Observatoire ARGA

INTERNATIONAL ANALYTICAL REPORT

Study #22

Personal Data Anonymization in the Age of Powerful AI Models

Risks of re-identification, array linking, and independent proof of effectiveness

Organization: Observatoire ARGA

English version

21 rue de l'Aviation, 64600 Anglet, France

info@argaobservatory.org | www.argaobservatory.org

October 22, 2026

Contents

Summary for management....5
1. Subject of the study, questions and limits of conclusions....5
2. Legal boundary: personal, pseudonymized and anonymous data....6
3. EDPB Draft 02/2026: New architecture for anonymity assessment....7
4. Contextual assessment and the recipient's perspective....8
5. No Record Isolation Criteria: Record Isolation....9
6. No Linkage Criterion: Linking to External Arrays....9
7. No Inference Criterion: Inference of Sensitive Information....10
8. Legal regime of the anonymization operation itself....10
9. Special categories, criminal procedure and sanctions information....11
10. EU case law and the test of factual identifiability....12
11. Threat model and capabilities of a motivated attacker....12
12. Direct and indirect identifiers: event data....13
13. Geolocation, time and movement trajectories....14
14. Relationship graphs, text embeddings and vector bases....15
15. Suppression, Generalization, and Randomization: What Works and Where It Breaks Down....15
16. k-anonymity, l-diversity, and t-closeness: model boundaries....16
17. Differential privacy: formal guarantee and meaning of parameters....17
18. Accounting for the DP budget, composition and quality of published results....18
19. Synthetic data and simulation sets....18
20. PETs, Federated Learning, and Secure Computing....19
21. Can a trained AI model be anonymous?....19
22. Memorization, regurgitation, and retrieval of training examples....20
23. Attacks on the belonging of a record to the training corpus....21
24. Model inversion and attribute inference....22
25. Test organization: test design and examiner independence....22
26. Statistical uncertainty and erroneous test conclusions....23
27. Re-checking after updating data and models....24
28. Publishing Rules: Open, Restricted, and Closed Environments....25
29. Documenting data provenance and transformation chain....25
30. International comparability: EU, UK and US....25
31. Protection of subjects and implementation of their rights....26
32. Security Breaches and Risk of Unauthorized Disclosure....27
33. Application in international search and extradition cases....28
34. Application in sanctions and AML/KYC compliance....29
35. Migration, asylum and decisions on vulnerable persons....30
36. Methodology for independent anonymity audit for Observatoire ARGA....31
37. EDPB Open Consultation: Proposals from the Observatoire ARGA....32
38. Pilot verification program for international research projects....33
39. Institutional guarantees of independence and prevention of conflicts of interest....33
40. Limitations of the study and conditions for applying the findings....34
41. Conclusions and institutional recommendations....35
Appendix 1. Glossary of key terms for interdisciplinary verification....35
Appendix 2. Primary Legal Decision Tree....36
Appendix 3. Array passport before processing....36
Appendix 4. Register of legal grounds and special restrictions....37
Appendix 5. Matrix of threats and offender competencies....37
Appendix 6. Map of supporting data and external sources....37
Appendix 7. No Record Isolation Working Protocol....38
Appendix 8. No Linkage Protocol and False Match Control....38

Appendix 9. No Inference Protocol for Individual Attributes....39
Appendix 10. Numerical example of k-anonymity: two different shares....39
Appendix 11. Differential Privacy Budget Journal....39
Appendix 12. Quality chart of statistical result with noise....40
Appendix 13. Checking the synthetic set before transmission....40
Appendix 14. Testing the AI model for reproducing training fragments....40
Appendix 15. Calculating the positive predictive value of a membership inference attack....41
Appendix 16. Interpretation of negative tests....41
Appendix 17. Vector Indices and RAG Checklist....41
Appendix 18. Evaluation of Remote Identifiers in Free Text....42
Appendix 19. Evaluation of trajectories and geographic data....42
Appendix 20. Control of Articles 9, 10, 11 and 25 of the GDPR....42
Appendix 21. AI model passport for inferring personal status....43
Appendix 22. Due diligence of an external anonymization provider....43
Appendix 23. Form of decision on release of anonymized result....43
Appendix 24. Anonymity Recheck Triggers....44
Appendix 25. Map of the incident of erroneous anonymization....44
Appendix 26. Form of a person's request for possible identification....45
Appendix 27. Training Scenario: International Search....45
Appendix 28. Training scenario: sanctions false positive....45
Appendix 29. Training Scenario: Asylum and Migration Report....46
Appendix 30. Draft comments for EDPB consultation 02/2026....46
Appendix 31. Charter of Auditor Independence....46
Appendix 32. Final Editorial Review Checklist....47
Appendix 33. Detailed case: a small group and the disclosure of a medical characteristic....47
Appendix 34. Biometric features, photographs and voice....48
Appendix 35. Multilingual Corpora and Identifier Extraction Errors....49
Appendix 36. Composition of Aggregated Publications and Differential Disclosure....49
Appendix 37. Anonymization of the financial relations graph....50
Appendix 38. A hypothetical example of the difference in perspectives between the controller and the recipient....51
Appendix 39. Contextual risk of models with open weights and closed API....51
Appendix 40. Removing data from the model and the limits of machine unlearning....52
Appendix 41. Protocol for preliminary assessment of suitability of open publication....53
Appendix 42. Protocol for reconstructing the linking of several publications....54
Appendix 43. Technical test for data extraction from the language model....54
Appendix 44. Membership inference: Interpreting precision in a rare event....55
Appendix 45. Risk register for vector bases and retrieval-augmented generation....56
Appendix 46. Comparative Protocol for the EU, the UK and the US....56
Appendix 47. Principle of minimum disclosure in litigation and regulatory disputes....57
Appendix 48. Test of Consistency of Independent Assessors....58
Appendix 49. Post-Publication Re-Identification Response Plan....58
Appendix 50. Contractual matrix for independent anonymity testing....59
Appendix 51. Managing Differential Privacy in a Series of Publications....60
Appendix 52. Evaluation of small samples and secondary suppression....61
Appendix 53. Management of source evidence and scientific reproducibility....61
Appendix 54. Escalation and Re-Check Threshold System....62
Appendix 55. Appeals of individuals and independent review of controversial findings....62
Appendix 56. Methodology for assessing the quality and usefulness of an anonymous result....63
Appendix 57. Complex scenario: multilingual judicial corpus....64
Appendix 58. Integrated Scenario: A Series of Monthly Statistical Bulletins....64
Appendix 59. Complex scenario: synthetic database of banking transactions....65

Appendix 60. Complex Script: Photo, Audio, and Metadata.....66
Appendix 61. Complex scenario: scientific access to rare medical and migration events.....66
Appendix 62. Complex Scenario: A Contractor Who Promises “100% Anonymity”.....67
Appendix 63. Complex Scenario: Reusing the Result for AI Training.....68
Appendix 64. Final map of the implementation of the ARGA methodology for 12 months.....68
Sources and methodological basis.....69

Summary for management

The key finding of this study is that anonymity is a contextual, verifiable, and reviewable legal and technical conclusion, not a name for a procedure for removing personal identifiers. Neither k-anonymity, nor differential privacy, nor synthetic generation, individually, replace testing for reasonably plausible ways of extracting, linking, and inferring data about an individual. This is particularly important for large language models, vector indices, forensic narratives, and small groups. [1][2][3]

As of 22 October 2026, EDPB Guidelines 02/2026 on Anonymisation remain a draft for public consultation, ending on 30 October. The mandatory requirements remain based on the GDPR and applicable national law; the EDPB draft is used for methodological analysis, but not as a new regulation that has already entered into force. The processing of personal data during anonymisation itself requires an independent legal assessment. [1][2]

A six-stage review process is proposed for Observatoire ARGA: defining the purpose and target audience; inventorying the initial data; constructing a threat model; technical and legal assessment; independent peer review; and documented release decision with triggers for re-review. A negative conclusion, where open publication is inadmissible, should be considered a valid professional outcome, not a research failure.

The primary red flags include a unique sequence of procedural events, the possibility of linking to public court records, secondary recovery of suppressed statistical cells, the storage of data in RAG indexes, and unsubstantiated claims by the developer that the model is completely anonymized. In all such cases, the conclusion is limited by the factual context and available evidence.

Practical route

Management is advised to establish the roles of data owner, independent technical reviewer, and legal expert; use access-controlled versions of documents; not publish real individual cases solely for the purpose of illustrating methodology; regularly review external sources for changes; and document any release denials. This report contains model procedures and training scenarios but does not claim to represent the results of actual customer testing or current accreditation of the research organization.

1. Subject of the study, questions and limits of conclusions

Anonymization of personal data has become a key requirement for the responsible use of large data sets for statistics, scientific research, inter-organizational exchange, and training of artificial intelligence systems. However, a modern AI model is capable of comparing disparate features, inferring probable attributes of a person, and reproducing fragments of training material. Therefore, the question is no longer simply whether a name, passport number, and address have been removed from a table. It is necessary to establish whether the individual remains identifiable by means that can reasonably be used by specific participants in the processing.[1][2][4]

This study answers four interrelated questions: when can data be considered anonymous in the legal sense; how to assess the residual risks of extraction, linking, and inference; what evidence should the developer or owner of the dataset provide; and how does the assessment change when moving from an anonymized table to a trained model, vector index, synthetic dataset, or published statistics. The analysis is normative and methodological in nature: the examples provided are hypothetical validation models, not test results from real organizations.

The work was prepared as of 22 October 2026. Of particular importance is the draft EDPB Guidelines 02/2026 on Anonymisation, open for public comment until 30 October 2026. It represents the interpretative position of the European Data Protection Board, but is not yet the final version of the guidelines and does not amend the text of the GDPR. Where the report proposes its own thresholds, forms, or procedures, these are expressly designated as recommendations of the Observatoire ARGA, not as existing legal obligations.[1][2]

The practical outcome of the study is a draft methodology for independent anonymity assessment, a set of control questions, matrices, test protocols, and procedures for disclosing results. Particular attention is paid to sensitive ARGA contexts: international investigations, extradition, sanctions and financial compliance, immigration procedures, and public reporting. In these areas, errors in record linking and disclosure of even small amounts of information can have disproportionately severe consequences.

2. Legal boundary: personal, pseudonymized and anonymous data

Article 4(1) of the GDPR covers information relating to an identified or identifiable natural person. Article 4(5) defines pseudonymization as processing after which the data can no longer be attributed to a specific data subject without additional information stored separately and under appropriate safeguards. Pseudonymization remains a method of processing personal data and does not in itself place it outside the scope of the GDPR. Anonymous information is information that is not attributed to an identified or identifiable natural person, taking into account the applicable conditions and means of identification.[1][10][14]

Recital 26 of the GDPR requires that all means reasonably employable by the controller or another person be taken into account, including costs, time, available technology, and technological developments. Therefore, the legal question is not tantamount to the mathematical impossibility of identification under all conceivable circumstances. However, formally removing identifiers, replacing a name with a code, and the existence of a confidentiality agreement are also not sufficient proof of anonymity.[1][3]

A practical classification must encompass not only the "dataset as a thing" but also the recipient's perspective, the moment of transmission, the auxiliary data available to them, the access rules, and the consequences for the individual. A publicly accessible table, a closed statistical result, and a dataset disclosed to a single trusted researcher all have different threat models. It's impossible to transfer an anonymity conclusion from one publication model to another without further analysis.

Category	Legal conclusion	Key check
Identifiable information	Personal data; GDPR applies where applicable	Direct or indirect identification
Pseudonymized information	Typically remain personal in the context of the controller	Availability of key, keychain and additional information
Anonymous information	Outside the scope of the GDPR as information about a subject that is not identifiable	Reasonably probable means of identification for relevant participants
Mix of data	GDPR requirements apply to the personal section	Isolation of components and impossibility of reverse coupling

Additional expert assessment

Practical disputes typically begin with the incorrect question: "Can the owner themselves discover the identity of the individual?" This question is too narrow. What matters is the ability to identify not only the owner but also any other relevant party, taking into account all means reasonably plausible in the given context. Both the data and the controlled release must be analyzed: is the recipient able to access a separate key, comparable commercial registries, event logs, search

engines, and local information not available to the original controller? Verifiability of the output requires explicitly stating whose perspective is being considered, what means are available to them, and why other means are deemed unreasonably plausible. [1][2]

Three operations must not be confused: pseudonymization protects an attribute's connection to a person, anonymization seeks to eliminate identifiability for relevant parties, and data deletion limits the continued existence of specific copies. A deleted surname guarantees neither anonymity nor complete deletion if fragments of the original document remain in indexes. Moreover, a legal conclusion cannot be determined solely by the filename 'anonymous.csv' or the contractor's certification. [1]

The processing chain should distinguish between the states *before*, *during*, and *after* the transformation. When an anonymous result is created, an organization may process personal data and is obligated to comply with the applicable rules for this operation. Only subsequently, if the result truly does not allow for identification of an individual under the contextual legal test, may it be exempt from the relevant personal data protection regime. Documenting the distinction between these stages prevents erroneous retroactive "zeroing" of obligations.

3. EDPB Draft 02/2026: New architecture for anonymity assessment

Guideline 02/2026, version 1.0, was adopted for public consultation on 7 July 2026 and published on 8 July. It updates the approach of the 2014 Article 29 Working Group and uses a more detailed assessment framework for modern processing tools. The draft proposes starting with two questions: does the information relate to an individual and can that individual be identified or distinguished in a way that allows for different treatment. Only then does the researcher proceed to technical criteria.[2][3]

The draft distinguishes between a contextual approach, which takes into account the capabilities of each relevant participant, and a simplified approach, which establishes a more conservative assessment without a detailed distinction between the parties. The former is closer to the contextual legal test but is more difficult to document. The latter may be useful for an organization that prefers a uniform, high level of caution, but may lead to excessive recognition of information as personal in situations where it is objectively anonymous for an individual recipient. Neither approach eliminates the need for reasonable, substantiated analysis.[2]

The technical core of the project consists of three criteria: No Record Isolation, No Linkage, and No Inference. When met together, these criteria provide a strong basis for concluding that anonymity is achieved. Violating any one criterion does not automatically mean that the data constitutes personal data: further assessment is required to determine whether the discovered capability actually allows for the identification of an individual in a specific context. This distinction is crucial, as "failing the test" and "establishing a violation of the GDPR" are not synonymous.[2]

For Observatoire ARGA, the 02/2026 draft is primarily useful due to its auditable procedure. When preparing an independent opinion, it is recommended to separately document the legal object of the assessment, the list of relevant participants, the source of supporting information, the selected test, its actual result, and the limits of applicability of the conclusion. Such an opinion allows for external review and revision following the emergence of new identification technologies.

Additional expert assessment

The EDPB project distinguishes between a contextual and a simplified approach. Three criteria—the impossibility of isolating a record, linking information, and deriving information—serve as a practical framework for assessment. However, they cannot be turned into a universal mathematical axiom. If the test of one criterion reveals a risk, further assessment of its materiality and legal

context is required; if all three tests are passed, the result confirms the examined configuration, but does not confirm the existence of an absolute guarantee for an indefinite period. [2]

This leads to two requirements for an independent report. First, each test describes the attacker, their initial knowledge, the information they obtained, and their level of confidence. Second, a negative result is explained through the test's detection capabilities: which attack classes are covered, which are excluded, how many attempts were made, and whether the original settings are available for verification. Only with this data will another specialist be able to assess the adequacy of the research. The recommendation lies not in a formal "three-tick" report, but in a hierarchy of evidence. [2][15]

For the EDPB's commentary, it's worth offering the example of a legal narrative. A publication may not contain unique strings suitable for automated search, but it can still unambiguously lead to a specific case thanks to a combination of rare dates, procedural events, and interstate movements. In such a situation, semantic matching poses a threat in its own right, requiring manual expert testing alongside statistical metrics.

4. Contextual assessment and the recipient's perspective

A single data set may be personal to an organization that maintains a table of codes and names, but non-personal to a recipient who objectively cannot obtain a match by lawful and reasonably probable means. It is precisely this relativity that must be analyzed, rather than reduced to a universal thesis that "pseudonyms are always anonymous to third parties." In case C-413/23 P *EDPS v SRB* of 4 September 2025, the Court of Justice of the EU separately considered the recipient's perspective and simultaneously emphasized that the controller's obligation to inform subjects is assessed at the time of data collection from its own perspective. The case was remanded to the General Court; the Court's decision does not constitute an abstract indulgence for any transfer of pseudonymized data sets.[7]

A contextual review must establish the legality and practical feasibility of access to auxiliary information. Open databases, leaks, official authority, contractual rights to request data, and the ability to obtain information through an intermediary are assessed separately. The mere existence of an identifier registry in the world does not make it reasonably probable for every observer to access it. However, a formal contractual prohibition without technical and organizational measures does not prove the impossibility of linking.[1][2][8]

In ARGA projects, it's especially important to distinguish between five perspectives: the original controller, the anonymizing contractor, the closed recipient, the public user of the published dataset, and the adversary with access to the data outside the intended route. If the material is planned for open publication, an analysis of a broad and long-term audience is used, not just the current user. For closed access, additional confirmation of the reality of contractual, technical, and organizational restrictions is required.

Relevant participant	Available funds	Preliminary conclusion
Controller with source registry	Key, internal events, change logs	High probability of attribution if the link is preserved
Processor with limited API	Functional access, possible logs and metadata	Side channel testing is mandatory
Recipient of anonymized extract	Publications, own database, allowed queries	The conclusion depends on the objectively available context
Uncertain audience	Open Sources, New Services, and Future Leaks	A more conservative test is required.

5. No Record Isolation Criteria: Record Isolation

Record isolation implies the ability to distinguish information about one individual from others, even without knowing their name. A combination of a precise date, a small town, a rare profession, and an unusual event sometimes constitutes a de facto identifier. In the EDPB draft 02/2026, the No Record Isolation criterion checks whether an array contains a unique combination of attribute values related to a single individual. Increasing the dimensionality and precision of the data generally increases the probability of uniqueness, but does not replace the test with an automatic presumption.[2]

For example, a table of trips with age, exact border crossing time, and route direction might highlight a single record without a name. Summarizing time to a month or region may reduce uniqueness but deprive the research of some of its usefulness. The auditor's task is to compare possible transformations with the publication's purpose and their residual risks. The evidentiary protocol should record the processing rules, the distribution of equivalent group sizes, the proportion of unique records, and the dynamics after updating the set.

The legal risk of identifying someone solely based on the mathematical criterion of uniqueness cannot be considered established without assessing the ability to distinguish the individual in a real-world context. However, a high proportion of unique profiles warrants in-depth verification and, in some cases, denial of public disclosure. The experience of WP29 and the CNIL shows that mechanically replacing fields does not ensure robustness to extraction and matching.[3][5]

For independent audits, it is recommended to calculate not only the total number of unique combinations but also their distribution by risk class, such as rare events, sensitive geographies, and decision types. Information on international searches or health status can reveal more than simple administrative indicators. Therefore, a single threshold for all contexts is not recommended.

6. No Linkage Criterion: Linking to External Arrays

Linking can occur without shared formal identifiers. A few timestamps, a route, a rare job title, and age can sometimes allow two records from different databases to be combined. Sources can include public court documents, corporate registries, news materials, social media, and official archives. The EDPB project identifies No Linkage as a standalone test: it should assess the ability to connect information about an individual in a given database with other data in such a way as to create a distinct relationship to the same individual.[2]

One of the characteristic risks arises from the sequential publication of aggregated sets of data. Today, a regional table of court decisions is published, tomorrow – data by age and case type, then – a narrow statistical report on extradition dates. Each table may appear secure on its own, but the overlap between them reduces the groups to one or two individuals. The auditor should check the planned disclosure in conjunction with the controller's own previous publications and the most accessible external datasets.

It's practical to describe the "matcher's capabilities": which fields match, how robust they are, whether there are publicly available events, what errors are possible, and how the false-match rate was estimated. Not only the percentage of successfully linked records is important, but also the confidence in the accuracy of the attribution. Incorrectly assigning a criminal or sanctioned status can cause independent damage, even if the attacker hasn't identified the real subject of the original dataset with high probability.

When transferring data to a secure research environment, risk can be mitigated through access to pre-approved requests, upload limits, export auditing, independent key control, and rules for publishing derived results. These mechanisms manage risk but do not automatically anonymize the original data. [12][15]

7. No Inference Criterion: Inference of Sensitive Information

Even if a specific record isn't isolated or linked to another database, publication can allow for highly reliable inference of information about an individual. EDPB draft 02/2026 considers No Inference as a third criterion, focusing on the accuracy and applicability of such inference. Unlike simple correlation at the level of a large group, inference about an individual or a very narrow group, which allows for special treatment, can be dangerous.[2]

For example, an anonymized report from an institution shows that on a specific day, a single applicant of a rare nationality received a specific medical report. Even without a name, the combination of institution, time, and external information can reveal information about the individual. Another example is the publication of sanctions match probabilities for a small sector, from which a counterparty can deduce who was subject to an in-depth review. A coherent review of criteria is needed, not three independent "checkboxes."

A mathematical prediction in itself does not equal true knowledge. Therefore, the protocol must distinguish between the verified truth of an attribute, probabilistic inference, false positives, and the subjective confidence of the observer. High classifier accuracy on an unbalanced test set does not eliminate dangerous errors for small groups. A single accuracy number should not be used as sufficient proof of privacy protection.

ARGA's recommendation: test scenarios in which the attacker has independent facts about a person, as well as scenarios where they only know the category membership. For each test, the initial knowledge, source, inference improvement, and potential harm are recorded. The legal opinion should describe this risk in a way that is understandable to a lawyer, not just an ML specialist.

8. Legal regime of the anonymization operation itself

When the original material is personal data, its transformation into anonymized form generally constitutes processing of personal data. Therefore, the principles of Article 5, the appropriate legal basis of Article 6, and, in the case of special categories of information or data concerning criminal offences, the additional requirements of Articles 9 and 10 may apply to the operation. Achieving anonymization in the future does not retroactively legitimize the unlawful collection, creation of intermediate copies, or transfer of the original material to a contractor.[1]

The controller must specify the purpose of processing, assess the need to use original identifiers, establish retention periods, and define the boundaries of the processor's authority. Even a temporary pseudonym mapping table can be a critical personal resource. Its deletion must be documented, taking into account backups, logs, downloaded models, cloud objects, and physical destruction deadlines. An unverifiable promise of "key deleted" is not a sufficient procedure.

In the field of research, Article 89 of the GDPR allows for special safeguards and may interact with national regulations and exceptions. However, "scientific interest" does not provide a universal, independent authorization to collect any information about an individual. An analysis of the specific purpose, legal basis, proportionality, and admissibility of subsequent transfer is required. Potential to limit a subject's rights is determined by applicable law, not by the contract between the research commissioner and the research provider.[1]

Stage	The main question	Trace of the document
Getting the original array	On what basis was the data obtained?	Processing register, origin, purpose
Transformation	Who sees the source data and keys?	Access scheme, DPA, operation log
Stability check	What attacks and roles are assessed?	Test protocol, limitations
Disclosure of the result	What audience is the conclusion valid for?	Release decision, set version

9. Special categories, criminal procedure and sanctions information

Anonymizing data about health, beliefs, biometric characteristics, or vulnerable individuals requires greater care, not only because disclosure could cause serious harm. The operations themselves on the original data must have a legal basis for processing special categories of data, within the meaning of Article 9 of the GDPR. Materials on criminal convictions and other criminal offences fall under a separate Article 10, which relates to official controls or appropriate safeguards provided by Union or Member State law. These provisions cannot be mechanically combined.[1]

Sanctions lists, information on wanted persons, procedural decisions, and investigations may simultaneously contain personal data and information about an individual's legal status. Their treatment depends on the origin, nature of the data, and the application of specific regulations. The existence of a public record does not guarantee the legitimacy of a new indexation, mass association with a commercial profile, or the publication of an individual risk score. The purpose of reuse, the accuracy of the information, the date of status expiration, and the ability to correct erroneous matches are of importance.

For ARGA, a "maximum protection against reverse attribution" principle is proposed: before publishing analytical examples, reduce the detail of events, change rare combinations of features, exclude materials with a clear individual profile, and separately test comparability with open procedural documents. If the material is used in a legal case, it may require a different handling logic: preserving the original in a closed evidentiary environment while publishing only a secure aggregated assessment.

Deleting the identifying link should not undermine the rights and possibilities of proof in a specific case, if the corresponding obligation to preserve the data is provided for by law. Therefore, anonymization for scientific publication and the storage of appropriate evidence in a legal file are separated architecturally, in terms of retention periods and access rights. This is not a compromise of words, but two distinct processing operations.

Additional expert assessment

Information about health, religious beliefs, biometric characteristics, and other special categories requires analysis under Article 9 of the GDPR when it falls within categories established by law. Information about criminal records and criminal offenses is treated separately under Article 10. It cannot be mechanically combined into a single list of "sensitive data" without distinguishing the grounds and restrictions. Information about sanctions or international wanted lists may not automatically fall under the same terminological regime, but may carry a high risk of harm and, depending on the circumstances, include data protected by special regulations. [1]

In a public case of denied extradition, it's not just the name and case number that are dangerous. Sometimes, a combination of date, citizenship, country of detention, and trial date is sufficient. If there's only one news item about similar events in the archive, readers can associate the report with a specific individual in minutes. Therefore, authors must check for semantic uniqueness before publication. Simply removing a name from a lawyer's case file in such an environment is pseudonymization, not reliable anonymization.

Coordination between the interests of research and protection of individuals must occur before typesetting. Otherwise, the editor will discover at the last minute that a key story contains unique details and will be forced to either remove the supporting evidence or publish the dangerous material. A prudent strategy provides two lines of evidence in advance: a closed original for authorized recipients and a public, methodologically independent case model.

10. EU case law and the test of factual identifiability

The EU Court of Justice's decision in the Breyer case (C-582/14, October 19, 2016) emphasizes the importance of accessible legal means to obtain additional information: a dynamic IP address can be considered personal data for a website operator when there are means of contacting the individual through the provider. The Court did not find that any digital identifier is always and for everyone personal, regardless of the circumstances.[8]

In the case of *IAB Europe* (C-604/22, March 7, 2024), a string recording a user's consent preferences was examined. The court emphasized that machine-readable code may relate to an identifiable individual if it can be linked to other information by appropriate means. An organization that sets the processing rules may also, under certain circumstances, act as a joint controller, even without direct access to each data set.[9]

The case of *EDPS v SRB* (C-413/23 P, 4 September 2025) demonstrates the need to distinguish the classification of data for the recipient from the controller's initial duty of transparency at the time of data collection. The court overturned the General Court's decision and remanded the case for a new trial. Therefore, the formula "after pseudonymization, the third recipient is exempt from the GDPR" is legally incorrect: actual evidence is required to demonstrate the impossibility of attributing the data to a given recipient, and the obligations of the initial controller are analyzed independently.[7]

These decisions form a methodological framework: first, determine whether the information pertains to a specific individual, then establish possible means of identification from the perspective of the relevant participant, and then separately examine the responsibilities of each party and each stage. This report does not replace an analysis of national regulations and sector-specific regimes; judicial principles are applied to the factual circumstances, not as abstract platitudes in the margins of an anonymization form.

Additional expert assessment

In Case C-413/23 P of 4 September 2025, the Court of Justice of the EU considered the transfer of pseudonymised comments to a third party in the context of Regulation (EU) 2018/1725. The significance of the decision is not that all pseudonymised data is always automatically anonymous or personal for any recipient. The Court drew attention to the need for a legally correct analysis of identifiability, taking into account the specific party and the structure of the transfer. [5]

To apply this logic to practical assessment, the auditor creates separate capability maps: for the original controller, the technical contractor, the research institute, the public, and the potential bad-faith recipient. If one participant has a compliance key and the other does not, this is a material fact, but not a complete answer: other reasonably plausible means of identification must be examined. Contractual restrictions, actual access, and supporting public data that may be publicly available are critical.

A decision on a specific dispute does not replace the text of the GDPR and does not exempt one from assessing other obligations: the lawfulness of the original processing, transparency, security, and the rights of data subjects. Transforming a single judicial argument into a general formula, "the recipient doesn't have the key, therefore the data is anonymous," would be a methodological and legal error. This is especially true in international cases, where recipients differ in their capabilities and legal obligations.

11. Threat model and capabilities of a motivated attacker

Anonymity testing begins with determining who might attempt to identify a participant and what they actually have access to. The British ICO recommends the concept of a "motivated attacker": a sufficiently competent individual with publicly available resources and the ability to purposefully

exploit publicly available and other reasonably accessible information. This model doesn't require unlimited resources, nor does it allow for the use of a random user uninterested in matching records.[15]

The recommended adversary matrix includes four categories: the average public user, a professional researcher or journalist, a recipient with their own commercial base, and an insider with privileges. For each category, it's necessary to describe the available data, computing resources, motivation, legal barriers, and the realistic possibility of breaching these barriers. If high-value financial data is involved, assumptions about motivation and resources should be stricter than when publishing general traffic statistics.

Threats are categorized by their lifecycle: source file disclosure, lookup table compromise, attribute recovery based on indirect evidence, linking to adjacent publications, data extraction from the model, leakage through the query interface, and re-identification after the introduction of a new external database. The report should indicate which scenarios were tested, which were excluded, and why. A test cannot be called "independent" without access to sufficient material to reproduce the main tests.

To measure risk, a registry of attempts is compiled: attack hypothesis, source of information, allowed request budget, actual attempts, achieved result, uncertainty, and impact on subjects. Refusal to conduct dangerous attacks on real people should be accompanied by a secure replacement with synthetic or consistent test data, rather than hiding behind the "no risk" label.

Additional expert assessment

When assessing an attacker's capabilities, one should not consider only a highly qualified security specialist. In some cases, a local journalist, colleague, or relative with knowledge of the context of the event may be more easily identified. Conversely, sophisticated machine learning attacks require computing resources and access to the model that are unavailable to the average recipient. The test is based on several attacker profiles: a public observer, a researcher with legitimate access, an insider, a commercial data aggregator, and a party specifically targeting an individual. [15]

For each profile, the supporting information available legally or factually, the expected motivation, the cost of a matching attempt, and the consequences of success are recorded. Time is also a factor: the ability to find a person in a public registry in one minute and a hypothetical attack requiring access to the private data sets of several countries are not equivalent. Both can be discussed, but they should be assessed separately and not reduced to the abstract formula "theoretically possible."

The threat model should be revised according to the release of new tools, changes in record availability, partner data leaks, and increased computing power. The reliability of the report is confirmed by its honest listing of threat classes unavailable for testing and the identification of events that require reassessment.

12. Direct and indirect identifiers: event data

In practice, the most resilient identifiers aren't just passport numbers and full names. A sequence of rare events, a description of a profession, the time of a surgery, a geographic route, a procedure number, or an unusual set of requests can all provide a stronger clue. For audit purposes, it's useful to classify attributes by role: direct identifiers, quasi-identifiers, sensitive properties, technical metadata, and quality control fields.

The transformation of quasi-identifiers must take into account their interrelationships. Generalizing ages to decades is of little help if the precise date of the event and a single job title in a small organization remain. Removing the case number does not solve the problem if the textual plot and trial date are clearly found by the search engine. Narrative fields are particularly dangerous: they can carry unique information that is invisible to automatic counting of table row repetitions.

The operational procedure involves semantic analysis of free text, searching for indirect references to individuals and small groups, checking links to unique public documents, and then testing the integration of processed attributes. The use of a personal data recognition model at this stage is useful, but requires verification of its completeness in the language and subject area of the specific dossier. An algorithmic filter should not replace expert analysis in highly sensitive materials.

Separate control is necessary for attachments: PDFs with copyright metadata, screenshots, EXIF timestamps, file names, comment IDs, and hidden layers can re-identify subjects after careful processing of the main text. The protocol should cover the entire completed package, not just its visually visible portion.

13. Geolocation, time and movement trajectories

Movement data creates the risk of identity reconstruction through "home-work-rare event" combinations, even if the coordinates are rounded. The higher the spatial and temporal resolution, the less likely it is that two independent trajectories will be identical. Location logs cannot be automatically anonymized by replacing the device identifier with a random string: the sequence of stays can itself act as a persistent fingerprint.

To reduce risk, aggregation over relatively broad areas and time intervals, truncation of rare trajectories, publication of probability distributions instead of individual routes, and restriction of access to the raw log are used. However, each operation can distort the research results: when investigating a border violation or a chain of extradition movements, the loss of time stamps can invalidate the evidence. Therefore, it is necessary to separate the closed evidentiary original and the published copy.

The audit must consider sequential releases: publishing data weekly allows for reconstructing behavior from overlapping time windows, even when each release is tested separately against small groups. An attacker can subtract and compare aggregates. The organization should maintain a common registry of releases with their filters, periods, and metadata, rather than evaluating new files in a vacuum.

Recommendation for ARGA: use the minimum necessary resolution for geographic reporting, record the permissible target's precise coordinates, check rare routes, and make an independent decision on publication if the report concerns individuals at risk of persecution. This applies to both open materials and client-accessible dashboards.

Additional expert assessment

Spatial and temporal cues often have high identifying power. A combination of a few route points can single out an individual regardless of their name; furthermore, geographic risk varies with population density and the availability of local publications. The general recommendation to "round coordinates to two digits" is unreliable without taking into account the size of the area, the frequency of events, and the number of individuals within the relevant range.[12][15]

The correct procedure involves comparing several resolution options: exact date/month/quarter, locality/region/country, full route/generalized sequence. For each option, the unique combinations, the possibility of linking to independent data, and the impact of the coarsening on the research result are assessed. If the subject of the work is compliance with procedural deadlines, excessively aggregating the time may conceal a violation. In such a case, it is wiser to provide the exact deadlines to the authorized reviewer, but publish only the final legal assessment without the rare combination of details.

Particularly dangerous are small groups with unique events: a single extradition denial in a region or a rare transfer from one facility to another. With a limited number of public cases, even a wide range of dates doesn't necessarily protect the individual. A hypothetical case, constructed with an

independent narrative, is preferable to an attempt to mechanically "dilute" a clearly recognizable reality.

14. Relationship graphs, text embeddings and vector bases

A vector representation of text doesn't contain a name in the traditional sense, but it can preserve content attributes. If embedding allows one to locate the original document, identify its topic, or distinguish a small group of people, the issue of anonymity requires separate proof. It's a mistake to assume that converting a string to a numeric array automatically destroys the personal nature of information. It's important to determine what queries vector search accepts, whether reversible links to the original documents are available, and what features can be extracted from the results.

Graph databases add another channel of identification: the role of a node and the configuration of its edges can be unique. Removing the names of participants in a financial transaction may not eliminate the ability to identify the participants based on amounts, transfer times, and counterparty structure. It's enough for one observer to possess some of the external connections to reconstruct the rest through a matching subgraph. Therefore, separate verification of each node is no substitute for verification of the graph.

For Retrieval-Augmented Generation, the entire data path must be considered: the source text, index, document metadata, access rights, search result, model context, generated response, and query logs. Anonymity in one layer does not imply anonymity in another. If a model receives a hidden identifier along with the text, its absence in the final response does not confirm that the entire processing process is exempt from GDPR.[4][6]

An independent auditor's review may include a set of queries to recover rare phrases, adjacency testing of vector representations, comparison of results with public corpora, and verification of fragments from restricted documents. These tests should be performed in a controlled environment, without publishing the personal information found in the report.

15. Suppression, Generalization, and Randomization: What Works and Where It Breaks Down

Classic methods for reducing identifiability remain useful when applied to a specific threat model. Suppression removes individual values or strings; generalization reduces the precision of attributes; randomization alters the ability to link an observation to a specific individual. However, none of these methods provides a universal guarantee. For example, address removal does not help when the case is unique; rounding the transaction amount does not eliminate counterparty rarity; and random permutation of values can destroy correlations important for research.[3][5][12]

When designing a transformation, it's important to determine what exactly needs to be preserved: frequencies by region, timing distribution, the presence of rare errors, or the ability to verify the validity of an individual decision. If only the percentage of erroneous matches in the overall sample is needed, publishing line reports is unnecessary. If a demonstrable reconstruction of a specific error is needed, anonymization to aggregates may be unacceptable, and a closed review under legal guarantees should be chosen.

The auditor examines not just the number of removed fields, but also changes in the probability of extraction, linking, and output, as well as the loss of data utility. The report should include the original characteristics in the closed protocol, the transformation method, before and after metrics, a list of unresolved risks, and a reasoned decision on the acceptability of the selected access mode.

Method	Typical effect	Typical residual vulnerability
Suppression of rare lines	Reduction of obvious isolated cases	Conclusion on the remaining units
Generalization of age, time, place	Growth of affinity groups	Linking by other attributes

Method	Typical effect	Typical residual vulnerability
Randomization of individual values	Reduced certainty of attribution	Reconstruction by statistical averaging
Synthetic publication	No straight lines with a successful generator	Memorization of training observations

16. k-anonymity, l-diversity, and t-closeness: model boundaries

k-anonymity requires that each record be indistinguishable from at least k−1 other records by the chosen quasi-identifiers. This is a useful diagnostic metric, but its meaning depends on the selected set of fields. If, in a group of five rows, all five contain the same sensitive value, a group match can reveal this value given the individual's membership in the group. This raises the problem of homogeneity of the sensitive attribute.

l-diversity attempts to mitigate the problem of homogeneity by accounting for the diversity of sensitive values within groups. t-closeness further links their distribution to the distribution across the broader population. These constructs are methodologically important, but they do not guarantee protection against well-informed observers, consistent publications, and contextual confounding. They may also require overgeneralization for small or uneven groups.[3][12]

Calculation example: if a dataset is divided into 120 groups, 15 of which include only one row for the selected quasi-identifiers, the share of such groups is 12.5%. This is not equivalent to the share of unique individuals in the entire dataset: the latter requires the weight of each group based on the number of rows. If the total number of rows is 600, the "15 unique rows out of 600" indicator is 2.5%. Reporting only 12.5% would be statistically incorrect.

ARGA recommendation: use k-, l-, and t-metrics only in conjunction with explicitly specified quasi-identifiers, an external information model, and a utility check. The conclusion "k = 5, therefore the GDPR does not apply" is legally unfounded. The feasibility of deriving legally significant attributes in small groups should be assessed separately.

Additional expert assessment

Group indistinguishability models provide useful tests, but they operate within the selected set of quasi-identifiers. If the set is supplemented by external publication, the equivalence class may disintegrate. Furthermore, if all individuals within a group share the same sensitive feature, the group size itself does not preclude the inference of this feature. Therefore, k-anonymity is not a universal replacement for No Inference analysis. [12]

A telling example: in a group of twenty people of the same age and region, the names are hidden, but all twenty share the same diagnosis. Even if it's impossible to determine which row belongs to which individual, identifying their group membership allows for a logical diagnosis. In another case, the distribution of diagnoses may be varied, but the time sequence remains unique and facilitates linking to a newspaper article. Different measures protect against different classes of threats.

Therefore, the auditor documents the full list of attributes used, the rationale for their selection, the impact of suppression on statistical quality, and tests with modified sets of quasi-identifiers. If the criterion is achieved only by excluding key cases, the study should uncover the resulting bias. The k value is not set arbitrarily out of habit, but is considered in the context of harm and available identification tools.

17. Differential privacy: formal guarantee and meaning of parameters

Differential privacy (DP) describes the constraint on the change in the distribution of a mechanism's outputs when replacing adjacent datasets that differ in the contribution of a single subject in the adopted neighborhood model. In its common form, a mechanism M has (ε,δ)-differential privacy if, for all neighboring datasets D and D′ and measurable events S, P[M(D)∈S] ≤ e^ε P[M(D′)∈S]+δ. This mathematical formulation applies to a correctly implemented mechanism with clearly defined assumptions and does not imply a complete lack of information about a person as a result.[13]

The parameter ε characterizes the limit of probability variation and depends on the mechanism and the adopted neighborhood model; δ allows for additional residual risk in the formal guarantee. A smaller ε, all else being equal, generally indicates a stronger privacy guarantee, but there is no universal legal "safe epsilon." Estimating it requires understanding the contribution of a single individual, the size of the audience, the number of releases, and possible side channels. NIST SP 800-226 (final version March 2025) warns that claiming DP without analyzing the parameters, implementation, and environment can create false confidence.[13]

This example is for illustrative purposes only: for a simple sequential composition of two mechanisms with parameters (0.5; 0) and (0.7; 0), the conservative upper bound on the total cost is ε = 1.2 and δ = 0 under appropriate assumptions. This is not a measure of the actual privacy of the ARGA model or a certificate of anonymity for the result. For repeated adaptive queries, budgeting and specific composition rules must be taken into account.

For an independent audit, it is necessary to disclose the neighborhood definition (row or person level), the maximum contribution of a subject, the noise type, the generation mechanism, the security of the random generator, the list of published issues, and the accumulated budget. Proving the correctness of a mathematical formula without verifying the code and processing chain is insufficient.

Additional expert assessment

Differential privacy formulates a statistical guarantee of the robustness of the result to the participation of the protected unit under a certain comparison of neighboring sets. The parameter epsilon is not equal to the reidentification probability, and delta cannot be described without a mechanism, a neighborhood model, and composition conditions. Even with a mathematically correct definition, an implementation can violate the guarantee with unprotected logs, a deterministic noise generator, unaccounted for repeated requests, or a side channel. [13]

In a legal opinion, it is necessary to distinguish between the mathematical fact and the legal conclusion. A mechanism may satisfy the chosen definition of DP, but the organization is still obliged to establish the lawfulness of the original processing, the adequacy of security measures, transparency, and the applicability of the rights of subjects. There is no known universal epsilon threshold that automatically converts all results into anonymous data for the purposes of the GDPR. [1][13]

The assessment methodology should require a report on the parameters, attacker model, protected unit, release composition, correctness tests, and statistical quality. Whenever possible, the units being tested and a description of the guarantee are published, while the source data and dangerous demonstration queries remain in a secure environment. This level of detail allows an external reviewer to challenge the conclusion without needlessly obtaining information about the individuals involved.

18. Accounting for the DP budget, composition and quality of published results

One of the main practical mistakes is reusing a single array with independent "every release is protected" promises. With repeated requests, privacy loss can accumulate, and aggregating outputs allows for more precise values. The system must account for budget expenditures per subject or group according to the adjacency model and technical implementation. It's impossible to simply sum up the budget ε if the mechanism makes other compositional guarantees, but it's also impossible to ignore the history of publications.

DP requires an analysis of the accuracy of the result. If the noise in a small group is comparable in magnitude to the group itself, the report may become useless or misleading. In a report on rare cases of illegal extradition, a gross distortion of a few observations can completely change the interpretation. Therefore, the criterion of usefulness must be established in advance, and a mechanism should not be selected simply for the sake of an attractive ε value.

Publishing charts and dashboards should take into account probability ranges, rounding methods, the possibility of negative noisy counters, and handling of zeros. Normalization, disclosure thresholds, and post-processing themselves can change the interpretation of the results. The phrase "differentially private" should be accompanied by a description of the metric, budget, and goals, and not serve as a decorative seal of approval.

The proposed DP release log contains the source population version, target, mechanism owner, requested statistics identifier, privacy parameters, testing evidence, accuracy, and budget cutoff date. This makes it possible to stop the publication of new slices when requests accumulate dangerously.[13]

19. Synthetic data and simulation sets

Synthetic data is created by software or a statistical generator and can sometimes reduce the need to transmit real observations. However, a generation model trained on a custom corpus can reproduce rare strings or preserve strong correlations that allow for the recovery of individual features. Therefore, the term "synthetic" describes the method of creation, not the guaranteed legal status of the result.[4][12]

Validation includes, at a minimum, searching for near-exact matches with the training set, testing record membership in the training set, attempts to reconstruct rare combinations, and analyzing statistical similarities for small groups. Generation errors should be separated from the risk of disclosure: a synthetic dataset that reproduces non-existent criminal events may be unsuitable for training a legal classifier, even with a satisfactory level of confidentiality.

The scope of application is particularly important. Synthetic questionnaires for demonstrating the database's structure can be generated from a pre-designed framework without accessing real files. However, synthetic court materials designed to replicate the frequency of procedural violations will require a convincing assessment of their representativeness and risks. Simulated personal profiles should not be published as data on real individuals; signatures, origin, and restrictions must be present in all versions.

Recommendation: Whenever possible, use completely synthetic test cases for training internal interfaces, not derived from confidential client data. If statistical closeness to real data is required, conduct a separate assessment of the synthesis mechanism and minimize access to the originals.

Additional expert assessment

Synthetic datasets are useful for testing programs, teaching methods, and publishing examples, but the word "synthetic" itself does not establish anonymity. If the generation algorithm has memorized and reproduced rare records, the result may retain a personal connection. This risk is especially noticeable in corpora where there are many standard records, but individual unique stories can be copied almost verbatim. [3][12]

The audit should include nearest neighbor searches relative to the original corpus, comparisons of rare combinations, membership inference tests, and an assessment of attribute disclosure. However, the overall similarity of distributions does not equate to proof of the absence of memorization. The auditor separately considers small groups, misclassifications, and the impact of generation on legal conclusions. If the synthetic sample reproduces too few negative cases, the analyst may erroneously conclude that there are virtually no discriminatory risks.

For publication purposes, it's reasonable to separate the two types of materials. The "training case" is manually created as a fictitious example and is not intended to be statistically representative. The "synthetic dataset" is generated using real data and may retain statistical properties, therefore requiring a separate disclosure test and methodological description. These formats cannot be interchanged for editorial convenience.

20. PETs, Federated Learning, and Secure Computing

Privacy Enhancing Technologies (PETs) encompass various threat mitigation approaches: secure multi-party computation, trusted computing environments, federated learning, cryptographic protocols, controlled research spaces, and differential privacy. These mechanisms address different challenges. For example, federated learning reduces the need to transmit source data to a central server, but model updates can reveal information if improperly configured. The presence of a secure transmission channel does not guarantee the anonymity of the results.

Cryptographic protection at rest and in transit reduces the risk of unauthorized access; the legal status of decrypted data for the authorized recipient remains a separate issue. A trusted environment restricts access and output methods, but does not eliminate the need to verify the security of its configuration, logs, and export rules. ENISA views advanced pseudonymization methods, including cryptographic ones, as risk mitigation tools rather than as a universal transformation into non-personal information.[14]

Technological choices should be guided by the threat model, not vendor trends. For joint investigations of wanted lists, secure aggregated queries without merging individual registries are possible. Independent algorithm audits may require access to a test sample within a controlled environment. For public reporting, an aggregated publication with secure thresholds and disclosure verification through release overlap is often sufficient.

The final decision should weigh legal admissibility, source data security, result quality, risk of residual extraction, and external verifiability. If a design reduces risk but does not legally guarantee anonymity, this should be clearly stated: the use of PETs and GDPR compliance are related, but not identical, issues.

21. Can a trained AI model be anonymous?

In Opinion 28/2024 of 18 December 2024, the EDPB stated that a model trained on personal data is not automatically anonymized. The assessment is conducted on a case-by-case basis. To justify anonymization, it is necessary to consider both the likelihood of obtaining personal data about training subjects directly from the model and the likelihood of obtaining it through requests, provided that reasonably probable means are used. In doing so, supervisory authorities should analyze the characteristics of the data, the architecture and training process, the deployment context, and the range of potential users.[4]

In its July 22, 2025, publication, the CNIL specifically proposes a method for providers to assess a model's GDPR status. The key risk is the memorization of fragments of the training corpus and their retrieval. Therefore, a model cannot be de-identified simply because it lacks an explicit text table or because the data is distributed across neural network parameters.[6]

For a scientific report, it's important to distinguish between "a model as a repository of probabilistic patterns" and "a model as a possible source of information about individuals." Even with high-quality tests, the absence of successful extraction in a fixed set of attempts does not prove the absolute impossibility of new attacks. Changing the access method, system prompt, fine-tuning, retrieval layer, and hardware environment can alter the probability of detection. The conclusion should include a verified version of the model and the access surface.

The proposed minimum proof package includes a description of the training corpus and its origins, a process for excluding high-risk categories, anti-memory measures, testing attacks according to a pre-registered plan, reporting of failed and successful attempts, access control to the weights, and a re-evaluation procedure after model modifications. This package does not replace the legal basis for the initial data processing.

Additional expert assessment

Concluding whether an AI model is anonymized requires an analysis of its actual capabilities, not just the structure of its training files. In Opinion 28/2024, the EDPB emphasizes the need to assess the likelihood of direct extraction of personal data, including probabilistically, and the feasibility of obtaining such information from queries to the model, taking into account reasonably probable means. The specific outcome depends on the architecture, training, access to weights, interface, and additional systems. [3]

For example, a closed interface may have filters that limit the number of queries and the return of long fragments; open weights allow the researcher to change the generation mode or use their own analysis tools. Therefore, a single conclusion about anonymity cannot be automatically transferred between the two dissemination methods. RAG additionally introduces a search layer that can return fragments from a closed corpus even if the base model lacks this capability.

Testing is carried out iteratively: acceptable scenarios are generated, attack success is measured, false recovery cases are analyzed, and limitations are established. Without access to the training corpus, the auditor must clearly state which memorization hypotheses cannot be tested. A convincing conclusion does not conceal unknown quantities behind the statement "no signs of leakage detected."

22. Memorization, regurgitation, and retrieval of training examples

Regurgitation occurs when a generative model reproduces material similar to the training data, either intentionally, accidentally, or as a result of a targeted request. The CNIL explicitly considers this mechanism as one of the significant risks when assessing the status of an AI model. However, the degree of similarity between the response and the original text does not in itself determine whether personal data has been disclosed: what matters is the presence of an identifiable individual and the nature of the reproduced information.[6][17]

Memory is affected by fragment repetition and rarity, corpus size, model complexity, training configuration, and inference characteristics. Tests should include not only obvious prompts for text repetition, but also contextual variations, alternative wording, partial prompts, and acceptable multi-step interface interaction. Separately, it is necessary to monitor whether the system allows for the extraction of personal information from a hidden search index or tool logs.

Risk mitigation measures include deduplication, exclusion of rare sensitive examples, training regularization, technical interface restrictions, and output monitoring. The CNIL cites such measures as possible ways to reduce memorization or retrieval in the context of AI development. However, the predefined response filter does not eliminate the presence of information in the model parameters and does not guarantee protection when releasing open scales.[17][20]

A professional opinion should not describe a successful test as instructions for reproducing a breach. Confidential evidence of a vulnerability must be separated from public disclosure of the risk class. For publication, it is sufficient to state the category of information extracted, the conditions, frequency, and the corrective actions taken; the personal data itself must remain secure until authorized persons make a decision.

23. Attacks on the belonging of a record to the training corpus

Membership inference aims to determine whether information about a specific individual was used to train the model. This inference can be sensitive even without reconstructing the full document: inclusion in a database of complaints, medical records, or criminal prosecution materials itself reveals facts about the subject. The evaluation method must compare the model's performance on specially selected examples from the training and validation sets and account for differences in distributions that could create a false signal.

The initial membership probability is important: even a test with high sensitivity can yield many false positives if the actual proportion of target group members is small. Therefore, the auditor must report not only the precision and recall on a balanced laboratory sample, but also an analysis of applicability to the likely conditions of a real attack. A technically sound experiment with artificial class balance does not allow for direct assessment of the risk to the population without adjusting the assumptions.

In the hypothetical scenario, it is known that the proportion of records of a certain type in the training set is 1%. With a test sensitivity of 90% and a specificity of 95%, the positive predictive value is 0.009 / (0.009 + 0.0495), or approximately 15.4%. This attack does not provide sufficient confidence to assert that a specific individual is included in the set, although its feasibility requires further investigation. The numbers here are hypothetical and are provided only to demonstrate the importance of the base frequency.

The auditor should record the access pattern, false positive rate, confidence intervals, overlap rates between test sets, and data versions. For sensitive areas, it is recommended to additionally model the actions of an attacker with knowledge of a single trip, hospitalization, or procedural event.

Additional expert assessment

For membership inference, the membership frequency in the study population is crucial. If it is low, a high sensitivity score does not guarantee that a positive signal most likely indicates a real training participant. Therefore, in practice, it is necessary to demonstrate at least true positives, false positives, true negatives, and false negatives, as well as the positive predictive value and the base frequency used. [3][13]

Attacks are often developed and evaluated using special shadow models. Their distribution may differ from the distribution of the actual model, meaning the metric values do not necessarily transfer to the target system. Leakage is additionally checked through repeated requests and threshold adjustments after reviewing the test set. To avoid optimistic bias, the protocol should fix thresholds before independent evaluation and separately account for exploratory attempts used in attack development.

The obtained results cannot be used to make unfounded personal accusations. If an attack incorrectly assigns a person to an academic institution, the public statement itself could cause reputational or legal damage. An independent monitor publishes aggregated indicators and restrictions, and confirming the membership of individual records requires separate proof of data origin.

24. Model inversion and attribute inference

Model inversion and attribute inference are aimed at reconstructing or inferring characteristics of observations used in training or accessible through the model. They differ from simply reproducing source strings: an attacker can obtain a probable attribute or prototype, which, combined with external data, allows the result to be attributed to a person. For legal assessment, the reliability of the attribution and the possibility of individual impact are essential, not just the visual similarity of the image or the formal similarity of the text.

When assessing risk, it's important to distinguish between what the model tells us about a group and what can reasonably be inferred about an individual. For example, predicting the probability of a disease based on a demographic set doesn't necessarily identify an individual; however, the combination of a closed procedure code and a rare event can effectively create an individual profile. If the system is used to score applicants, probabilistic inference can have real-world consequences even without certainty about the exact identity.

The test protocol records the attacker's initial assumptions, the model or API, the queries used, anti-hallucination controls, and independent confirmation of the recovered attributes. Results are measured separately by subgroup; risks for a rare vulnerable group should not be averaged out against a large number of less sensitive profiles. The appropriateness of test data preparation should also be assessed, and dangerous experiments on client materials without appropriate safeguards should be avoided.

The outcome of the assessment should be a limited statement: given the described assumptions and resources, it was or was not possible to obtain data of a certain category. Statements such as "the model does not store personal data" and "it is impossible to determine whether a person participated in the training" require stronger evidence than a simple set of negative tests.

25. Test organization: test design and examiner independence

A high-quality audit requires a pre-established plan: the subject of the study, the data version and model, the roles of the participants, the set of threats, the measurement methods, and the termination criteria. Without this, the researcher may inadvertently tailor tests to the desired outcome and declare security based on tests that were never capable of detecting a significant leak. The test plan must be approved before the auditor has access to the final metrics and include a procedure for recording deviations.

In an independent framework, it is advisable to separate the data preparation team from the risk assessment team. The developer has the right to communicate the architecture and necessary technical constraints, but should not solely determine which errors found are considered significant. The auditor is obligated to disclose financial, organizational, and contractual dependencies, as well as any access restrictions to scales, logs, and the training corpus. Having a management system certificate does not replace the substantive testing of a specific model.

The replay procedure includes a control copy of the source materials, file hashes, the environment version, the execution log, secure storage of results, and the ability to independently recalculate metrics. If parts of the corpus are inaccessible for legitimate reasons, the expert must indicate which conclusions can be drawn within the granted access and which cannot be confirmed. A properly limited conclusion is preferable to a confident but unprovable certificate.

For ARGA, it is recommended to prepare a two-part report: a public analytical note and a closed technical report. The first describes the method, boundaries, and risk status; the second contains evidence, including classified parameters and potentially personal examples, with limited access and retention periods.

Additional expert assessment

An independent test should not be limited to testing the safest examples pre-selected by the client. The protocol includes sampling rules, exclusion criteria, number of attempts, protection of source material, and fixed success metrics. Some scenarios are agreed upon with the developer to prevent harm, but others are created by an independent team to prevent the test from becoming a demonstration of already known positive results. [2][3]

It's especially important to separate the development of an attack from its final evaluation. If a specialist tests dozens of query variants on a single test set and then reports only the best result, the final metric doesn't reflect the expected effectiveness on independent data. A public report should show the number of configurations examined, the final threshold selected, and the performance on the deferred group. For scenarios where real data cannot be shared with an external expert, a secure environment with role-based access and an action log is used.

Before the test begins, participants agree on a procedure for immediate termination upon detection of a living person or other sensitive information. The detection is recorded in a confidential report, but publication of a self-identifying sample is not required. Independence is considered documented if the client cannot unilaterally rule out a negative result, and the conflict of interest is disclosed before the start of the test.

26. Statistical uncertainty and erroneous test conclusions

A negative test result does not prove zero risk. If no successful leaks are detected in n independent tests, the approximate upper 95% bound on the probability of the event under binomial assumptions is often estimated using the 3/n rule. For example, with 0 successes out of 300, this is approximately 1%, not zero. For a different test setup, dependent queries, or an adaptive adversary, this estimate may not be applicable. The calculation method and assumptions must be disclosed.

If 8 out of 200 attacks are successful, the empirical rate is 4%. However, this figure alone does not indicate whether the sample reflects the actual risk. It is necessary to evaluate the uncertainty intervals, the composition of the tests, the proportion of rare examples, the quality of the control, and the method of query selection. The "average success rate" may conceal a 40% vulnerability in a single, particularly sensitive subset.

Utility and privacy metrics should be reported separately. For example, a reduction in the risk of linkage after removing precise dates could simultaneously reduce the usefulness of extradition time statistics. A scientifically sound conclusion would allow for either a deterioration in both metrics or the need for a different access regime than publishing a detailed set. Tables should include denominators and the method for defining "success."

It is recommended to publish a minimum specification: population, test sample, event measure, point estimate, interval, significant systematic limitations, and the difference between the tested version and the planned product. A pre-approved acceptance matrix can be used for internal decisions, but its thresholds should be explained as a risk management policy, not a mandatory GDPR standard.

Additional expert assessment

A zero number of successful attacks detected does not imply a zero probability of success. In a simple independent trial model, with zero successes out of n attempts, an approximate upper bound of 3/n is often used for the 95% one-sided interval. For example, with zero successes out of 300, this is about 1%, but only under the assumptions of comparable and independent trials. If queries are correlated or the attacker adapts strategy after each response, the simple formula may not be applicable. [3]

Therefore, the report should show not only the "0/300" fraction, but also the sample design, the number of unique targets, the experiment duration, the adaptation mechanism, confidence limits, and the reasons for termination. Results from a single interface, language, or group of people cannot be reliably generalized to the entire lifecycle of the model. An estimate of the number of multiple hypotheses is also needed: with thousands of comparisons, some random matches are expected even in the absence of actual detection.

A separate scale is used for legal conclusions: experimentally confirmed fact, statistical evidence, technical feasibility, reasonably probable risk, and untested hypothesis. This scale prevents both overcomplacency ("not found, therefore impossible") and exaggeration ("one matching fragment proves a massive leak").

27. Re-checking after updating data and models

The anonymity assessment is not permanent. A new public registry, a leak of a related database, changes in search engine performance, API expansion, or the release of a new fine-tuning tool could transform a previously innocuous combination into an identifying one. The ICO recommends reviewing the assessment based on technological developments, new sets, and changes to the disclosure model. A similar approach is reflected in the EDPB draft document 02/2026.[2][15]

Triggers for revisions should include not only calendar dates but also changes in conditions: transfer to a new counterparty, public offering of a previously closed data set, release of additional statistical sections, data expansion to new years, merging with another system, a security incident, or the introduction of a model extraction method. The revaluation report should reference the previous analysis and explain which assumptions are no longer valid.

For AI systems, it's advisable to version control not only the model, but also the loading pipeline, training corpus, search index, postprocessing, and logging system. The same set of weights added to a new search layer can significantly change the actual risk of obtaining customized data. In this case, the notation "model unchanged" is misleading.

The recommended procedure includes a designated person responsible for monitoring new threats, a library of reproducible tests, an exception log, and a mechanism for suspending access in the event of risk escalation. The decision to republish should be made after testing is complete and documented separately from the product's production release.

Additional expert assessment

The verification report must list specific versions. Changes to weights, retraining data, tokenizer, filtering policy, RAG database, chunking procedure, or logging infrastructure can alter the probability of disclosure without visible changes to the end user. Therefore, the developer is responsible for identifying events that no longer apply previous results. This is especially important for constantly updated research platforms. [3]

The regression suite retains tests for unique narratives, control string extraction, indirect linking, and sensitive attribute inference. A separate section is designed for new attacks not used in the initial check. Threshold values are set based on the risk being investigated; previous values are retained for comparison purposes. A decrease in detections after an update does not necessarily prove the cause has been fixed if testing settings are simultaneously changed.

The responsible person maintains a log of changes and decisions on republishing. For materials concerning criminal prosecution or asylum, the relevant section may be frozen until the review is complete. The regression report must be linked to the previous act so that the reader can see the conditions in effect during the previous approval.

28. Publishing Rules: Open, Restricted, and Closed Environments

The method of information disclosure also determines the practical threat model. Public release does not allow for effective revocation of already distributed copies. A closed research environment with controlled requests reduces the number of recipients but preserves the legal processing of the original data and the risks within. Transferring through a trusted intermediary can be useful, but requires a contract, authority control, and a breach-in-case procedure.[12][15]

When choosing a data management system, one should consider not only whether anonymization is possible, but also whether publication of the data is necessary at all. In many cases, public accountability is achieved by publishing aggregated indicators and method descriptions, while retaining the primary dataset with an authorized owner. Where scientific verifiability requires reproducibility, controlled access with logs and a prohibition on arbitrary inference can be provided to reviewers.

When a disputed disclosure request arises, it's useful to divide the decision into two stages: the right to access the original information and the suitability of the resulting data for transfer to a specific recipient. Even if the public interest is high, this does not automatically eliminate the need to comply with legality, data minimization, and data protection requirements during preparation and transfer.

Mode	Advantage	Vulnerability/Control
Public open-data set	Verifiability and reusability	Irrevocable, broad binding
Closed access for specialists	More detailed research	Insider, export, contractual restrictions
Managed statistical queries	Issuance control	Differential attacks, repeated requests
Publishing only aggregates	Low detail about people	Rare cells, release intersections

29. Documenting data provenance and transformation chain

It's impossible to convincingly prove the effectiveness of anonymization without knowing what data was transformed, what was excluded, and what versions were obtained by contractors. A provenance registry must describe the original source, the purpose of collection, the permitted user base, the set of identifiers, the history of aggregations, and the derivatives created. For each operation, the configuration, the responsible party, the input and output checksums, and the decision on the subsequent fate of intermediate copies are recorded.

If data is imported from multiple countries or institutional systems, differences in legal regimes and specific restrictions of the original source must be taken into account. The technical status of an anonymized result does not negate the need to legally obtain the original source and contractual confidentiality obligations. In international research, this problem is particularly noticeable when the anonymizer is effectively given the entire data for preprocessing before their role and security measures are formalized.

When training AI models, a link between versions of the training corpus and model versions should be maintained, sufficient to allow for controlled risk checking and error response. However, the provenance system itself should not become a new centralized identifier; access to the link should be minimal, logged, and restricted. The tradeoff between traceability and privacy requires an explicit management decision.

The practical result is a "passport of origin": a brief public section with source types and processing conditions, and a private section with specific files, hashes, permissions, and legal grounds. A record without a date, role, reviewer, or version is unacceptable as a complete evidentiary trail.

30. International comparability: EU, UK and US

In the EU, the key criterion is the identifiability of a natural person under the GDPR and its Recital 26, as clarified by the Court's case law and EDPB guidance documents. In the UK, the ICO applies its own

guidance to the UK GDPR and analyzes identifiability taking into account the "motivated infringer." The ICO guidance, published in March 2025, contains a warning for revision as of October 2026 in light of the Data (Use and Access) Act. It cannot be described without reservation as a literal implementation of EU regulations.[1][2][15]

In the US, industry-specific regimes and technical guidelines exist. For example, HIPAA provides specific de-identification methods for protected health information, and NIST SP 800-188 describes methods and controls for de-identification in government data sets. These approaches can provide useful engineering tools, but certification of compliance with one regime does not automatically equate to anonymization under the GDPR.[12]

For international research exchange, a requirements comparison matrix is recommended: applicable law, identifiability verification method, permissible recipient, original legal regime, transferred artifacts, possibility of further disclosure, and error notification procedure. Separately, it is necessary to check which data are subject to cross-border transfer before anonymization. If the transfer concerns personal data, the GDPR rules on transfers to third countries are analyzed independently of the intended outcome.

An organization working with both European clients and international technical suppliers should avoid a single, global marketing claim, "fully anonymized." A more precise formulation specifies the specific version, purpose, access model, and jurisdiction for which the assessment was conducted. This makes the conclusion narrow, yet honest, and more legally robust.

Additional expert assessment

International publications often compare data from organizations with different anonymization criteria and procedures. If some sources consider de-identification to be the removal of direct identifiers, while others require verification of indirect linkages, mechanically combining the results creates a false impression of comparability. The methodological protocol must describe the actual transformation operation and its verification, not just the legal name stated in the source. [1][15]

For each dataset, the applicable law, the status of the source data, the party determining the purposes and means, the final recipients, and any additional restrictions are specified. Particular attention is paid to future use: submitting data for the calculation of a single statistic does not in itself grant permission to train a model, publish a corpus, or freely integrate it into commercial datasets. If some data fall under the US industry-specific regime, its specific requirements should be established, rather than relying on some abstract "US anonymization law."

The research task can be completed in stages: first, the criteria are reported on, then a new check is conducted for a consistent methodology, and finally a statistical comparison with uncertainty intervals is constructed. This is slower than a simple summary table, but reduces the risk of an incorrect international conclusion.

31. Protection of subjects and implementation of their rights

Even when an organization seeks to create anonymized results, the initial processing of personal data may affect rights to information, access, correction, erasure, restriction of processing, and objection. The applicability of each right depends on the legal basis, circumstances, and statutory exceptions. These rights cannot be declared irrelevant simply because the research result will be published anonymously in the future.[1]

Article 11 of the GDPR provides special provisions for situations where the purposes of processing do not or no longer require the identification of the data subject: the controller is not obliged to retain or obtain additional data solely for the purpose of identification. However, this provision does not permit the arbitrary destruction of records and thereby circumvention of legitimate requests. It is necessary to assess whether identification is truly no longer necessary for the purposes, and whether it is possible to

link the data subject's request to the retained information without creating a new, disproportionate accounting system.

If it is discovered that a supposedly anonymous data set does allow for identification of an individual, the organization must review the legal classification, limit the risk of further disclosure, and assess applicable data protection obligations. Operational steps may include suspending the public API, correcting documentation, requesting deletion from controlled recipients, notifying affected individuals where required by law, and conducting a further independent assessment. A technical recall of all downloaded copies of the public data set cannot be promised.

ARGA should establish a separate procedure for responding to complaints from people who believe they are recognized in a publication. There should be a clear contact method, a secure way to provide confirmation, a timeframe for internal review, a prohibition on unjustified requests for excessive documentation, and a means for independent escalation of a disputed decision.

Additional expert assessment

An organization sometimes claims that it cannot identify the individual in a published result, but retains the original personal data and matching key. This situation requires a separate analysis of each transaction. Article 11 of the GDPR regulates processing where identification is not necessary for the controller's purposes, but does not constitute a general exemption from the obligation to respond to requests where there is sufficient information to establish contact. [1]

The legal regulations must include a procedure for appealing to prove a connection to the publication without unreasonable requests for additional documents. The response must include information on the actual possibility of identification, the applicable regime, the adopted restrictions, and the supervisory appeal procedure. It is unacceptable to claim that an individual lacks rights simply because the dataset is designated "anonymous" in the technical data sheet. If data recovery is identified, the legal status and response must be reconsidered.

Research projects may have legally mandated special features and limitations on certain rights, but these depend on the actual purpose and relevant regulations; simply citing scientific research without a legal basis is insufficient. Ensuring rights and data minimization should be part of the research design, not an additional obligation following a public dispute.

32. Security Breaches and Risk of Unauthorized Disclosure

Erroneous anonymization and a cyber incident are distinct, although they can overlap. Incorrect legal classification of published data can lead to the unlawful disclosure of personal information without a hack. A breach of a secure repository of original data can affect personal information even if the public copy is correctly anonymized. Therefore, organizations need separate frameworks for assessing privacy and information security.[1][19]

If reverse identification is discovered, it is necessary to collect the facts: what data was accessed, from when, who accessed it, how many records were affected, what information was actually disclosed, and what the consequences were. If the incident qualifies as a personal data breach, the terms and conditions for notification are determined by Articles 33–34 of the GDPR and applicable law. Not every technical defect automatically triggers a public notification obligation, but the absence of proven exploitation does not always mean there is no risk to rights and freedoms.

Preserving digital evidence must be balanced with minimizing secondary damage. Logs, checksums, and screenshots of results should not be posted in a publicly accessible appendix to the report; a separate, limited archive should be prepared for internal investigation. If a test accidentally extracts a client's real name or case number, further copying of the example in the presentation escalates the incident.

The control system must include a designated person, criteria for suspending distribution, the legal team's competencies, a documented decision on the need for notifications, and corrective measures for republishing. The role of an independent auditor is to establish the fact, risk, and verifiability of the correction; decisions regarding procedural notifications are made by the appropriate controller with legal expertise.

Additional expert assessment

Erroneous anonymization and breaches of personal data security must be classified separately. If the controller discovers that published material allows information to be linked to an individual, they determine the status of the data and the circumstances of the disclosure. They then assess whether a personal data breach has occurred within the meaning of the GDPR, what risks it creates, and what legal obligations follow from the facts. A 72-hour notification period cannot be automatically promised for every statistical signal, but it also cannot be ruled out without a documented assessment. [1]

The technical plan includes preserving evidence, suspending access if necessary, identifying all copies and associated sets, securely correcting, and retesting. For publicly accessible content, it's important to consider the residual risk: archived versions, search indexes, and downloaded copies may exist independently of the original publisher's actions. Therefore, the corrective action fairly separates the organization's actions from the outcome, which cannot be fully guaranteed.

An internal event log records the time of detection, observed indicators, risk assessment, decision, and responsible personnel. This helps prove the reasonableness of the response and simultaneously allows for the review of false alarms, without turning every technical incident into a public crisis.

33. Application in international search and extradition cases

International wanted lists often consist of unique combinations of facts: citizenship, travel route, year of the procedural decision, circumstances of detention, judicial authority, and the language of the documents. This combination can identify a person more quickly than a name. Publishing a case study with the last name redacted does not guarantee anonymity if any reader can find the ruling or news article using the preserved details.

For ARGA analytical work, a two-tier model is recommended. The private tier stores the original evidentiary package under appropriate legal authority and confidentiality. The public tier describes a typical legal problem without reproducing a single set of events. In some cases, it is safer to present a generalized, composite training scenario, clearly labeled as hypothetical, than a "cosmetically anonymized" real case.

A public case review should mimic a standard search in official registries and open sources using a combination of the remaining attributes. If the individual is located outside their country of origin, additional information about their place of residence and the schedule of procedures may increase not only the privacy risk but also the physical risk. Therefore, the criterion for admissibility of disclosure cannot be reduced to a numerical reidentification percentage.

The opposite problem is overgeneralization, which obscures the legal significance of the violation. In this case, the researcher should choose a different publication format: a comparison of procedures and standards, aggregated statistics, or a closed legal review, rather than creating a false impression of a real court case by altering its essential facts.

Additional expert assessment

In cases involving international fugitives, rare combinations of legal facts are dangerous. Court decisions, age, date of arrest, country of asylum, and specific charges can create a unique profile

without a last name. If the research relies on a well-known public case, simply replacing geographical names is often insufficient: the sequence of events will remain recognizable through searching or semantic matching. [1][4]

For internal legal analysis, it is permissible to preserve the entire chain of evidence, provided it is properly justified and subject to limitations. For the public version, it is necessary to decide whether an individual narrative is required at all. Periodic summary indicators, methodological frameworks, and hypothetical cases often allow one to demonstrate a systemic problem with a lower risk of disclosure. If individual circumstances are necessary for public discussion of a violation, the legality of publication and the possibility of obtaining legal consent or other grounds should be separately considered, without substituting the term "anonymization" for this analysis.

Oversight and editing are separated. The legal expert is responsible for the correctness of procedural formulations, the privacy specialist for identifiability, and the independent reviewer for the repeatability of the conclusion. These different competencies prevent a situation in which a formally correct legal story causes further harm to its participant.

34. Application in sanctions and AML/KYC compliance

Sanctions screening, AML/KYC, and the processing of suspicious activity reports create specific risks of false attribution. A match based on a common name and date of birth should not automatically become a public statement of an individual's sanctioned status. Anonymization of test cases for screening algorithms should conceal both direct identifiers and rare payment routes, company combinations, timestamps, and the reasons for internal investigations.

When publishing false positive/false negative indicators, it is recommended to indicate the sample boundaries and the method for determining the true status. If the sample includes cases involving criminal offenses, the applicability of Article 10 of the GDPR and national regulations should be separately analyzed. The presence of "de-identified training data" from an external provider does not prove that the bank has the right to transfer its compliance logs to them without legal and contractual assessment.[1]

The risk matrix must consider two independent harms: the erroneous disclosure of a genuine sensitive connection and the creation of a false similarity, in which an innocent person is mistakenly linked to a separate sanctioned entity. When preparing a synthetic set for algorithm testing, fictitious records must not be transferred to the operational database without special markers and isolation.

The recommended method for independent verification is to test the model on a controlled set prepared according to a consistent design, with anonymized results by error category. If original disputed matches need to be examined, they should be submitted specifically to authorized experts rather than included in a public report. The integrity of the original evidence and the safety of publication are assessed separately.

Additional expert assessment

Sanctions checks and AML/KYC systems create the risk of false matches, where a person is attributed to the activities or status of another person. Anonymization of the research report should prevent not only disclosure but also false association with a specific counterparty. Even if the name is redacted, the amount, industry, blocking date, and name of the financial instrument may still be sufficient to identify the organization and individual. [1]

The researcher checks how many possible individuals match the published profile, whether there are isolated transactions, whether aggregation allows for the reconstruction of a specific event, and the consequences of error in reader perception. If a case yields a false positive, repeating the old

false accusation in a form that will continue to spread through search engines and training corpora should be avoided. At the same time, the conclusion about a systemic defect itself must remain verifiable through a protected set of evidence.

The relationship between the legal entity and the beneficiary is additionally assessed. Transaction graphs may remain identifying even when names are replaced with random codes if the structure of large transfers is publicly known. A separate test verifies the structural uniqueness of the graph and the ability to link to beneficiary registries.

35. Migration, asylum and decisions on vulnerable persons

Materials on international protection and migration are particularly sensitive to external linkage: the specific route, place of detention, time of application, circumstances of political or other persecution, and information about family members sometimes create a unique fingerprint. Even if the published text does not reveal the name, it may allow interested parties to identify the applicant or their relatives. Therefore, a statistical assessment of anonymity should be complemented by an analysis of security threats and potential harm.

The internal data set for analyzing migration algorithms should be minimized based on the legal issue being tested. If the rate of decision overturns on procedural grounds is being studied, detailed descriptions of personal stories are often unnecessary. If a comparative analysis of individual motivations is required, a closed research regime with reliable access control and publication of anonymized findings verified by external sources is more appropriate.

A hypothetical case: a publication reports age, the specific month of evacuation, the transit country, a rare profession, and the date of extradition denial. Each characteristic may seem general, but taken together, it can identify a single individual. Without checking external registries and news reports, the conclusion of anonymity would be unfounded. Additional verification includes mentions of children, aid providers, and other third parties who are not the applicants.

ARGA recommends applying a heightened level of editorial oversight to materials about individuals at risk of persecution: independent text approval, verification of supporting identifiers, refusal to publish dangerous details, and a procedure for urgently removing materials upon a substantiated complaint. Consent to publish an individual fact does not replace an assessment of the legality and safety of the entire set.

Additional expert assessment

In migration and humanitarian research, the consequences of disclosure can be particularly severe. Rare histories of persecution, departure routes, medical records, and family composition make it possible to identify individuals even without direct identifiers. However, excluding such cases from statistics leads to the systematic underrepresentation of the most vulnerable groups. Consequently, the problem cannot be solved simply by removing all small categories. [1][15]

A robust procedure combines limited access to evidence, more general public indicators, and an independent assessment of the extent to which anonymization has distorted the true picture of violations. The study identifies which types of cases have become undetectable due to protection, and whether this can be compensated for by high-quality methodological descriptions without disclosing identities. It is especially important not to publish combinations of country of origin, unique trial date, and rare disease if these indicators are already present in local news.

For children and vulnerable adults, increased caution applies, but the specific legal consequences are determined by the applicable regime and circumstances. Guaranteeing safety based on a single

statistical test is unacceptable. A publication decision may be negative even with satisfactory aggregate metrics if the predicted individual harm is disproportionate to the research value.

36. Methodology for independent anonymity audit for Observatoire ARGA

Six sequential stages of independent verification are proposed. The first is to define the object: a specific file, statistical mechanism, embeddings, model, or aggregate system. The second is to establish the legal framework: applicable regulations, objectives, data categories, roles, and the disclosure model. The third is to map the actors and supporting information. The fourth is to select and register No Record Isolation, No Linkage, and No Inference tests, supplementing them with model-specific attacks. The fifth is to assess the residual risk, uncertainty, and utility. The sixth is to accept and document the limited conclusion with conditions for reconsideration.[2][4]

The report must contain at least the following fields: the identifier of the verified version, the date the object was frozen, the legal basis for access, the list of test areas, adversary models, test parameters, sample characteristics, a results table, expert limitations, and a decision on disclosure. A publication limited to the words "data reliably anonymized" is not suitable for independent verification. A reproducible trace that does not reveal protected data in the open section is required.

Auditor's decision	When is it applied?	Re-evaluation condition
Reasonably anonymous in a given context	There is a verifiable package of evidence	Change of object, addressee or threats
Personal/pseudonymized data	Identifiability is preserved	Successful additional processing
There is insufficient data to draw a conclusion	No tests or recipient information available	Providing missing evidence
Publication is not permitted in the form stated.	Significant unresolved risk	Different access mode or new version

An independent auditor does not replace the data protection authority and does not issue formal authorization for processing. The Observatoire may provide research opinions and risk mitigation recommendations, but the legal decision on GDPR compliance and enforcement measures rest with the authorized entities. Commercial interests in the results must be disclosed, and the methodology must undergo internal scientific review.

Additional expert assessment

The methodology must contain at least six stages: legal definition of the scope of the audit; source inventory; construction of a threat model; technical testing; independent review; and a publication decision with the obligation to re-evaluate. The absence of any stage is reflected in the report as a limitation, rather than hidden in general statements about a successful audit. [1][2]

The auditor does not assign the universal "anonymous" label unless they have information about all relevant recipients or the model training. Instead, a contextual conclusion is formulated: what result, for what purpose and environment, and under what identification means was it examined. Evidence of reproducibility and access limits are listed separately. If the data is only suitable for a secure environment, this conclusion is considered a valid audit result, not an audit failure.

ARGA considers it important to institutionally separate research, consulting, and commercial support for specific cases. A person involved in a client's defense may legally provide source material, but should not, without an independent procedure, unilaterally approve the publication safety of their own case. The decision to publish a negative result should not be dependent on further commercial interests.

37. EDPB Open Consultation: Proposals from the Observatoire ARGA

Since Guidelines 02/2026 are currently in the consultation phase in October 2026, ARGA may provide targeted comments on the methodology without confirming that its provisions are finally adopted. Justified input should be related to the subject matter of the project: the identifiability test, the contextual approach, the three criteria, additional information available to third parties, and documentation methods. The proposals in this section constitute research proposals and do not constitute submissions to the regulator.[2]

The first proposal is to supplement practical examples with high-dimensional text and graph data, as identifying a single individual can result from a combination of narrative and structured attributes. The second is to specify the procedure for evaluating subsequent publications and external updates so that the controller does not confirm the anonymity of a new release regardless of previously published statistics. The third is to recommend separately reflecting available state and court registries if the publication concerns criminal procedure or sanctions information.

Fourth proposal: Include an example of evaluating an AI model trained on personal data, distinguishing between the anonymity of the initial dataset, model weights, retrieval index, and final response. Fifth: Provide a minimum specification of evidence for anonymity claims, including the version, threat type, test composition, negative results, and a revalidation mechanism. Sixth: Clarify how to account for the risk of false attribution, when publication leads not to the disclosure of a true identity, but to the erroneous attribution of a sensitive fact to a person.

Proposals should not be submitted as a closed critique of the regulator. Draft 02/2026 already contains significant methodological framework; ARGA's goal is to propose practical scenarios and reproducibility criteria for specific areas. When participating in the consultation, it is necessary to check the current submission window against the requirements of the official form, save the submitted version, and retain the registration confirmation.

Additional expert assessment

The proposed comments should be formulated as an exploratory position, not as a description of a mandatory test already conducted by the regulator. The draft 02/2026 under discussion remains in consultation until the end of the open period, and the final version may differ. In the examples, it is important to clearly distinguish between the legal standard of identifiability and the supporting quantitative indicators; no single metric replaces an actual analysis of the means reasonably probable for a particular participant. [2]

In terms of content, the Observatoire can offer three clarifications. The first is illustrative examples of semantic linking of legal narratives, where a unique sequence of events reconstructs identity without direct identifiers. The second is a specific methodology for releases consisting of multiple versions and overlapping aggregates, including secondary suppression. The third is a description of how technical testing of the AI model is separated from the legal conclusion regarding the personal nature of the data. These proposals should be presented as open questions for consultation, not as assertions of gaps in the current mandatory regulation.

Before submitting comments, the submission form itself is reviewed, the legality of metadata disclosure is verified, and ARGA's official position is agreed upon. Public presentation of research opinions should not include actual client information without a separate analysis of the necessity and legality.

38. Pilot verification program for international research projects

It is advisable to divide the pilot into three independent datasets: a table dataset with quasi-identifiers, a corpus of brief legal texts, and a test retrieval/response generation system. For the initial stage, it is recommended to use synthetic data and materials pre-cleared for analysis; real client files are not acceptable test material by default. Each dataset should have known initial properties so that risk assessment does not rely on assumptions about real individuals.

The table-based dataset tests group features, string rarity, and linkage to an open artificial registry. The text-based dataset tests the ability to detect unique plots and hidden metadata. The model-based dataset tests fragment memorization and retrieval, possible inferences about ownership, and the risk of retrieval leaks. The tests are pre-registered, the results are collected separately, and then applicability to legal requirements and research objectives is assessed.

Criteria for transitioning from a pilot to real data include an independent legal assessment, the appointment of a data protection officer (DPO) if necessary, a DPIA where required, a contractual framework, and verification of the technical team's competencies. In the absence of such conditions, the organization must limit itself to synthetic materials and methodological publications. A "successful pilot" does not guarantee permission to scale without re-evaluating the new environment.

The roadmap is provisional: Stage 1 – preparation of methodology and experimental setup; Stage 2 – reproducible testing; Stage 3 – independent review; Stage 4 – publication of a limited report and correction of weaknesses. For each stage, a solution owner, product, risk, and reason for stopping are assigned. Deadlines are set separately after resources are confirmed; the report does not simulate existing research.

Additional expert assessment

It is recommended to begin the pilot with a limited number of heterogeneous datasets: tabular statistics, a text corpus, a dataset with temporal and spatial features, a RAG system, and a publication case study with a small group. The choice should take into account the availability of legitimate test materials, the measurability of the results, and the availability of independent reviewers. The program should not begin with the most sensitive cases, where an error alone would create a high probability of harm. [1][2]

During the first stage, a technical specification with success and failure criteria is developed. During the second stage, a test of the basic anonymization methods and a comparative assessment of changes in data usefulness is conducted. During the third stage, an independent review and publication decision are made. For each dataset, an open methodological passport and closed evidence are maintained; their availability is adjusted according to the risks. If a link to a living individual is identified during the pilot, the trial is terminated, the information is not included in the public report, and authorized specialists evaluate subsequent measures.

The success of a pilot is assessed not by the number of positive certificates issued, but by the quality of identified limitations, the completeness of the documentation, the reproducibility of the assessments, and the ability to make an informed decision not to publish. This criterion reduces the incentive to formally confirm the customer's desired outcome.

39. Institutional guarantees of independence and prevention of conflicts of interest

When Observatoire ARGA provides scientific expertise and has an institutional link to consulting activities, it is necessary to ensure a distinction between the commercial commission and the research conclusion. The presence of a paid client does not invalidate the conclusion per se, but it does create the

risk of selecting a methodology that favors the client's interests. The document must provide for disclosure of funding sources, an independent reviewer, and the right to publish critical results within permissible legal limits.

The proposed governance system includes a ban on pre-test promises, registration of the methodology prior to testing, independent approval of protocol amendments, confidentiality controls, a mechanism for recusal of an expert, and a separate complaints procedure. In technical assessments, it is useful to maintain a registry of access rights and versions to prevent the participant requesting the test from surreptitiously replacing the object after the assessment.

The publication policy should provide for three types of documents: a methodological review, a test report, and a legal opinion on the specific context. These cannot be summed up simply as "certification." If ARGA is not an accredited conformity assessment body for the relevant scheme, the publication should not create the opposite impression. A separate legal review should verify both the conclusion and the description of the legal status of the procedure itself.

The sustainability of the system is achieved not by the size of the final PDF, but by the verifiability of decisions: it is clear what has been verified, by what means, in what context, what remains unknown, and who is responsible for its continued use. This principle applies to each attached act and brief public announcement.

Additional expert assessment

The declaration of independence must be supported by specific procedures. The editorial board discloses funding sources, establishes rules for disqualification of participants, documents prior contractual arrangements with organizations whose systems are being evaluated, and provides a channel for challenging factual findings. Independence from a government agency does not equate to freedom from dependence on a private client. [2]

If a research organization simultaneously provides paid services in a related field, the possibility of altering the scientific conclusion for the sake of commercial gain must be excluded. Reviewers involved in the preparation of the legal case being defended should not unilaterally decide on the anonymity of its published version. Positive and negative results are documented equally, and the right to publish methodological criticism cannot be entirely dependent on the consent of the reviewed party.

The process for correcting an erroneous public finding is also part of independence: the organization publishes a correction, stating the subject and date, but does not disclose new personal information in the correction itself. Quality assurance is not the absence of errors, but the ability to detect and correct them without causing undue harm to individuals.

40. Limitations of the study and conditions for applying the findings

This report presents an analysis of regulations, official recommendations, and a potential independent verification methodology. It does not include the results of access to closed AI models, testing of company systems, or research into actual international wanted lists. All numerical examples in the analytical section are illustrative and should not be cited as enforcement statistics or actual detection rates in any industry.

The results of a technical anonymity test depend on the threat model, time, environment, and volume of information available to the expert. The failure of a successful attack in a pilot experiment does not mean that future data extraction is impossible. Recommendations from NIST, CNIL, and EDPB have different legal force and purpose; they cannot be combined into a single list of direct legislative obligations.[2][4][12][13]

A particular uncertainty surrounds the status of Guidelines 02/2026 and Guidelines 01/2025, which are available on the EDPB's official resources as draft versions of the relevant consultation procedures. Before using the report in a submission, it is necessary to check whether the final versions have been released and what changes they represent. The legal analysis is based on available sources as of 22 October 2026.

If the project is to be used in litigation or interaction with a national authority, additional research into the applicable procedural law, the specific jurisdiction, and the facts is necessary. This general methodology is not intended to be individual legal advice and does not replace a qualified assessment of evidence in a specific case.

41. Conclusions and institutional recommendations

The first conclusion is that anonymization is not a name-swapping operation; it is a verifiable legal and technical assessment of factual identifiability in a given context. The second is that the EDPB 02/2026 approach is useful for constructing a reproducible verification process, but remains a draft at the consultation stage. The third is that pseudonymization, encryption, data synthesis, federated learning, and differential privacy mitigate different risks and do not yield the same legal results.[1][2][4]

Fourth, powerful AI models, embeddings, and retrieval systems require their own verification of data extraction, memorization, attribution, and association, rather than transferring training set status. Fifth, public publication should be analyzed alongside previously released data and accessible external sources. Sixth, in sensitive legal contexts, an additional criterion for choosing an access mode is the severity of the potential harm, including false attribution of criminal or sanction status.

Observatoire ARGA is recommended to approve a standard for an evidentiary anonymization passport, an independent protocol form, a registry of versions and external releases, a threat matrix, a secure testing procedure, and a revalidation process. For implementation on real-world materials, a separate decision is required regarding the legal basis, the role of the independent auditor, data protection, and access to the primary file. No internal template should be referred to as a "GDPR certificate."

It is proposed to prepare a separate public position on EDPB Guidelines 02/2026 with specific scenarios for judicial, sanctions, and migration data. The priority is reproducibility and proportionality: protecting individuals while maintaining the possibility of high-quality international analysis, without masquerading formal procedures as proof of the absence of risk.

Appendix 1. Glossary of key terms for interdisciplinary verification

The terms below are used as working vocabulary for this study. They should be compared with the GDPR definitions and the context of the specific regulatory document; the translation of a technical term does not always create an independent legal category.[1][2]

Term	Significance in research	Warning
Identifiability	The ability to distinguish and associate information with a person by reasonably available means	Depends on the context
Pseudonymization	Separating attribution to a person from the main set with separate protection of the bundle	Not equal to anonymity
Record isolation	Highlighting a unique post or profile	Does not automatically prove true identity
Linkage	Linking different sets and observations	Possibly without a name
Inference	Inferring information about a face from other features	Accuracy and harm are assessed separately
Regurgitation	Close reproduction of the training material	Not every repetition contains personal data

Term	Significance in research	Warning
Membership inference	Conclusion about the participation of a specific record in training	Depends on the base frequency
DP budget	Accounting for mathematical privacy loss by the mechanism	Not a universal legal threshold

For external documents, it is recommended to cite the original English term at the first use. If the definitions in EDPB draft 02/2026 are changed following the consultation, the terminology dictionary and the test matrix should be revised.

Appendix 2. Primary Legal Decision Tree

This tree is intended for initial screening of questions and does not replace a legal opinion. It requires a documented answer at each branch, especially if the qualification changes after disclosure to a third party.[1][2]

Step	Security question	Next step
1	Does the information relate to an individual by content, purpose or effect?	If no, record the reasons; if yes, then
2	Is the person identified or can he/she be identified?	Estimate the reasonable probable means
3	Who owns the additional information and can the recipient obtain it?	Create a participant and access map
4	Does the kit pass No Record Isolation, No Linkage and No Inference?	Record the results separately
5	If the test fails, does this lead to identifiability?	Additional contextual assessment
6	For which mode and date is the withdrawal valid?	Limited wording and revision triggers

The decision to "anonymize" must be accompanied by a version of the object and an indication of the verified audience. If the documents do not allow for the identification of a model of external information, the acceptable conclusion is "insufficient data to confirm anonymity," rather than an automatic negative or positive answer.

Appendix 3. Array passport before processing

Completing the passport precedes the transfer of initial information to cleanup specialists. It allows for the assessment of legal prerequisites and the determination of which technical procedures are generally permissible. For datasets at risk of harm, it is important to specifically identify vulnerable groups and existing evidence preservation obligations.

Passport field	Required entry
ID, version, date	Immutable code and responsible group
Source categories	Origin, jurisdiction, validity period
Purpose and basis	Specific operation and applicable standard
Categories of subjects	Including third parties and small groups
Special categories / Article 10	Separate legal assessment
Future Access Mode	Publishing / Closed Environment / Aggregates
Additional identifiers	Keys, reference books, external links
Destruction deadlines and archive	Primary original, intermediate copies

The passport must not reproduce actual personal data; it must be accompanied by pointers to secure storage. A change in the publication purpose requires updating the passport, not simply changing the name of the resulting PDF.

Appendix 4. Register of legal grounds and special restrictions

The object of registration is each significant processing operation, not the "project as a whole." For a single research program, collection, transformation, testing, and publication may have different recipients and legal conditions. The fields below must be completed on a case-by-case basis.[1]

Operation	Subject of legal assessment	Mandatory certificate
Primary collection	Articles 5 and 6, source, purpose	Description of legality and transparency
Transformation	Need for access to originals	Operation regulations, roles, security
Use of special data	Article 9 and applicable conditions	Justification for exclusion/reasons
Processing of data on offenses	Article 10 and national regulations	Relevant authority and guarantees
Export to contractor	Roles, contract, cross-border transfer	Data Route Agreement and Assessment
Publication	Anonymity of the object for the audience	Final protocol, owner's decision

The internal reviewer flags cases where publication is legally possible, but the security risk to the subject remains unacceptable. For special evidence retention regimes, automatic destruction cannot be used as a means of circumventing the obligation to preserve.

Appendix 5. Matrix of threats and offender competencies

The matrix should specify the capabilities, not the characteristics, of actual suspects. This supports the reproducibility of tests and reduces the risk of arbitrarily assuming a "safe recipient." The motivated attacker model is based on ICO recommendations, with mandatory adjustments for the project's jurisdiction.[15]

Category	Example of information	Verifiable risk	Stopping condition
Regular user	News and search results	Search for a unique event	A direct match was found
Specialized Analyst	Public registers, industry databases	Graph linking	Sensitive profiles are highlighted
Recipient under the agreement	Own client logs	Combining fields	The permission allows attribution
Internal employee	Directory, logs, key	Bypassing the demarcation	Reproducible access escalation
The intruder after the leak	Part of someone else's base	Historical connection	A new identifiability has emerged

For each scenario, the protocol captures not only potential access but also its actual reasonable probability. Mentioning a hypothetical omnipotent adversary without estimating the costs does not meet the required contextual approach.

Appendix 6. Map of supporting data and external sources

One of the typical shortcomings of the assessment is the lack of a list of available third-party datasets. For an independent audit, it is necessary to compile a list of open registries, archival publications, the controller's own datasets, and possible sources provided by the partner, along with the access conditions. Specific client links are not included in the public version.

Source category	Risk of combination	Verification method
Judicial and administrative decisions	Unique plot and date	Search for rare combinations
Corporate information	Position and relationship graph	Mapping roles and events
Geographical and temporal publications	Route and presence	Time accuracy analysis
Old reports of the organization	Differential disclosure	Release Intersection Analysis
Open media materials	Photos, quotes, events	Semantic matching
Commercial directories	Brand identifiers	Availability by recipient role

A change in the availability of any key source triggers a review. Even a recent report can be considered obsolete if, after its preparation, a new public document has become available that allows for the reliably linked, previously anonymized records.

Appendix 7. No Record Isolation Working Protocol

The protocol is designed for static structured datasets. For narrative, graph, or multimodal data, it must be supplemented with specific tests. The choice of fields and observation units is recorded before calculating the results.

Indicator	Method of fixation	Interpretation
Observation unit	Line / person / event	Discuss multiple recordings
Set of quasi-identifiers	Field list and precision	To capture the meaning of the choice
Single combinations	Number and proportion of lines	Reason for checking the connection
Small equivalent classes	Distribution of group sizes	Not a universal legal threshold
Share of lost utility	According to pre-selected tasks	Compare with publication objectives

A researcher should not publish a list of unique combinations, as this could facilitate an attack. A public report should suffice with aggregated metrics, a method, and a list of exceptions; specific rare strings are stored in a closed evidentiary appendix.

Appendix 8. No Linkage Protocol and False Match Control

It is proposed to test two layers: linking to internal published materials and to external sources reasonably accessible to the recipient. The results distinguish between true matches, probable matches, false matches, and cases that cannot be verified. Without tagging, matching names and dates cannot be considered proof of identity.

Test stage	Action	Proof
Definition of an attack	Describe the enemy's knowledge and access	Scenario and date
List of sources	Fix versions and access conditions	Register of auxiliary information
Combining fields	Run a test on permitted data	Closed result log
Quality control	Calculation of precision and false matches	Labeled control sample
Solution	Assess real identifiability	Auditor's justification

If false matches have potentially serious legal consequences, they are classified as a separate category of harm. For example, an incorrectly linked sanctions profile can be dangerous even without successfully revealing the original identity.

Appendix 9. No Inference Protocol for Individual Attributes

The test is designed to assess the ability to draw a sensitive conclusion about a specific individual from a publication, given the specified initial knowledge. The test should measure changes in information, not just the algorithm's ability to classify general categories. The report separately identifies the trueness of the attribute, the accuracy of the model, and the harm caused by false attribution.[2]

The element being tested	Documentation
Subject of conclusion	A specific attribute type, not a personal name
Attacker's Basic Information	What was known before the disclosure
Result after access	What new information has been revealed?
Reliability	Precision, errors, uncertainty interval
Effect on the subject	The possibility and consequences of different treatment
Solution	Correction, other mode, or refusal to disclose

A violation of a specific criterion of the EDPB 02/2026 draft does not automatically result in a final legal conclusion; a check is made as to whether the information obtained is capable of attributing the result to a person in the relevant context.

Appendix 10. Numerical example of k-anonymity: two different shares

The conditional set includes 600 records grouped by three quasi-identifiers. Of the 120 equivalent groups, 15 have one row each. The share of small groups among all groups is 15/120 = 12.5%, and the share of unique rows among all rows is 15/600 = 2.5%. These metrics answer different questions and should not be used interchangeably.

Size	Meaning	Calculation
Number of lines	600	Original set
Number of groups	120	By selected quasi-identifiers
Single groups	15	Empirical counter
Proportion of single groups	12.5%	15 / 120 × 100
Percentage of single lines	2.5%	15 / 600 × 100

The k value for the entire array here is 1. This does not automatically prove that each unique row truly identifies a person; tests of available external tools and sensitive inferences are necessary. The arithmetic result cannot be presented as an experiment on customer data.

Appendix 11. Differential Privacy Budget Journal

The registry is designed to prevent the unaccounted for composition of multiple statistical releases from a single population. The use of ε and δ requires a pre-selected neighborhood model, a description of the mechanism, and a validated technical implementation.[13]

Release	Example parameters	Status
Section A	$\varepsilon = 0.5$; $\delta = 0$	Just a hypothetical example
Section B	$\varepsilon = 0.7$; $\delta = 0$	Just a hypothetical example
Conservative simple composition	$\varepsilon \le 1.2$; $\delta = 0$	Subject to the following conditions being met:
Next cut	Specify new parameters	Requires decision before launch

The work log should also record the scope of participation of a single person, the relationship between sets, accuracy, the overall disclosure plan, and the responsible party. No numerical epsilon in this appendix is proposed as universally acceptable under the GDPR.

Appendix 12. Quality chart of statistical result with noise

Before using DP or other data perturbation methods, it's important to document the potential management error that the distorted indicator can cause. This is especially important in studies of rare process events: publishing noise without specifying the interval can turn a zero or low frequency into a false conclusion.

A question of quality	Entry in the card
What indicator is being protected?	Numerator, denominator, unit
Why is it being published?	A specific solution or study
Tolerable error	Pre-agreed accuracy range
Noise addition method	Formula, generator, parameters
Post-processing	Truncation of negative values, rounding
Message to the reader	Interval, limitations and prohibition of individual conclusions

If usefulness is not achieved with an acceptable level of protection, it is permissible to refuse publication of the string and switch to controlled research access. The public interest does not require presenting distorted statistics as accurate facts.

Appendix 13. Checking the synthetic set before transmission

The synthetic set undergoes a separate technical and legal review. The goal is to identify near-exact copies of the training data, the risk of membership, and the possibility of recovering rare combinations. At the same time, it is verified that the generator does not create false evidence that could be used in production.[4][12]

Trial	What is recorded	Possible measure
Match with training strings	Proportion of close couples and type of closeness	Changing the generator
Rare sensitive patterns	Number of reproduced cases	Removing risk sources
Membership inference	Metrics with base frequency	Refining the training
Quality of distributions	Discrepancies in target objectives	Limitation of applicability
Actual marking	Indication of artificial origin	Prohibition of confusion with the register of persons

If a set is derived from originals, a secure trace of origin must be maintained within the organization. Real matches found should not be published in an external application, otherwise the verification itself will become a new leak.

Appendix 14. Testing the AI model for reproducing training fragments

The test is conducted on a separate, agreed-upon test base in a controlled environment. The objective is to measure the potential for personal information to be revealed through standard and complex queries, without distributing the test data. The protocol requires the model version, access method, presence of a retrieval layer, maximum test query budget, and a decision on publication safety.[4][6]

Type of verification	Security Question
Regular user	Is information disclosed when asked neutral questions?
Restatement	Does the result change with acceptable variation?
Partial context	Is it possible to restore a rare fragment?
Retrieval index	Is it possible to access documents outside of a user role?
Response filter	Doesn't it hide the risk of accessing parameters/index?

Negative results cannot be generalized to open scales or the new version. Personal information detected is kept separate; the public report only provides the category, risk level, and corrective measures.

Appendix 15. Calculating the positive predictive value of a membership inference attack

The conditional model has a sensitivity of 90% and a specificity of 95%; the actual proportion of members of the target population is 1%. With 10,000 individuals tested, 100 training participants are expected. The test will correctly flag 90 of them. Of the 9,900 non-participants, 5% will be falsely flagged, or 495 individuals. 585 positive results are expected, and the true rate among them is 90/585 ≈ 15.38%.

Category	Forecast under conditional parameters
True participants	100
True positives	90
False positives	495
All positive	585
Positive predictive value	15.38%

These numbers demonstrate the need to consider the baseline frequency and cost of false attribution; they do not represent the actual results of the attack. Under different conditions, training dataset size, and model characteristics, the values will change.

Appendix 16. Interpretation of negative tests

In a conditional independent binomial experiment with zero successful events in 300 trials, the approximate upper bound on the probability of an event at the 95% level according to the 3/n rule is 1%. The exact one-sided bound is approximately 1 − 0.05^(1/300) ≈ 0.9936%. The interpretation is applicable only if the assumption of identical independent trials is correct.

Indicator	Meaning	Limitation
Number of tests	300	Conditional sampling
Successful events	0	Doesn't prove zero probability
Approximate boundary	1.0%	The 3/n rule
The exact border	≈ 0.994%	One-sided binomial

If the queries are adaptive, the distributions differ, or some of them occur through different channels, this interval may be incorrect. The report must specify the event type, the selection method, and the conditions for applying the statistical model.

Appendix 17. Vector Indices and RAG Checklist

Using retrieval doesn't eliminate the need to distinguish between the statuses of the original document, the index representation, and the model response. One of the most dangerous scenarios is when the aggregated result references a protected document or reproduces its rare content. Below is a minimal object-based check.

Component	Control	Unacceptable situation
Source documents	Classification, terms, basis	Illegal transfer to index
Embeddings	Rare Text Search Test	A recognizable fragment of the dossier
Metadata	Removing redundant identifiers	Subject name in a hidden field
Model context	Role-based access and permission boundaries	Mixing closed text
Logs	Access, term, masking	Collecting new sensitive stories

Component	Control	Unacceptable situation
Result	Regurgitation test	Quoting classified information

The assessment must encompass the entire model environment. A separate "embedding anonymity certificate" does not substantiate the app's security if the search connects it to non-anonymized sources.

Appendix 18. Evaluation of Remote Identifiers in Free Text

Texts require a two-stage review: first, a search for obvious identifiers, then an analysis of narrative patterns. Named entities, dates, institutions, and geographic features should be checked along with the content of individual legal events. For classified materials, two independent reader evaluations and a search compatibility check with open sources are recommended.

Category	Typical marker	Test reception
Direct identifier	Name, document, contact	Automatic search + reading
Procedural	Case number, exact date	Comparison with the official register
Plot	A rare combination of facts	Search for a combination without a name
Family	Kinship and a unique episode	Third-party verification
Technical	File author, EXIF, comments	Metadata analysis

Verification should not involve mass transmission of confidential text to an unknown external AI service. If a model is used for preliminary marker search, its processing mode and storage conditions are separately assessed.

Appendix 19. Evaluation of trajectories and geographic data

The purpose of the matrix is to prevent uniqueness of routes that persists after names are removed. The check is performed separately for exact coordinates, sequence of visits, and public events to which the subject may be linked.

The feature being tested	Potential risk	Risk reduction
Exact coordinates	Attachment to home or work	District aggregation
Minute timestamp	Coincidence with a public event	Interval enlargement
A rare sequence of places	Behavioral imprint	Reduce detail
Route from a small town	Recognizability of the subject	Suppression of rare cases
Overlapping publications	Differential disclosure	General registry of all releases

If the transformation deprives the material of evidentiary accuracy, open publication may be the wrong goal. In that case, it's necessary to keep the original closed and release only the methodological conclusions.

Appendix 20. Control of Articles 9, 10, 11 and 25 of the GDPR

The register is intended for lawyers coordinating independent research. It serves as a reminder that the technical result does not override the rules applicable to the original processing. Article 25 is supplemented by the current EDPB guidelines on data protection by design and by default; draft Guidelines 01/2025 on pseudonymization should not be considered definitive.[1][10][11]

Legal norm	Critical question
Article 9	Is there a condition for processing special categories?
Article 10	Are there any special requirements for reporting offences?
Article 11	Is identification really not required for the purposes?

Legal norm	Critical question
Article 25	Are the relevant measures already built into the design?
Article 32	Is the source code and keys secure enough?
Article 35	Is a data protection impact assessment required?

The conclusion separately reflects the legislation of the Member State if the application of Article 10 or the study conditions depends on national regulations. The absence of a jurisdictional analysis should be noted as a limitation.

Appendix 21. AI model passport for inferring personal status

A model report cannot be limited to the family name and the number of parameters. To assess the extractability of personal information, it is necessary to specify the version, training period, access type, associated indexes, and significant memory reduction measures. Separately, it is necessary to record who can call the model and what secondary sources it has access to.[4][6]

Passport section	Minimum content
Object of verification	Scale version, date and control ID
Educational resources	Categories and exceptions, without publication of originals
Access	Closed API, internal loop, open weights
Additional resources	RAG, tools and external databases
Tests	Extraction, membership, inference, regurgitation
Measures taken	Deduplication, filters, rights control
Restrictions	Unverified modes and accesses

The anonymity statement applies only to the described object and context. The publication of a new version or the inclusion of a different retrieval index requires separate verification.

Appendix 22. Due diligence of an external anonymization provider

The controller should not transfer the original customer base to the supplier based on the marketing slogan "GDPR compliant." It is necessary to establish the actual roles, the legality of the transfer route, subcontractor access, and security measures. Having a technical specification for the transformation is helpful, but without proof of its implementation, it is insufficient.[1][12][14]

Request to supplier	What to demand in return
Who decides on goals and means?	Roles and contractual scheme
Where is the raw data processed?	Regions, subprocessors, transfers
Are the matching keys saved?	Possession and Destruction Scheme
Are there any tests for binding and withdrawal?	Protocols and conditions of reproduction
What steps should be taken if a leak is detected?	Response procedure and deadlines
Is it possible to conduct an independent audit?	Access, restrictions, right of inspection

Transfer of original data should be minimal. If executing code on the customer's infrastructure is sufficient for the task, this option should be compared with fully offloading the data to an external environment.

Appendix 23. Form of decision on release of anonymized result

The form is intended for a signed decision, not an automatic certificate. The object, audience, brief justification, and review conditions are filled in. If the lawyer and technical auditor disagree, the disagreement is recorded rather than concealed by an average "green" status.

Solution field	Content
ID and version	Unique artifact code
Purpose of the release	Why is this level of detail needed?
Legal classification	Anonymous / personal / not enough data
Verified audience	Public / Private / Specific Recipient
Main residual risks	Conditions under which the risk is returned
Release decision	Allow / restrict / modify / deny
Responsible	Controller, technical reviewer, lawyer
Next revision	Date or specific trigger

Empty assumptions and constraints fields should block document approval. Conditional permission to publish aggregates only does not grant permission to publish raw rows.

Appendix 24. Anonymity Recheck Triggers

The regulations must provide for immediate and scheduled reviews. The frequency is determined by the nature of the dataset, but the date of the next audit alone does not replace event-based triggers.[2][15]

Trigger	Mandatory reaction
New external base	Repeat the binding test
Change of recipient	Update identity context
Opening an additional section	Check for overlapping publications
New model release	Check extraction and inference
Changing API mode	Review access and request budget
Security incident	Suspend and investigate if necessary
Complaint by a potentially identifiable person	Address verification and documented solution

Each revised conclusion references the identifier of the previous one. If new information changes the status of the result, a correction is published with a clear description of the boundaries of the previous conclusion, without revealing any new personal information.

Appendix 25. Map of the incident of erroneous anonymization

If there is a suspicion that published data is identifiable, fact-finding must be immediately separated from public communication. Removing the material may be a reasonable temporary measure, but it cannot be assumed that all previously downloaded copies have been destroyed. The legal qualification of a potential breach should be based on the provisions of the GDPR and the actual risk, not on the designation of an internal vulnerability.[1]

Temporary stage	Action	Responsible
Detection	Register signal and file version	Incident Officer
Containment	Limit further spread	Service owner
Facts	Set data, recipients and period	Technical group
Legal assessment	Check the status of personal data and notifications	Lawyer/Controller
Corrective measures	Re-anonymization or refusal to publish	Data owner
Closing	Independent retest and report	External reviewer

A closed appendix to the report stores technical evidence; the public report does not repeat personal data found during testing. Any notifications to regulatory authorities are prepared in accordance with applicable law and actual deadlines, not the provisional SLAs in this template.

Appendix 26. Form of a person's request for possible identification

An accessible way to challenge a publication is necessary even if an organization is confident in its methodology. The applicant need only state the material and the reason for their concerns; they should not be required to provide a full passport, medical history, or procedural file in advance if a smaller amount of information is sufficient for the initial analysis.

Field	Procedure for applying
Link to the published material	URL, version, or citation of the secure fragment
Nature of the risk	Isolation, binding, withdrawal or other harm
Contact for response	Minimum required means of communication
Proposed measures	Correction, limitation, clarification
Temporary measures	Decision to block for the duration of the investigation
Summary and reasons	Documented conclusion and escalation procedure

An organization should not automatically delete evidence of a request. A sufficient but minimized trace of the review should be maintained, with limited access and defined retention periods.

Appendix 27. Training Scenario: International Search

Hypothetical situation. A researcher wants to publish a brief description of a successful challenge to a cross-border search warrant. The name has been replaced with "Applicant A," but the text retains the year of detention, rare citizenship, transit country, specific court, and the child's age. The publication can be compared with a court decision and media report.

Examination	Possible outcome	Solution
No Record Isolation	A rare combination of events	Summarize the plot
No Linkage	Similarity to a public case	Remove exact court and dates
No Inference	Family status is derived from the case	Exclude details about the child
Utility	The principle of refusal of extradition remains in force	Publish a legal opinion, not a story

The scenario is designed for the purpose of methodology and does not describe a real client. The final document should indicate the hypothetical nature of the example and prohibit its use as evidence of fact regarding a specific case.

Appendix 28. Training scenario: sanctions false positive

Hypothetical situation. The service checks matches against sanctions lists. To assess quality, ARGA receives an anonymized table of errors: month, country, industry, name type, match explanation, and transaction amounts. In a small sector, a unique combination of large transaction and region has been preserved, allowing one to find a specific company and its manager.

Risk	Evaluation measure
Linking to the corporate registry	Search for a unique combination
False accusatory impression	Checking the explanations and marking of the outcome
Indirect identification of an individual	Separate the company and its officials
Loss of utility when generalized	Save error type statistics
Re-releases	Control of intersecting slices

A publicly accessible table can show error rates by algorithm type without disclosing the amount of a specific transaction or the exact month. The original disputed matches remain in a closed loop with legally valid access rights.

Appendix 29. Training Scenario: Asylum and Migration Report

Hypothetical situation. The migration decision report provides the country of origin, occupation, month of evacuation, route through two countries, and the outcome of the case. Even after the name is removed, the combination may be recognizable in local publications. Unique information about the child and place of residence may create an additional security risk.

Control	Question	Action
Plot uniqueness	Is the case being searched externally?	Change the level of detail
Vulnerable third parties	Is it possible to find out relatives?	Exclude additional information
Original evidence	Does the law of conservation require it?	Separate closed archive
Publication purpose	Do you need a real case?	Use a composite case study

The purpose of anonymization is to protect individuals, not to merely replace names and titles. If preserving legally significant circumstances is incompatible with secure publication, the individual case should be abandoned.

Appendix 30. Draft comments for EDPB consultation 02/2026

This text is a draft of the substantive position, not a submitted document. The initial version must be further agreed upon at the ARGA level and comply with the current EDPB format. The consultation period for the draft is 8 July – 30 October 2026.[2]

Offer	Motivation	Expected clarification
An example for legal narratives	The rarity of the plot is not visible in the tables	Semantic Linking Test
Consecutive publications	Individual units are revealed at intersections	Mandatory release history description as a best practice
AI models and retrieval	One object does not cover the processing chain	Multilayer inspection passport
Vulnerable and persecuted persons	The harm may be disproportionate	Risk-based examples
Evidence-based reproducibility	The conclusion is difficult to independently verify.	Minimum protocol form

Proposals should be supported by references to the draft and documented methodology, avoiding the claim that every request is already a GDPR requirement. Submission and confirmation of receipt are documented separately.

Appendix 31. Charter of Auditor Independence

Independent expertise should not guarantee a commercially desirable outcome. The draft charter establishes a register of interests, prohibits the editing of negative findings under pressure from the client, provides a transparent description of limitations, and requires separate approval of the public version, taking into account data protection.

Principle	Practical guarantee
Unconditionality of inference	The reward does not depend on the rating "anonymous"
Conflict management	Disclosure of connections and the mechanism of withdrawal
Methodological independence	Pre-registration of tests and changes
Technical reproducibility	Archive of versions, hashes and protocols
Responsible publication	Closed section for found personal data
Bug fixes	Public history of significant corrections

The Charter does not transform the Observatoire into a supervisory body or an accredited conformity assessment organization. The legal status of any external declaration must be clearly defined.

Appendix 32. Final Editorial Review Checklist

The editor performs a formal and substantive review of the report before publishing. Noting the presence of an appendix does not replace its quality control; empty templates that are not relevant to the project's scope are excluded from publication.

The requirement being verified	Sign of completion
Dates and versions	A cross-section of the law and the EDPB project is indicated
Status of documents	The law is separated from the consultation project
All factual statements	There is a confirmed source or a disclaimer
All mathematical examples	Denominators and assumptions are shown
Personal information	Not included in the public report
Internal links	The table of contents matches the pagination
External links	Leads to official documents
Header and footers	Matches series #11–21
Displaying characters	No squares, cuts or damage to tables

The review concludes with a separate entry confirming that the document is suitable for internal review. Official approval requires a specialized legal and scientific review, including verification of the applicability of the conclusions to the actual materials.

Appendix 33. Detailed case: a small group and the disclosure of a medical characteristic

Initial hypothetical situation. The research center publishes a report on the impact of the migration authority's decision on a group of 28 individuals. After removing names, the table contains the age range, month of application, region of residence, and access to medical care indicator. Six combinations of quasi-identifiers contain one entry each. One combination contains the rare sequence "resettlement - surgery - payment denial," which a local media outlet described in an article without mentioning the medical diagnosis.

The first test should answer the question of No Record Isolation: whether a single record is distinguished by its preserved characteristics. The second is No Linkage: whether a regular reader or interested party can find a matching story in a local publication. The third is No Inference: whether the linked table allows one to infer a medical condition previously unknown to the person who found the article. The combined score is significantly stronger than any single test. [2][3]

Possible technical measures include combining multiple months, excluding unusual medical indicators, and switching from individual tables to statistics that suppress the number of observations for very small groups. However, if the desired analytical goal is precisely to study a rare error, then suppressing all rare cases could systematically conceal the most important violations. The decision to publish should consider the tradeoff between privacy and usefulness, rather than automatically seeking to increase the size of the groups.

Draft independent audit result: The original publication is rejected in its current form; the reviewer is asked to retain the closed evidentiary version, publish aggregates without unique events, and support the conclusions with a high-quality legal analysis without connecting details. This

conclusion is conditional: without real data and verification of external sources, it does not constitute a conclusion about a specific individual.

Publication option	Privacy risk	Scientific usefulness
String table with dates	High probability of binding	High detail
Generalized quarter and region	The risk is lower, but it remains rare.	Average
Aggregated indicator without rare groups	Reduced for external audiences	Limited
Closed environment with an inspector	Managed access risk	High with guarantees

Appendix 34. Biometric features, photographs and voice

Facial, vocal, and motion data are not always amenable to traditional identifier removal methods. Facial blurring can preserve identifying clothing contours, tattoos, backgrounds, silhouettes, and surrounding text. Voice timbre changes may not obscure specific events or consistent speech patterns. Automatic translation and retelling of interviews can preserve the unique narrative that identifies a person in external sources.

When analyzing biometric data, it is necessary to distinguish between biometric data in the technical sense and special categories of data under the GDPR, which are relevant, in particular, for special processing for the purpose of uniquely identifying a natural person. The applicability of Article 9 depends on the legal classification of a specific transaction; it is not possible to automatically declare all images to be special categories or, conversely, to ignore the biometric risk in a standard photo file. [1]

Three sets of tests are recommended for auditing multimodal material. The first is the removal of direct annotations: faces, names on badges, case numbers, and voiceovers. The second is an analysis of contextual identifiers: location, dates, unique backgrounds, accompanying people, and file metadata. The third is a comparison test with publicly available publications and reverse searches under permitted conditions. The auditor's use of recognition tools should be accompanied by an assessment of the legality of their use and limitations of the results.

Conditional scenario: In a public video, an expert talks about an anonymous witness. The face is blurred, but the unique courtroom building is visible on the screen, and the hearing time is indicated in the caption. In this case, technical manipulation of the face does not eliminate the risk of identification. The story and visual environment should be reworked, rather than increasing the blurring and declaring the material anonymous.

Level	What is being tested?	Typical measures
Media file	Video, audio, subtitles	Editing, generalization of context
Metadata	EXIF, geotags, author names	Cleaning the container
Plot	A rare event, participants	Reformulation and elimination of details
External search	Comparability of frames/phrases	Decision on the inapplicability of the publication

Appendix 35. Multilingual Corpora and Identifier Extraction Errors

ARGA legal corpora can include Russian, English, French, and local languages, sometimes in a single document. The personal data search system, tested only in English, is capable of omitting transliterations, patronymics, variable surnames, organization names, case forms, and identifiers with unusual structures. A simple name pattern search is insufficient to prove high-quality anonymization.

The verification process begins with the creation of a language map of the set. For each language, a representative test sample is selected, with direct and indirect identifiers marked. The automatic detection recall rate should be measured separately for each language and identifier type. Overall recall calculated on a predominantly English corpus may conceal weaknesses in international search documents in a rare language.

Machine translation poses a distinct risk. While translation can normalize rare wording, it can also introduce new interpretations and distort legally significant facts. In publication reports, it's important to distinguish between human-informed editorial summaries and uncontrolled text generation. For legal examples, risk mitigation should not create false procedural circumstances; if anonymization requires significant changes to the facts, the example is labeled as composite and for educational purposes.

A possible test is to first compare the anonymized Russian-language text with open Russian sources, then with translations into other languages. If a unique passage is meaningful after translation, anonymity cannot be confirmed solely by the absence of a literal match. The technical protocol records the query languages, translator versions, and control restrictions; external AI services are used only under legitimate conditions for processing the source text.

Examination	The error that is revealed	Document
Stratified markup	Uneven completeness across languages	Quality table
Transliterations	Skipping name variations	List of rules
Semantic search	Recognizability of the plot	Protocol without publication of the original
Translation check	Substitution or loss of legal facts	Editorial Review Log

Appendix 36. Composition of Aggregated Publications and Differential Disclosure

A hypothetical example. The organization publishes the number of complaints to a specific administrative body for the entire quarter, and then for each month except one. Formally, each table contains only aggregates. However, the reader can subtract the sums and reconstruct the number for a hidden month. If the relevant month includes one externally known complaint, the additional disclosure could reveal a sensitive fact about the individual. This is a classic risk of overlapping releases, which cannot be assessed solely from individual files.

For each new publication, a map of statistical dimensions should be maintained: population, time window, geography, set of filters, classifications used, previously published overlapping totals, and access conditions. If the system allows the user to manually change filters, it is necessary to assess subtraction attacks and sequential queries. Limiting the table to a single table with small cells is ineffective if the sum and adjacent slices accessible to the user reveal a hidden value.

Measures vary in purpose and threat: suppression of correlated results, controlled responses to queries, privacy budgets, reduction in precision, and limiting the number of available measurements. A universal threshold of "not revealing values below five" is insufficient; differences in results may reveal values below the threshold. The choice of mechanism requires an assessment of scientific utility and the actual user model.[12][13]

The internal protocol must include at least one simulated reverse-calculation test on specially created aggregates. If the final interface does not allow row dumping but allows an unlimited number of arbitrary filters, it cannot be considered secure without analyzing the aggregated information output.

Published	Hidden	Possibility of recovery
Quarter = 120	March	If there are January and February
January = 53	March	Subtracting a partial amount
February = 66	March	120 − 53 − 66 = 1
March = 1 (output)	Personalized communication	In the presence of external knowledge of the complaint

Appendix 37. Anonymization of the financial relations graph

A graph set is created from organizations, individuals, addresses, and payments. Even if names are replaced with random identifiers, the relationship structure can remain unique: a central node with several subsidiaries, a single foreign recipient, and a rare transfer sequence can match a public corporate registry. However, using the graph for AML analytics often requires preserving the anomalous structure, otherwise the subject of the study disappears.

It is proposed to separate the proof graph and the publication graph. The former stores the original connections with restricted access and legal basis. The latter reveals only node categories, network statistics, aggregated features of suspicious patterns, and, where necessary, carefully modified training structures. A scientific publication should disclose that the training graph is synthetic or modified, rather than pretending to be an unmodified, real transaction set.

The No Record Isolation test covers the uniqueness of node degree combinations, link types, and transaction sequences. No Linkage is tested against available corporate records and public financial events. No Inference verifies what sensitive inferences can be made about an official or beneficiary when accessing a partial graph. The graph metrics used sometimes reveal the presence of a rare participant, especially if the publication contains a small sector.

For ARGA, the risk of false positives is equal in significance to the risk of a true leak: a matching structure does not always prove that a company participated in a prohibited transaction. In plaintext, it is necessary to avoid language that transforms a statistical suspicion into a statement of established wrongdoing.

Graph object	Risk of disclosure	Publication version
Unique central node	Identification of the company and the manager	Generalized class of roles
Rare payment edge	Linking to a transaction	Range and Aggregation
Time chain	Event recovery	Educational synthetic structure

Graph object	Risk of disclosure	Publication version
Open registry	Linking to the beneficiary	Context checking and exceptions

Appendix 38. A hypothetical example of the difference in perspectives between the controller and the recipient

Organization A conducts a survey of 2,000 participants and assigns a random code to each questionnaire. It stores the table of codes and contact information in a secure location for future clarification. Organization B receives only the cleared table of responses and the analysis rules. The agreement prohibits re-identification, but this prohibition alone does not prove that the information is anonymous to B. It is necessary to determine whether the responses contain rare identifying features and external sources accessible to B.

From Perspective A, data remains personal if it can be linked to a person; the obligations of the GDPR apply. From Perspective B, the assessment requires specific facts about the possibility of communication, the access device, and the information obtained. Following the *EDPS v SRB* case, the differences between the participants cannot be ignored or used to circumvent the initial controller's obligation to provide information when collecting data. [7]

If B receives additional request logs or can send queries to A to link responses, the initial assessment changes. If the questionnaire contains a specific, rare profession and a single region, even the absence of a correspondence table does not guarantee the impossibility of identification. Therefore, the protocol should include a list of acceptable queries between A and B and a check for meaningful uniqueness, rather than a single line about a contractual prohibition.

Conditional result:A contextual inference of anonymity for B is permitted only if there is a verified absence of a reasonably probable attribution in its actual environment. A's own obligations remain in place; the legality of the original collection is analyzed separately, and B's conclusion does not legitimize A's previous mistakes.

Question	Organization A	Organization B
Correspondence table	Yes	No
Possibility to attribute the code to the face	Direct via private key	Requires factual assessment
Applicability of the GDPR	In the presence of personal processing	Depends on the qualification of the data obtained
Information responsibilities during collection	At A	Not eliminated by status B

Appendix 39. Contextual risk of models with open weights and closed API

The same model may pose different risks to the user of a limited API and to the person receiving the weights, the ability to further train, and unlimited local computation. Through the API, the owner can impose request limits, filters, and retrieval monitoring. However, a software response filter does not prove that personal information is not included in the parameters; when weights are transferred, some security measures become inapplicable. [4][6]

The evaluation must distinguish between "model as a hosted service," "model weights," and "system with external knowledge." For a public API, retry limits, adaptive behavior, and log

protection are additionally important. Open weights require stronger verification of direct extraction and recognition of test budget limitations; after publication, it is technically impossible to prohibit local experiments.

If a developer considers weights to be anonymous, such a conclusion must be based on evidence of robustness to reasonably plausible means for the intended set of weight holders. It's not enough to test the chat interface under a strict filter and then declare equivalent anonymity for the distributed parameters file. Changing the access mode should trigger a new evaluation cycle.

It is proposed to disclose separate lines in the report for APIs, local weights, fine-tuning, retrieval, and the interaction log. If any mode is not investigated, it should be marked as outside the scope of the review. A conclusion of "everything is safe for all access types" is inconsistent with the methodology being reviewed.

Access mode	Type control	Significant limitation
Closed API	Limits, filters, log	Adaptive queries are possible
Open weights	Direct extraction test	Poor post-release control
Fine-tuning	Data and model version	Possible new memorization
RAG	Isolation of documents and roles	The leak is not from the scales, but from the index

Appendix 40. Removing data from the model and the limits of machine unlearning

A data subject's request to delete personal data poses a complex question for the developer: whether the original files, derivative copies, search indexes, training test cases, and information potentially stored by the model have been deleted. The applicability of the right to erasure depends on the specific terms of Article 17 of the GDPR and the exceptions provided by law. It is impossible to promise the automatic disappearance of all information from the scale simply by deleting a record from the original table. [1][18]

Technical strategies include excluding relevant material from future training, changing the index, rebuilding derived sets, using specialized unlearning methods, or retraining when justified. However, their effectiveness requires testing: can the model retain a general pattern, reflect data in logs, or be deployed to inaccessible copies? The evaluation must distinguish between the fact of removal from the managed repository and the likelihood of subsequent retrieval from a specific model.

The legal report must describe what operations were performed, what systems were covered, which copies cannot reasonably be recalled, and what verification was performed after the correction. The wording "record deleted" should not conceal the fact that the original material remained in an archive uncontrolled by the partner, or that the results of previous publications have already been disseminated to external users.

For ARGA, the priority of procedural fairness applies: the sole evidentiary original cannot be destroyed if applicable law requires its preservation, and at the same time, the need for an archive cannot be used as a basis for the indefinite preservation of all derivative copies. The legal regime and the risk of the published version must be addressed separately.

Component	Possible measure	Verification method
Original string	Removal/legal restriction	Operation log
Archive and backup	Managed lifespan	Destruction plan
Retrieval index	Segment Exclusion	Search test
Model parameters	Unlearning / retraining if necessary	Post-change extraction test
Public copy	Correction and notification	Accounting for limited recall

Appendix 41. Protocol for preliminary assessment of suitability of open publication

Publicly posting a proposed anonymized dataset requires a different vetting process than providing the same dataset to a single researcher under contract. In an open environment, the pool of potential recipients cannot be reliably limited, and the emergence of new open sources and matching systems significantly changes the risk of re-identification. Therefore, research decisions should not be made on the principle of "the data is cleaned, therefore it can be published," but rather after a separate determination of the purpose and terms of use. [1][2][15]

The first stage concerns the target audience: who will have access, whether files can be copied, indexed, automatically compared with registries, and whether findings can be reproduced in commercial systems. The second stage describes threats that are relevant to a static data set: unique time patterns, event sequences, rare relationships between participants, and indirect indicators of profession, location, or judicial status. The third stage considers dynamic threats that arise following new publications, leaks of a neighboring data set, or changes in search algorithms. [2][15]

A separate rule applies to sensitive cases: even a statistically low probability of identifying an individual may be unacceptable if disclosure could lead to persecution, loss of access to financial services, or a threat to personal safety. The harm from false data linkage is separately assessed. If a negative decision is made, an alternative must be offered: partial publication of aggregates, limited access, results upon request, or a fully synthetic example with proven independence from real cases. [1][4]

Release level	Access and control	Admission conditions	Limit of imprisonment
Open	Unlimited copying	Checking for potential external linking and inference	The conclusion is valid only for the described state of knowledge and threats
Controlled	Established recipients and agreement	Authorization check, anti-redisclosure, and logging	The agreement does not anonymize personal data.
Safe research environment	Supervised computing	Control exports, roles, query logs, and recalculations	The environment reduces risk, but does not guarantee anonymity.
Unpublished	Internal analysis only	Checking the legality of the original processing and minimization	The absence of a public release does not negate the GDPR

The solution is documented in a report with the date, the array version, the signatures of two independent specialists, and a list of the adopted limitations. The absence of a written solution means the test has not been completed, even if the developer verbally reports a successful test.

Appendix 42. Protocol for reconstructing the linking of several publications

Publications should be viewed as a single sequence, not as independent documents. Let's say the first table shows the number of asylum applications by region, the second the number of rejected applications by age, and the third the distribution by specific months. Each table alone can only contain groups of sufficient size. However, the intersection of three filters sometimes forms a group of one or two individuals about whom the external press already contains additional information. This mechanism is called compositional unfolding and is especially significant for small areas. [4] [12]

The auditor reconstructs a catalog of all related releases for the agreed-upon period, including archived versions and partner documents. Operations on permitted aggregates are then simulated: category intersections, subtraction of adjacent time periods, comparisons before and after record deletion, and queries based on combinations of attributes. It is clarified whether the results could be accessed by the same recipient, including through a search engine, editorial archive, or machine interface. Exact and approximate matches are separately noted; the inability to automatically link does not preclude manual recovery. [2][15]

When a disclosure is identified, it is necessary to establish a minimum set of published elements sufficient to replicate the finding and document it without disclosing the protected entity in the public portion of the act. To correct, simply changing the latest release is insufficient: subsequent revisions will be compared to previous copies, and caching does not guarantee the global destruction of previously accessible information. The registry retains the dates of correction, the notified recipients, and the residual risk. [1]

Step	Proof	Typical failure
Inventory	All versions, formats and external copies	Only the current PDF was checked.
Intersections	List of compatible filters and test queries	Ignored the subtraction of adjacent periods
External comparison	List of available publications and access level	The open archive was considered inaccessible
Reaction	Correction, restriction of access, retest	The source was removed, but the index was left.

Appendix 43. Technical test for data extraction from the language model

The goal of the test is not to prove the absence of any personal data within the model parameters, which is practically impossible to achieve with a simple test, but to evaluate specific extraction mechanisms in a given environment. Before testing, the scale version, system instructions, availability of external tools, context windows, output filters, and interfaces are recorded. A model with a closed API and the same model with published scales present different attack surfaces. [3] [12]

The test team generates secure control strings known in advance only to the auditor and developer within authorized access limits. No genuine leaks are permitted: real document numbers, client addresses, and harassment information are not placed in public test prompts. Synthetic markers, control secrets, and material legally accessible to the team are permitted; their connection to

training is documented, not guessed. An additional set of tests utilizes paraphrasing, incomplete quotes, variations in language, formatting, context length, and uninformative noise.[3][17]

Results are classified by accuracy: verbatim recovery, partial match, probabilistic paraphrase, answer previously available in a public source, and false match. If an external source could have provided the same string, the result cannot be considered a memorization of the training corpus without additional justification. The protocol must store the hypothesis, query sequence, generation parameters, responses, logs, and success criteria selected before the test. [3]

Measurement	Definition of the result	Limitation
Exact match	Match a control string according to a predefined rule	Doesn't show how the line got into the model
Partial match	The proportion of matching fragments by normalized metric	Easily inflated with clichéd phrases
Targeted extraction	The result of a query targeting a specific record	It is not possible to scale the output to all subjects
Untargeted extraction	Finding unexpected matches in a series of queries	Requires separate analysis of multiple comparisons

The conclusion necessarily states the query budget and test bounds. The lack of a successful attack after 300 trials does not prove zero risk. It is merely the result of a limited experiment, which must be interpreted taking into account statistical uncertainty and the adversary's model.

Appendix 44. Membership inference: Interpreting precision in a rare event

A training set membership attack attempts to determine whether a particular record was used to train a model. It should be assessed not by overall accuracy alone, but by the baseline prevalence of training set members in the target population, sensitivity, and false positive rate. If only one percent of the individuals being tested are actually in the training corpus, even a relatively good attack can produce more false positives than correct inferences.[3][13]

Let's consider a hypothetical experiment: among 10,000 records being tested, 100 are members of the corpus. The attack detects 70 of these 100 (70% sensitivity) and mislabels 5% of the remaining 9,900 (495 false positives). This yields a total of 565 positive signals, of which only 70 are correct. The positive predictive value is 70/565, or approximately 12.39%, despite a sensitivity of 70%. This example is illustrative, does not refer to a specific model, and does not replace independent testing.

This calculation demonstrates why a legal opinion should not confuse a "positive attack result" with a person's actual membership in the database. Research on real people requires independent evidence of true labeling and a process for appealing errors. The report includes confidence intervals, a threshold calibration procedure, the split between training and testing data, the retry rate, and cases where a result was obtained only after retroactively selecting successful queries. [1][3]

Indicator	The Importance of Academic Calculation	Interpretation
Base rate	100/10,000 = 1%	Rare accessory
True positives	70	Of 100 members
False positives	495	5% of 9,900 non-members
Precision / PPV	70/565 ≈ 12.39%	Most positive signals are false.

In a real-world case, the conclusion that the material was in the training set is only permissible after independent verification of the material's origin. Otherwise, the organization risks creating new false claims about the individual during the anonymization verification process.

Appendix 45. Risk register for vector bases and retrieval-augmented generation

In RAG systems, personal data can be stored in multiple objects at once: the original file, extracted text, fragment, vector representation, metadata, search index, query log, and model response. Removing an identifier from the original PDF does not automatically clear the previously created search index or cache. The check covers the entire creation, retrieval, and redistribution chain, not just the original storage. [1][17]

The auditor issues a test query to find a specific case by anonymous attributes: meeting number, jurisdiction combination, timeline, and asset amount. Testing is conducted in a permitted environment on synthetic or anonymized case models to avoid revealing real clients. An assessment is made of whether the system returns the original fragment, suggests a name, or provides sufficient consistent details for external linking. Technical similarity of embeddings alone does not establish identity, but when combined with metadata, it can increase risk.[2][15]

The report records user rights, indexer version, split function, retention policy, similarity threshold, and the presence of filters. After a test deletion of the document, indexing and queries are repeated against all related copies. If the index is deleted but the response cache continues to return a fragment, the corrective action is not complete. The public report provides the overall vulnerability category, not the text of the sensitive document found.

Layer	What is being checked?	Confirmation of elimination
Source	Copies of the document and metadata	Inventory of objects and removal
Chunk	Fragments and quotes	Re-search for unique markers
Vector	Index and backup replicas	Control of reconstruction/search after deletion
Logs	Prompts, answers, and routes	Retention settings and access restrictions
Interface	Issuance by indirect signs	Regression tests of roles and search queries

Appendix 46. Comparative Protocol for the EU, the UK and the US

There is no universal international standard that automatically deems any data set anonymized across all jurisdictions. In the EU, this determination depends on the definition of personal data and Recital 26 of the GDPR, as well as the specific context of processing. In the UK, the UK GDPR is in effect, with its own ICO enforcement mechanism, including a practical approach to "motivated infringers." In the US, federal and industry regulations use different concepts of de-identification, and individual regimes provide their own criteria and exceptions. [1][15][19]

Therefore, an international project must begin with a data flow matrix: where the data comes from, who performs the anonymization, who receives the results, where they will be stored, whether they will be reused for model training, and whether the receiving party can link them to their own registries. The assessment is not automatically transferred between jurisdictions, even if the

transformation technique remains the same. A legal advisor will separately consider the responsibilities of the original controller, the transfer of personal data prior to anonymization, and the possible status of the result at a specific recipient.

A comparative report does not replace national expertise. This is especially true for criminal procedural information, immigration materials, biometrics, and medical data, where special restrictions may apply. The formula "data de-identified according to the American standard" does not prove compliance with the requirements of the European GDPR. Likewise, a conclusion of low risk for a research partner does not guarantee the safety of global online publication. [1][15]

Law and order/level	Subject of inspection	What you can't assume
EU, GDPR	Personal status, means of identification, legality of the anonymization operation	The presence of a single regulatory numerical risk threshold
UK, UK GDPR/ICO	Practical identifiability and the motivated offender	Completely identical to the EU test in every detail
USA	Applicable federal, state, or sectoral regime	A unified legal status for all types of data
International project	Transfers, roles of the parties, further use	Extraterritorial applicability of one certificate

Appendix 47. Principle of minimum disclosure in litigation and regulatory disputes

Legal work often requires preserving individual facts: the date of arrest, the court, the route of travel, the details of the court decision, or the name of the representative. These elements may be necessary to prove a violation of rights, but they also pose a risk of identifying the individual in a public analytical report. The decision to publish is made separately from the decision to preserve the material for the lawyer, the court, or the authorized body. [1]

Within a case, two versions should be maintained: an evidentiary version, protected by access control and procedural safeguards, and a public version, containing only the minimum necessary facts. Each transformation is recorded: the replacement of direct identifiers, the aggregation of dates, the generalization of geography, the exclusion of rare circumstances, the creation of a synthetic case. To ensure verifiability of conclusions, an independent reviewer may be granted limited access to the original on legal grounds; the mere fact of such a review does not require publication of the original. [1][2]

The auditor separately checks the so-called uniqueness of the narrative: a combination of known circumstances can unambiguously lead to a specific individual, even if the last name is missing. Automatic name recognition tools cannot be used alone. Sources of indirect identification are separately described: public court records, press releases from agencies, company registries, international wanted notices, and old publications with personal photographs. [4][15]

Layer of work	Example of a permitted action	Red line
Protected evidence base	Storage of the full version subject to grounds and restrictions	Unauthorized publication of the original
Analytical verification	Plot Uniqueness Test	Experimental disclosure of a living

Layer of work	Example of a permitted action	Red line
		face by an external tester
Public report	Aggregation or conditional case	Assertion of anonymity without assessment of indirect evidence
Judicial discovery	Procedurally necessary transfer	Automatic equivalence with open publication

Appendix 48. Test of Consistency of Independent Assessors

Two experts may evaluate the probability of identification differently: a technical expert focuses on the quality of the classifier, a lawyer on the accessibility of information for a specific party, and a security specialist on the capabilities of a motivated attacker. To prevent the dispute from devolving into a vote on authority, the methodology must pre-define the criteria, the uncertainty scale, and the procedure for independently examining the same scenario. [2][15]

Each expert receives the same data description but conducts an initial assessment independently, without knowledge of their colleague's conclusion. Discrepancies are then classified according to the following reasons: different factual assumptions, different assessments of technical feasibility, different understanding of the legal standard, or incomplete information. Agreement is permitted only with a written argumentation protocol. In the event of a serious discrepancy, the release decision is postponed, and the group requests additional data or engages a third reviewer. [2]

The quality of the work is measured not by the proportion of identical conclusions at any cost, but by the accuracy of procedural discrimination. Agreement between two experts using the same incorrect set of facts does not guarantee the correctness of the conclusion. Therefore, control includes reference scenarios with a pre-determined hidden connection, assessments of false positives and negatives, and repeating the test after a predetermined time. [3]

Type of discrepancy	Example	Required response
Actual	One expert did not take into account the public court register	Update the source data and repeat the analysis
Technical	Varying success rates of linkage attacks	Lock down code, environment, and reproducibility
Legal	Other relevant recipient	Please describe the applicable context separately.
Unresolved	Lack of access to a key source	Don't make an absolute conclusion about anonymity

Appendix 49. Post-Publication Re-Identification Response Plan

Anonymity is not an irreversible characteristic of a data set, something that can be confirmed once and then forgotten forever. External sources, machine matching methods, and the publishing environment itself change over time. If new information emerges that allows one to identify the person to whom a previously published profile pertains, this event must be considered a signal for an urgent review of the data's status and a possible security incident. The consequences under the GDPR depend on the specific circumstances; not every signal automatically triggers the notification procedure provided for in Articles 33–34, but it cannot be ignored. [1][2]

The first step is to stop further dissemination, if technically feasible, preserve evidence of the event in a secure environment, and describe the actual information that was made available to third parties. In the second step, the data protection officer assesses whether the disclosed material constitutes personal data in the given context; whether there was a security breach; whether there are risks to the rights and freedoms of individuals; and what notification and documentation obligations arise. The third step involves correcting the page, restricting API access, working with indexers, and notifying known recipients where required by applicable regulations. [1]

For particularly sensitive cases—information about persecution, asylum, medical conditions, and transactions—the risk of secondary victimization should be separately considered. A public announcement that "person X has been re-identified" could increase harm. An internal event map contains the actual connection to the individual only within the scope of authorized access; an external report describes the type of error without redisclosing details.

Period	Responsible	Action	Confirmation
Primary reaction	Publication Ownership and Security	Suspension of further issuance, preservation of traces	Restrictions log and status copy
Legal qualification	DPO / lawyer	Determining the status of data and applicable regulations	A reasoned act with a legal basis
Correction	Publishing Team	Change, revocation, notification of addressees according to the conditions	Release and Confirmation Registry
Follow-up control	Independent reviewer	Regression test and residual risk assessment	Conclusion with restrictions

Appendix 50. Contractual matrix for independent anonymity testing

External auditing creates new risks, as the auditor may need access to data intended for anonymization. The agreement must describe the parties' roles for each transaction. It is impossible to declare the auditor solely as a "processor" or "independent controller" without analyzing the purposes and means of this determination. The agreement must be consistent with the actual relationship and applicable GDPR requirements. [1]

The terms of reference include the version of the test suite, the purpose of the test, the scope of permissible attacks, methods for accessing auxiliary data, and restrictions on further use of the information. Launching unauthorized attacks against living individuals and attempting to obtain their real data solely for the purpose of demonstrating risk is prohibited. A documented legal basis, minimization of initial access, and the ability to terminate the test when the test could create additional disclosures are required.

Publication of results is separately regulated. The auditor's commercial dependence, the client's ability to edit inconvenient conclusions, or remuneration only for positive findings undermine the independence of the audit. The contract must provide for documentation of negative results, a procedure for correcting factual errors, a dispute resolution procedure, and the ability to honestly communicate the boundaries of the test. [2][15]

Element of the contract	Minimum requirement	Checking execution

Object	Array, Model, and Environment Versions	Identifiers and hashes
Right of access	Goals, roles, processing location, prohibition of external transfer	Role Matrix and Session Log
Acceptable tests	Methods, stops, prohibition of real damage	Approved test plan
Destruction	Deadlines and confirmation for deleting working copies	Destruction and backup copies certificate
Independence	The customer has no right to conceal a negative result	Charter and Protocol of Conflicts

Appendix 51. Managing Differential Privacy in a Series of Publications

Differential privacy describes the mathematical robustness of a mechanism to the participation of an individual subject in adjacent datasets. It does not create unconditional anonymity for any source file, does not remove the requirement for the original lawfulness of processing, and does not constitute a standalone permission to publish sensitive information. The reliability of the promise depends on the formal definition of the neighborhood, the choice of mechanism, the correct randomness generator, privacy settings, and composition conditions. [1][13]

For the practical log, a unit of account is introduced: the identifier of the protected unit (person, episode, or record), the release period, the declared epsilon and delta, and the mechanism and method of composition. Guarantees cannot be arbitrarily summed or averaged, especially if several departments publish statistics on the same people. Ancillary unprotected releases, metadata, and logs can circumvent the claimed benefit of DP and should be considered separately. [13]

A case study: mechanism A has a budget of epsilon = 0.4, mechanism B has an epsilon = 0.6. In a simple sequential composition with a compatible definition of the protected unit, the upper bound on epsilon might be 1.0; the actual composition rules and the overall delta require separate calculation. This is not an estimate of the 1% re-identification probability, nor is it a legally established threshold for acceptable risk. Before approval, each mechanism is tested for correctness of implementation, including noise generation, seed protection, availability of retrials, and leakage through exceptions.[13]

Question	**Minimum entry**	**Typical mistake**
What are we protecting?	Person/Recording/Episode and Neighborhood Model	The report uses different units
What is the guarantee?	epsilon, delta, mechanism, date	Calls epsilon the "leakage percentage"
How many releases?	Unified Request Log and Budget	Each team only takes into account its own publication.
Is the usefulness retained?	Error, bias, confidence limits	Noise leads to a wrong legal conclusion

Appendix 52. Evaluation of small samples and secondary suppression

In reports on sanctions, searches, and immigration decisions, groups are typically small. Even if a researcher decides to suppress a cell containing one person, its value can be reconstructed from published results. For example, if it is known that a unit handled 12 cases in total, and 5 and 6 are distributed among the open categories, the hidden third category is calculated as 1. Therefore, primary suppression should be accompanied by an analysis of secondary clearances.[4][12]

The methodology involves constructing a system of accessible equations based on the results of tables, checking for uniquely recoverable hidden values, and searching for information revealed by differences between successive reports. For statistical series, a minimum publication frequency and acceptable precision limits are specified. If a researcher publishes a grand total and several sensitive cross-sections, it is necessary to check whether the final figure reveals all hidden rows. [12]

Suppression should not be applied mechanically. Sometimes, generalizing about geography or time is more useful than suppressing a specific column; sometimes, the only safe measure is to refrain from publishing a narrow cross-section. If a measure is used to analyze potential discrimination, the consequences of losing statistical power should also be assessed. The methodology should clearly explain which conclusions remain possible and which are no longer valid after suppression. [1]

Published	Meaning	What can be restored?
Total number	12	The outcome is known
Category A	5	Part is known
Category B	6	Part is known
Category C is hidden	small group	12 − 5 − 6 = 1; hiding is ineffective

Appendix 53. Management of source evidence and scientific reproducibility

Legal verification of anonymity requires reproducibility, but fully disclosing the source data could undermine the very purpose of data protection. Therefore, the evidence system must be multi-layered. The public section contains the method, aggregated metrics, risk justification, and a list of limitations. The private section stores control samples, attack configurations, test results, and descriptions of the supporting material used. Access to the private section is granted only on legal grounds and for specific verification purposes. [1][2]

Internal evidence should be preserved with data and script versions, fixed parameters, test times, responsible party names, and cryptographic checksums. A checksum helps ensure that certain bytes have not been altered, but it does not prove the veracity of the content or justify its acquisition. Integrity, provenance, legitimacy, and authenticity are different issues requiring different proofs.

Negative evidentiary assessments are particularly important. If a test fails to find successful linking attempts, it is necessary to document the scenario coverage, computational budget, initial assumptions, time, and limitations of the research environment. A report of "no attacks detected" should not be translated into a statement of "re-identification impossible." A separate reviewer evaluates not only the test results but also the risk of missing the test due to the chosen methods. [3]

Material	Open part	Protected part
Methodology	Attack types and assumptions	Detailed requests if they increase risk
Results	Aggregates and uncertainty intervals	Line-by-line logs and control marks

Material	Open part	Protected part
Initial data	Description of category and volume	Minimum required sampling for legal access
Correction	Public history of editorial offices	Journal of Personnel and Technical Operations

Appendix 54. Escalation and Re-Check Threshold System

The Independent Observatoire should not establish a single, numerical "acceptable risk" for all data. The decision depends on the severity of the potential harm, the attacker's capabilities, the subsequent distribution channels, and the feasibility of control measures. A high-risk report may be permitted for a strictly limited circle of authorized individuals, but is not permitted for public release. Conversely, a very general aggregate may be published without retaining redundant identifiers in the working set. [1][2]

To manage decisions, escalation levels are established: a standard review, an enhanced review by an independent specialist, a decision by a committee with legal representation, and a temporary release ban. The threshold triggered depends on the unique record identified, probable identification, the ability to recover a particularly sensitive attribute, the lack of a legal basis for the original processing, or an uncontrolled transfer to a third party. Any deviation from the recommended action requires a written justification from the responsible official.

A re-check is not performed on the formal anniversary of the conclusion, but rather upon the occurrence of events: a new open database, a change in publication platform, the addition of results from another organization, further training of the AI model, a change in index, or the introduction of a mass comparison tool. A regular calendar is still necessary to detect slow changes. [2][15]

Trigger	Response level	Expected decision
New external source with overlapping data	Enhanced verification	Linkage test and status review
Confirmed identification of a living person	Critical	Restriction of access and legal qualification of the event
Updating the model/index	Technical	Regression attacks and a new version of the act
Change of purpose or recipient	Legal	Re-evaluation of the grounds and context

Appendix 55. Appeals of individuals and independent review of controversial findings

Anyone who believes that a published data set allows them to be identified or to infer sensitive information about them must have a clear channel of recourse. Requiring a complainant to publicly disclose their identity in support of a complaint is dangerous and often excessive. It is sufficient to provide a secure initial assessment process, in which detailed information is accepted only to the extent necessary and is accessible to a limited number of specialists. [1]

The process has two independent branches. The first establishes whether there is a genuine connection to the individual and what means of identification are actually available; the second verifies the correctness of the initial decision to anonymize and whether correction is necessary. If the published content constitutes personal data for the relevant controller, the applicable rights of the subject and the timeframe for their exercise are then determined. The specific set of rights and exceptions is assessed according to the GDPR, not the internal regulations of the research center. [1]

The organization must avoid conflicts of interest: the specialist who prepared the disputed publication cannot review the complaint unilaterally. The decision includes the facts, the scope of the investigation, the measures taken, the procedure for contacting the competent supervisory authority, and the procedure for correction. For statistical publications where an individual link cannot be established, independent vulnerability analysis is permitted without unnecessary identification of the complainant.

Stage	Requested information	A risk to avoid
Initial contact	Link and description of concern	Collection of additional passport data without necessity
Technical inspection	Minimum features and test scenario	New publication of a sensitive coincidence
Legal assessment	Role of the parties and legal grounds	Automatic denial of the applicability of the GDPR
Correction	Technical measures and communication	The promise of complete removal of already distributed copies

Appendix 56. Methodology for assessing the quality and usefulness of an anonymous result

Strong protection does not necessarily mean a useful result. If all rare events are excluded during the preparation of a report, the study will systematically underestimate violations in small groups. For example, excluding isolated cases of unlawful detention can distort the assessment of the vulnerabilities of precisely those individuals whose rights require special protection. The auditor's task is not to prove that publication poses zero risk, but to transparently compare the level of protection with the research objective and select a less risky release method. [1][12]

Quality assurance begins with a list of the initial research questions: period comparison, anomaly detection, geographic distribution, and algorithmic error assessment. For each question, the required precision, acceptable information loss, and consequences of erroneous inference are specified. The impact of the transformation is then measured: changes in distributions, missing categories, mean shift, changes in proportions, and interval widths. If these indicators deteriorate significantly, either another anonymization method or limited research access to a more detailed version is required.[12] [13]

When using synthetic data, not only average values are compared, but also rare tail events, dependencies between features, and possible traces of training records. High statistical similarity does not prove the absence of leaks. A separate check is made to ensure that the synthetic dataset does not become a means of reproducing individual histories. A publicly available solution should

describe the residual scientific limitations to prevent readers from interpreting protected aggregates as a complete picture of reality. [3]

Purpose of the study	Utility metric	A typical compromise
Event frequency	Difference in shares and error limits	Suppressing small cells changes the score
Dynamics over time	Trend stability when aggregating	Wide intervals hide short peaks
Regional differences	Distortion of regional ranks	Generalizing geography reduces detail
Errors of AI systems	Preserving the false positive/negative distribution	Rare bugs may disappear from the release

Appendix 57. Complex scenario: multilingual judicial corpus

This research project examines extradition decisions in five jurisdictions. The original documents are presented in Russian, French, English, and two national languages; along with official translations, the researcher uses automatically generated summaries. Direct names have been removed from the corpus, but airport names, charge statements, bail amounts, and rare sequences of court dates remain. External archives contain some of the same stories, transliterated in various ways.

During a standard check, searching for the exact name returns nothing, but the semantic system combines two versions of a single solution and then displays the associated arrest message. The problem isn't a flaw in the regular expression used to search for the last name, but in the corpus structure: a rare combination of actual names remains recognizable. The sample is created solely for training purposes; it doesn't reflect any specific client's materials. For replication, the auditor creates a dictionary of transliteration variants, time windows, a set of commonly known events, and separate translation equivalence tests.

The release decision entails two products. For internal legal analysis, a protected original is retained, with access granted by role; for the public report, aggregated findings and synthetic trajectories are used, which do not reproduce unique details. The loss of some detail is openly acknowledged. If a researcher cannot justify why a rare combination is no longer identifying, they do not claim that the corpus is fully anonymized. [1][2][15]

Stage	Examination	Result criterion
Language normalization	Transliterations, names and time formats	No unaccounted coincidence due to language
Binding	Court archives and accessible registers	The uniqueness of events was measured
Model	Search for similar fragments	Semantic match described
Release	Comparison of benefits and harms	The limitation or aggregation is confirmed by the act

Appendix 58. Integrated Scenario: A Series of Monthly Statistical Bulletins

The organization publishes monthly bank account denials across four geographic zones. Each table aggregates events for at least five individuals, so the editors consider the publication safe. However, the March bulletin contains the cumulative total since the beginning of the year, while the April

bulletin presents the same figure for a new date. Subtracting these figures yields a monthly change for a group of one individual. If a single local report of a bank account being blocked is also available, the analytical conclusion becomes specific to a specific individual.

An audit requires collecting not just a single bulletin, but a sequence of all published files, including original versions and editorial corrections. A model of the available linear combinations of aggregates is then constructed, and it is verified whether hidden cells can be uniquely reconstructed from the total sums. This is not just a statistician's task: a legal expert must explain when a reconstructed value can refer to an identifiable individual. [1][12]

Response measures include aligning publication periods, presenting small groups in ranges, secondary suppression, and storing detailed calculations separately. The impact of the methodology change on time series should be documented: the published average before the defense may not be comparable to the figure afterward. The auditor should not conceal this effect, otherwise readers may mistake the discontinuity in the graph for a change in the banks' actual practices.

Source of risk	Why Simply Removing Names Isn't Enough	Control
Difference in accumulated totals	Restores the minor period	Checking all time sections
Local news	Associates an event with a person	Retrospective search of external publications
Archive corrections	Store different versions of amounts	Accounting for published versions
Change of method	Creates an artificial trend	Documenting statistical comparability

Appendix 59. Complex scenario: synthetic database of banking transactions

A developer creates a training set of transactions to test a suspicious transaction detection system. They report that the data is synthetic, as the names and identifiers are generated. However, the generator was trained on a narrow set of real, large transactions, and the original amounts, dates, and counterparty relationship graph are highly unique. When another researcher compares the synthetic graph with public corporate ledgers, a plausible link to the real beneficiary emerges.

Verification should consider not only the similarity of averages and frequencies, but also the structural uniqueness of the graph, the nearest neighboring records, the repetition of rare sequences, and the possibility of attribute inference. A legally formed closed test sample and independent metrics are used for verification. If a definitive link to a real person is found, the information remains in a protected document and does not become an illustration in a public article. [3][12]

The decision depends on the intended purpose of the dataset. For internal testing of the algorithm under controlled access, more detailed structures may be retained, provided there is an appropriate legal basis. For unrestricted distribution, the developer must justify the lack of identifiability to relevant parties; if this is impossible, they must modify the generation, reduce the unique features, or replace the material entirely with a conventional one. The terms "synthetic" and "de-identified" do not constitute proof of legal status.

Component	Sign of a leak	Measure
Amounts and dates	Repeat of a rare deal	Proximity control and generalization
Connection graph	A unique chain of legal entities	Structural attacks and topology

Component	Sign of a leak	Measure
		changes
Text of notes	Verbatim repetition of the original	Memorization test and generation cleanup
Final publication	Unlimited aggregation by external users	Separate evaluation of the open release

Appendix 60. Complex Script: Photo, Audio, and Metadata

The legal report is accompanied by a photograph of the conference and a brief video interview, which intentionally omits the names of the participants. However, it is possible to extract a face, badge, computer screen, and a distinctive interior from the image; metadata may include coordinates and the time of recording. Voice, accent, and job description can serve as additional clues. Cropping the image or removing the audio track can reduce the risk, but does not always eliminate it completely. [1][4]

To audit media content, a separate list of possible identifiers is compiled: faces and outlines, signs, badges, vehicles, on-screen captions, file metadata, subtitles, and automated transcripts. Links to public photos of the event, performance schedules, and broadcast archives are assessed. The most dangerous combinations of attributes are: an anonymized face can be reconstructed by a rare position and performance time, and a deleted caption can be read from subtitles. [15]

Processing images and voices does not automatically translate into processing of special-category biometric data: legal classification depends on the characteristics and purpose of the processing, including the issue of unique identification. This does not eliminate the need to assess the personal status of the material and the risks to individuals. If the potential for harm is high, it is preferable to discard the original media and replace it with a diagram or illustration not based on the protected individual.

Object	Test	Possible solution
Photo	Matching image search, badge OCR	Remove the original or significantly change the content
Video	Scene, voice, subtitles, participants	Replace with neutral graphics
Metadata	GPS/EXIF, date, author	Clean and recheck the export
Archival copies	Primary media search	Assess the impossibility of a complete recall

Appendix 61. Complex scenario: scientific access to rare medical and migration events

The researcher needs to assess whether health conditions are associated with specific procedural delays in asylum cases. The regional sample includes several rare diseases and unique combinations of marital status, hearing date, and resettlement route. Public disclosure of the data would create a significant risk of identification; completely excluding rare cases, on the contrary, would weaken the scientific validity of the findings. [1]

The project provides for separate levels of access. The method, secure summary indicators, and uncertainty interpretation are publicly released. A more detailed table remains in a secure research

environment; researchers are granted limited access after reviewing the basis, purpose, and permitted procedures. Articles 9, 10, and 89 of the GDPR are analyzed separately, depending on the actual data category and the legal regime of the study; the general label "for science" does not replace the required analysis. [1]

When publishing, it's important to avoid a false sense of accuracy. If categories are combined due to a small number of papers, the reader should understand that a specific risk may have remained hidden. An external reviewer examines why a particular level of generalization was chosen, why the same goal cannot be achieved with less detailed data, and what measures will prevent further reuse of the information. A negative conclusion regarding the possibility of open publication does not preclude legitimate, protected research.

Target	Wednesday	Proof of control
Public review	Units	Small Group and External Data Linkage Test
Checking statistics	Limited access	Role and Operations Log
Review	Protected materials	Independence and legal basis for access
Archiving	Controlled storage	Removal timeline and plan

Appendix 62. Complex Scenario: A Contractor Who Promises “100% Anonymity”

An international organization receives an offer from an external provider that guarantees complete anonymity of any data transferred to it after automated processing. The agreement does not describe the methods, conditions for access to source files, or a list of saved technical copies. The provider offers a certificate without providing testing and promises that any data processed is not subject to GDPR. This approach cannot be considered sufficient either technically or legally. [1][2]

Before transferring data, it is necessary to verify the basis on which the provider obtains personal information, its role, and whether the contract allows for training its own models on the client's corpus. Documented transformations, residual unique combinations, verification of reasonably probable external links, processing locations, and deletion rules are then requested. If the provider refers to a specific standard, the auditor must establish its scope and boundaries: an information management standard is not the same as proof of the anonymity of a specific release.

A refusal to provide substantial evidence should not automatically conclude that the data is identifiable in any environment, but neither should a positive conclusion about anonymity be made. The correct conclusion is insufficient evidence and a decision not to apply the proposed procedure to sensitive data until necessary testing has been conducted. This is an independent professional conclusion that must be recorded in the report and contractual correspondence.

Supplier's statement	Security question	Insufficient answer
"100% anonymity"	For which environment, which party and what means of identification?	"Our algorithm removes names"
"Guaranteed by the standard"	What exactly is certified?	General ISO reference without subject
"We don't have any copies."	Where are the logs, indexes and backups?	Oral assurance without an act

Supplier's statement	Security question	Insufficient answer
"GDPR is no longer in effect"	How are the source data processed?	Lack of analysis of the operation itself

Appendix 63. Complex Scenario: Reusing the Result for AI Training

Publishing a statistical dataset for a scientific report and subsequently using it to train a model are different activities with potentially different risks. Even if the initial reviewer assessed the data as anonymous in a given environment, a subsequent developer may combine it with other sources, add new labels, and train a system capable of inferring hidden attributes. Therefore, the release document should describe the intended reuse and limitations of the initial review.[1][3]

The verification process is conducted in two stages. The first establishes whether the personal data existed at the time of transfer and what obligations arose. The second examines the actual capabilities of the new recipient, including their external datasets, access to weights, retraining method, and probable ability to extract the data. A contractual prohibition of linking can reduce the risk of misuse, but does not provide mathematical proof of anonymity. A conclusion regarding the status of the data should be formulated separately for each context examined.

A research report should not prohibit any reuse in advance. Instead, it should require a reassessment if the purpose, audience, and means of comparison change. Where necessary, controlled research sites and public aggregates with greater resilience to reuse should be specified. This allows for scientific work without the erroneous assumption that the original positive act applies to all future processing methods.

Change	New risk	Required verification
External casing added	Linkage	Funds and details of the new recipient
The model receives open weights	Extraction	Different budget and class of attacks
The target has been changed	Violation of source processing restrictions	Roles, bases and compatibility
Public API	Bulk requests	Control of withdrawals and limits

Appendix 64. Final map of the implementation of the ARGA methodology for 12 months

The proposed program does not represent a completed accreditation, audit, or recognition by EU bodies. It is a project for the gradual development of research competence. The first stage involves the approval of an independence charter, a system of roles, a legal model for working with source materials, and a protocol for protecting test data. The second stage involves the development of training tests and basic sets of reproducible scenarios. The third stage is devoted to independent peer review, a public methodological report, and an error correction procedure. [1][2]

It is not recommended to immediately publish individual real-world cases simply to demonstrate the Observatory's capabilities. It is better to conduct initial pilots on legally and technically secure conditional data, and then on legally accessible materials in a controlled environment. Public release of results is permitted after verifying the initial legal basis, the independence of the reviewer, and the possibility of correction. The ability to abandon a risky release, not the number of documents with positive conclusions, is an indicator of maturity.

At the end of the cycle, an external methodological audit should be conducted: randomly repeat the tests, compare the decisions of different reviewers, check active references to regulatory sources, evaluate the subjects' submissions, and assess the effectiveness of the response to errors. If significant gaps are not addressed, the expansion program is suspended. This mechanism protects research reputation better than a quantitative publication plan.

Stage	Conditional period	The result being tested
Management and Law	1–3 months	Roles, charter, access, legal basis
Methodology and tests	4–6 months	Control kits and test repositories
Limited pilot	7–9 months	Independent acts and corrections
External review	10–12 months	Reproducibility and scale decision report

Sources and methodological basis

[1] Regulation (EU) 2016/679 (GDPR), in particular Articles 4–6, 9–11, 25, 32–35, 89 and recital 26

https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Status and purpose: Mandatory EU law; applicability to a specific transaction is assessed separately.

[2] EDPB Guidelines 02/2026 on Anonymisation, version 1.0, 7 July 2026

https://www.edpb.europa.eu/public-consultations/guidelines-022026-on-anonymisation_en

Status and purpose: Public consultation draft until 30 October 2026; not final version.

[3] Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP216)

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Status and purpose: Historical methodological reference point, updated by EDPB draft 02/2026.

[4] EDPB Opinion 28/2024 on certain data protection aspects related to AI models

https://www.edpb.europa.eu/documents/opinion-of-the-board-art-64/opinion-282024-on-certain-data-protection-aspects-related-to_en

Status and purpose: EDPB opinion of 18 December 2024 on model anonymity and other GDPR-related issues.

[5] CNIL. L'anonymisation de données personnelles

https://www.cnil.fr/fr/technologies/lanonymisation-de-donnees-personnelles

Status and purpose: Practical criteria for the effectiveness of anonymization.

[6] CNIL. IA: Analyzer le statut d'un modèle d'IA au regard du RGPD (22 juillet 2025)

https://cnil.fr/fr/ia-analyser-le-statut-dun-modele-dia-au-regard-du-rgpd

Status and Purpose: Method for assessing a model's status under the GDPR.

[7] Court of Justice of the EU. EDPS v SRB, C-413/23 P, judgment of 4 September 2025

https://juris.curia.europa.eu/juris/document/document_print.jsf?docid=303863&doclang=EN

Status and purpose: Decision on the prospects of identifiability and the initial duty to inform.

[8] Court of Justice of the EU. Breyer, C-582/14, judgment of 19 October 2016

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A62014CJ0582

Status and purpose: The meaning of reasonably available means of obtaining additional identifiers.

[9] Court of Justice of the EU. IAB Europe, C-604/22, judgment of 7 March 2024

https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A62022CJ0604

Status and purpose: Machine-readable strings and the concept of a shared controller.

[10] EDPB Guidelines 01/2025 on Pseudonymisation, version 1 (public consultation)

https://www.edpb.europa.eu/public-consultations/guidelines-012025-on-pseudonymisation_en

Status and purpose: Published for consultation; not designated as final.

[11] EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, final version

https://www.edpb.europa.eu/documents/guideline/guidelines-42019-on-article-25-data-protection-by-design-and-by-default_en

Status and purpose: 2020 Definitive Guidance on Data Protection by Design.

[12] NIST SP 800-188. De-Identifying Government Datasets: Techniques and Governance

https://csrc.nist.gov/pubs/sp/800/188/final

Status and purpose: Technical guidance, September 2023; does not establish EU law.

[13] NIST SP 800-226. Guidelines for Evaluating Differential Privacy Guarantees

https://csrc.nist.gov/pubs/sp/800/226/final

Status and purpose: Final version March 2025; DP assessment methodology.

[14] ENISA. Data Pseudonymisation: Advanced Techniques and Use Cases

https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases

Status and purpose: 2021 Report on technical measures and limits on pseudonymization.

[15] UK ICO. How do we ensure anonymization is effective?

https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/

Status and purpose: ICO Guidance, with notice of revision following the Data (Use and Access) Act.

[16] UK ICO. Pseudonymisation

https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/

Status and purpose: Technical and legal limitations of pseudonymization under the UK GDPR.

[17] CNIL. Développement des systèmes d'IA: recommandations pour respecter le RGPD

https://www.cnil.fr/fr/developpement-des-systemes-dia-les-recommandations-de-la-cnil-pour-respecter-le-rgpd

Status and purpose: Official guidelines for AI development.

[18] CNIL. IA: Informer les personnes concernées

https://cnil.fr/fr/ia-informer-les-personnes-concernees

Status and purpose: On the transparency of processing, including the risks of memory and regurgitation.

[19] NIST Privacy Framework, Version 1.0 (January 2020)

https://www.nist.gov/privacy-framework/privacy-framework

Status and purpose: Voluntary risk-based instrument, not a mandatory right.

[20] CNIL. Mobiliser la base légale de l'intérêt légitime pour développer un système d'IA

https://www.cnil.fr/fr/base-legale-interet-legitime-developpement-systeme

Status and purpose: Measures to reduce memorization and extraction in the context of model development.