Response to the public consultation on Guidelines 02/2026 on Anonymisation

Respondent: Llevant Group Pty Ltd (Australia), developer of Nonimo, a rule-based detection engine for personal data in free text. **Date:** 5 August 2026

Who we are, in three lines

We build a deterministic detection engine that identifies personal data inside free-flowing text before that text is sent to a third-party AI assistant. The engine runs locally on the user's device, contains no network code and uses no machine learning model. We work in Spanish, Catalan and Australian English, and we measure our results against real administrative and judicial documents.

We are not established in the EU. We are responding because the practical problem the Guidelines describe is the problem we work on every day, and because four of our measured findings bear directly on the framework.

1. The Guidelines are written for datasets. The fastest-growing case is a paragraph of text

The three criteria are defined over records and attribute values. The No Record Isolation criterion (paras 54 to 57) is expressed as a property of "a unique combination of attribute values that relate to a single individual", and is to be tested by "looking at, for example, multiple lines in a table or database table".

In the text of the draft, "dataset" appears 61 times and "record" 68 times. The terms "free text" and "unstructured" do not appear at all.

This matters because of what people are actually doing right now. A caseworker pastes three paragraphs of a complaint into an assistant. A lawyer pastes a client interview. A council officer pastes a resident's email. In each case the data is a short passage of natural language, not a table.

Applied literally to that passage, the criteria give a degenerate answer. A narrative paragraph is, by construction, a unique combination of attributes, so No Record Isolation fails on essentially every document that is worth writing. The framework therefore returns "not anonymous" for the entire category, without distinguishing between a passage that has been carefully reduced and one that has not been touched at all.

That is not wrong as law. It is simply not usable as guidance by the controller who has to decide, today, whether a given redaction is good enough.

Suggestion. Add a short subsection acknowledging unstructured text, and state how a controller should reason about it. Even a statement that the criteria are not intended to be applied passage by passage, and that free text should be assessed at the level of the corpus or the disclosure, would remove a real ambiguity. A worked example of a redacted document would be worth more than any amount of general language.

2. Anonymisation quality is not evenly distributed across people, and the gap falls on minorities

This is the finding we would most like the Board to consider. We found it in our own engine, it was invisible until we went looking, and we have since corrected it. We set it out because the shape of the problem generalises well beyond our product.

Our list loader applied what looked like a hygiene rule: discard candidate surnames shorter than three characters, on the grounds that two-character tokens generate noise. The rule does not remove noise evenly. It removes **Li, Wu, Ng, Ho, Vo, Yu, An, Ha and Na**. A formatting condition with no demographic intent switched off protection for Chinese, Vietnamese and Korean names while leaving protection for Spanish and Anglo names fully intact. It was the third time in our own codebase that a condition of that shape had silently disabled one origin.

The correction is more interesting than the defect, and this is the part we would ask the Board to weigh. We could not simply restore these names to the same treatment as the others. A two-character token cannot be removed silently from running text without an unacceptable rate of false positives, because those same two characters occur constantly as ordinary words. Our fix therefore admits the short surnames but caps their confidence, so they are surfaced to the user for confirmation rather than removed automatically.

That is the honest outcome, and it is asymmetric. The names are protected, but they are protected at a lower tier, which means the person handling a document containing them gets a slower and more manual experience than the person handling a document full of Spanish surnames. **Equal protection was available. Equal protection at equal cost was not.**

The consequence for the framework is this. "Anonymous" is presented throughout the draft as a property of the given data. In practice, for any real detection system, it is a property that **varies from data subject to data subject**, and the residual risk concentrates on the people whose names, addresses and document formats are least represented in whatever the system was tuned on. A dataset can be 99% anonymised and still leave a specific minority fully identifiable.

The draft's own logic supports taking this seriously: under No Inference (paras 66 to 69), an inference is meaningful if it is "liable to have an effect on the data subject's rights and interests". An anonymisation process whose failures cluster on one ethnic group has exactly that effect, and it also engages Article 9 and the non-discrimination principle.

Suggestion. State that where anonymisation is performed by an automated detection system, the assessment should consider whether residual identifiability is **distributed unevenly across groups of data subjects**, and not only whether an aggregate error rate is acceptable. An average is the wrong summary statistic for a fundamental rights guarantee.

We would add that the assessment should extend to the **cost** of protection and not only its presence. Where a system protects one group automatically and another group only through additional human effort, the practical outcome under time pressure is predictable, and it is not equal.

---

3. An anonymisation assessment does not transfer across languages

The Guidelines will apply across 24 official languages and 27 legal systems. We have hard evidence that an assessment carried out in one language says nothing about another, even for the same engine and the same rules.

We measured the same detection engine, unchanged, against two categories of real official document:

Measured against	Coverage before the language-specific work
Spanish official documents (state gazette, municipal ordinance, council forms)	1.0000
An Australian tribunal decision (Fair Work Commission)	**0.2857**

The same rules that found everything in Spanish found roughly a quarter of what they should have found in English. The causes were mundane and entirely structural: an Australian post-nominal collided with a Spanish company-form abbreviation; a tribunal job title sits adjacent to the surname in a pattern that does not occur in Spanish; a Spanish council form splits a person's name across three separate fields and places two fields on one printed line.

We also recorded zero false positives across 37,126 words of real administrative text, so the gap is not a tuning artefact of an over-cautious engine.

The rule we adopted internally, and which we would offer to the Board, is: **a number measured in one language says nothing about another. The two are measured separately, always.**

Suggestion. Where the Guidelines discuss assessing the effectiveness of re-identification techniques (para 83 onwards), state that an assessment is specific to the language, jurisdiction and document conventions it was carried out in, and does not automatically extend to others. Otherwise a controller will reasonably validate once, in one Member State, and deploy across the Union.

---

4. The anonymisation process is itself an inference channel

Para 82 recognises that specific inferences can be made "by querying (or, in the case of AI models, prompting) the given data", and para 67 notes that inferences may be based on "information about the anonymisation process".

We would add a case that we believe is not yet covered, and that we have confirmed against our own product rather than against someone else's.

A deterministic redaction system can be probed. An adversary who can submit text and observe what comes back can recover, term by term, the vocabulary the system uses to decide. Each probe is one bit, and enough probes reconstruct the dictionary. We have concluded that this cannot be prevented in a system that gives the user immediate feedback, which is the only kind of system people actually use.

This has a consequence that runs the other way from the usual concern. It is not that the redacted output leaks. It is that **the redaction tool leaks, and what it leaks is a map of what it fails to catch**, which is precisely the information an attacker needs to target the residual.

Suggestion. Extend the discussion of inference to note that where the given data has been produced by an automated anonymisation system, the properties of that system may themselves be discoverable and should be treated as information "reasonably likely" to be available, particularly where the system is commercially distributed and can be exercised by anyone.

---

5. A tension worth naming: suppressing an attribute can destroy the purpose

We work with text from migration and asylum casework. In that setting the attributes that make a record unique are frequently nationality, clan, ethnicity, language and district. Those attributes are special category data under Article 9 and, at the same time, they are the substance of the claim. Remove them and the document no longer says anything.

Our own design decision is that this class of attribute is never suppressed. We consider that the right answer for our use case, and we would not present it as the right answer generally. But the tension is real and the draft does not address it: maximising No Record Isolation on this kind of document destroys its usefulness, and there is no guidance on how a controller should weigh that.

Suggestion. Acknowledge that suppression is not always available, that in some contexts the identifying attributes and the purpose of the processing are the same attributes, and give controllers a way to reason about it that does not amount to "anonymise less and hope".

Summary of what we are asking for

1. Say something about unstructured text, even if only to place it outside scope.
2. Require the assessment to look at whether residual identifiability falls unevenly across groups of data subjects, not just at an average.
3. State that an anonymisation assessment is specific to its language and jurisdiction and does not transfer.
4. Treat the anonymisation system itself as a discoverable source of inference.
5. Name the conflict between suppression and purpose where the identifying attributes are the substance of the document.

We would be glad to provide our measurement methodology, our test corpus structure or a worked redacted example if that would be useful to the Board.