European Tech Alliance

Guidelines on Anonymisation

EUTA's feedback

OCTOBER 2026

EUTA welcomes the European Data Protection Board (EDPB)'s update of Opinion 05/2014 and supports its core: anonymity is assessed in context, from the perspective of each entity, and the standard is an insignificant likelihood of identification, not zero risk. However, several provisions risk pushing the test back towards zero risk in practice. We ask the EDPB to revisit the all or nothing treatment of mixed datasets, to set clear criteria for the "means reasonably likely" test, to give proper weight to organisational and contractual safeguards, and to provide practical guidance on privacy-enhancing technologies (PETs), statistical thresholds, AI models and synthetic data. The final guidelines should also be aligned with the outcome of the Digital Omnibus.

Clear guidelines help companies allocate their compliance resources. Where the General Data Protection Regulation (GDPR) does not apply, companies should not have to apply its obligations "just in case" (e.g. data protection impact assessments (DPIAs), records of processing activities, contractual negotiations with business partners, management of data subject rights). Those resources would be better spent on data processing within the scope of the GDPR.

The guidelines rely mainly on hypothetical use cases and focus largely on what does not work. Legal certainty would come from setting out what does: the technical measures, privacy-enhancing technologies (PETs), thresholds and tests that companies can rely on. We invite the EDPB to draw on the "From pseudonymous to anonymous data" report, which sets out five real-world use cases on pseudonymisation and anonymisation from leading European-born and headquartered tech companies across sectors.

Anonymisation also matters well beyond the GDPR: the Digital Markets Act (DMA), the Digital Services Act (DSA), the Data Governance Act, the Data Act and the Artificial Intelligence (AI) Act all refer to pseudonymisation or anonymisation. We encourage the EDPB to consult the relevant regulators so that the guidelines work consistently across the *digital acquis*.

1. Timing: aligning with the Digital Omnibus

Anonymisation and the line between personal and non-personal data are central questions the co-legislators are currently working through in the Digital Omnibus, including the clarification of Article 4(1) GDPR, the new Article 29a and the recognition of PETs under Article 41a. Guidelines finalised while those provisions are still being negotiated would leave companies with two moving reference points instead of one, adding legal uncertainty rather than reducing it.

In our view, getting this right matters more than adopting it quickly. We therefore encourage the EDPB to align the final text with the outcome of the Digital Omnibus.

2. What we support

EUTA supports the core of the guidelines. In particular, we welcome:

- the confirmation that the legal standard is an insignificant likelihood of identification in reality, not absolute certainty or zero risk; and
- the contextual approach, under which anonymity is not a static property of the data but is assessed from the perspective of each entity holding or receiving it. The same data can therefore be anonymous for one entity and personal for another. This reflects the case law of the Court of Justice, notably *Breyer* (C-582/14) and *SRB* (C-413/23 P).

These principles are the right foundation. However, the draft does not yet fully reflect the call and practical guidance made in the cross-industry joint letter of November 2025, which EUTA co-signed.

Our concern is that several other parts of the draft, taken together, risk pushing the test back towards zero risk in practice: the treatment of mixed datasets, the breadth of actors the assessment must cover, the decision to set motivation aside, and the limited weight given to organisational and contractual safeguards.

3. Mixed datasets: an all or nothing rule that goes beyond the case law

Section 2.5.1 (paragraph 36) establishes that:

- a dataset is anonymous as a whole only if anonymisation is effective for every individual it contains;
- the likelihood of re-identification must remain insignificant for each of them; and
- where a dataset mixes personal data with anonymous data, the entire dataset must be treated as personal data, unless the separate parts can be processed independently.

To support this position, footnote 31 draws an analogy with *Meta v Bundeskartellamt* (C-252/21), in which the Court held that a dataset mixing special category and other personal data must be treated collectively as special category data where the parts cannot be processed separately. The analogy does not hold. In *Meta*, all the data was already personal data within the scope of the GDPR; the only question was which obligations applied. Mixing personal and anonymous data raises a different question altogether: whether the GDPR applies at all under its material scope. Transposing a rule on the level of protection to the scope of the Regulation goes beyond what the case law supports, and we ask the EDPB to remove or substantially qualify this reasoning and to clarify that GDPR obligations apply only to the personal records within a dataset, not to the dataset as a whole.

Separately, paragraph 38 states that anonymisation is itself a processing operation requiring a legal basis under Article 6 GDPR (and an Article 9 exemption where relevant). This creates a risk that anonymisation for secondary analytics is challenged as incompatible further processing. We ask the EDPB to confirm, as the Article 29 Working Party did in Opinion 05/2014, that anonymisation carried out to protect privacy, including for secondary business analytics, is compatible with the initial purpose of collection under Article 6(4) GDPR.

4. Outliers in large-scale datasets

The guidelines identify outliers, meaning records with unique combinations of attributes, as a primary source of anonymisation failure. Section 3.4.3 (paragraph 76) notes that inferring record-level information from aggregate statistics is "particularly true for outlier data". Outliers are indeed more exposed to singling out and inference.

Turning this into an all or nothing legal rule, however, creates a serious barrier for any organisation processing data at scale. In a dataset of millions of transactions there will always be outliers: a customer who bought an unusual combination of products, or one ordering from a very rare postcode. Under the current draft, the presence of such records would require an entire dataset to be treated as personal data, even where the overwhelming majority of records cannot be linked to anyone. Outliers should instead be treated as a residual risk to be managed through technical, organisational and contractual safeguards, rather than as a reason to bring the whole dataset within scope.

The exception for parts that can be processed separately does not fully address this. In complex data engineering environments such as AI model training, dynamically isolating individual outlier records is not feasible in practice.

Examples 16 and 17 show how aggregate data can fail the "no inference" test, but offer no guidance on thresholds. Companies publishing aggregate commercial reporting, or open datasets for research, therefore lack clarity on when data is sufficiently aggregated. We invite the EDPB to set out concrete statistical safe harbours (such as minimum k-anonymity values), which Data Protection Authorities (DPAs) would also apply in enforcement.

The draft also extends the "no inference" test to AI models and synthetic data (paragraphs 66 to 82 and footnote 45). Treating model weights or synthetic outputs as personal data because of theoretical membership inference risks would hold back AI development and the use of PETs in Europe. We recommend that the guidelines confirm that AI models and synthetic datasets are anonymous unless individual training records can be reconstructed with high accuracy and confidence.

5. "Means reasonably likely": the test needs clear criteria

Whether data is anonymous depends on the means reasonably likely to be used to identify a person (Recital 26 GDPR). The term "reasonably" appears 57 times in the draft, yet the guidelines offer no criteria for applying it and the glossary does not address it. We are not asking for a rigid definition. We are asking for a practical toolbox that allows controllers and DPAs to distinguish realistic scenarios from purely theoretical ones. Without it, the outcome of an assessment will depend largely on the individual assessor, and the contextual approach the EDPB rightly endorses will lose much of its practical effect. As a result, companies may abandon innovative data and AI projects for fear that a DPA will reach a different conclusion on anonymisation than their own.

Key questions remain open. Does "reasonably" refer to what is objectively possible, or to what is probable? Can current experience and practice confirm or rule out that a scenario is reasonable? What role, if any, does an entity's presumed good or bad faith play?

Two features of the draft make these questions more pressing:

- **Relevant entities.** Paragraph 30 states that relevant entities may include "anybody directly or indirectly receiving the data, including unauthorised and malicious actors". Taken literally, this requires controllers to account for every conceivable actor, which moves the test back towards absolute risk.
- **Motivation.** Paragraph 31 cautions against using a lack of motivation as a factor. Yet an entity's incentives, and the legal, contractual, commercial and reputational consequences it would face, are central to what is reasonably likely in reality. These factors should not replace an assessment against a competent and motivated intruder, but should sit alongside it: setting motivation aside altogether departs from the realistic assessment called for in Breyer and SRB.

To illustrate: company X shares a pseudonymised dataset with Y, a long-standing analytics partner. The partner does not hold the means to reverse the pseudonymisation, is bound by

confidentiality obligations and a contractual prohibition on re-identification, and has a clear commercial and reputational interest in maintaining the relationship. Re-identification may be objectively possible, for instance by paying a third party to attempt it, but nothing suggests it is probable. The guidelines should make clear whether, and under which conditions, such a scenario falls outside the means reasonably likely to be used, in which case the data is anonymous from Y's perspective and falls outside the scope of the GDPR.

Paragraph 30 also calls for proportionate rules on intra-group data sharing. Corporate groups with several entities share aggregated or tokenised data under strict access controls, including to manage joint controllership risks. Treating affiliates of the same group like adversarial third parties would make anonymisation very difficult to use in practice. The guidelines should clarify that intra-group and affiliate recipients operating under documented technical and contractual access controls form a distinct, lower-risk category.

Paragraph 21 provides that where a recipient may pass data on to a party able to identify individuals, the data must also be treated as personal for the transferring entity. This exposes companies sharing keyless data to open-ended liability if a downstream recipient suffers a breach or transfers the data unlawfully. The guidelines should clarify that anonymity is assessed on the basis of the means reasonably foreseeable at the time of transfer, and that originators are not liable for unauthorised downstream actions.

Under paragraph 15, where data is personal from the controller's perspective, service providers processing it on the controller's behalf are processors under Article 28 GDPR, even if they only receive keyless, tokenised data they cannot re-identify. Cloud and analytics providers would then face the full set of Article 28 obligations, regardless of their actual technical position. We ask the EDPB to allow such providers to be assessed from their own perspective, or to set out a simplified framework for providers with no means of re-identification.

6. Organisational and contractual safeguards

Paragraph 34 accepts that contractual terms may affect the means reasonably likely to be used, but only as a complement to technical measures. It also states that a contractual prohibition "should not be treated as a prohibition by law". In practice, access restrictions, internal policies and legally binding non-reidentification commitments are, together with technical measures, what makes re-identification unlikely in reality.

This sits uneasily with the case law in Breyer and SRB, and with the GDPR's own reliance on contractual safeguards such as standard contractual clauses and Article 28 data processing agreements.

The guidelines should recognise enforceable, monitored and legally binding contracts as valid barriers to re-identification and give these safeguards their proper weight, in particular where they address the residual risk posed by a small number of records.

Finally, paragraph 35 requires controllers to monitor technological developments and reassess re-identification risks over time, without saying what should trigger a reassessment. Open-ended monitoring obligations create lasting legal uncertainty, for instance after onward transfers, the deletion of source data, or third-party breaches that create new linkable data. We ask the EDPB to set clear criteria for when a reassessment is required.

7. Privacy-enhancing technologies

The draft does not address PETs such as differential privacy, synthetic data or secure multi-party computation, which European companies already deploy and which the Digital Omnibus proposes to recognise. Guidance on how these techniques are assessed would considerably increase the practical value of the guidelines and support their uptake. Practical examples could show how PETs contribute to anonymisation and which standards (such as ISO standards) companies can rely on.

8. Recommendations

EUTA calls on the EDPB to:

- **Align with the Digital Omnibus:** finalise the Guidelines in light of the outcome of the negotiations.
- **Replace the all or nothing rule in paragraph 36 with a risk-based approach:** where state-of-the-art techniques make identification insignificant for the overwhelming majority of records, the dataset should retain its anonymous status as a whole, provided the residual risk from outliers is mitigated through technical, organisational and contractual safeguards. These could include restricted access, internal policies prohibiting the singling out or targeting of outliers, and binding non-reidentification agreements with any recipient of the data. At a minimum, clarify that GDPR obligations apply only to the personal records within a dataset, not to the dataset as a whole.
- **Remove or qualify the analogy with Meta v Bundeskartellamt** *in footnote 31, which concerns the level of protection, not the material scope of the GDPR.*
- **Provide criteria for "means reasonably likely"**, for instance in the glossary, making clear that "reasonably" should be read close to "probable" rather than merely "possible", and that theoretical scenarios are not sufficient.
- **Clarify the evidence needed to show that an identification scenario is unrealistic**, and explain how the wording "cannot be ruled out" fits with the acceptance that zero risk is not required.
- **Limit relevant entities to those realistically able to access the data**, and allow incentives, motivation and legal or contractual constraints to be taken into account alongside, not instead of, an assessment against a competent and motivated intruder.
- **Recognise organisational and contractual safeguards** as full elements of the assessment, not only as a complement to technical measures, and provide practical examples showing how access restrictions, contractual commitments and compliance checks work together in the assessment.
- **Add dedicated guidance on PETs**, including differential privacy, synthetic data and secure multi-party computation.
- **Confirm the compatibility of anonymisation:** recognise that anonymisation is compatible with the initial purpose of collection under Article 6(4) GDPR.
- **Provide safe harbours:** set out statistical thresholds for aggregate data, and confirm when AI models and synthetic data are anonymous.
- **Clarify the position of recipients:** address intra-group recipients, transferring entities and service providers receiving keyless data, in line with the contextual approach.
- **Define triggers for reassessment:** set out when anonymity must be reassessed over time.

EUTA and its members remain at the EDPB's disposal to discuss these comments and to share practical examples from their data practices.

About the European Tech Alliance

EUTA represents leading European tech companies that provide innovative products and services to more than a billion users[1]. Our 41 EUTA member companies from European countries are popular and have earned the trust of consumers. As companies born and bred in Europe, for whom the EU is a crucial market, we have a deep commitment to European citizens and values.

With the right conditions, our companies can strengthen Europe's resilience and technological autonomy, protect and empower users online, and promote Europe's values of transparency, rule of law and innovation to the rest of the world.

The EUTA calls for boosting Europe's tech competitiveness by having an ambitious EU tech strategy to overcome growth obstacles, making a political commitment to clear, targeted and risk-based rules, and enforcing rules consistently to match the globalised market we are in.

Our members



Contact

Visit us at www.eutechalliance.eu.

The European Tech Alliance, Square de Meeûs 35, 1000 Brussels, Belgium

[1] It reflects users, consumers and business customers from EUTA member companies, per year. It includes overlaps but illustrates the reach and impact of our services.