# Comments on EDPB Guidelines 02/2026 on Anonymisation

Submission to the EDPB public consultation

Florian Feltz
Universität zu Lübeck

Prof. Dr. Esfandiar Mohammadi
Universität zu Lübeck

30 October 2026

## Executive summary

The EDPB's Guidelines 02/2026 on Anonymisation provide a welcome and timely framework for assessing whether data can be considered anonymous under the GDPR. In particular, the distinction between the contextual and simplified approaches, together with the criteria of *No Record Isolation*, *No Linkage* and *No Inference*, offers a useful structure for technical and legal analysis. At the same time, the contextual approach can make outcomes difficult to predict and reproduce where assessments depend on uncertain assumptions about relevant entities, auxiliary information and foreseeable developments. Further guidance is needed to make the reasoning behind such assessments sufficiently transparent and to strengthen legal certainty, especially in borderline cases.

To make the framework easier to apply consistently in practice, we recommend further clarification in four areas:

1. explain how assumptions underlying the *means reasonably likely to be used* (MRLTBU) assessment should be justified and documented, including assumptions about access to auxiliary data and onward transfers;

2. provide practical guidance on evaluating probabilistic record linkage and the phrase "certainty or high likelihood", without imposing a universal numerical threshold;

3. clarify the practical operation of the reassessment envisaged in paragraph 35 when re-identification risks change over time, particularly where data have already been made publicly available; and

4. explain how formal technical privacy guarantees, including differential privacy, may support an assessment under the three criteria without being treated as automatically equivalent to legal anonymity.

These recommendations are not intended to replace the contextual legal standard with rigid rules or to require controllers to test every attack that is merely conceivable. Their purpose is to make assessments more transparent, reproducible and proportionate, while preserving the flexibility required by the variety of data and processing contexts.

## 1. Recommendations

### 1.1 Make contextual assumptions transparent and explain how they are justified

The contextual approach necessarily depends on assumptions about the relevant entities, their access to data, the availability of auxiliary information, and the methods they are reasonably likely to use. The Guidelines themselves acknowledge that this approach can be complex and may lead to false positives where an assessor is unaware of a relevant entity's means (paragraph 47). Context is indispensable to the legal standard, but it also means that the outcome may depend on assumptions that are uncertain, difficult to verify or open to disagreement. Without clearer guidance on how those assumptions should be justified and bounded, controllers may struggle to predict whether similar cases will be assessed consistently. This raises a practical question of legal certainty, particularly in borderline cases; it is not an argument for replacing the contextual standard with a rigid checklist. The Guidelines already identify objective factors relevant to the assessment and emphasise the importance of documenting the anonymisation process. More concrete guidance on the content and reasoning of that documentation would improve consistency across assessments.

We recommend that the EDPB provide a practical documentation schema, proportionate to the complexity and risk of the case, covering at least:

- the entities and perspectives considered relevant, including entities that may receive or obtain the data indirectly;
- the auxiliary datasets, technical capabilities, access routes and re-identification methods considered;
- the factual basis for treating a particular method or source of additional information as available or unavailable, or as reasonably likely or unlikely to be used;
- the significance of legal, contractual, organisational and technical access restrictions, including the evidence supporting the expected effectiveness of those restrictions; and
- material uncertainties, plausible alternative assumptions and the date or circumstances to which the assessment applies.

For example, a controller may assume that a potential attacker cannot access database X. The assessment should make clear why this is a justified assumption in the circumstances, rather than merely state it as a premise. Illustrative examples of sufficient and insufficient reasoning would help practitioners distinguish evidence-based assumptions from unsupported optimism. The guidance should not require certainty about every possible attacker; it should require a reasoned and reviewable explanation of the assumptions that materially affect the conclusion.

The Guidelines already address chains of means and the possibility that an entity may obtain help from a third party. The remaining practical question is how far an assessment should extend along a chain of recipients and what makes a subsequent action reasonably foreseeable. Consider a recipient B that commissions third party C to process data under access restrictions, after which C publishes the data publicly contrary to those restrictions. It would be helpful to clarify how the original assessment should treat this possibility: for example, which facts about C, the contractual and technical safeguards, the surrounding context and the likelihood of onward disclosure are relevant at the outset, and how a genuinely unforeseeable later event should affect subsequent assessments. This is particularly important because public availability may materially expand the set of entities able to access and link the data.

We do not propose that every theoretically possible onward transfer must be included. Rather, the Guidelines should give practical examples showing how to determine the relevant boundary of the chain under the MRLTBU standard.

### 1.2 Clarify the assessment of probabilistic record linkage

Paragraph 60 describes *No Linkage* by reference to records that can be linked with "certainty or high likelihood" to the same individual. Paragraph 64 also recognises that linkage may remain possible when records are inaccurate or have been deliberately modified, and paragraph 83 refers to sophisticated combinations of probabilistic techniques. These are important acknowledgements; however, practitioners would benefit from more concrete guidance on how uncertainty affects the assessment.

We recommend that the EDPB explain, with examples, how to evaluate probabilistic matching in

cases where attributes are noisy, change over time or are measured differently. Examples might include physiological measurements recorded at different times or with different devices, where a match is possible but not exact. Guidance should address how the reliability and precision of a proposed match, the possibility of false matches and missed matches, and the quality of the available auxiliary information should inform the conclusion under *No Linkage*.

The aim should not be to define a universal probability threshold for "high likelihood". A single numerical threshold is unlikely to be appropriate across all contexts, datasets and potential consequences. Instead, the Guidelines should explain which methodological considerations and forms of evidence are relevant, and provide boundary examples illustrating when residual uncertainty is or is not material to the assessment. This should also cover profiling or behavioural matching, where a record may be associated with a person through a collection of probabilistic signals rather than a stable direct identifier.

The guidance would further benefit from distinguishing the reliability of identifying a record as belonging to a particular individual from the possibility of an erroneous attribution. This distinction matters when assessing the effectiveness of a linkage method and when communicating the limitations of the resulting assessment. It need not settle separate questions of harm or liability arising from a mistaken attribution, which are outside the narrow technical issue addressed here.

### 1.3 Clarify reassessment when re-identification risks change over time

Paragraph 35 appropriately recognises that re-identification may become more likely as technology advances and additional information becomes available. It states that data previously considered anonymous should again be considered personal if the likelihood of identification rises to a level that is no longer insignificant, and recommends periodic reassessment where possible and appropriate.

Consider an online grocery retailer sharing historical purchase records with a research partner for scientific purposes. Before sharing, direct identifiers such as names, email addresses and customer account identifiers are removed. The partner receives only the transaction data, including purchased products, quantities, product categories and the day and time of each purchase. Access to the original customer database and the corresponding identity mappings is effectively restricted, and the partner has no access to other suitable datasets containing identified purchase histories of the same individuals. In this context, and in the absence of reasonably accessible auxiliary information for matching the records to named individuals, identification may initially appear unlikely from the recipient's perspective. The purchase records do not directly reveal the identity of the customers, and conventional record linkage would appear difficult without sufficiently comparable auxiliary information.

However, the situation may change over time. For example, the recipient may later gain access to identified purchase histories covering a different time period, whether through a new data-sharing arrangement or the emergence of an additional data source. Persistent individual purchasing patterns may then allow the records in the original dataset to be linked to identified customers, even though the two sets of records do not cover overlapping periods. Research by Tournier and de Montjoye (2022) demonstrated this possibility using a grocery shopping dataset of approximately 85,000 individuals, achieving correct identification in 65.2% of cases using auxiliary information from a non-overlapping period.

This example illustrates how data that may initially be considered anonymous from the perspective of a particular recipient can become identifiable as the availability of auxiliary information changes. It raises the question of how the initial assessment under the MRLTBU standard should account for the potential future availability of such information, particularly where the relevant data sources do not yet exist or are not reasonably accessible to the recipient. It also illustrates why the temporal validity of an anonymity assessment may depend on changes in the surrounding information environment, even where the released dataset itself remains unchanged.

Recital 26 GDPR requires the assessment to take account of available technology at the time of processing and technological developments; paragraph 29 of the Guidelines refers more specifically to reasonably foreseeable technological developments. These provisions make the temporal dimension relevant, but further guidance would help clarify how far into the future an initial anonymity assessment should look. The appropriate time horizon may depend on the nature of the data, the purpose and context of processing, and the intended lifetime and dissemination of the dataset. Some attributes may lose identifying significance as circumstances change, whereas other data, including genetic data, may remain informative over long periods and may also reveal information about biological relatives or descendants. The Guidelines should clarify how such differences should inform a proportionate assessment horizon,

without prescribing one universal period or requiring controllers to predict developments that could not reasonably have been foreseen.

The temporal scope of an initial assessment should also be distinguished from the question of when a later reassessment is warranted. We recommend that the EDPB provide risk-based examples of events that may warrant reassessment, such as:

- the publication or discovery of new auxiliary datasets;
- material advances in re-identification techniques or in the accessibility of such techniques;
- a change in who can access the data, including a new recipient or onward transfer; or
- a security incident that undermines an assumption on which the original assessment relied.

The Guidelines should also explain how entities can determine an appropriate reassessment interval where no specific trigger is known, while recognising that a single fixed period will not fit every use case. Such guidance should take account of the nature of the data, the intended use, the expected lifespan of the dataset and the practical likelihood of changes in relevant capabilities or auxiliary information. Clarification would be particularly useful on whether, and on what basis, an assessment may reasonably assume that the relevance of certain data will diminish over time, as opposed to data whose identifying or inferential potential may persist for a much longer period.

A particular difficulty arises where data have already been published publicly and cannot realistically be recalled. In such cases, later reassessment may identify a materially increased risk but may not offer a simple means of reversing the disclosure. The Guidelines should discuss this situation explicitly and explain how paragraph 35 should be understood in relation to processing that takes place after circumstances change, without prejudging questions of responsibility for earlier processing. Practical guidance on monitoring, mitigation and response would be especially useful for publicly released data.

### 1.4 Position formal privacy guarantees within the assessment

Paragraph 71(b) proposes comparing the outcome of an inference based on the output of an anonymisation process applied to the original dataset, $A(D)$, with the outcome based on a modified dataset, $A(D')$, from which the relevant individual's record has been removed. If the inference is available only from the first output, this indicates that it relies on the individual's data. This counterfactual comparison is a useful way of examining whether an inference depends on a particular record.

We recommend that the EDPB clarify how this comparison relates to formal privacy guarantees, particularly differential privacy. Differential privacy provides a quantitative bound on how much the output distribution of a mechanism can change between neighbouring datasets, such as datasets that differ in the inclusion of one individual's data. It may therefore offer relevant technical evidence about the influence of an individual's data on released outputs.

The Guidelines should explain whether, and under what conditions, such guarantees can support the analysis under *No Inference* and how they relate to the other criteria. They should also clarify the limits of that evidence, including the importance of the mechanism, the definition of neighbouring datasets and the chosen privacy parameters. A formal guarantee should not, by itself, be presented as an automatic determination that data are anonymous under the GDPR; the legal assessment and the relevant processing context remain important.

A concise explanation of the relationship between the Guidelines' individual-record comparison and formal guarantees would help practitioners understand when a rigorous technical method can strengthen an assessment, which risks it addresses, and which questions still require contextual analysis.

## 2. Conclusion

The Guidelines establish a valuable framework for assessing anonymity. The clarifications proposed here are intended to make that framework more usable in practice by improving the transparency and predictability of contextual assessments, making probabilistic linkage assessments easier to substantiate, clarifying the temporal scope of initial assessments and subsequent reassessment, and explaining how formal technical guarantees can contribute to the analysis.

The common objective is not maximal formalisation for its own sake. It is a proportionate approach in

which important assumptions and uncertainties are visible, conclusions can be reviewed by others, and technically grounded evidence is used consistently without losing sight of the legal context.

## References


[1] European Data Protection Board (2026), *Guidelines 02/2026 on Anonymisation*, Version 1.0, adopted 7 July 2026, in particular paragraphs 21, 27–35, 41, 43, 47, 60–64, 71, 83 and 89–90. Available at: EDPB Guidelines 02/2026 webpage.

[2] Regulation (EU) 2016/679 (General Data Protection Regulation), Recital 26. Available at: EUR-Lex.

[3] Cynthia Dwork and Aaron Roth (2014), *The Algorithmic Foundations of Differential Privacy*, Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. doi:10.1561/0400000042.

[4] Tournier, A. J., and Y.-A. de Montjoye (2022). *Expanding the attack surface: Robust profiling attacks threaten the privacy of sparse behavioral data*. Science Advances, 8(33), eabl6464. https://doi.org/10.1126/sciadv.abl6464.


## Signatories

The following individuals endorse this submission in their personal capacity. Their affiliations are provided for identification purposes only and do not imply institutional endorsement.

| **Name** | **Affiliation** |
|---|---|
| Prof. Dr. Esfandiar Mohammadi | Universität zu Lübeck |
| Florian Feltz | Universität zu Lübeck |