Response to the EDPB consultation on draft Guidelines 02/2026 on Anonymisation

Alfred Rossi[1] *and Sophie Stalla-Bourdillon.*[2]

The Draft Guidelines 02/2026 on Anonymisation[3] provide a structured assessment framework and help explain the practical implications of *EDPS v SRB*.[4] Our central concern is that they do not clearly state what a finding of anonymity promises. Is it a protective guarantee that limits what a recipient can learn about a person, or a legal classification reached through an assessment process that determines whether the data are personal for a particular holder and which obligations follow? Following that process does not necessarily establish the protective guarantee, particularly where the assessment rests on incomplete knowledge of the recipient.

The Guidelines appear to pursue both objectives without fully explaining their relationship. Several tests produce different answers depending on the assumed protection, the recipient's knowledge and the meaning of a record. The following seven recommendations seek to make those choices explicit. Paragraph references are to the consultation version adopted on 7 July 2026.

1. Specify whose knowledge counts and the sender's due diligence.
2. Explain the evidential role of contracts and the oversight required.
3. Define records and the complete or partial groupings to be tested.
4. Clarify whether pseudonyms count as attribute values.
5. Define singling out's object, protected secret and attribution conditions.
6. Specify the scope, significance and assessment of person-specific inference.
7. Explain the relationship with other regulatory guidance.

1. Whose knowledge counts

In the sender–recipient scenarios discussed here, "sender" means the anonymising controller making the disclosure. When a sender assesses whether data will be anonymous to a recipient, it must judge another entity's capabilities based on incomplete information. Paragraphs 30 and 47 require attention to relevant entities and their respective means, but leave questions about the knowledge on which the sender may reasonably rely. Example 21 illustrates the dependence on assumptions about access to additional information. An incorrect assumption may change the anonymity verdict.

> **Recommendation 1.** Specify whose knowledge the assessment assumes, the due diligence expected of the sender, the supporting evidence it should retain, and the consequences of mistaken assumptions or later changes. Clarify whether a motivated-intruder test can help define relevant attackers and capabilities, consistently with paragraph 31's caution against relying on a lack of

[1] Senior Research Fellow, Brussels Privacy Hub, VUB, & Distinguished Research Scientist, Immuta.
[2] Co-Director of the Brussels Privacy Hub, LSTS, VUB.
[3] European Data Protection Board, *Guidelines 02/2026 on Anonymisation*, version 1.0, adopted 7 July 2026 for public consultation. Unless otherwise indicated, paragraph and example references in this response refer to this version.
[4] Court of Justice of the European Union, Case C-413/23 P, *European Data Protection Supervisor v Single Resolution Board*, judgment of 4 September 2025, ECLI:EU:C:2025:645.

motivation. Clarify whether, and under what conditions, evidence of segmentation within a recipient's systems permits information to be excluded from the linkage assessment.

2. The role of contracts

Contracts can restrict linkage and onward disclosure, require security measures and impose consequences for breach. They can therefore affect which means are reasonably likely to be used. They do not, by themselves, remove information the recipient already knows or establish that the sender has correctly assessed its capabilities.

Paragraph 34 accordingly requires that contractual measures be reliable, verifiable and enforceable, and complement technical measures. Its caution about contractual obligations contrasts with paragraph 89's apparent acceptance of appropriate security controls as a basis for excluding unauthorised actors from the access assessment. Those controls also depend on implementation and continued maintenance; their continued operation is not guaranteed. Paragraph 90 requires an assessment beyond the mere presence of technical, organisational or contractual measures. The Guidelines should explain the basis for this difference in treatment, what evidence justifies reliance on each type of measure, and how that reliance should be maintained over time.

Recommendation 2. Explain whether contractual obligations concerning data security, purpose limitation and the prohibition of re-identification reflect a duty of due diligence inherent in the GDPR, particularly in light of Recital 26 and its requirement to consider technological developments.[5] Specify what further due diligence, including verification and continuing oversight, is expected beyond inserting contractual clauses to ensure that the sender appropriately assesses the recipient's knowledge and means. Distinguish evidence of restricted access from a promise not to use information already available.

3. Record isolation and additional knowledge

Paragraph 48 presents the simplified approach as cautious: it accepts the possibility of treating anonymous data as personal to reduce the risk of wrongly treating personal data as anonymous. Assuming greater capabilities should therefore make it harder to conclude that data are anonymous. The record-isolation test raises a difficulty because additional knowledge can change what counts as a record in the first place.

A prescription table may contain several entries about the same patient. Paragraph 54 defines a record as the values relating to one individual; paragraph 57 and the glossary refer to values *known* to relate to that individual. Paragraph 57 also says that record isolation can generally be tested using the given data alone. The example below examines the reading under which entries that a recipient knows concern the same patient are assessed together as one record, including where that knowledge comes from additional information. In this example, that record is the **patient history the recipient can assemble**. The recipient need not know the patient's name.

[5] Regulation (EU) 2016/679 of 27 April 2016 (General Data Protection Regulation), [2016] OJ L 119/1, recital 26.

Consider the following release. Each pharmacy uses its own patient codes, so the codes do not indicate which entries from different pharmacies pertain to the same patient. The entry numbers are labels added here to explain the example.

Entry	Pharmacy	Patient code	Medicine	Sex	Birth year	District
1	X	X17	Metformin	F	1955	North
2	X	X83	Metformin	F	1955	North
3	Y	Y42	Insulin	F	1955	North
4	Y	Y09	Insulin	F	1955	North

In fact, entries 1 and 3 concern one patient, and entries 2 and 4 concern another. Both patients take both medicines. Those connections are absent from the release. The first two recipients below are also not told that there are exactly two patients, each with two prescriptions.

For this illustration, patient codes are used only to connect entries and are excluded when comparing histories for uniqueness; section 4 addresses that choice. The test asks whether any assembled history contains a combination of information that is unique to the histories being compared. Entries left separate may still concern the same patient; the recipient lacks the connection.

The following roles illustrate how recipients might acquire different information. All three receive the same prescription table. The additional information links pharmacy codes but provides no names or additional patient attributes. No other information about these patients is assumed.

Recipient	Additional information	Patient histories used in the test	No Record Isolation
1. Outside researcher	Receives only the released table.	Four separate entries: two for Metformin and two for Insulin.	**Pass:** each entry has another with the same compared values.
2. Researcher running a pilot study	Receives confirmation that codes X17 and Y42 belong to the same patient, connecting entries 1 and 3.	One history containing both medicines (entries 1 and 3), plus entries 2 and 4 separately.	**Fail:** only one assembled history contains both medicines.
3. Coordinator of the combined study	Receives the same confirmation, plus confirmation that codes X83 and Y09 belong to the same patient, connecting entries 2 and 4.	Two histories, each containing both medicines: entries 1 and 3 together, and entries 2 and 4 together.	**Pass:** the two histories have the same compared values.

Paragraph 102 calls for linked information to be tested again. The example shows why that re-test must specify which patient histories it examines.

The result is **pass, fail, pass**, although each recipient knows everything the previous one did. The study coordinator can still examine the same entries and use any identification procedure available to the pilot researcher. What changes is the comparison: the history that appeared unique now has a matching counterpart. Greater knowledge therefore does not automatically make this isolation test more cautious.

The Guidelines could explain how to preserve paragraph 48's cautious approach as the recipient's knowledge increases. Tests can be designed so that protection established for a more informed recipient also holds for recipients with less information.

This finding concerns the isolation test alone. A unique combination of medicines does not necessarily allow the recipient to attribute that history to a particular person, and the remaining singling-out, linkage, and inference tests still matter.

> **Recommendation 3.** Define which entries must be assessed together as one person's record and whose knowledge determines that grouping. Say whether the test must examine the complete grouping by patient, each partial grouping available to a relevant recipient, or both. Paragraphs 48 and 101–103 should explain how the simplified assessment remains cautious under that rule, including when connections between entries are recovered within the released data.

4. Pseudonyms as attribute values

Paragraphs 19 and 55 leave a consequential question about pseudonyms: are they values used to test uniqueness, or only devices for grouping rows? If the distinct pseudonyms count as compared values, each fill in the pharmacy example is unique from the start. If they are excluded, the pass–fail–pass result remains possible. Both readings appear defensible under the present wording, but they can yield different results for the same release. Stable person-specific pseudonyms raise the corresponding question at the level of grouped records.

This should be reconciled with Pseudonymisation Guidelines 01/2025, paragraphs 115–121, which distinguish pseudonym scopes and favour transaction pseudonyms where appropriate.[6] Removing a visible link and removing the underlying relationship between rows are different operations.

> **Recommendation 4.** State whether pseudonyms count as attribute values for record isolation, and explain the treatment of stable and transaction pseudonyms consistently across the two Guidelines.

5. What constitutes singling out

Paragraphs 98–100 require further analysis when record isolation fails. Three choices need to be explicit. The first is the object of the test: does a unique matching record suffice, or must the recipient be able to attribute it to the person? The second is the protected secret: the person's connection to a record, their participation in the dataset, or an attribute about them? The third is the anchor supporting attribution: independent knowledge of participation, knowledge that nobody else in the relevant population fits, or either route?

These choices have concrete consequences. One matching record can fit two people. Conversely, knowing that a person participated can support attribution without establishing that their attributes are unique in the wider population. The test must also distinguish actual relationships from relationships the recipient can establish. In the pharmacy example, both patients take both medicines, even though the second recipient can assemble only one history containing both.

[6] European Data Protection Board, *Guidelines 01/2025 on Pseudonymisation*, adopted 16 January 2025 for public consultation, paras 115–121.

The ambiguity in these terms gives rise to eighteen interpretations of record isolation and sixteen of singling out.[7] These illustrate the effect of definitional choices without exhausting the possible readings.

> **Recommendation 5.** Specify the object, protected secret and attribution anchor, and whether each condition concerns facts or what the recipient can establish. Make clear what a negative singling-out result leaves unresolved under the linkage and inference criteria.

6. Inference beyond record matching

Record-based screens can pass while information about a person is disclosed. In a group of eleven patients with identical diagnoses, a recipient who knows that a person belongs to the group can learn the diagnosis without locating their particular record. Eleven visits by one person can also appear to provide a group of eleven if rows are mistaken for distinct people. Information about one person can also support inferences about another through a family, household or other relationship, even where the other person has no separate record. A relative's genetic data or a known cohabitant's address illustrates this possibility. Partial linkage creates the additional difficulty described in section 3.

These examples concern the limits of record-based screens. The Guidelines separately address inference from aggregates (paragraphs 76–78 and Examples 16–17), nonparticipants (paragraph 69), membership (paragraph 79) and attributes shared by several matching records (paragraphs 80–81 and Example 19). The unresolved issue is how these provisions define a consistent protection across release models. Membership, a person's connection to a record and attribute disclosure are distinct secrets; a unique-record matching test cannot, by itself, protect all three.

A useful statistical release cannot promise that nothing will be learned about anyone. If it helps someone make an informed decision about a particular person, someone else with comparable background knowledge can use the same information without linking that person to a released record. Population findings, such as evidence about smoking, can change predictions about particular people, including nonparticipants. A workable account must distinguish that learning from disclosure attributable to a person's contribution. One quantity that can be bounded is the additional information released because that person's data were included.

Paragraph 71(b) and footnote 43 already compare inferences drawn from releases produced with and without all data relating to the individual. Footnote 44 refers to a significantly different "posterior belief" and connects that difference to possible action. This supplies a qualitative indication of significance, but leaves unclear what is being measured and how it determines whether the source data are personal.[3]

---

[7] The eighteen record-isolation readings arise from reading "relate to" in three senses: a connection by content alone; a connection by content, purpose, effect or what processing reveals; or a connection recognised only where the person's combined values are unique. "Record" is read as rows grouped by the people they concern, by shared pseudonyms, or by shared pseudonyms together with links recovered within the data. "Attribute values" are read as including or excluding pseudonyms when testing uniqueness. These choices give three meanings of relation, three meanings of record, and two treatments of attribute values.

The sixteen singling-out readings arise from reading "records" as rows grouped by person or by visible links; "the individual's corresponding attributes" as demographics alone or any recorded attribute known to the recipient; and "distinguished from others" in four senses: a unique record match alone; such a match supported by independent knowledge of participation; a matching isolated record supported by population uniqueness established through a complete source; or a unique record match supported by either form of knowledge. These choices yield two meanings of "record," two sets of matching attributes, and four conditions for distinguishing the person. The two counts arise from these separate combinations of meanings, rather than from multiplying the object, secret, and anchor questions above.

Pufferfish privacy provides a model for stating the protected secrets, the alternatives an attacker should not distinguish, and the assumptions about data and background knowledge.[8] Differential privacy offers a recipient-independent bound on how much a release mechanism's output distribution changes when one person's contribution changes. To cover the multiple-visit example, that contribution must encompass all of the person's rows.[9] This is not a promise to prevent every inference arising from population patterns or data about other people to whom the individual is connected.

Paragraph 36 distinguishes anonymity of particular records (without further specifying this concept) from that of the data set and calls for anonymity to hold for every included individual. Paragraph 87 states that a technique can succeed against a single individual, subject to sufficient accuracy and confidence; Example 20 excludes a result where the entity cannot tell whose salary was correctly inferred. The Guidelines should also clarify the baseline and release model for assessing whether the release enabled the inference. A demonstrated disclosure about one person may be decisive. A correct guess, without evidence that the release enabled it, does not establish the same result. The distinction matters for aggregate, synthetic and randomised outputs as well as for released microdata.

> **Recommendation 6.** Map "specific and meaningful" inference explicitly to the protected secrets, and explain when membership and attribute inference must be mitigated. Explain how to assess separately whether an inference concerns an identified or identifiable individual and whether it depends on data relating to that individual. Clarify how paragraph 67's reliance on the given data relates to paragraph 71(b)'s reliance on the individual's records, in light of Example 13. Clarify what makes a change in inference significant: the additional person-specific information disclosed, its capacity to support decisions about the individual, or its potential consequences. Explain how that criterion determines whether the source data are personal, and provide appropriate baselines and operational thresholds.
>
> Clarify which data must be removed under footnote 43 where information about an individual can be inferred through their relationship to someone represented in the data, including whether this covers information contained in that other person's records. Paragraph 16(b) recognises that information about an associate or relative may relate to the individual. If footnote 43 requires removing such information from another person's records, a guarantee covering only the individual's own records would not suffice. Distinguish such inferences from population learning. Use Pufferfish as a model for specifying the protection sought, and differential privacy where a bound on an individual's contribution is required for any recipient. Failure to establish such a guarantee should not, by itself, constitute a legal finding that the data are personal.

7. Relationship with other guidance

The simplified approach invites comparison with clinical-data disclosure frameworks. EMA's 2025 Policy 0070 guidance recommends aggregation and k-anonymity and identifies 0.09 as a conservative risk threshold.[10] Health Canada's 2019 guidance recommends the same threshold and associates it with a

[8] Daniel Kifer and Ashwin Machanavajjhala, "Pufferfish: A Framework for Mathematical Privacy Definitions" (2014) 39(1) *ACM Transactions on Database Systems*, article 3, pp 3:1–3:36, section 3; author manuscript.

[9] Cynthia Dwork and Aaron Roth, *The Algorithmic Foundations of Differential Privacy* (2014) 9(3–4) *Foundations and Trends in Theoretical Computer Science* 211–407, chapter 1 and section 2.3, especially Definition 2.4 and Theorem 2.2 (group privacy), DOI: 10.1561/0400000042. The protected contribution must be defined at the person level; a guarantee for changing one row alone does not give the same bound for changing all of a person's rows.

[10] European Medicines Agency, *External guidance on the implementation of the European Medicines Agency Policy 0070 on the publication of clinical data for medicinal products for human use*, EMA/90915/2016, version 1.5, 14

target cell size of eleven patients.[11] The fact that these frameworks draw on related concepts does not, in itself, make them as equivalent to the Draft Guidelines.

EMA allows considerable contextual judgment. Its guidance states that l-diversity and t-closeness "may not be recommended as they limit inferences significantly". It also treats patient-ID redaction or recoding on a case-by-case basis according to the impact on risk, and allows applicants to propose and justify which indirect identifiers require transformation.[10] Those choices bear directly on the treatment of pseudonyms and inference discussed above.

The contextual approach would also need a clearer relationship with the ICO's March 2025 guidance and Ontario's October 2025 guidelines. The ICO expressly uses the motivated-intruder test and recognises contractual controls in restricted releases.[12] Ontario combines recipient assessment, due diligence and contractual protections, and distinguishes identity, membership and attribute disclosure.[13] These differences concern both assessment methods and the protection being assessed.

> **Recommendation 7.** Explain where the Draft Guidelines agree with these frameworks, where they depart, and the practical consequences. Otherwise, controllers operating across borders or working in clinical research must reconcile their approaches. In case of doubt, they are likely to follow the most granular guidance, even where its assumptions differ from those of the intended release.

May 2025, sections 3.5.3.2 (anonymisation techniques), 3.5.3.2.1 (patient IDs), 3.5.3.2.2 (indirect identifiers) and 3.5.4, step 4 (risk threshold).

[11] Health Canada, *Public Release of Clinical Information: Guidance Document*, version 1.0, 12 March 2019, section 6.2, Step 2 (measuring re-identification risk).

[12] Information Commissioner's Office, *Anonymisation*, version 1.0.21, 28 March 2025, section "How do we ensure anonymisation is effective?", especially "What is the 'motivated intruder' test?" and "Does the type of data release matter for the motivated intruder test?".

[13] Information and Privacy Commissioner of Ontario, *De-Identification Guidelines for Structured Data: Updated and Expanded*, October 2025, sections 2.8 (disclosure types), 4.3 (third-party sharing) and 5.4–5.6 (recipient assessment, controls and their verification).