be recommended.

3.3 Combination of approaches

One of the novel insights into anonymisation is that both approaches can be combined.[55] A controller might begin with the simplified approach and determine whether identification is theoretically possible. Here, the substantive threshold of reasonable likelihood is not of much importance, as the controller is more interested in whether re-identification may be obtainable under more demanding assumptions of possibility, and not of probability.[56] If no possible means of re-identification is envisaged, the analysis can end, and the data might safely be regarded as anonymous.[57] However, where a possible method, or many possible methods, are found, the controller can move to the contextual approach and ask whether those methods are realistically available and reasonably likely to be used by the relevant

[51] Ibid, para 48.
[52] Ibid.
[53] Ibid.
[54] ISO/IEC 27559:2022, Section 7.3.2; IPC Guidelines, Sections 6.6.1, 5.4 and 5.8
[55] Ibid, para 49.
[56] EDPB (n 1) paras 45, 48-49; Fontanillo Lopez (n 18) 469-470.
[57] Ibid, para 49.

entities.[58] By contrast, the controller might start with a contextual approach because they know that many re-identification means are possible but are irrelevant in the given context. In that case, beginning with an assessment of theoretical privacy attacks would add little value, and the controller can instead go directly to the contextual approach and discount those methods because they are not reasonably likely to be used.[59]

A combined approach entails additional costs in that the risk assessment process would need to be performed more than once under different assumptions. While potentially an informative approach from a theoretical perspective, it seems more plausible that organisations will focus their efforts on the approach that is likely to provide the most utility at a lesser cost so that they can maximise the value from their data, which is the contextual approach.

4 The standard of substantive law

The standard of substantive law governing identifiability is whether the natural person can be identified, directly or indirectly, by *means reasonably likely to be used* by the controller or by another person.[60] The term *means* is interpreted broadly.[61] It may involve anything from reading a document or conducting an internet search to acquiring additional information, engaging a third party, decrypting data or deploying AI-assisted analysis. The EDPB draws on CJEU case law and argues that the phrase *means reasonably likely to be used* currently has no definition in law, but could be approached negatively instead, that is, by considering the circumstances in which a reasonably likely identification will not occur.[62] This may happen when the re-identification risk is insignificant, where it involves disproportionate effort in terms of time, cost, and labour, or where a legal prohibition to re-identify the data exists.[63] Whether means are reasonably likely to be used is something which must be analysed through objective factors, which evoke more broadly the concept of the processing entity's *capabilities*, introduced previously. Such objective factors include:

- the nature of the data itself, e.g., its properties, accuracy and granularity;

[58] This is in line with Fontanillo Lopez (n 18) 469-470 on the two-step tests, which EDPB guidelines seem to endorse.
[59] EDPB (n 1) para 49.
[60] GDPR, art 4(1) in relation with rec 26.
[61] EDPB (n 1), para 28.
[62] Ibid, para 26.
[63] Ibid.

- the processing context;
- restrictions on data access and control;
- the availability of additional data and how likely it can be used;
- the time and cost required for identification;
- existing technology at the processing time; and
- reasonably foreseeable technological developments.[64]

The EDPB also addresses whether a *prohibition by law* to re-identify data, understood as an objective factor within the reasonably likely means test, may support a legal presumption of anonymity. This is an important insight because the CJEU already indicated in *Breyer* that such a prohibition imposed by law may justify excluding a particular means of re-identification from the assessment.[65] That proposition was subsequently repeated in *OC*[66] and *SRB*.[67] The EDPB accepts the relevance of such legal prohibitions but qualifies what could become an overly formalistic interpretation of the case law, namely, that a statutory prohibition automatically results in anonymity. It argues instead that *Breyer*'s doctrine constitutes, at most, a rebuttable legal presumption, which may be displaced if the prohibition is not effectively monitored, enforced, or sufficiently dissuasive; if there is evidence that the gains of breaking the law outweigh its costs; if it is poorly protected; or if it is easily circumventable.[68] It adds that contractual restrictions do not have the same

[64] EDPB (n 1) para 29. On a side note, the EDPB cautions against relying on motivation as an objective factor to assess whether information can be considered identifiable. Motivation, it argues, is difficult to assess or demonstrate objectively; it may change over time and may not account for identification caused by negligence, accident or external pressure. This approach is different than in other jurisdictions where motivation was explicitly considered. Cf. IPC Guidelines, 13, 20, 74, 94. However, the potential value of the information may nevertheless be relevant where it offsets the costs and effort involved in re-identification, and the organisational risks associated with the personal data breach should also be factored in. Cf. EDPB (n 1) para 31. On the other hand, arguably the potential value of information is also hard to assess because value will depend on the nature of the processing entity and its business. And even if the information has theoretical value, whether that value is realisable will depend on the motivations and other capabilities of the entity. Furthermore, the value of information can change over time and value will change if there is an increase in the information's identifiability caused by negligence and accident. Therefore, a valid question is whether value is in the same category of uncertainty as motivation, or whether they should both be considered as valid criteria for decision making.

[65] *Breyer* (n 3) para 46.

[66] *OC* (n 3) para 51.

[67] *SRB* (n 3) para 82.

[68] EDPB (n 1) paras 32-33.

legal status as prohibitions imposed by law and that they should only be used to complement technical measures.[69] [70]

Finally, and like previous opinions,[71] the EDPB suggests that anonymity is dynamic, and that the likelihood of re-identification increases progressively due to advances in technology and increased availability of information.[72] This implies that anonymity is a defeasible legal claim which might produce regulatory re-entry as time passes. The Guidelines do not, however, claim that all anonymous data will become personal data in the long run.[73] Their use of the term *typically* instead of *inevitably* or analogous terms leaves room for anonymisation to remain effective over time.[74] Nevertheless, the underlying assumption is that, *ceteris paribus*, the likelihood of re-identification increases with the passage of time to a level at which, at some point, the risk can no longer be regarded as insignificant.

5 The three identifiability criteria

The technical framework retains and refines the three identifiability criteria introduced in the 2014 Opinion, though it renames them as *No Record Isolation*, *No Linkage* and *No Inference*. The EDPB confirms, again, that identifiability does not follow a *zero-risk approach*,[75] as it did in its previous 2007[76] and 2014 guidance.[77] Thus, an insignificant or residual likelihood of identification will not lead to the data becoming identifiable. The robustness of an anonymisation technique is assessed against these three criteria, yet the following factors will also count: (1) the nature of the data, whether it is aggregated or record-level data; (2) the dimensionality of the data; (3) the resolution of the data; (4) the diversity of the data; (5) the number of

[69] Ibid, para 34.
[70] Accounting for foreign intelligence and criminal activity, as suggested by the EDPB, in the risk assessment is not always going to be practical. Practical. The EDPB should acknowledge the illustrative character of these restrictions and nuance that these requirements might have difficult implementation in general settings that are not concerned so much with intelligence agencies or law enforcement, while they might play a bigger role in areas like telecommunications or spaces where the role of these entities is more prevalent.
[71] A29WP (n 10) 15; A29WP (n 2) 4.
[72] EDPB (n 1) para 35.
[73] Ibid.
[74] Ibid.
[75] Ibid, para 22; Fontanillo López and Elbi (n 30) 7.
[76] A29WP (n 10) 15.
[77] A29WP (n 2) 23.

individuals in the dataset; and (6) the additional information that is available for combination with the data.[78]

5.1 No Record Isolation

The notion of record isolation is tied to uniqueness in a given context.[79] However, what a *context* entails is not defined by the EDPB. The criterion is satisfied where the data does not single out the individual from the rest. Singling out becomes easier where the individual record contains unique identifiers or attributes that are non-unique but which result in a unique combination.[80] The larger the record and the greater the number of attributes it contains, the more likely it is that a record isolation may happen.[81] One may add that the resolution of the record, or how granular its information is, also has an impact on the criterion. Record-level datasets are therefore generally more vulnerable than aggregated data, which generally pass this criterion, because their granularity is much lower and no individual records can, in principle, be isolated. The EDPB's terminology for referring to this criterion is noteworthy. The older *singling out* criterion is now renamed as *No Record Isolation*, bringing the legal notion closer to a technical understanding of identification in datasets.[82]

A key limitation of the definition of this criterion is that it only considers uniqueness of attribute combinations within the given dataset. This would be strongly affected by multiple factors, including the sample size. For example, a dataset of two records with only two attributes, age and sex, that are unique in combination in that dataset may seem to fail the no isolation criterion. But in the population that these two records were drawn from, their attribute values are very common. In the disclosure control community, it is well established that the uniqueness must also consider uniqueness in the population and not just in the sample. The Guidelines themselves later recognise this limitation by permitting data to remain anonymous despite

[78] EDPB (n 1) para 85.
[79] Ibid, para 55.
[80] Ibid, paras 23-24 and 54-56.
[81] EDPB (n 1) para 56.
[82] Importantly, some authors have noted that the notion of singling out as isolation alone cannot provide a complete theory of singling out. In particular, they have noted that some probability of isolation exists as a statistical baseline, even where no data is disclosed. Otherwise, one would need to conclude that either anonymisation is mathematically impossible or that anonymisation allows for a relatively high likelihood of singling out. Cf Micah Altman, Aloni Cohen, Kobbi Nissim and Alexandra Wood, 'What a Hybrid Legal-Technical Analysis Teaches Us About Privacy Regulation: The Case of Singling Out' (2021) 27 *Boston University Journal of Science & Technology Law* 1, 20–22.

violation of No Record Isolation where the isolated record cannot actually be used to single out an individual. Therefore, it would be beneficial to make the population uniqueness requirement more explicit from the outset in the definition of the No Isolation criterion.

5.2 No Linkage

The No Linkage criterion asks whether a record can be connected, with sufficient certainty or a high likelihood, to information concerning the same individual in another dataset or in multiple datasets.[83] Linkages, therefore, increase the amount of information about an individual, which increases, in turn, their re-identification risk. Linkages may rely on a common identifier or on combinations of attributes.[84] Although data that satisfies No Record Isolation will be more likely, but not necessarily, satisfy No Linkage (because records that cannot be isolated are more difficult to link), the two criteria remain conceptually distinct and must be assessed separately.[85] Records that are not fully accurate because they have undergone masking through privacy-enhancing techniques (PETs) may notwithstanding be linkable, provided that they can still be matched with sufficient accuracy or probability.[86] The availability of additional information is critical in these cases. For instance, information that has not been recorded elsewhere and is not correlated to similar information about the same individual in another context is more likely to satisfy the No Linkage criterion.[87] Finally, if linkage occurs without producing identifiable information, the resulting combined dataset must be reassessed against all three identifiability criteria.[88]

To make the concept concrete, there are two types of linkage that would be helpful to distinguish. In the first situation, if two datasets are anonymised separately and then linked, then the linked dataset may or may not be considered anonymised. That would have to be determined through an additional risk assessment on the three criteria. In the second scenario a dataset is linked with identifiable information about the individuals, and in that case the linked dataset is not anonymous any more.

[83] EDPB (n 1) para 60.
[84] Ibid, para 62.
[85] Ibid, para 63.
[86] Ibid, para 64.
[87] Ibid, para 65.
[88] Ibid, paras 101-102.

5.3 No Inference

Of the three criteria, No Inference has undergone the most substantial development in comparison with Opinion 05/2014. The earlier Opinion defined inference as the possibility of deducing, with significant probability, the value of one attribute from a set of other attributes.[89] The literal interpretation of that definition would mean that if information can be learned from data through modelling or statistical analysis, then an inference would have occurred, making the data personal information. Such a definition would create significant difficulties with most data analytics workloads.[90]

The new guidelines construct a much more elaborate test. The criterion is satisfied, in particular, where no *specific* and *meaningful* inference can be drawn from the data under assessment, either alone or together with additional information.[91] An inference is *specific* when it concerns a single identified or identifiable individual.[92] It is *meaningful* when the inference can have an effect on the data subject's rights and interests, it derives from the data under assessment, and the same conclusion could not be obtained from general knowledge or population-level information.[93] This latter requirement introduces the important distinction between *baseline belief* and *posterior belief.* Baseline belief can be considered a sort of general knowledge that could already be inferred about a person without relying on the person's contribution to the dataset under assessment. An inference about an individual, therefore, is meaningful only if it results in a materially different posterior belief, after their data are taken into account.[94] According to the Guidelines, the data should generate a significantly different posterior belief, capable of affecting an action concerning the individual, before it can be established that the given data meaningfully relates to that person.[95] The relevant question is, hence, whether the inclusion of an individual's data makes a difference to the inferred conclusion.[96]

The definition of *significantly different* is important here as that determines whether the criterion is triggered. It is to be expected that inclusion of someone in a dataset

[89] A29WP (n 2) 12.
[90] 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, 81–82.
[91] EDPB (n 1) para 67.
[92] Ibid, para 67.
[93] Ibid.
[94] Ibid, para 71 and fn 43 and 44.
[95] Ibid.
[96] Ibid, fn 44.

would result in some incremental knowledge gain, but how much of that makes it significant is not specified in the Guideline. Further clarification of this would make the criterion easier to operationalise.

The framework includes the counterfactual logic similar to differential privacy.[97] Where it is not possible to say whether the data relates to an individual, the EDPB proposes comparing the inference produced from the original dataset with the inference produced from the modified dataset from which the information about that person has been removed. If the inference operates in the same way in both cases, it is attributed to population-level information instead of being attributed to the individual's own data.[98] In that case, it does not violate the No Inference criterion. This approach performs an important limiting function. Without it, almost any dataset containing statistical information could become personal data in relation to every person to whom its generalisations might be applied (as in the earlier definition in the 2014 opinion). The EDPB illustrates this point through the example of a bank dataset showing that people with a specific combination of attributes are likely to default on their loans.[99] Applying that inference to an individual produces personal data for the applicant, but it does not make the dataset from which it was derived a personal dataset in relation to that person because they were not represented in the dataset.[100]

The criterion, however, adopts a partial conception of identifiability. It asks whether marginal contributions result in material differences to the inferred conclusion, and not so much whether an inferred conclusion could be drawn about a person. These questions produce different answers. If a dataset contains, say, 100.000 smokers who subsequently developed cancer, removing one record of any one participant may leave the statistical relation essentially unchanged. Indeed, the marginal contribution of an individual to the posterior inference may not materially change anything if the relevant reference class is sufficiently large and statistically stable.[101] Under a marginal contribution concept of identifiability, applying this inference to an individual does not violate No Inference, yet the resulting population-level

---

[97] Cynthia Dwork, 'Differential Privacy' in Michele Bugliesi and others (eds), Automata, Languages and Programming: ICALP 2006 (LNCS 4052, Springer 2006) 1–12.
[98] EDPB (n 1) para 71 and fn 43.
[99] Ibid Example 14.
[100] Ibid.
[101] Ke Wang, Peng Wang, Ada Waichee Fu and Raymond Chi-Wing Wong, 'Inferential or Differential: Privacy Laws Dictate' (arXiv:1202.3686v1, 16 February 2012) 1, 3; Cynthia Dwork and Aaron Roth, *The Algorithmic Foundations of Differential Privacy* (Foundations and Trends in Theoretical Computer Science vol 9, nos 3–4, Now Publishers 2014) 6.

relationship may still support accurate and consequential predictions about that participant as well as about smokers who never contributed to the dataset.[102] EDPB's conceptualisation of the No Inference criterion can therefore protect the individual against additional learning attributed to their participation; but this does not prevent sensitive inferences based on patterns produced by collective datasets. This, of course, is not a failure of the adopted reasoning model of differential privacy in relation to its own objective, on which No Inference seems to be anchored.[103] However, it may be a failure in relation to the more general understanding of what an interference with the rights to personal data protection and privacy might entail. On the other hand, and to use the same example, if drawing a conclusion about the relationship between smoking and cancer is deemed to interfere with the rights of individuals, then that interpretation of the No Inference criterion would halt much learning from data and surely not what is intended here.

This reveals a structural tension in the proposed framework. Population-level data can be used to generate personal data when its generalisations are applied to specific individuals in a meaningful way. However, the dataset from which such generalisations derive may remain anonymous, notwithstanding the effects that the resulting inferences may have on individuals and, often, regardless of whether those individuals contributed data to the source dataset. The criticism, therefore, is that this arguably creates regulatory disparity. Data production aims to generate population-level insights that can be applied beyond the original data subjects, while EDPB's conception of No Inference remains centred on the relationship between the

---

[102] On the other side, a traditional objection to a broader notion of privacy, such as one that would prevent inferences at all, is that it may be impossible to achieve because any statistical release provides information and can be combined with auxiliary data to reveal something. Cynthia Dwork and Moni Naor, 'On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy' (2010) 2(1) *Journal of Privacy and Confidentiality* 93, 93, 95–96.

[103] Although EDPB's No Inference criterion should not be per se assumed to equate differential privacy, in particular its formal guarantees, its counterfactual reasoning resembles it. Differential privacy has several advantages. It is a technical mechanism which composes securely, meaning that multiple data releases remain differentially private when considered jointly, although increased data releases must account for a cumulative privacy loss. That enables differentially private mechanisms to be analysed modularly. Differential privacy is said to not make assumptions about auxiliary information or background knowledge in relation to whether the individual's data was included or not in the original dataset. See Dwork and Roth (n 89) 41 ff; Shiva Prasad Kasiviswanathan and Adam Smith, 'On the "Semantics" of Differential Privacy: A Bayesian Formulation' (2014) 6(1) *Journal of Privacy and Confidentiality* 1. However, on criticism about the claim that differential privacy operates without substantive assumptions, in particular its allegedly lack of external assumptions, see Daniel Kifer and Ashwin Machanavajjhala, 'No Free Lunch in Data Privacy' in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data* (Association for Computing Machinery 2011) 193–204; Wang et al (n 89) 3-4.

source data and marginal individual contributions. It may, therefore, fail to represent the collective interests and distributive effects associated with that data production.[104] In that sense, the framework cannot guarantee that an adversary will learn nothing about an individual, a limitation that the EDPB seems to acknowledge.[105] This criticism is, however, open to a counterargument. One may say that the No Inference criterion is not designed to prevent all learning about individuals, but to determine whether the data under assessment relates to them in a legally relevant way. Requiring an anonymous dataset to support no inference whatsoever would approximate an unattainable zero-learning standard, something which runs counter to the very release of useful statistical information,[106] and would imply bringing aggregate data within the scope of the GDPR.[107]

One final concern relates to the Guidelines' definition of inference. Similar to A29WP's former opinion,[108] they define inference narrowly, as the process of reaching a conclusion by *deduction*.[109] Induction, abduction or analogduction seem to be left outside the definition, yet these modes of inference arguably contribute as well to re-identification risks.[110] For example, a system may infer a general pattern from repeated observations, identify the most plausible explanation for an observed fact, or predict an individual's characteristics by comparing them with similar cases. The Guidelines' own examples appear to adopt this broader understanding, encompassing statistical, probabilistic and algorithmic conclusions,[111] yet the formal

[104] This, however, might be a logical consequence of data protection being an individual right, and the Guidelines may, therefore, be justified in not accounting for them. On similar discussions on the privacy regime and regulatory disparities, see Salomé Viljoen, 'A Relational Theory of Data Governance' (2021) 131 Yale Law Journal 573, 582; Alicia Solow-Niederman, 'Information Privacy and the Inference Economy' (2022) 117 Northwestern University Law Review 357, 387.

[105] EDPB (n 1) para 18, where it says that the goal of anonymisation is often 'to have an anonymous dataset which lets you learn something about the people in question (i.e. information which is still related to those individuals) but does so in a way that keeps their identities private (i.e. they are not identified or identifiable).'

[106] Authors have argued that such a strong anonymity goal is unattainable because it gets close to Dalenius' privacy goal of zero knowledge. See Dwork and Naor (n 90) 93–96; Tore Dalenius, 'Towards a Methodology for Statistical Disclosure Control' (1977) 15 *Statistik Tidskrift* 429.

[107] A further countercriticism might be that collective interests and distributive effects may indeed require regulation, but this does not necessarily justify expanding the concept of personal data to aggregate data. One may also add that the decision about which anonymisation model and identifiability criteria to choose should not only be technical but also reflect normative judgments about which informational interest the law should protect and ultimately be determined through democratic decision-making.

[108] A29WP (n 2) 12.

[109] EDPB (n 1) para 66.

[110] Fontanillo Lopez (n 18) 473.

[111] EDPB (n 1) Examples 13-15.

definition of inference does little justice to these examples.[112] In contrast, the EDPB gets it right when acknowledging that an inference which does not initially violate the No Inference criterion may subsequently be combined with the given data or with additional information to generate a further inference that is specific and meaningful and therefore does violate the criterion.[113] The definition should accordingly reflect inference as a broader process of deriving, estimating or predicting information from available evidence.

6 Other takeaways

Finally, three other key points of the new guidelines deserve particular attention.

6.1 Identifiability criteria as indicators

The EDPB confirms the non-determinative character of identifiability criteria. It argues that the criteria are indicators, not conditions whose individual violation automatically turns a dataset into personal data.[114] Where all three criteria are satisfied, the data may safely be regarded as anonymous. Where one or more criteria are violated, further analysis is required.[115] For example, a record may be unique within a dataset but still not allow any relevant entity to identify, distinguish, or meaningfully act upon the person concerned.[116] Similarly, two records may be linkable without the combined information enabling identification.[117] The criteria, therefore, act as support tools for the final inquiry of whether data can be deemed personal. This concretisation is important, as A29WP's 2014 opinion frequently presented the three criteria as basic requirements of effective anonymisation,[118] leaving some uncertainty as to whether each criterion constituted an independent

[112] This may hold true unless the EDPB severs the two stages of the process, i.e., first, the derivation of a generalisation from the source data, and second, the application of that generalisation to a particular individual. Yet even so, this distinction does not resolve the definitional problem, because one may still say that not all inferences come from aggregate data, e.g., if a person's location history shows that they visit an oncology clinic every Tuesday, one might infer that the person is receiving cancer treatment. In this case, the observer has leveraged on abductive reasoning. The EDPB is, of course, aware of this point. See, to this extent, EDPB (n 1) Example 15. Yet, the definitional problem still lingers.

[113] EDPB (n 1) para 72.

[114] EDPB (n 1) para 52.

[115] Ibid, paras 46 and 52.

[116] Ibid, para 99 and Example 24. This can also occur if a dataset has sample uniques that are not population uniques, and therefore would not satisfy the more meaningful uniqueness test.

[117] Ibid, para 101 and Example 25.

[118] A29WP (n 2) 24.

and sufficient benchmark, such that the failure of any one of them would render the information at stake personal data. The new guidelines dispel that ambiguity by making clear that failure of a criterion does not result in an automatic conclusion.

At the same time, one may argue that the non-determinative character of the three criteria does not imply that they occupy an equivalent position within the identifiability analysis. A possible functional hierarchy may be reconstructed among them. At its apex, arguably, would lie record isolation, since uniqueness of a particular person appears to be a precondition for regarding that person as identified or identifiable. This also seems to be suggested by Recital 26 GDPR, which emphasises record isolation as a relevant identifiability criterion. On the other side, information linkages and inferences would seem ancillary to record isolation in the sense that their importance often depends on whether they ultimately facilitate such isolation. Linking records may assemble a sufficiently distinctive combination of attributes that may be used to ultimately distinguish one person from the relevant candidate set of possible persons. Inference may perform a similar function by producing additional attributes that may ultimately narrow the candidate set, reduce the similarities among individuals, or make their remaining differences more expressive. In either case, linkage and inference increase identifiability to the extent that they create or reinforce a profile capable of singularising the person concerned.

This hierarchy would also explain why the violation of No Linkage or No Inference does not automatically render the data personal.[119] Linkage may combine records without producing a uniquely distinguishable profile,[120] while an inference may add information without reducing the relevant candidate set sufficiently to single anyone

[119] On a side note, one may argue that the violation of No Record Isolation may sometimes not result in a specific individual's differentiated treatment if the criterion is not understood *as such*, i.e. as the mere ability to distinguish a record from others, without further qualifiers. For instance, a dataset may contain a unique record associated with a particular smart TV. The record may include the device's location, viewing history and application usage. The device record may therefore be isolated. However, the television may be shared by five members of a household, none of whom uses an individual account, and the available information may not reveal which person generated any particular viewing event. In that case, the record isolates one device, but the relevant candidate set still contains five natural persons. Record isolation, therefore, might distinguish an object or, at most, a household, but arguably, not a specific individual who can be treated differently with respect to others. This seems to be the main reason why, in some instances, the CJEU's case law has required, in addition, the revelation of the civil identity of the individual for holding a particular data processing as personal data processing. See, to this extent, fn 18. One may extend this argument to information about vehicles which are shared by, potentially, several individuals. See, *mutatis mutandis*, *Scania* (n 3) paras 46- 49.

[120] EDPB (n 1) para 101.

out.[121] EDPB's further qualification of identifiability nevertheless remains significant. If differentiated treatment constitutes an independent operational requirement for identifiability,[122] even successful record isolation would remain only an indicator instead of a legally sufficient condition for identifiability. In other words, as has already been introduced in Section 2.2, if this particular reading of identifiability is endorsed, it may suggest that the fact that a person can be singled out may not, by itself, establish identifiability unless that distinction is made somehow operational in the relevant context.

In this regard, we could argue, again, how at odds this qualification would be with the preventive risk rationale of data protection law. If the regime is intended to intervene before informational capabilities materialise into concrete effects, identifiability should not depend on whether record isolation is already liable to affect the individual at a particular moment.[123] On this view, an effects-based conception of identifiability should be rejected. The violation of the No Record Isolation criterion, insofar as it gives practical expression to identifiability,[124] should itself be sufficient to trigger the data protection regime whenever the relevant entity can single out the person through means reasonably likely to be used.[125] Linkage and inference may then be understood as ancillary mechanisms whose significance depends on whether they contribute to that person's singularisation.

6.2 Tainting effects

Remarkably, the Guidelines introduce what might be called a *tainting effect* in relation to mixed datasets. This implies that the presence of a legally riskier data category within a dataset affects or *taints* the classification of less risky data categories within the set it is part of. The effect evokes a worst-case classification principle under which the most protective legal status represented in the dataset governs its collective treatment. This happens at least with respect to personal data and sensitive data.[126] For instance, where anonymised data exhibits a high risk of re-

[121] See, in this regard, ibid, para 75. Note that, in so arguing, I am discounting any possible violation of *any* identifiability criteria on the basis of inaccurate results. More information about the accuracy of re-identification and its problematic is provided in Section 6.3.
[122] Ibid, Executive Summary and para 19.
[123] And as noted earlier, this is quite difficult to operationalize a priori in practice.
[124] GDPR, rec 26.
[125] This interpretation, of course, would result in false negatives in some cases. This would be especially so where the isolation of a record does not ultimately single out an individual in the real world. See, to this extent, the previous fn 111.
[126] EDPB (n 1), para 36.

identification for some records, while the others enjoy a negligible risk, the status as personal data of the high-risk records has a tainting effect over the rest of the dataset, and it as a whole must be treated as personal data because the worst case governs the classification rule.[127] A possible exception to this rule might arise, however, where the identifiable records can be severed from the dataset in a way that their separate processing does not affect the non-identifiable records.[128] A comparable spillover occurs where sensitive and non-sensitive personal data are combined in the same dataset, in which case the dataset as a whole must be treated as containing special categories of data.[129] This latter case seems to follow from the case law of the CJEU.[130]

Tainting effects, however, have their limits, and such limits seem to depend on the level of protection of the dataset records. For instance, a dataset may be considered anonymised even if its records do not all enjoy the same level of protection, provided that the likelihood of identifying any individual remains insignificant.[131] This might happen if certain records in a dataset are more likely to be re-identified than others because they have more attributes, more detailed information or more unusual combinations than the rest, yet their re-identification risk remains sufficiently low and below a pre-specified threshold. In such a case, the unequal protection across the records does not, by itself, prevent the dataset from being considered anonymous.[132]

6.3 Alethic conceptions of personal data

Finally, a last point of the Guideline merits consideration. As many of us know, data protection law adopts an alethically neutral conception of personal data; that is, information is not required to have a specific alethic value to be a matter of the law.

---

[127] Ibid.
[128] Ibid.
[129] Ibid, fn 31.
[130] See *Meta Platforms* (n 3) para 89.
[131] EDPB (n 1) para 36.
[132] EDPB (n 1) para 36. It should also be noted that the concept of average risk applied in other jurisdictions [ISO standard and Ontario] means that some individuals can have a risk value that is higher than the threshold and some lower, and if the average risk is below the threshold then the data would be deemed non-identifiable overall (i.e., all individuals are protected). Residual risk to the individuals at the tail of the distribution is managed by the additional layers of protection, which would include additional administrative, technical, and contractual controls that would be placed on the data recipient. Therefore, rather than relying only on the vulnerability to re-identification in the data only, these models take a broader perspective of risk where the additional controls also count.

Both true and false information are of concern because both may result in adverse effects to the rights and freedoms of individuals.[133] A29WP made this explicit in its Opinion 4/2007, observing that personal data encompasses objective information, subjective opinions and assessments, and that information need not be *true or proven* to qualify as such.[134] The Court's judgment in *Nowak* reflects the same position.[135] A candidate's examination answers might be incorrect, and not for that reason they may fail to be personal data.[136] *Deldits* confirmed this same view by holding that inaccurate data concerning gender identity in a public register had to be capable of rectification.[137] Indeed, the right to rectification would have little purpose if inaccurate information ceased to constitute personal data.

The Guidelines do not expressly restate this principle, but they implicitly proceed with it, in particular when it comes to information about already identified or identifiable individuals.[138] However, a key difference seems to appear in relation to information that may or may not be about identified or identifiable individuals. For instance, consider the two propositions *p = 'Bob has cancer'* and *q = 'Bob is identifiable from dataset D through means M'*. Both are, in principle, truth-apt,[139] but they play a different legal role. Provided that Bob is, at least, identifiable, the truth of *p* is irrelevant for data protection law because a false statement may still harm identifiable individuals. The proposition *q*, however, expresses an antecedent

---

[133] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (Version 2.0, adopted 20 October 2020) paras 77–78.
[134] A29WP (n 10) 6.
[135] *Nowak* (n 3) para 34.
[136] Ibid, paras 53-56.
[137] Case C-247/23 *Deldits* EU:C:2025:172, paras 23–27 and 38.
[138] For instance, in Example 13, the inferred statement that 'Connor likes green' is treated as personal data because it relates to an identified person, although its truth is not established.
[139] We can assume, for these purposes, a correspondence theory of truth; that is, one in which a proposition, statement or belief will be true when it sufficiently corresponds to a fact, object, or state of affairs in the external world. For instance, a proposition attributing a particular characteristic to a person for the purpose of identifiability will be true where that characteristic sufficiently corresponds to an attribute the person actually possesses in a way that makes them identified or identifiable. Similarly, a claim of record isolation will be true where the isolated record will in fact sufficiently distinguish the individual in the external world whose information it purports to isolate. In such cases, propositions, statements and beliefs concerning identifiability function as truth-bearers, while the external facts or states of affairs to which they correspond function as their truth-makers. On the correspondence theory of truth, see Marian David, 'The Correspondence Theory of Truth' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Summer 2025 edn, Metaphysics Research Lab, Stanford University 2025). However, please observe that this assumption is more dependent on a strict concept of identifiability because specific individuals must be ultimately singularised in the external world. A record isolation *as such* would have little grounding on this account.

condition for bringing the information within the material scope of the GDPR. For these purposes, a merely speculative or inaccurate assertion that Bob is identifiable cannot by itself defeat anonymity, the Guidelines tell us.[140] Thus, at the jurisdictional threshold, the alethic value of information is not legally neutral because identifiability claims must be sufficiently grounded to declare personal data status.

This does not mean that identifiability must be demonstrated with *perfect accuracy* or *absolute certainty*, yet as the EDPB notes, identifiability claims must nevertheless produce results that are 'at least sufficiently close and provide at least a sufficiently high confidence' in the circumstances at hand.[141] The Guidelines reinforce this in various ways. First, they note that re-identification techniques 'should be assessed according to their ability to generate accurate results'.[142] Second, the objective factors for assessing the personal nature of the data include the properties of the data itself, such as 'the accuracy and precision of the information'.[143] Third, the No Linkage criterion is satisfied where records cannot be connected 'with certainty or high likelihood'.[144] Fourth, even if records altered through PETs may remain linkable despite not being fully accurate, some sufficient level of accuracy would be implied.[145] Fifth, when an inferential method fails to attribute its result to a particular individual with sufficiently high confidence, it will not violate the No Inference criterion.[146]

The resulting position, hence, might be described as one of alethic neutrality of informational content about identified and identifiable individuals, and one of veridicality for informational content subject to the material scope of data protection law. The Guidelines, therefore, seem to operate with two conceptions of information depending on their legal function. Again, although identifiability claims are not required to show perfect accuracy or absolute certainty, they still must produce

[140] EDPB (n 1) para 87, Example 20. See also paras 26, 29 and 52.
[141] Ibid, para 87.
[142] Ibid, Executive summary and para 87.
[143] Ibid, para 29.
[144] Ibid, para 60. See also para 64.
[145] EDPB (n 1), para 64.
[146] Ibid, Example 20 and para 87. Note also that the notion of inference has changed in the new Guidelines. According to the former A29WP's 2014 opinion, inferences had to be deduced *with significant probability* to qualify as such. Cf A29WP (n 2) 12. The new guidelines abandon that formulation and instead define a relevant inference by reference to its *specificity* and *meaningfulness*. Cf EDPB (n 1) para 67. Despite this, the assessment of identifiability remains sensitive to the *accuracy*, *precision*, and *reliability* of the data and methods involved. Cf EDPB (n 1) Executive summary and paras 29, 64, and 87.

results that are sufficiently *precise*, *reliable* and *confident* for the law to apply.[147] An epistemic threshold for accurate attribution, though not fully specified, seems therefore to underlie EDPB's identifiability reasoning.

7 Conclusion

The EDPB's Guidelines provide a more sophisticated account of anonymisation in comparison with previous opinions. Their central insight seems to be that anonymisation is *contextual, entity-relative, and dynamic*. The same information may be anonymous for one entity while remaining personal for another, depending on the *relevant perspective* for identification, the *capabilities* of the relevant entities, and the *means reasonably likely to be used*. Anonymity may also change over time as additional information becomes available and technological capabilities develop. The Guidelines also reject both absolute anonymity and the proposition that any theoretical possibility, however remote, necessarily brings information within the GDPR.

Two questions are crucial for anonymisation: whether the information relates to a natural person, and, if so, whether the person is identified or identifiable. The order of these questions is conceptually important. Relatedness defines the broader domain within which identifiability operates. The Guidelines, nevertheless, make relatedness increasingly contextual and entity-relative, yet they do not explain how the contextuality of relatedness differs from that of identifiability, or how the two requirements can be prevented from collapsing into one another. Treating relatedness as dependent on an entity's capabilities may exclude information from the relevant domain before future means of identification can be considered. Also, the boundaries between purpose and effect relations remain undertheorized in the Guidelines.

The EDPB also recognises two approaches to assessing anonymity: the *contextual* and the *simplified*. The contextual approach reflects the legal standard more faithfully by taking account of the relevant entity's perspective and capabilities, but may result in false positives, while the simplified approach adopts a more conservative position which may go beyond what the law strictly requires, but reduces the risk of misclassification. Both approaches can be combined to ensure effective data protection. However, the Guidelines seem to create incentives that may lead

[147] Ibid, Executive summary and paras 29, 64, and 87.

controllers to adopt the simplified approach and accept some degree of over-inclusion in exchange for greater legal certainty.

The Guidelines further clarify that identification does not require the discovery of a person's civil identity. What matters is whether the person can be *distinguished* from others within the *relevant context* in a manner that makes *differentiated treatment* possible. This conception captures contemporary processing practices such as tracking, profiling and targeting, which may operate without names. Yet, it also raises the question of whether distinguishability should be insufficient unless it can be operationalised as to affect, target, or otherwise treat the individual differently. If that is the case, this may narrow the material scope of the law, while if the relation is merely explanatory, it would not alter it. The future Guidelines should clarify which interpretation is intended.

Identifiability continues to be subject to the three traditional identifiability criteria, now termed *No Record Isolation, No Linkage,* and *No Inference*. The Guidelines correctly clarify that these criteria are indicators rather than strict conditions. Passing all three offers strong support to consider the data anonymous, but the violation of one criterion does not per se determine the legal status of the data. Nevertheless, the criteria may not occupy equivalent positions within the analysis. Record isolation appears to perform a central function because the singularisation of a person is ordinarily a precondition for identification. Linkages and inferences are often ancillary mechanisms whose importance lies in their ability to create or reinforce a profile capable of isolating the natural person.

Of the three criteria, No Inference has undergone the most substantial development. It is still fleshed out narrowly, as its definition does not contemplate the various modes of inference. In addition, it is now understood in functional terms. The inference should be *specific* and *meaningful.* It must concern a single identified or identifiable individual, rely on the data under assessment, not merely reproduce what could already be concluded from general knowledge or population-level information, and be liable to affect that person. In doing so, No Inference performs a limiting function, preventing aggregate datasets as such from becoming personal data. However, this raises important questions about the ultimate normative goal of privacy and data protection law.

The Guidelines introduce other important developments. Their treatment of mixed datasets reflects a form of tainting effect or worst-case classification, which affects both personal data and special categories of data. Moreover, they reveal a nuanced

relationship between the alethically neutral status of information *in the law* and the emphasis placed on accuracy for the information that is *subject to the law*. Data protection applies to both accurate and inaccurate data, yet (re-)identifiability claims should have a sufficient degree of accuracy to hold. These propositions may be compatible, but the relation between them is not fully articulated in the Guidelines.

Despite these unresolved issues, the Guidelines move the anonymisation debate in a constructive direction. They offer a more structured and specific account of identifiability, grounded both in the CJEU's legal reasoning and technical expertise, and their numerous examples offer valuable and practical guidance for controllers and other processing entities. Their ultimate message is that anonymity is an evidenced, documented and revisable legal conclusion. The consultation process now provides an opportunity to clarify the remaining boundaries and concerns.