PUBLIC CONSULTATION — COMMENTS

Comments on the EDPB Guidelines 02/2026 on Anonymisation (Version 1.0, adopted 7 July 2026)

Submitted by Hoferer GmbH, a tax advisory and statutory audit firm in Oppenau, Germany, as a practitioner contribution. Contact: Christian Hoferer, Managing Director. Date: 18 October 2026.

1 · Who we are and why we comment

Hoferer GmbH is a tax advisory and audit firm with around 20 employees, bound by professional secrecy under § 203 of the German Criminal Code (StGB), which also covers the third parties and organisations named in our files. Our interest in these Guidelines is practical. We have used an external AI service for research and analysis on client documents since 2026, and we have built and operate our own on-premises pipeline that pseudonymises those documents before anything leaves the firm. We write as users of the standard the Guidelines describe, not as lawyers. We have no commercial interest in anonymisation technology; we do have an interest, which we state openly, in a clear standard for the architecture described below, because it determines how we document our processing and how we assess the risks of using a provider outside the EU/EEA.

We welcome the draft. Its perspective-dependent test (paras 6–7), its cautious treatment of contractual prohibitions (para 34) and of access restrictions (para 90), and its warning against relying on a supposed lack of motivation (para 31) match our own experience: what protects our clients is the technical decoupling of the data from the key, not the contract with the recipient. Our comments concern five situations that the draft does not yet address and that we expect to become common among small controllers.

2 · Our use case in brief

The documents we work with are unstructured text: contracts, tax assessments, correspondence, financial statements, payroll records, scanned letters. Before a document is given to the AI service, our local pipeline (a) recognises identifying content with deterministic rules and a locally run recognition model that operates **offline**, (b) replaces each identified value with a consistent token per engagement — for example `[PERSON_001]`, `[IBAN_001]`, `[STRASSE_001]` — so that the same person or account carries the same token across all documents of that case, (c) keeps the token map exclusively on our premises, and (d) has a staff member check the result before release. Only the tokenised text is transmitted to the AI service provider, a commercial provider established outside the EU/EEA, under an Article 28 data processing agreement with standard contractual clauses and a contractual prohibition on training. Results come back tokenised and are re-identified locally. Every run is documented, the quality of each run is fed back into the pipeline, and the provider never receives the key.

Under the draft Guidelines this is, from our perspective, pseudonymisation: we hold the key, and we continue to treat the data as personal (para 37). From the provider's perspective the data are — by para 15 — personal data as well, because the provider processes them on our behalf. We agree with both results. What we lack is guidance on the questions that follow from them.

3 · Comments

C1 — The hand-over point with a locally held key (section 2.2, paras 11–15, Examples 3 and 4)

Example 3 comes close to our case: a controller transmits excerpts of records it can identify to an agency acting on its behalf, and the controller's perspective applies to both. Example 4 is the opposite case of an independent recipient. What neither example addresses is the question that matters in practice for the case in between — **a controller that pseudonymises locally and hands tokenised text to a processor that never obtains the key** — namely whether the processor itself could identify anyone. We would welcome an example for this case and a statement on two points. First, that the processor's status as processor of personal data (para 15) does not prevent the controller from assessing, under the contextual approach (para 45), the processor's actual capability to identify individuals without the key. Such an assessment is needed in particular for a transfer impact assessment under Chapter V, which is where our own interest lies. Second, how this case relates to the EDPB Guidelines 01/2025 on pseudonymisation, which the draft does not reference.

C2 — Unstructured text (section 3.4, paras 54–82)

The draft defines "record" broadly (para 5, Glossary) and counts simply reading a document among the means of identification (para 28). The three criteria themselves, however — No Record Isolation, No Linkage, No Inference — are illustrated exclusively with tabular data, and no example covers free text. In our experience the risks in text lie

elsewhere than in attribute values: a person is isolated by a narrative (profession, family situation, locality), linked across documents and years by recurring content (the same amounts, dates and facts reappearing in a later engagement), and inferred from context rather than from a field. We suggest two additions. The unit of assessment for text should be the engagement or case rather than the single document, because consistent tokens across documents are necessary for the text to remain usable and, at the same time, make the whole case the object of linkage. And the assessment of free text cannot be exhausted by removing identifiers; it requires a judgement on the identifying power of the remaining content, which is where a human check and a documented feedback loop belong. A short section or a worked example on unstructured text would help controllers in our position considerably.

C3 — An AI service as recipient (paras 12, 20–21, 77, 82, 92–93)

The draft considers AI as a re-identification tool (paras 92–93) and as an object of assessment (paras 77, 82), but not as the **recipient** of the data. Increasingly, however, the recipient is an AI service. The means it is reasonably likely to use are then of a different kind: the ability to infer from context, to link information across sessions and documents of the same account, memory and retention features, and the possibility that content is used to train models. We suggest that the Guidelines address the AI processor as a receiving entity in its own right and indicate which of these capabilities the controller should take into account — and that, consistently with paras 34 and 90, contractual restrictions on training and retention are to be treated as complements to technical decoupling, not as its substitute.

C4 — The anonymisation process itself as a channel of disclosure (paras 41–42)

Para 41 asks controllers to document the anonymisation process. We suggest adding that this documentation should cover the **data flows of the process itself**. If the anonymisation step is delegated to an AI system with network access — an assistant that may call tools, look things up or consult external services while it still has the clear text in front of it — it cannot be predicted with certainty which requests, containing which fragments of that text, the system will send out in order to perform its task. A disclosure of this kind would be a security incident in the sense of para 42, caused by the very step that was meant to prevent it. A process that handles identifiable data should therefore run offline, deterministically, or with a verifiable guarantee that it makes no external calls while identifiable data are within its reach. This does not concern locally run recognition models as such; it concerns agentic systems with connectivity. In our own build this was the decisive design choice, and we believe the Guidelines would help small controllers by naming it.

C5 — Proportionality and terminology for small controllers (paras 35, 40, 41)

Our processing is pseudonymisation, and paras 35 and 41 address anonymisation. They become relevant for us where a controller assesses whether the data are anonymous for the recipient under the contextual approach, and because we expect supervisory authorities to apply the same documentation discipline to a pseudonymising controller. Para 35 recommends periodic reassessment "wherever possible and as appropriate"; para 41 asks for documentation of the process "including for the testing of the supposedly anonymous datasets". For a controller of our size, with a standardised and repeated process, both are in our case achievable only at the level of the **process** — how the pipeline works, how it is tested on representative material, what it is known not to catch — combined with a per-case record of what was released, in which version and with which findings; not as a test of each dataset against the three criteria. We ask the Guidelines to say what "as appropriate" means for such controllers, whether testing the process on a representative corpus can stand in for testing each dataset, and to point to the EDPB's guide for small business.

Finally, para 40 rightly says, in the context of transparency towards data subjects, that data should not be described as "anonymous" where individuals remain identifiable. Controllers in our position are then left without an accepted term for data that are pseudonymised by the controller and not identifiable by the recipient without the key — and our data subjects include the clients to whom we must explain the processing. A recognised description, in the Guidelines or by cross-reference to Guidelines 01/2025, would let small controllers describe their processing to data subjects and supervisory authorities accurately, without either overstating ("anonymised") or obscuring what they do.

4 · Closing

We would be glad to describe our pipeline and its documentation in more detail if that is useful to the Board or to a supervisory authority. We thank the EDPB for the opportunity to comment.

Christian Hoferer

Managing Director, Hoferer GmbH
Steuerberatungsgesellschaft · Wirtschaftsprüfungsgesellschaft
Oppenau, Germany
Farnweg 2a · 77728 Oppenau · Germany
+49 7804 97990 · info@hoferer.de · www.hoferer.de