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

Patrick Brunner[*] ✉
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

Dr. Shlomi Hod[*]
Karlsruhe Institute of Technology

Thilo Gottschalk
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure

Dr. Simon Hanisch
CeTI, Technical University Dresden

Dr. Jan Krasni
Zentrum für Antisemitismusforschung, Technical University Berlin

Dr. Abhishek Kumar Mishra
Karlsruhe Institute of Technology

Stephanie von Maltzan
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
European Centre on Privacy and Cybersecurity, University of Maastricht

Dr. Kati Mozygemba
RDC Qualiservice, Universität Bremen

Prof. Dr. rer. nat. Jochem W. Rieger
Carl von Ossietzky University of Oldenburg

Prof. Dr.-Ing. Thorsten Strufe
Karlsruhe Institute of Technology

October 2026

* Equal contribution.
✉ Corresponding author. Email: patrick.brunner@fiz-Karlsruhe.de

Introduction

On 7 July 2026, the European Data Protection Board (EDPB) published draft *Guidelines 02/2026 on Anonymisation* for public consultation. The draft updates the Article 29 Working Party's 2014 Opinion and assesses anonymisation.

Overall, we support the EDPB for this effort, and we believe that **these guidelines are strong contributions towards scientifically-sound practices of anonymisation that transpose the corresponding GDPR requirements**.

In this response we focus on four main issues. First, we read the Guidelines as setting a **framework for assessing the risk of anonymized data**. We provide a detailed recommendation regarding the assessment of the identification risk, the selection of relevant entities, and the consideration of harms. Second, we discuss **the need to clarify core definitions of the Guidelines terminology** –"given data" and "treat them differently". Third, we analyse each criteria, and we provide an array of recommendations **to make the criteria technically and legally more sound**. Finally, we examine **the interplay of the Guidelines on scientific research**, and provide recommendations on how the Guidelines may be implemented in research domains.

We structure our response thematically around these ideas and identify how they relate to the paragraphs, examples and footnotes of the draft Guidelines.

Several of these issues were surfaced in a Red Teaming workshop[1] held as part of Dialectic Policymaking, a methodology we develop at KIT to fuse technical and legal expertise when stress-testing draft regulation. Participants acted as adversaries looking for cases that comply with the Guidelines yet leave data that is not effectively anonymous.

[1] Shlomi Hod, 'Red Teaming the EDPB's Draft Anonymisation Guidelines: A Workshop' (*Chair of Privacy and Security, KIT*, 8 October 2026) https://ps.tm.kit.edu/english/1094.php accessed 29 October 2026.

About the Authors

We are a multi-institution, multi-disciplinary group of German-based academic researchers who specialize in cybersecurity and privacy enhancing systems, data protection law, AI regulation and research data management. Our expertise spans designing privacy enhancing technologies[2][3][4], deploying privacy enhancing systems[5], their formal analysis[6][7], discovering vulnerabilities and risks in anonymisation[8][9][10] and deployed technologies[11][12], designing and analyzing techno-legal approaches to

[2] Shlomi Hod and others, 'Differentially Private Modeling of Disease Transmission within Human Contact Networks' (2026) <https://doi.org/10.48550/arXiv.2604.07493> accessed 29 October 2026.
[3] Tom Nicolai and others, 'QualiAnon - Qualiservice tool for anonymizing text data' (version 1.5.0, *University of Bremen*, *GitHub*, 2025) <https://github.com/pangaea-data-publisher/qualianon> accessed 29 October 2026.
[4] Christiane Kuhn, Martin Beck, Thorsten Strufe , 'Breaking and (Partially) Fixing Provably Secure Onion Routing' in 2020 *IEEE Symposium on Security and Privacy* (SP) (IEEE, 2020).
[5] Shlomi Hod and Ran Canetti, 'Differentially Private Release of Israel's National Registry of Live Births' in 2025 IEEE Symposium on Security and Privacy (SP) (IEEE 2025).
[6] Christiane Kuhn and others, 'On Privacy Notions in Anonymous Communication' (2019) 2019(2) PoPETs 105.
[7] Àlex Miranda-Pascual and others, 'SoK: Differentially Private Publication of Trajectory Data' (2023) 2023(2) PoPETs 496.
[8] Rachel Cummings and others, 'ATTAXONOMY: Unpacking Differential Privacy Guarantees Against Practical Adversaries' (2024) <https://doi.org/10.48550/arXiv.2405.01716> accessed 29 October 2026.
[9] Simon Hanisch and others, 'A False Sense of Privacy: Towards a Reliable Evaluation Methodology for the Anonymization of Biometric Data' (2024) 2024(1) PoPETs 116.
[10] Julian Todt, Simon Hanisch and Thorsten Strufe, 'Fantômas: Understanding Face Anonymization Reversibility' (2024) 2024(4) PoPETs 24.
[11] Clemens Deußer, Steffen Passmann and Thorsten Strufe, 'Browsing Unicity: On the Limits of Anonymizing Web Tracking Data' in 2020 *IEEE Symposium on Security and Privacy* (SP) (IEEE 2020).
[12] Julian Todt, Felix Morsbach, and Thorsten Strufe. 2025. 'BFId: Identity Inference Attacks Utilizing Beamforming Feedback Information' in *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security* (CCS '25) (Association for Computing Machinery 2025) <https://doi.org/10.1145/3719027.3765062> accessed 29 October 2026.

anonymisation[13][14][15][16][17], providing guidance on the interpretation of data protection law[18][19], and managing research data[20].

We have previously responded to other governmental requests for information, focusing on privacy enhancing technologies[21][22][23], general-purpose AI[24], and trained policymakers in responsible AI[25].

---

[13] Stephanie von Maltzan and others, 'How anonymous is anonymous? a techno-legal exploration' in Réka Markovich and others (eds) *Legal Knowledge and Information Systems: JURIX 2025: The Thirty-Eighth Annual Conference* (IOS Press 2025).
[14] Patrick Brunner and Stefan Bavendiek, 'Secure infrastructures and protection of personal data: Approaches for the development of the NFDIXCS Research Data Management Container' in Vincent Heuveline and others (eds) *E-Science-Tage 2025: Research Data Management: Challenges in a Changing World* (heiBooks 2025).
[15] Thilo Gottschalk 'European Union · The EU-US Data Privacy Framework (DPF) – A Blueprint for International Data Transfers?' (2023) 9 EDPL 448.
[16] Stephanie von Maltzan, 'A Data Protection-Compliant Framework for Wi-Fi-Based Location Tracking (in Law Enforcement)' in Kai Rannenberg, Prokopios Drogkaris and Cédric Lauradoux (eds) *Privacy Technologies and Policy: 11th Annual Privacy Forum* (Springer 2024).
[17] Stephanie von Maltzan, 'The Concept of Identifiability in ML Models' in Denis Bastieri and others (eds) *Proceedings of the 7th International Conference on Internet of Things, Big Data and Security 2022* (SciTePress 2022).
[18] Vasilka Stoilova and others, 'FAIR-Prinzipien und Datenschutz | Factsheet 1 - Einwilligung: FAIR und datenschutzkonform (1.1)' (*Zenodo* 2025) <https://doi.org/10.5281/zenodo.15912755>.
[19] Thilo Gottschalk, 'Article 11 - Automated individual decision-making' in Eleni Kosta and Franziska Boehm (eds) *The EU Law Enforcement Directive (LED): A Commentary* (Oxford University Press 2024).
[20] Prof. Dr. rer. nat. Jochem Rieger (*Carl von Ossietzky Universität Oldenburg*) < https://uol.de/en/angewandte-neurokognitive-psychologie/mitarbeiter/prof-dr-rer-nat-jochem-rieger> accessed 29 October 2026.
[21] Rachel Cummings and others, 'Comment on "NIST SP 800-226: Guidelines for Evaluating Differential Privacy Guarantees' (2024) < https://www.cs.umd.edu/~kaptchuk/publications/nist24-dp-public-comment.pdf> accessed 29 November 2026.
[22] Rachel Cummings and others, 'Response to the RFI on Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information' (2024) < https://thepalakjain.com/wp-content/uploads/2025/01/cai_rfi.pdf> accessed 29 October 2026.
[23] Mireille M. Sant and others, 'Response to the Public Consultation on the EDPB Guidelines 1/2026 on Processing of Personal Data for Scientific Research Purposes' (*EDPB*, 2026) < https://www.edpb.europa.eu/system/files/consultation_feedback/1495/feedback_1495_edpb_guidelines_1_2026_consultation_response.pdf> accessed 29 October 2026.
[24] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, 'Statement of the Consultation for the Commission Guidelines to clarify the scope of the obligations of providers of General-Purpose AI Models in the AI Act' (*FIZ Karlsruhe*, 2025) < https://www.fiz-karlsruhe.de/sites/default/files/FIZ/Bilder/FIZnews/GPAI_Consultation_FIZ-Karlsruhe.pdf> accessed 29 October 2026.
[25] Margaret Stanton, 'Demystifying AI to Mitigate its Negative Consequences' (*Rafik B. Hariri Institute for Computing and Computational Science & Engineering*) < https://www.bu.edu/hic/2023/08/30/shlomi-hod-leads-congressional-workshop-on-responsible-ai/> accessed 29 October 2026.

Table of Content

Introduction
- About the Authors

Table of Content

Framework for Assessing the Risk of Anonymized Data
- Guidelines prescribe a risk identification assessment; naming should reflect this
- Assessing the identification risk
- Constructing the Chain of Entities
- Data is not anonymised for the controller if they retain the original data
- Harms because of Anonymisation
- Purpose does not travel with the anonymised data

Definitions
- Given data
- “Treat them differently” as element of identification

The Criteria for the technical analysis
- No Record isolation
- No Singling out
- No Linkage
- No Inference

Scientific Research Data
- Resources and shared infrastructure
- Tiered access is established practice in research
- Qualitative and textual data requires the contextualised approach

Editorial comments

AI Disclosure

Acknowledgements

Framework for Assessing the Risk of Anonymized Data

Guidelines prescribe a risk identification assessment; naming should reflect this

Read as a whole, the guidelines require the anonymising controller to carry out a process whose structure is familiar from risk management and security engineering: to enumerate the entities for whom the data must be anonymous, to characterise the means each of them is reasonably likely to use, to identify the scenarios in which those means would succeed, and to revisit the conclusion as the data environment changes. This is a risk assessment of identification with threat modelling at its centre, and the framework the guidelines propose is therefore, at its core, a structured and well-established *governance mechanism* which we regard as the document's principal strength, and one that deserves to be stated and built upon rather than left implicit.

At present, however, the process is neither named nor presented as a process; its elements are distributed across the legal analysis of Section 2 and the technical analysis of Section 3,[26] and the reader is left to reassemble them.

We recommend that the Guidelines designate the process explicitly – "anonymisation risk assessment" – would be one natural choice – and restructure the document so that the process can be read out of it in sequence.

The case using this process as the organisational principle of the guidelines rests on the maturity of the surrounding discipline, which already possesses a standardised toolbox for precisely this exercise.[27] To demonstrate this as an inspiration, consider attack trees for instance. They were designed to represent exactly the multi-step, branching attack paths that the guidelines invoke when they require linked data to be re-tested against

[26] We note, without claiming completeness, the following correspondences with established practice, grouped by the phase of a risk assessment to which they belong. Processing characterisation (para 13). Adversary enumeration and characterisation (Section 2.2, para 9, para 23, 29, 30–33, 42, 45, 47–49, 51, Example 1). Attack catalogue (paras 28, 62, 64, 72, 77, 79, 82, 85, 102) Likelihood and impact (para 8, 16(c), 22, 36, 67, 70–71, 87, fn 42–43, 44). Controls and lifecycle (paras 34, 35, 41, 42, 44, 88–91, 95, Example 8).

[27] See for example: International Organization for Standardization, *ISO 31000:2018 Risk management: Guidelines* (2nd edn, ISO, 2018); National Institute of Standards and Technology, *Guide for Conducting Risk Assessments* (NIST SP 800-30 Rev. 1, NIST CSRC, 2012); Government Digital Service, 'The Five Safes Framework' (UK Government, 2025) < https://www.gov.uk/data-ethics-guidance/the-five-safes-framework> accessed 29 October 2026; 'LINDDUN: A framework for privacy threat modelling' (*LINDDUN*) < https://linddun.org/> accessed 29 October 2026; International Organization for Standardization and International Electrotechnical Commission, *ISO/IEC 27559:2022 Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework* (ISO 2022).

the three criteria (para 102), inferences that are not themselves specific and meaningful to be retained because they may enable further inferences (para 72), and chains of means spanning several entities to be evaluated in their entirety (para 28).

Viewed as a governance mechanism, the process also carries *organisational implications* that the guidelines ought to make explicit, the first of which is that anonymisation requires cross-functional teams of legal and technical expertise. Second, this mechanism should be integrated into a holistic GDPR compliance process.

Assessing the identification risk

Setting the thresholds for insignificant identification in reality

Identification risks should be addressed from the beginning, i.e., during the planning of the associated data processing, and in the context of a DPIA, if necessary. With respect to anonymous data under the GDPR, the risk of (re-)identification need not be zero. Instead, the risk of identification must be insignificant in reality.[28] The guidelines however use differing terminology regarding what the threshold is and what it refers to. Terminology used for what the threshold refers to includes: “likelihood [...] successfully distinguished” (para 22); “likelihood of identification”; “likelihood of their use is insignificant” (para 32); “likelihood of re-identification” (para 36). Terminology used to provide measurements for the threshold includes: "insignificant in reality" (para 22), "sufficiently close" and "sufficiently high confidence" (para 87), "certainty or high likelihood" (para 60), and a "significantly different posterior belief" (fn 44).

We recommend that the Guidelines use “risk of identification” when describing the point of reference for the threshold and to explain how differing terminology on the threshold aligns with the concept of an identifiable natural person in Article 4(1) GDPR as the CJEU has interpreted it and also with one another.

Furthermore, regarding identification that is probabilistic or uncertain, we recommend that the Guidelines explicitly denote the required confidence in the results.

We appreciate that the Guidelines do not provide general quantifications of the normative threshold. We believe that such thresholds are heavily dependent on the domain-specific aspect of a data processing and its particular context.[29]

We recommend that the EDPB facilitates, empowers and provides opinions on the creation of reliable thresholds for well-defined data processing contexts.

---

[28] Case C-413/23 P *EDPS v SRB* ECLI:EU:C:2025:645, para 82.

[29] Without judging their quality, such quantifications of the threshold for identification are already used in practice. See for example: European Medicines Agency, ‘External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use - Version 1.5’ EMA/90915/2016 Version 1.5 (15 May 2025), page 42.

In alignment with para 41, with respect to accountability, we further recommend that the Guidelines explicitly require the Controller to determine in advance and document the concrete threshold, and its justification, for the risk of identification against which the anonymisation of the given data with regard to relevant entities is being assessed.

Harm and impact in the risk assessment

Building on the recommendation above, we note that risk is conventionally understood as the combination of the likelihood of an event and the severity of its consequences, and that this is how established risk assessment methodologies, and the GDPR itself (Recital 76), approach it.

The guidelines take a step in this direction with the No Inference criterion, which asks whether an inference is "meaningful", that is, whether it is liable to affect the data subject's rights and interests (para 67). We welcome this, since it allows identification risk to be assessed in context rather than in the abstract, although the notion is at present confined to a single criterion.

We recommend that the EDPB bring it one level up, so that harm to individuals becomes a dimension of the anonymisation risk assessment as a whole.

Estimating harm in a reliable and valid way is a demanding exercise, and precisely because it is so central to the assessment,we recommend that the Guidelines provide guidance on how it is to be done.

Re-assessments of the residual identification risk (after anonymisation)

In processing situations where data is legally considered anonymised, the GDPR does not apply. However, the risk of identification cannot be completely ruled out. In this regard, the Guidelines mention that it is "good practice for entities processing anonymised data" to "periodically" (para 35) re-assess the likelihood of identification. We think this is insufficient to address the residual identification risk.

We recommend that the Guidelines explicitly require the controller that anonymises data only from the perspective of other entities (para 37), to continuously monitor whether the data is still anonymous

Furthermore, we recommend that the EDPB encourage and facilitate the creation and the use of codes of conduct and standards for the monitoring of identification risk by all entities that process anonymised data. If an identification risk materialises and the given data becomes personal again, any processing of the data must comply with the GDPR's rules. This requires periodic as well as incident-based (para 42) re-assessments of the anonymity status of the data.

We recommend that the Guidelines provide guidance on how to address identification risk on an ongoing basis. In particular, this includes guidance on how to choose concrete triggers for incidents that require a re-assessment (e.g., new public auxiliary data, published attacks on comparable data, further releases from the same source, leaks) as well as guidance on choosing appropriate time frames for periodic assessments. Guidance should also be provided on technical and organisational measures that can support re-assessments. In this regard, logs and output identifiers show what was released to whom, which makes notification and reassessment feasible.

F**urthermore, we recommend that the Guidelines do not suggest that anonymised data in accordance with Opinion 05/2014 does not require a periodic and incident-based re-assessment of the anonymity status (fn 3).** Instead, necessary re-assessments must comply with these new Guidelines that transpose the GDPR's requirements for anonymous data.

Composition of outputs from the same original data

When several outputs are produced from the same original data, the anonymity of each output on its own does not guarantee their anonymity taken together, because each further output adds information about the same individuals. This is a well-established risk that has materialised in practice.[30] The Guidelines recognise the mechanism in Example 16 and in para 77, which notes that inference often draws on multiple aggregated datasets produced from the same original data, but only as a method of inference available to an adversary.

Throughout the framework, the given data is a single output: the No Record Isolation criterion is tested on the given data alone (para 57), the No Linkage criterion concerns records from a different dataset (para 60), and the glossary defines given data as the data whose anonymity is to be assessed. Likewise, the grounds for re-assessing anonymity are technological progress, the increased availability of additional information (para 35) and security incidents (para 42). The controller's own further outputs from the same original data are not mentioned, although the controller knows better than anyone that they exist and who holds them.

We recommend that the Guidelines explicitly list other outputs produced from the same original data, whether released, concurrent or planned, among the objective factors in para 29, require that they be assessed together with the given data, and state in para 35 that each further output requires a re-assessment of the earlier ones.

Assume the anonymisation method is public

Para 67 allows inferences to draw on information about the anonymisation process, but does not assume such information is available, and does so only for No Inference.

[30] Aloni Cohen, 'Attacks on Deidentification's Defenses' in *31st USENIX Security Symposium (USENIX Security 22)* (USENIX Association 2022).

In line with well-established Kerckhoffs's principle, we recommend that the Guidelines instruct to assume that the anonymisation algorithm is known to every relevant entity, while elements drawn independently of the personal data, such as random values from a cryptographically secure pseudorandom number generator (CSPRNG), may be assumed to remain secret.

Make documentation as an inseparable part of anonymous data

Transparency about the anonymisation process supports accountability and monitoring, and allows entities receiving anonymous data to assess whether it remains anonymous from their perspective and to understand what it can reliably be used for. Documentation accompanying datasets is an established practice in AI, e.g. datasheets for datasets[31] and data cards[32]. Para 41, however, treats documentation only as an internal record of the anonymising controller. We note that the documentation is itself an output derived from the original data, and that identification risks should be assessed together with the anonymous data it accompanies (see the Composition section).

We recommend that the EDPB considers requiring documentation to accompany anonymous data, as an inseparable part of it[33], wherever it is passed on, and develop guidance on what it should include and under which conditions.

Constructing the Chain of Entities

Choosing the applicable perspectives

To construct a "chain of entities" (para 9) resulting from means reasonably likely to be used, an initial entity for the chain must be chosen. Fn 21 indicates the initial entity is the one "for whom [the data] is supposed to be anonymous". With reference to para 12, this would be the "applicable perspective(s)" for the assessment. Paras 13 and 14 extend these "applicable perspective(s)" to entities based on objective circumstances, such as data access, and normative requirements by the GDPR. Entities which will receive the data, be it directly or indirectly, must however also be included as means reasonably likely to be used (para 30).

We recommend the Guidelines provide a clearer distinction between the "applicable perspective(s)" for the assessment in paras 11 to 15 and relevant entities for the "means reasonably likely to be used" in paras 26 to 35.

[31] Timnit Gebru and others, 'Datasheets for Datasets' (2021) 64(12) CACM 86.
[32] Mahima Pushkarna, Andrew Zaldivar and Oddur Kjartansson, 'Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI' in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '22) (ACM 2022).
[33] Shlomi Hod and Ran Canetti, 'Differentially Private Release of Israel's National Registry of Live Births' in 2025 *IEEE Symposium on Security and Privacy* (SP) (IEEE 2025).

Data is not anonymised for the controller if they retain the original data

In para 10 of the guidelines, the EDPB refers to a broad concept of "anonymisation", as "any processing which successfully produces anonymous data under the GDPR, whether or not the original data is deleted". Recital 26 of the GDPR explains that the data protection principles should not apply to anonymous information. This contradicts the EDPB's notion of anonymisation that "does not contribute to storage limitation", which would mean that this data protection principle would still apply in relation to the "anonymous data".

We recommend that the Guidelines refer to processing which successfully produces data that is anonymous for relevant entities under the GDPR (para 11), which do not necessarily include the controller of the anonymisation process (as indicated by paras 37 and 39).

Providing a method to assess the means reasonably likely to be used

The Guidelines provide factors to be considered for the means reasonably likely used for identification (among others paras 20, 29 and 30) and discuss factors that can reduce or limit the likeliness of such means being used for identification (among others paras 26, 32, 33, 34, 89, 90, 91, 92 and 93).

Note that the CJEU treats "disproportionate effort in terms of time, cost and labour" only as an example of when identification is "impossible in practice", and not as an independent alternative.[34]

The Guidelines do however not provide a method for assessing the means reasonably likely to be used for identification by a relevant entity. This is a crucial step done for each relevant entity as part of the re-identification risk assessment process and threat modelling discussed above. Whether a means is available to, and likely to be used by, a relevant entity, is in substance a question of foreseeability.[35] In addition, the effectiveness of available means is also a question of measurement, which requires an evaluation protocol as well as standards to be implemented successfully (see the Setting the thresholds for insignificant identification in reality section).

We recommend that the Guidelines provide a clear and well-explained test method to assess the means reasonably likely to be used by a relevant entity in a specific data processing context.

---

[34] Case C-413/23 P *EDPS v SRB* ECLI:EU:C:2025:645, para 82.

[35] Peter Craddock, 'EDPB Guidelines on Anonymisation: where unforeseeable is reasonable?' (DATALAWS, 15 July 2026) < https://datalaws.net/2026/07/edpb-guidelines-on-anonymisation-where-unforeseeable-is-reasonable/> accessed 29 October 2026. However, we do not agree with Craddock's conclusion that unlawful means do not count as reasonable means of identification.

The chain of entities leads to the most capable realistic entity

We appreciate the chain-of-entities concept (para 28, fn 21). It captures what makes anonymisation hard in practice: an entity's means include the means of anyone to whom it can pass the data (para 20), so the assessment must follow the data wherever it can reasonably go. The guidelines already recognise that data rarely stays where it is put – Example 22 counts hackers, public authorities, accidental recipients and anyone who obtains the data from them as relevant entities for a dataset that is neither published nor shared, and Example 8 adds entities beyond the reach of EU law – and to these one must add the ordinary commercial routes of sale to a data broker or outsourcing to a cloud provider, which are lawful, routine and require no ill intent (motivation being in any case excluded, para 31). Whichever route the data takes, it arrives with an entity that holds both the technical means and the additional information needed to identify (para 29(c)).

What we believe is not sufficiently clear in the guidelines is that this is where their own principles lead: followed as required, the chain of entities ends at a highly capable entity for any data that can leave the controller's control. Assuming the most capable realistic entity is not an additional caution but the logical conclusion of the analysis the guidelines prescribe.

Following a chain to its conclusion yields several entities, each of which would require separate analysis. We argue that a chain is best reduced to a single entity combining the capabilities of everyone on it. This lowers the analytical burden and removes the need to assume anything about cooperation between entities on the chain, or about which entities to include – a selection that is otherwise open-ended (para 30, fn 25). Since different routes produce different chains, the result is a small collection of the most capable realistic entities – big tech companies, global data brokers, foreign intelligence services, and possibly others specific to the domain – rather than one.

The consequence is that the contextual approach, applied as the guidelines themselves require, converges on this collection of entities. For any data that can leave the controller's control, it is against these entities that the three criteria should be tested.

We recommend that the Guidelines instruct that any narrower set of entities is a claim about scope that the controller must justify, and the Guidelines should set out what such a justification requires.

Evidence that the prohibition in question is not respected by the entities involved

We agree that for a prohibition by law to be considered with respect to the likelihood of means being used for identification, the prohibition must effectively prevent the use of such means (para 32). The Guidelines however explicitly denote that all the referenced cases involved “actors who were properly limited by the rule of law”. The meaning of this addition for a considerable prohibition by law is unclear. Is it a limitation on the scope

with respect to certain actors (e.g., public-law entities), a distinct requirement regarding the existence of procedural rules, with relevance for the prohibition, or is it a question of practical enforcement?

We recommend that the Guidelines explain more clearly what a prohibition by law in this context requires.

As to the effectiveness of prohibitions by law, a controller should not have to assess the enforceability of the law and of certain prohibitions under the law in a particular country (para 33). We also find it extremely difficult if achievable at all for a controller to assess when the gains from breaking the law may outweigh the associated costs and risks. For example, the controller might not have the confidential information regarding organization revenue structure, available resources, and so forth.

We recommend that the Guidelines do not include these evidence factors or to provide specific (calculatory) examples.

We recommend that the Guidelines include another factor that there is evidence that the entities involved have previously breached or circumvented the law or required compliance measures for the prohibition in question and other related aspects (such as, cybersecurity).

Taking into account the factors presented in para 33 requires non-trivial legal and technical analysis.

We recommend that the Guidelines include an example that demonstrates the kind of reasoning that this analysis requires, as Example 8 does not demonstrate the interplay of these factors well.

Technical, contractual and organizational measures to constrain the chain of entities

As the reasonably likely means of identification to be considered for any entity are very far reaching in nature, especially through a likely chain of entities, measures to restrict the scope of means for identification become constitutive for the contextual approach. The Guidelines should more clearly explain the potential of using technical and organisational measures to constrain the chain of entities that must be considered.

We strongly support the approach of considering technical measures to support contractual measures. However, the role of technical measure in conjunction with contractual measures is unclear.

We recommend that the Guidelines better elaborate the role of technical measures in this interplay, and how to choose them for the purpose of limiting the reasonably likely means that must be considered.

Contractual measures must be "reliable, verifiable and enforceable", yet the meaning of these standards are not defined in the guidelines.

We recommend that the Guidelines explain the meaning of these standards more clearly and should explain how entities can comply with these.

We also recommend that the Guidelines consider the intended impact of contractual measures, i.e., whether their conceptual design effectively limits the means reasonably likely to be used. This can also be supported through (dis-)incentivizing measures, such as sensible penalties in the case of a breach of contract.

Consider the following as a use-case. A common means for this purpose is sharing the data only with selected entities that are willing to sign a Data Usage Agreement (DUA) that excludes re-identification attempts and making data available to third parties. However, para 34 calls this approach into question with the argument that a contractual partner may not be trustworthy. Secure Processing Environments (SPE)[36], when implemented correctly, can limit access to data as a technical measure. What is the interplay, for example, between DUA and SPE?

We recommend that the Guidelines include an example that stress-tests the role of DUA and SPEs according to the guidelines to illustrate how contractual and technical measures might work together

We encourage the EDPB to consider organisational measures (for example, a use and access committee) and certification (for example, in cybersecurity and risk management) as a supplement to the technical and legal access-limitation measures.

Harms because of Anonymisation

As discussed before, the assessment in the course of the anonymisation process focuses on identification risks, i.e., when anonymisation fails. However, there are also issues that may arise despite or because of an effective anonymisation and that should be considered during the ex-ante risk assessment of this process. Examples include:

- Aggregated results may introduce biases against certain groups, such as misrepresentations of minorities, that may for example "jeopardise the promotion of educational equity".[37]
- In medical research, there may be a need and an interest by patients to get informed on incidental findings about their health, requiring a way to learn or

[36] See Article 2(20) Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) [2022] OJ L 152/1; Article 73 Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 [2025] OJ L 5.3.2025.
[37] Paulo Fazendeiro, Paula Prata and Maria Eugénia Ferrão, 'Subtle biases introduced in equity studies through data anonymization' (2025) 20(10): e0332441 PLoS One.

receive information about relevant research results that have undergone anonymisation.[38]
- AI models trained on hateful content may reproduce it, thereby leading to re-dissemination issues.

We recommend that the Guidelines explain that the risks of anonymisation must not only be assessed and mitigated in view of identification, but also with respect to the purpose for which the anonymised data is intended to be used.

Purpose does not travel with the anonymised data

Purpose limitation binds personal data, and anonymisation is the one operation whose output is, by definition, no longer personal data; at that boundary the purpose declared at collection simply stops. The guidelines make the crossing free: anonymisation is defined by its result alone, "any processing which successfully produces anonymous data" (para 10), and para 38 presumes that its legal basis coincides with that of the preceding processing when it is part of the same activity and pursues the same purposes. The input is thus covered by an inherited basis, the output falls outside the GDPR, and the computation in between – where the useful result of any analysis (such as a dataset, a model, aggregates) is actually produced – is never examined.

A controller may therefore collect data with consent for analysis X, train an AI model on it and make sure it successfully passes the anonymisation assessment described in the Guidelines. The controller may even delete the original Data. Then, the controller (or another entity that the controller shared the model with) uses the model for a purpose Z for which consent would never have been obtained. Formally only anonymisation occurred, no purpose Z was ever declared or assessed, and purpose limitation was not circumvented but never triggered. Purpose limitation and the granularity of consent exist to prevent exactly this, and para 38 provides cover for the one step in which the refused analysis takes place.

We recommend that the Guidelines state that the purpose of anonymisation processing includes the intended use of its output, to be assessed for compatibility with the original purpose before the boundary is crossed.

Definitions

Given data

The glossary defines given data as "data whose anonymity is to be assessed", and the document then treats it, almost without exception, as a table of records: images are mentioned only in passing, and AI models only as a special case of inference, never in

[38] Henry S. Richardson and Mildred K. Cho, 'Secondary researchers' duties to return incidental findings and individual research results: a partial-entrustment account' [2012] 14 Genet Med 467.

the definition or in any worked example.[39] Yet a trained model is itself a data source: through memorisation and overfitting it retains information about the individuals in its training data, and its anonymity must be assessed on the same footing. The guidelines have already extended the criteria from records to aggregate statistics (paras 58, 81; Example 19); the same generalisation should be applied to given data itself.

We recommend that that Guidelines define "given data" as the output of any processing that is claimed to be, or relied on as, an anonymisation, whatever its form – such as but not limited to tables, images, audio and text, synthetic data, aggregate statistics, AI models and their parameter updates, vector embeddings, interactive systems, and quotations alike.

Para 77 notes that inference often combines several aggregate datasets produced from the same original data, but the guidelines do not require such outputs to be assessed together.

We recommend that the Guidelines instruct assessing all outputs from the same original data available to an entity as one body of given data.

"Treat them differently" as element of identification

In para 19 of the guidelines, the EDPB explains that "[t]o identify a natural person means to distinguish them from others within a given context, consequently making it possible to treat them differently from those other people". The second half of this defining sentence adds a new element to the definition of identification previously introduced by the WP29.[40] The Guidelines however do not explain the meaning of this new element for the identification requirement of personal data, which makes its use in the illustrative examples of the Guidelines unclear.

Treating natural persons differently as a result of data processing functionally explains when data relates to a natural person, e.g., by its effect.[41] *Stevens* also previously argued – in relation to singling out - that individuals are treated differently, "when 'singling out' would result in impact upon individuals", although acknowledging that the term "singling out" within the GDPR is disconnected from this understanding.[42] However, the new element to the definition of identification – because it follows the word

[39] Images appear in the definition of record (para 5; Glossary), as content relating to a person (para 16(a)) and in Example 5; AI models appear as objects from which data can be extracted (para 77), as data "aggregated to represent correlations" that can be prompted (para 82), and otherwise as tools of the attacker (paras 83, 92–93). The worked examples concern tables (Examples 10, 12, 19, 24), lists of records (Examples 11, 15), aggregate figures (Examples 16, 17) or single statements (Examples 1, 13).
[40] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' 01248/07/EN WP136 (20 June 2007) 12.
[41] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data' 01248/07/EN WP136 (20 June 2007) 11.
[42] Leslie Stevens 'The Proposed Data Protection Regulation and Its Potential Impact on Social Sciences Research in the UK' (2015) 1 EDPL 97.

"consequently" – could also be interpreted as a mere logical conclusion of the original distinguishability element, without adding any new substantive meaning.

To illustrate the unclear meaning of the identification requirement as a whole, we compare Examples 5 and 24 of the Guidelines. In Example 5, a natural person is considered identified in CCTV footage solely based on: "the person in a black suit who is also walking a dog". Conversely, in Example 24 of the guidelines, the likely unique survey answers can distinguish a respondent from others within that context. However, the survey answers alone are not considered enough to identify the respondent, as "there is no way to link the answers to their identity". How is the information in Example 5 sufficient to identify a natural person within the given context, but not the information in Example 24?

We recommend that the Guidelines explain more clearly the significance of the new element to the identification requirement and the role of the word "consequently". This would also help in resolving misunderstandings about how the identification requirement is used in the illustrative examples of the Guidelines.

The Criteria for the technical analysis

The three criteria for the assessment of anonymisation are explained using the simplified approach. This leads to all examples for the explanation of the three criteria being assessed without considering the means of relevant entities, which promotes the simplified approach. Section 3.5.3 gives only two examples of how the three criteria are applied in view of the contextual approach. It rather extends on the means reasonably likely to be used by relevant entities under the contextual approach, separating practical explanations from related considerations made in Section 2.4.2.

We recommend that the Guidelines provide a more detailed explanation of how the three criteria are to be applied under the contextual approach. This should include at least one realistic example for each of the three criteria (including one for "singling out") and one that assesses all of them holistically. Realistic examples may include longitudinal studies in research as well as probabilistic or unstable measurements of record attributes.

In general, we recommend that the Guidelines provide a structure for the assessment of the three criteria and for the means reasonably likely to be used at the beginning of Section 3.4.

No Record isolation

Record isolation does not seem to be a helpful criterion for the assessment. Many datasets will have records with more complex entries than Example 10 (e.g., records with continuous values or containing noise). On the other hand, qualitative data will usually contain unique records (Example 10 also does not capture qualitative data). Focusing only on the uniqueness of records within the given data also incentivizes

collecting personal data from more individuals than necessary to meet the criterion, which contradicts with the data minimisation principle. But even if the criterion is not met, this says nothing about the identifiability of an individual. Instead, one must then assess whether individuals can be singled out via the unique records (para 98).

It is not clear whether and how this criterion applies to given data that is not in the form of a collection of records. For aggregate statistics or an AI model, which we argued above should be treated as given data in their own right, it is not clear what a record is and how the criterion would apply.

We recommend that the Guidelines treat No Singling Out as a criterion instead of No Record Isolation. Furthermore, if No Record Isolation is being left in the Guidelines, we recommend that the Guidelines clarify the applicability of it.

No Singling out

Singling out is explicitly mentioned in recital 26 of the GDPR as means of identification, but the GDPR does not define it. The Guidelines reference singling out but provide no clear definition in the text, and it's unclear if this just refers to record isolation or to identification in general. Fn 37 promises a detailed explanation of singling out in Section 3.5.3, which does not deliver on it. Para 98 only describes it as matching isolated records to the individual's attributes, which is essentially linkage. The accompanying Example 23 asks whether individuals can be identified rather than singled out. This raises the question of how identification and singling out relate to each other.

We recommend that the Guidelines provide a clearer and distinctive definition of singling out as a means of identification.

No Linkage

The No Linkage criterion is defined too narrowly. Para 59 describes linkage as "linking information from one dataset to another", and para 60 requires that a record in the given data be matched to a record from a different dataset; "information", "additional information" and "dataset" are nowhere defined, although record, record-level data and aggregate data are. Standard linkage joins a record to a record, but the second source need not be a dataset at all: a record can equally be joined to a trained model, which through memorisation answers for the individuals it was trained on; or to a table of aggregate statistics, as the guidelines' own Example 19 shows when it recovers a diagnosis from a histogram and files the result under inference.

We recommend that the Guidelines define linkage broader as the combination of the given data with any other source of information relating to the same individuals (via content, purpose, or effect), whatever its form – such as but not limited to tables, images, audio and text, synthetic data, aggregate statistics, AI models and their parameter updates, vector embeddings, interactive systems, and quotations alike – so

that the No Linkage criterion covers these cases directly rather than leaving them to No Inference. The same recommendation applies to the definition of given data.

In addition, we recommend that the Guidelines clarify the intention behind the No Linkage criterion more clearly from the outset. Para 59 suggests that a violation of the No Linkage criterion can lead to a violation of the anonymity of the given data. However, paras 101 to 103 explain that a violation of the No Linkage criterion does not itself attest to whether an individual is identifiable via the given data. Instead, a violation of the No Linkage criterion requires re-testing of the "new (linked) information" against the three criteria. In essence this means (also because the No Record Isolation criterion is obsolete), that only the singling out and the No Inference criterion attest to whether an individual is identifiable via the given data.

No Inference

No Inference should be the primary criterion

We welcome the distinction in para 70 between inferences drawn from information about the population at large and inferences that rely on an individual's data. In particular, we see the counterfactual test in para 71(b) and its elaboration in fn 43 and fn 44 as the technically correct definition of leakage: anonymisation is broken when something can be learned about an individual that could not have been learned without that individual's data in the original dataset. In our view, this is one of the strongest parts of the guidelines.

Such an inference can also be considered as the only objective of an attack on anonymisation. What is learned lies on a continuum, from membership (a single bit) through attribute inference (including identity) and predicate singling-out to reconstruction of the full record.

We recommend that the Guidelines consider No Inference as the primary criterion.

Given the weight that para 71(b) carries under this recommendation, we also recommend that the Guidelines include guidance on how to conduct the test. For example, where the anonymisation process is randomised, a single run of A(D) and A(D') is not informative and the comparison has to be made over repeated runs.

"General knowledge" and "population at large" are not well defined

It is unclear whether para 70's definition, "a result of a generalisation from the original dataset", covers both terms or only the second, and whether the two differ. Furthermore, although fn 42 provides some scope of "population at large", it leaves much room for interpretation of definition of the "sample" such as its size and granularity.

We recommend that the Guidelines define each term explicitly, including in the glossary. We recommend that the Guidelines refine the scope in fn 42.

The general knowledge alternative is constructed in hindsight

Showing that an inference "could have been drawn from general knowledge or from information about the population at large" (para 71(a), 71(c)) is done with the inference already in hand. The given data told the controller what to look for: a population-level findings that reaches the same conclusion may be often available. The showing is a rationalisation, and the information that made it possible came from the given data. For example, a controller generates synthetic data from the original dataset and releases it, or locates synthetic data on similar populations already released by others. Some queries on it, such as certain histograms, reproduce the original data accurately, others do not. The controller then releases, as aggregate statistics calculated from the original data, only the histograms that are accurate enough to allow attributes of individuals to be inferred, and justifies them as general knowledge derivable from the synthetic data.

Para 70 describes such inferences as "a result of a generalisation from the original dataset represented by the given data." The only way to demonstrate that an inference is such a generalisation (rather than, for example, copying, memorisation or overfitting) is the counterfactual of para 71(b) and fn 43: fix the procedure, run it with and without the individual, compare. The Guidelines already contain the right tool. Fn 44 points in the same direction: it ties the prior belief to A(D'), the given data produced without the individual, which is the comparison of para 71(b) restated in terms of prior and posterior.

We recommend that the Guidelines do not include para 71(a) and 71(c), or amend them so that a general knowledge or information about the population at large showing must establish that it was reasonable to produce the inference without access to the given data, and argue for this. We believe claims about general knowledge or about the population at large that do not go through the para 71(b) approach should be held to a high standard and receive greater scrutiny.

In any case, we recommend the Guidelines require documentation, logs and an evaluation of the anonymisation process under para 71(b) at the time of anonymisation, retained so that the assessment can be verified and re-run later. This is especially critical where the controller chooses to delete the original data, since from that point para 71(b) can no longer be run.

The membership inference example understates the harm

We recommend that the Guidelines replace the membership inference example (Example 18) with one that demonstrates a more serious harm, such as revealing a medical condition or a political opinion, since being part of the dataset is in itself sensitive information.

Assessing whether an inference is meaningful

The first condition for an inference to be meaningful under para 67 is that the processing of the inferred information is liable to have an effect on the data subject's rights and interests. We discussed this condition in the section on Harm and impact above, where we recommend that harm become a dimension of the assessment as a whole.

With respect to the No Inference criterion, we recommend that the Guidelines provide guidance on how to assess whether an inference is liable to have such an effect, or refer to established methods for assessing impacts on data subjects.

Scientific Research Data

Research is one of the activities where the practical consequences of the Guidelines are most acute. The freedom of the sciences is protected by Art. 13 of the Charter; it is rightfully also considered a privileged data processing purpose under the GDPR. We acknowledge the importance of data protection in scientific research, both to protect the rights of individuals and to ensure they can trust that their data will be handled responsibly. In line with the data minimisation principle, researchers must anonymise personal data as soon as the research purposes allow it (Art. 89(1) GDPR). The diversity of research data and limited resources available in research projects require domain-specific and resource-sensible guidance on the anonymisation process. Furthermore, established practices of research data infrastructures should inform the Guidelines.

Resources and shared infrastructure

Many research institutions lack the resources and domain knowledge for a contextual approach and will default to the simplified approach. This may limit the scope of scientific research and will create inequality between well-resourced and other institutions. Research data centres and trusted research environments operated as shared services (for example within national research data infrastructures, EOSC or the common European data spaces) can carry the necessary expertise, documentation, accreditation and audits for many institutions.

We recommend that the EDPB intensify its deliberations on anonymization for scientific research, and in particular, to

- encourage the use of shared services to facilitate the anonymisation of personal data, including by providing codes of conduct under Art. 40 GDPR for operators of research data centres and with respect to secure processing environments;
- clarify how para 15 applies to repositories and environment operators acting on behalf of depositing researchers;

- provide more specific guidance and standards on how anonymisation can be achieved in unique research domains, including through consultation with research data experts working in those domains;
- cross-reference Guidelines 1/2026, so that a failed anonymity assessment is not read as a bar to research.

Tiered access is established practice in research

Research data centres, such as those accredited by the German Data Forum (RatSWD), provide access in tiers[43] [44], for example: (1) public use files; (2) scientific use files for registered researchers; (3) on-site or remote access to more detailed data under supervision. This long-standing practice embraces the contextual approach described in the Guidelines.

One measure of the toolbox to implement tiered access is Secure Processing Environments (SPE). They may reduce the likeliness of means reasonably likely to be used for identification, especially with regard to the entities that must be considered ("breaking the chain"). SPEs should therefore be considered, when implemented appropriately, as technical, contractual and organisational measures that support the contextualisation of the anonymity assessment, especially when working with highly sensitive data that has undergone anonymisation.

We recommend that the Guidelines acknowledge the tiered access, in which anonymisation methods are combined with technical, contractual, and organisational measures, in particular the secure processing environments (SPE) measure, as a way of implementing the contextual approach.

To facilitate discovery of anonymised research data available through tiered access, researchers should know about its existence. A standard way to do this is to share the data's metadata with researchers.

We recommend that the Guidelines provide guidance or examples that consider the impact of publicly available and domain-specific metadata on the identification risk (see the Composition section).

Qualitative and textual data requires the contextualised approach

The Guidelines and the examples therein focus mostly on structured given data that is subject to the anonymity assessment. In social sciences among others, qualitative and unstructured data, such as interviews, biographical-narrative data, text corpora and

[43] Micah Altman and others, 'Towards a Modern Approach to Privacy-Aware Government Data Releases' (2016) 30 Berk Tech LJ 1967.
[44] Rat für Sozial- und Wirtschaftsdaten, *Handreichung Datenschutz* (RatSWD Output 8 (6), 2nd edn, RatSWD 2020) <https://doi.org/10.17620/02671.50>.

social media are unique by default and can be linked to a single individual both by content as well as through style. Anonymising such data under the simplified approach will often make the data useless for research, while the contextual approach for such data remains under-developed.

We recommend that the Guidelines provide guidance and a realistic example on how to apply the contextual approach to the assessment of anonymity for qualitative and unstructured research data. This example should include technical, contractual and organisational measures such as tiered access.

Editorial comments

We found a few minor mistakes throughout the guidelines.

- Terminology (pp. 2, 16, 33): “contextual”/”contextualised” and “simple”/”simplified” are used interchangeably. One term for each should be used throughout.
- Para 53: “reintroduced sub-section 3.5” is missing a word.
- Para 71(c): “rebuttal presumption” should read “rebuttable presumption”.
- Fn 44: “footnote 40” and “footnote 41” should read “footnote 42” and “footnote 43”.
- Para 89: “a mean reasonably likely” should read “a means reasonably likely”.
- Para 97 table: “See paras. 98 to 00” should read “See paras. 98 to 100”.
- Example 24: “No Records Isolation” should read “No Record Isolation”.
- Flowchart: para 104 is referred but it does not exist.

AI Disclosure

General-purpose AI tools were used solely for proofreading and editing; the authors reviewed the output and take full responsibility for the document.

Acknowledgements

We appreciate the thoughtful input from all participants of the red-teaming workshop at KIT. We thank Dr. Dara Hallinan, Maria Rill, Dr. Kris Shrish and others for their valuable contributions. The workshop has thankfully been supported by the Topic Engineering Secure Systems of the Helmholtz Association (HGF) and KASTEL Security Research Labs

We would also like to thank Dr. Diana Dimitrova for her review of this response.

The contribution Patrick Brunner and Thilo Gottschalk was supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) under the National Research Data Infrastructure – [NFDI 52/1] – 501930651.

The contribution of Shlomi Hod was supported by the German Ministry of Research, Technology, and Space as part of the project SynTrace (16KIS2674K).

The contribution of Simon Hanisch was supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/2 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of TUD Dresden University of Technology.

The contribution of Jan Krasni was supported by ERIHAD.OS, funded by the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751.

The contribution of Abhishek Kumar Mishra was supported by H.W. & J. Hector Stiftung, Karlsruhe.

The contribution of Stephanie von Maltzan was supported by NFDI4DataScience (NFDI4DS) project (DFG project no. 460234259).

The contribution of Kati Mozygemba was supported by KonsortSWD-NFDI4Society funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) (funding number 442494171).

The contribution of Jochem W. Rieger was supported by DFG Device Center INST 184/261-1 and EU COST Action 24161.

The contribution of Thorsten Strufe was supported by H.W. & J. Hector Stiftung and the Karlsruhe Security Research Labs (Topic ESS of HGF).