Comments on EDPB Guidelines 02/2026 on Anonymisation (version 1.0)

Submitted by: Nonimo (Llevant Group Pty Ltd) — https://nonimo.ai
Consultation: Guidelines 02/2026 on Anonymisation, adopted on 7 July 2026; public consultation open until 30 October 2026

All names, identification numbers and companies used in the examples are fictitious.

Executive summary

We welcome the Guidelines and the clarity they bring to the notion of anonymous data. Our comments address three points:

1. **The AI provider's perspective.** We ask the EDPB to explain, with worked examples, how text pseudonymised on the user's own device, with the re-identification key retained on that device, should be assessed for an AI provider acting as a processor or as an independent recipient.
2. **Unstructured documents.** We ask the EDPB to make explicit that, for documents, data mapping (para 51) must cover every layer of the file and every variant of an identifier, and to add an example based on a document rather than a table.
3. **Testing anonymisation tools.** We propose a recognised, independent testing method for anonymisation and pseudonymisation tools, anchored in a code of conduct or certification mechanism under Articles 40–43 GDPR, and would welcome the opportunity to contribute.

1. About Nonimo

Nonimo is a desktop application (macOS and Windows) and a browser extension designed to detect and mask personal data **locally, on the user's own computer**, before a text or document is sent to an AI service or shared with others. Each masked item is replaced by a placeholder. The table linking placeholders to original values (the "key") is stored on the user's device and is not transmitted by Nonimo, allowing the user to restore the original values in the AI's response.

Declaration of interest: Nonimo commercially supplies software of the kind discussed in section 4.

2. The AI provider's perspective (section 2.2)

A now common scenario is as follows. A professional acting as controller (for example, a lawyer, an accountant or a clinic administrator) pseudonymises a document on their own device, retains the key locally and sends only the masked text to a general-purpose AI service. The AI provider never receives the key.

The Guidelines point to different outcomes depending on the provider's role:

- **Para 15 and Example 3:** where an entity processes information *on behalf of* a controller, the controller's perspective also applies to that entity. On one reading, where the AI provider acts as a **processor**, the masked text could therefore be regarded as personal data for the provider, because it is personal data for the professional who holds the key.
- **Example 4:** where the recipient acts **independently**, the assessment is made from the recipient's own perspective. Read together with paras 20 and 89, the key would be additional information held under the controller's own access controls, so that an AI provider with no means reasonably likely to be used to obtain it could, in principle, receive data that is anonymous *from its perspective*.

In both cases, para 37 (and Example 2) confirm that the masked text remains personal data for the professional, who must comply with all GDPR obligations; and para 21 notes that data is personal for a recipient that is reasonably likely to pass it on to an entity able to identify the individual.

Questions for the EDPB:

- **Q1.** How should para 15 apply where the AI provider acts as a processor but never has access to the key? In either case, could the final version recognise the value of local pseudonymisation (Article 4(5)

and Recital 26 GDPR; see also EDPB Guidelines 01/2025 on Pseudonymisation) as a measure under Articles 25 and 32(1)(a) GDPR, and in the risk assessment under Articles 33–34 GDPR following an incident at the provider?

- **Q2.** A provider that reuses inputs for its own purposes (for example, model training) is not acting purely on the controller's instructions. How does this affect the role analysis under para 15 and Example 4, and, where the provider is an independent recipient, how should residual narrative content be assessed, such as a unique combination of role, place and event (para 24 and Example 5)?
- **Q3.** How should placeholders used across successive submissions to the same provider be assessed in light of para 50, para 62 and Example 6?
- **Q4.** Could the final version include at least one worked example based on an unstructured document (a contract, a letter or a medical report) sent to an AI service? Most current examples concern tables and record-level data.

3. Unstructured documents: industry-wide challenges

Anonymising real-world documents raises challenges common to all tools and methods, whether manual or automated. A frequently cited difficulty is that a direct identifier may survive somewhere in the file, so that the case falls under para 50 ("immediately apparent that information is not anonymous") before the three criteria even come into play. Since a dataset is anonymous only if anonymisation is effective for all individuals (para 36), and a technique can be successful even if it succeeds against only a single individual (para 87), a single surviving occurrence is generally enough.

The fictitious contract excerpt below illustrates typical patterns:

	Original text (fictitious)	**What a complete approach would require**
Party in capital letters	BETWEEN: LUCÍA VALMORQUIZ ARDENTE, holder of identity number Y0000000Z…	BETWEEN: [PERSON_1], holder of identity number [ID_1]…
Defined alias	…TRANSPORTES VALMORQUIZ SA (hereinafter, "VALMORQUIZ")…	…[COMPANY_1] (hereinafter, "[COMPANY_1_ALIAS]")…
Repeated item, other format	Clause 9: notices to l.valmorquiz@example.com; signed: L. Valmorquiz	Clause 9: notices to [EMAIL_1]; signed: [PERSON_1]
Hidden layers	Document properties: Author "L. Valmorquiz"; tracked change by "LVA"; scanned annex showing the identity number in the image	Metadata, revision history and the image layer each need to be addressed, not only the text layer

Each row illustrates a typical failure of an incomplete approach. Names written in capital letters are often not recognised as names. A defined alias derived from a surname keeps identifying the person, as identifiers may relate to "other individuals, groups or objects which are otherwise linked to them" (para 25). The same individual appears in several formats (full name, initials, email address), which together form a single "record" within the meaning of the glossary. Metadata, revision history, comments, hidden text and the image layer of scanned pages are invisible on screen but remain in the file. Scanned documents add a further difficulty: optical character recognition may read *Y0000000Z* as *YOOOOOOOZ*, so a pattern search on the text misses a number that any reader can still see in the image.

Question for the EDPB:

- **Q5.** Could the final version state explicitly that, for documents, the data-mapping step in para 51 must cover every layer of the file (visible text, image layer, metadata, revision history, comments and hidden content) and every occurrence and variant of an identifier (capitalisation, defined aliases, initials, OCR variants)?

4. Proposal: a recognised method for testing anonymisation tools

Controllers increasingly rely on software tools to anonymise or pseudonymise documents, yet have no common way to compare how well those tools perform. Para 41 asks controllers to document the testing of supposedly anonymous data, and para 84 encourages "the development, dissemination and application of

scientifically grounded expertise". A shared testing method would help controllers meet both expectations. We propose that the EDPB encourage such a method through a code of conduct under Article 40(2)(d) GDPR, which expressly names pseudonymisation, or through certification under Article 42 GDPR of processing operations that use such tools. The method could include the following elements:

- **What is measured.** For each layer of the file (visible text, image layer, metadata, revision history and hidden content): (a) direct identifiers remaining in the output, by category of identifier (the para 50–51 stage); and (b) for the residual content, whether No Record Isolation, No Linkage and No Inference are met, following section 3.4.
- **Test sets.** Realistic reference documents for each Member State and language, reproducing real document formats (contracts, invoices, medical reports, scanned identity documents, office files with metadata) and populated only with fictitious data. They should include the patterns described in section 3: capital letters, defined aliases, repeated occurrences, OCR variants and hidden content. Part of each set should remain undisclosed to prevent tools from being tuned to it.
- **Who performs the tests.** Independent bodies with no commercial interest in anonymisation tools, such as certification bodies accredited under Article 43 GDPR, working with academic institutions. Tool providers would contribute methodology and feedback but would not evaluate their competitors.
- **What is published.** The method and its version; the composition of the test sets (countries, languages, document types and layers); a summary of certification status and scope for each certified tool version; and the date of the test, with periodic re-testing as techniques evolve (paras 35 and 88).
- **Limits.** Certification of a tool should not be presented as proof that a particular output is anonymous. The contextual assessment of each disclosure remains the controller's responsibility.

Nonimo would be pleased, on a non-binding basis, to share its approach to measuring document masking across countries and languages, and to contribute to any working group or pilot that the EDPB or national supervisory authorities may establish.

We thank the EDPB for the opportunity to comment and for guidance that will make anonymisation both safer and more practical.

Nonimo (Llevant Group Pty Ltd) — https://nonimo.ai