# Response to EDPB Guidelines 02/2026 on anonymisation

Comments on the draft guidelines, version 1.0, adopted 7 July 2026, submitted to the public consultation (8 July to 30 October 2026).

| RESPONDENTS | ORGANISATION | CONTACT |
| --- | --- | --- |
| Alex Murphy and Man Yong Toh, as named individuals for Cadren | Cadren, a foundation being set up in the Netherlands | info@cadren.org<br>cadren.org |

## Summary

- **Attack tests can mislead without a reference point.** A widely used open-source attack tool rated a release that holds no information about anyone at about 13% singling-out risk. A synthetic release scored the same, so it looked risky while being no riskier than knowing nothing. The reverse mistake, reading a test artefact as safety, is just as easy.
- **We propose wording the EDPB can use.** Residual likelihood should count as insignificant where, for each relevant recipient, the upper 95% bound of attack success above a reference release stays below a threshold set for the context. A six-point checklist for such evidence is in the annex.
- **We offer to help test it.** We will repeat the test on a public dataset the EDPB chooses, or within a data protection authority's sandbox, and share the code and signed record.

## Who we are

Cadren is a foundation being set up in the Netherlands to publish an open standard for testing whether data releases and AI decision systems hold up under attack. We sell no anonymisation tool. The respondents are also setting up Cadren Labs, which will offer testing services. Our proposals would benefit testing providers, including us. To keep the method neutral, we publish it as section S1 of the Cadren Open Standard, free for anyone to use, and any qualified tester can apply it.

Respondents: Alex Murphy (CEO) and Man Yong Toh (CCO), responding as named individuals for Cadren. Contact: info@cadren.org.

## Our position

We support the contextual approach: anonymity judged against what a specific recipient could know and do. It matches how re-identification works in practice. To make it workable, we ask the EDPB to:

1. **Recognise documented adversarial testing as evidence.** A test run per recipient, against all three criteria in paragraph 52 (no record isolation, no linkage, no inference), compared with a control baseline and reported with confidence intervals, should count as evidence that residual likelihood

is "insignificant". This matters in practice for analytics processed on the server, where one data stream goes to several recipients (an analytics provider, an advertising platform, a data buyer) and anonymity can hold for one recipient but not another.

2. **Say what good evidence looks like.** Give a method, or worked examples, for showing that likelihood is "insignificant", including how to treat the uncertainty of the test itself, and a reference release that holds no information about anyone, run through the same attacks, so that a test's own bias is not read as risk or as safety.
3. **Make synthetic data and trained models testable.** Confirm that they are judged on the same three criteria and say which tests are expected.
4. **Set triggers for re-testing.** Name the events that should prompt a new assessment: a new recipient, new auxiliary data, a new release, or cheaper attacks.
5. **Recognise independent testing.** Testing by a party that did not produce the release adds weight, as it does in security and financial audit. It should be encouraged, not required.

We also ask that the final text stays consistent with the motivated intruder test used by the UK ICO, so that firms working across both regimes can use one method.

## Evidence: the tests work, but they need a reference point

We attacked three common "anonymised" releases of the same public benchmark dataset (UCI Adult census data, 8,000 records, no real identities), plus a reference release that holds no information about any individual (each column sampled independently). Each criterion was tested for three recipient profiles: an insurer, an employer or HR data broker, and a neighbour. The table shows the worst result per criterion, at 2,000 attacks per test.

| Release | Singling out | Linkage (worst recipient) | Inference (worst recipient) |
|---|---|---|---|
| A: names removed, attributes kept | 99.9% (99.8-100%) | 99.9% (99.8-100%) | 79.9% (77.1-82.6%) |
| B: generalised (age bands, top-coding) | 99.9% (99.8-100%) | 19.8% (17.8-21.8%) | 41.9% (37.4-46.4%) |
| C: fully synthetic (CTGAN) | 13.8% (12.0-15.5%) | 1.7% (0.0-3.7%) | 1.7% (0.0-7.2%) |
| Reference: no information about any individual | 13.1% (11.4-14.9%) | 0.1% (0.0-0.8%) | 1.0% (0.0-5.3%) |

Note on singling out: in this method the singling-out test has no working baseline, so chance guesses show up as risk. That is why the reference release, which holds no information about anyone, still scores about 13%. The tool's singling-out attack is closest to the No Record Isolation criterion together with the singling-out check in paragraphs 98-100; it is not a direct count of unique records (paragraph 57). By "recipient" we mean a relevant entity in the sense of paragraph 30. For inference, the baseline from held-out records serves the test in paragraph 71: an inference counts only if it could not be drawn from data about the population at large.

Risk is the attacker's success rate above a baseline guess made from held-out control records, with 95% confidence intervals; 2,000 attacks per test. A run at 500 attacks and a second random sample

of 8,000 records gave the same picture (on singling out, Release C 15.0% against 15.7% for the reference in the second sample), and a reference built from the generalised columns scored similarly (14.8-16.8%). Method: the open-source Anonymeter framework (Giomi et al., PoPETs 2023), fixed random seed, results in a signed record we can share with the EDPB.

What this shows for the guidelines:

- **A test needs a reference point.** The singling-out test scored the reference release at 13-16%, although that release holds no information about any person, and the figure did not fall with more attacks. The synthetic release scored the same. Without the reference, Release C would have looked like a release that lets one in seven attempts single someone out; in fact, on singling out it shows no measurable advantage over a release that knows nothing. Guidance should expect a reference release, or another calibration, next to the control set, so that test artefacts are not read as risk, or as safety.
- **Generalisation can look safe and still fail.** Release B cut linking to about 20%, but singling out stayed near 100%, and an employer could infer marital status well above chance (32-42% across two random samples, against 1-2% on the reference).
- **Synthetic data is not automatically anonymous.** In separate runs on the same dataset, a synthesizer built from decision trees with small leaves copied 13-25% of its training records exactly, against under 1% for the reference release. Showing no advantage on one attack is not proof of anonymity; membership and copy tests are needed too.
- **The result depends on the recipient.** The same release scored differently for each profile, which supports the contextual approach and argues for testing per recipient.
- **Uncertainty matters.** Even where a point estimate is close to 0%, the upper bound reaches up to about 14% at 2,000 attacks. Guidance on "insignificant" should say whether it applies to the estimate or its upper bound, and how many attacks are enough.

**Limits of this evidence.** This is a demonstration on one public benchmark dataset with one open-source attack tool, not a finding about any real dataset or controller. Other tools and data may behave differently, and we invite others to repeat the test; the code, seeds and signed record are available on request. We would also repeat it on a public dataset the EDPB chooses, or within a data protection authority's sandbox.

## Specific comments

Paragraph numbers refer to the draft Guidelines 02/2026, version 1.0, adopted 7 July 2026.

| Topic | Paragraph | Comment and proposed change |
|---|---|---|
| "Insignificant" residual likelihood | 22, 26, 36 | Add that a controller may show insignificance through documented adversarial tests per recipient, against all three criteria, with a control baseline and confidence intervals. Proposed wording: "Residual likelihood may be considered insignificant where, for each relevant recipient, the upper 95% confidence bound of attack success above a reference release remains below a threshold appropriate to the context." |
| AI lowering the cost of re-identification | 92-93 | Welcome. Footnote 3 rightly does not require earlier assessments to be redone. We suggest naming a material fall in attack cost, including from AI-driven tools, as a reason to reassess under paragraph 35, and that tests include automated and agent-driven attackers. |

| Topic | Paragraph | Comment and proposed change |
|---|---|---|
| Synthetic data and trained models | 77, 82 | Confirm they are assessed on the same three criteria as any release, and say that membership inference and copy tests are expected; one synthesizer in our tests copied 13-25% of its training records exactly. |
| Contextual versus simplified approach | 30, 45-49, 89 | Support. Clarify which recipient profiles must be tested when a release goes to many recipients or to the public. |
| Contracts and organisational measures | 34, 90 | Agree contracts alone are not enough. Say that measured attack results can show how much a contract actually reduces likelihood. |
| Anonymisation as processing | 38 | Note that testing on the original data is also processing; confirm a legal basis can cover testing done for this purpose. |
| Periodic review | 35, 41-42, 88 | Name triggers for review: new recipient, new auxiliary data, new release version, a material fall in attack cost. |
| "One successful identification" | 87 | Define a successful identification relative to a baseline. Read literally, every release fails, because any attack makes some correct guesses by chance, even against data that holds no information about anyone (our reference release). |
| Small datasets | General | Say how controllers with small datasets can show that likelihood is insignificant. On a public 1,000-record loan dataset, our confidence intervals were wider than any reasonable threshold, so no release could be shown to pass or fail. |

## Annex: what evidence of insignificance should contain

1. **Recipient:** who receives the data, and what they could already know or obtain (auxiliary data, means, motives).
2. **Criteria:** results for all three criteria in paragraph 52 (no record isolation, no linkage, no inference), not one alone.
3. **Attacks:** the attacks used, how many, and whether they include automated and AI-driven attackers.
4. **Control:** a baseline guess from held-out records, so success by chance is not counted as risk.
5. **Reference:** a release that holds no information about anyone, run through the same attacks, so the test's own bias is visible.
6. **Decision:** the upper 95% confidence bound of success above the reference, compared with a threshold stated in advance, plus the triggers for re-testing.

Sources: EDPB consultation page; Anonymeter on GitHub.

Cadren Foundation (in formation) · info@cadren.org · cadren.org. This response may be published by the EDPB. The demonstration uses public benchmark data; results are statistical statements about the attacks listed, not legal conclusions.