Contribution to the Public Consultation on EDPB Guidelines 02/2026 on Anonymisation

(Version 1.0, adopted 7 July 2026)

Contributor: Thiago Francis Alves da Silva

Capacity: Independent software architect and founder; designer of privacy-engineering and regulatory-compliance software for small and medium-sized organisations (records of processing under Article 30 GDPR, cryptographic erasure, audit-trail integrity, regulatory reporting). Based in Nice, France.

Contact: thiagofrancisalves140@gmail.com

Date: 4 August 2026

General remarks

I welcome the adoption of Guidelines 02/2026 and the decision to update Opinion 05/2014 in light of CJEU case law (notably C-413/23 P EDPS v SRB) and a decade of advances in re-identification research. As a practitioner who implements data-protection controls for small and medium-sized organisations, I find the three-criteria framework — No Record Isolation, No Linkage, No Inference — conceptually sound and, importantly, empirically testable. My four comments pursue a single goal: making the assessment the Guidelines require **operational, reproducible and affordable**, so that rigorous anonymisation does not become a capability reserved to the largest controllers.

1. Specify the evidentiary content of the empirical test the Guidelines already imply

Paragraph 41 requires documentation of the anonymisation process, “including for the testing of the supposedly anonymous datasets”. Paragraph 71(b), with footnote 43, goes further and describes a precise empirical method for establishing whether an inference is meaningful: comparing the outcome of an inference drawn from data produced from the original dataset A(D) against data produced from a modified dataset excluding the target's records A(D'). This differential design is exactly the structure of documented, peer-reviewed attack-based evaluation methodology already in production use (main attack vs. control vs. naive baseline, for singling-out, linkability and attribute-inference attacks). Paragraph 84 encourages “the development, dissemination and application of scientifically grounded expertise”.

Request: that the final Guidelines take the natural next step and specify the minimum evidentiary content of such a test report: (i) the attacker model assumed, ideally drawn from a standard taxonomy based on the archetypes in paragraph 30; (ii) the auxiliary information assumed available; (iii) the attack classes run against each of the three criteria; (iv) the number of attacks and measured success rates against control and baseline; and (v) the resulting confidence assessment. Naming methodological requirements — not products — would give controllers of every size a common, auditable evidence format and give supervisory authorities comparable submissions.

2. Provide non-binding quantitative anchors for residual risk

The Guidelines necessarily retain qualitative thresholds: likelihood that is “insignificant in reality” (paragraph 22), results “sufficiently close” with “sufficiently high confidence” (paragraph 87), “adequate safety margins” (paragraph 88). In implementation, however, every engineering decision is quantitative: a k value, a suppression threshold, a differential-privacy budget, a measured attack success rate. Without any indicative anchors, identical datasets risk being judged differently across Member States, and smaller controllers cannot predict whether a design will pass.

Request: a non-binding annex of worked numerical examples — dataset archetypes with parameter choices, measured residual-risk ranges, and the reasoning for accepting or rejecting them — drawing where possible on existing supervisory practice. Clearly marked as illustrative, such anchors would materially reduce divergence without freezing the state of the art.

3. Make the periodic-reassessment obligation proportionate and bounded

Paragraph 35 rightly notes that re-identification likelihood increases over time, recommends periodic reassessment as good practice, and states that once the likelihood is no longer insignificant the controller “will be accountable for any processing of that personal data”. Three clarifications would greatly help practice: (a) **triggers and cadence** — reassessment tied to identifiable events (publication of significant new auxiliary datasets; emergence of new attack classes; material capability advances in widely available models, consistent with the Guidelines' own observations on agentic AI in paragraphs 92–93) and proportionate to the sensitivity of the release and the controller's size and means; (b) **released data** — where data was lawfully assessed and released as anonymous and later ceases to satisfy the criteria, what is realistically expected of the original controller, given that public dissemination cannot be recalled; (c) **continuity** — footnote 3 helpfully grandfathers assessments made under Opinion 05/2014; analogous clarity for assessments made in good faith under the final Guidelines, later overtaken by technique advances, would encourage rather than penalise documented testing.

Request: indicative trigger events, an explicit proportionality statement for SMEs, and express treatment of the release-and-forget scenario.

4. One methodological bridge for synthetic data, AI training data and trained models

Paragraphs 77 and 82, with footnote 45, already extend the analysis to AI models and synthetic data and cross-reference Opinion 28/2024. Practitioners today face three adjacent assessments with unclear articulation: whether a synthetic dataset is anonymous; whether a training corpus is anonymous; and whether a trained model contains personal data. All three are, at bottom, answerable by the same attack-based evidence (membership inference, attribute inference, singling out), and paragraph 71(b)'s differential method — testing with and without the target's records — is precisely the reference test for meaningful inference in generative settings.

Request: a dedicated sub-section confirming that synthetic data and model outputs are assessed under the same three criteria and the same evidentiary standard as any other transformation, designating the paragraph 71(b) differential comparison as the reference method in generative contexts, and cross-referencing Opinion 28/2024 so that one documented evaluation can serve both the dataset-level and the model-level assessment. Coherence here would remove a significant source of duplicated cost.

Closing

The relative, context-dependent approach the EDPB has adopted is, in my view, the correct one. Its success will depend on whether small controllers can actually perform — and afford — the assessment it requires. I would be glad to contribute further practitioner feedback.