Contribution to the 01/2025 guidelines on pseudonymisation

Date	14/03/2025
Author	Dr. Ricard Martínez Martínez
Job 1	Lecturer at Constitutional Law
Job 2	Researcher on different European projects related to research in the field of health (BigMedylitics, BodyPass, CHAIMELEON, EUCAIM)
Entity	Universitat de València
Contributor	Ruben Ortiz Uroz
Job	Data protection officer
Entity	University of Barcelona

Signed: RICARD JOSEP|MARTINEZ|MARTINEZ Firmado digitalmente por RICARD JOSEP|MARTINEZ|MARTINEZ Fecha: 2025.03.14 13:18:16 +01'00'

Index

1. Purpose of the allegations

This document has been created taking into account the experience gained in the Cancer Image Europe Project (EUCAIM). The Project is addressed to build a pan-European digital federated infrastructure of cancer-related images, which will be used for the development of AI tools toward Precision Medicine. We hope that this infrastructure will provide the means to develop AI tools that will be able to enhance the (cancer) diagnosis procedure, treatment and the identification of the need for predictive medicine benefiting patients across Europe. The project cooperates with other actors (EOSC, TEDHAS, UNCAN.eu, Quantum Project) in the development of the European Health Data Space.

In our experience, defining the concept of anonymization is strategic for the establishment of the EHDS. The generation of anonymized data sets requires a functional approach that, without compromising the rights of data subjects and the security of processing environments, allows the use of high-quality data sets.

In this project, Dr. Ricard Martínez assumes the responsibilities relating to regulatory and ethical compliance, together with the privacy and security research team in IRTIC at the University of Valencia (University Institute for Research in Robotics and Information and Communication Technologies) and Dr. Janos Meszaros (Centre for IT & IP Law at the KU Leuven University). Ruben Ortiz Uroz collaborates as Data protection officer of the University of Barcelona, research partner entity within EUCAIM.

The aim of these allegations is to consider the impact on the position adopted by the European Data Protection Board in the following areas:

- Concept of anonymisation in the General Data Protection Regulation.
- Interrelation with Opinion 05/2014 on Anonymisation Techniques
- Foreseeable effects in the future Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space and on research with electronic health data.

The allegations will be made on the following paragraphs of Guidelines 1/2025:

> 22. Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person and is therefore personal. **This statement also holds true if pseudonymised data and additional information are not in the hands of the same person.** If pseudonymised data and additional information could be combined having regard to the means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal. **Even if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data becomes anonymous only if the conditions for anonymity are met**.
>
> 42. If a controller or processor wants to use pseudonymisation to reduce confidentiality risks from some or all unauthorised third parties, they will include those third parties in the pseudonymisation domain and assess the means they are reasonably likely to use for attribution. Relevant third parties include not only cyber-crime actors, but also employees or maintenance service providers acting in their own interests rather than on instructions from the controller. Taking due account of contextual elements and the circumstances at hand, it is recommended to consider both actions in good faith, and those executed with criminal intent.

43. For instance, pseudonymisation may be performed prior to transmission of the data to a processor or third party that ensures only a level of security that would not be appropriate for the processing of the original data but is appropriate for the risk connected with the processing of data that cannot be attributed to data subjects. In this case, all means available to unauthorised parties that might access the pseudonymised data while the (authorised) recipient of that data processes them need to be considered.

51. For external processing, i.e. processing under instruction by a processor or transmission to an independent controller, more extensive measures and risk assessment may be necessary to prevent attribution to data subjects. In particular, all intended recipients of the pseudonymised data need to demonstrably assure that the pseudonymised data are not disclosed to unauthorised recipients beyond the defined domain. For processors, additional tools under Art. 28 GDPR such as audits are available to support this assurance.

Example 5: Secondary use for research

2. Preliminary remarks: the concept of anonymisation

Neither Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, nor any of the subsequently approved regulations (Data Act, Data Governance Act) or in the process of being published (Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space) define the concept of anonymisation. The only reference is in recital (26) of the GDPR:

> (26) The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. **To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments**. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

The reference document for data protection authorities is the Opinion 05/2014 on anonymization techniques issued by the Article 29 Working Party. Despite the fact that this document does not affect the GDPR, and therefore does not appear in EDPB Endorsement 1/2018, it has obviously inspired the practice of data protection authorities and legal doctrine. This document considers the Recital (26) of Directive 95/46/EC:

> (26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person; **whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable**; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered

anonymous and retained in a form in which identification of the data subject is no longer possible;

According to this document, the minimum requirements that an anonymization procedure should meet have two dimensions, the legal one and the dimension relating to the techniques applied by the organisation.

From a legal point of view, anonymization is linked with the original legal basis for processing at source and it is considered as a natural process for data storage and reuse. This entails the need to:

▪ Justify the existence of a basis that legitimises the anonymised reuse of data:

- consent given unambiguously,

- performance of a contract to which the data subject is party or for the implementation of pre-contractual measures taken at the request of the data subject, or

- compliance with a legal obligation to which the controller is subject,

- necessary to protect the vital interest of the data subject,

- necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed, or

- satisfaction of the legitimate interest provided that the interest or fundamental rights and freedoms of the data subject are not overriding.

▪ Ensure conditions of coherence with the purpose:

- relationship between the purposes for which the personal data were collected and the purposes of further processing (purpose limitation);

- ensure that in the context in which personal data were collected, the reasonable expectations of data subjects as to their further use were met;

- be appropriate to the nature of the personal data and the impact of further processing on the data subjects;

▪ The controller must implement and demonstrate the safeguards adopted to ensure proper processing and to prevent any undue negative impact on data subjects.

Ensure the storage of data in an identifiable format during the legal time limits as a guarantee of access (CJEU case C-553/07):

- information on the recipients or categories of recipients to whom the data are disclosed,

- to the content of the information communicated.

It is therefore advisable to bear in mind that:

- The possibility of reversibility should be assessed in advance

- The technological means "are not only those of the controller" but rather "those available".

- Linking must be impossible both for the controller and for any third party.

- It is not just a matter of retrieving first name, surname, address, if there is the slightest possibility of potential identifiability by singling out, linkability or inference: **data protection law applies**.

- If the context (technology, relationship of data with other subsets of data...) involves risks: **data protection law applies**.

From the point of view of the anonymization process, the Article 29 Working Party, or the European Data Protection Board, establishes a procedure based on a risk analysis approach. The first step is to establish whether we face any of the three key risks:

▪ Singularisation: the possibility of extracting from a data set some records (or all records) that identify a person.

▪ Linkability: the ability to link at least two records of a single data subject or a group of data subjects, either in the same database or in two different databases.

- If the attacker can determine (e.g. by correlation analysis) that two records are assigned to the same group of persons but cannot single out the persons in this group, then the technique is resistant to singling out, but not to linkability.

▪ Inference: the possibility of deducing with significant probability the value of an attribute from the values of a set of other attributes.

Three key ideas should be retained from this document:

▪ Anonymisation is a processing operation. Therefore, legality requirements must apply, including the conclusion of a processor contract when it is carried out by a third party.

▪ Pseudonymisation is not the same as anonymization.

▪ Anonymisation is conceived as irreversible or equivalent to erasure.

It is therefore recommended that a variety of complementary techniques be used to achieve the goal of anonymization (see the Opinion 5/2014 categorisation).

2. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 (EHDSR)

The EHDSR expressly states the preference of anonymization as a strategy for the processing of data for research purposes. Pseudonymisation will be reserved for those cases, such as rare diseases (whatever the case may be) in which it is more complex to anonymise and always upon justification of the need to process pseudonymised data.

(53) Electronic health data used for secondary use can bring great societal benefits. The uptake of real-world data and real-world evidence, including patient-reported outcomes, for evidence-based regulatory and policy purposes as well as for research, health technology assessment and clinical objectives should be encouraged. Real-world data and real-world evidence have the potential to complement health data currently made available. To achieve that goal, it is important that datasets made available for secondary use pursuant to this Regulation be as complete as possible. This Regulation provides the necessary safeguards to mitigate certain risks involved in the achievement of those benefits. **The secondary use of electronic health data is based on pseudonymised or anonymised data, in order to preclude the identification of the data subject.**

(…)

(65) (...) They **(Health Data Access Bodies) should apply tested state-of-the-art techniques** that ensure electronic health data are processed in a manner that preserves the privacy of the information contained in the data for which secondary use is allowed, **including techniques for pseudonymisation, anonymisation, generalisation, suppression and randomisation of personal data**. (…). When relevant, the Commission should set out the procedures and requirements, and provide technical tools, for a unified procedure for pseudonymising and anonymising electronic health data.

(…)

(72) Given the sensitivity of electronic health data, it is necessary to reduce risks for the privacy of natural persons by applying the data minimisation principle. Therefore, non-

personal electronic health data should be made available in all cases where the provision of such data is sufficient. **If the health data user needs to use personal electronic health data, it should clearly indicate in its request the justification for the use of that type of data and the health data access body should assess whether that justification is valid. The personal electronic health data should only be made available in pseudonymised format. Taking into account the specific purposes of the processing, personal electronic health data should be pseudonymised or anonymised as early as possible in the process of making data available for secondary use**. It should be possible for pseudonymisation and anonymisation to be carried out by health data access bodies or by health data holders. As controllers, health data access bodies and health data holders should be allowed to delegate those tasks to processors. When providing access to a pseudonymised or anonymised dataset, a health data access body should use state-of-the-art pseudonymisation or anonymisation technology and standards, ensuring to the maximum extent possible that natural persons cannot be re-identified by health data users. Such technology and standards for data pseudonymisation or anonymisation should be further developed. Health data users should not attempt to re-identify natural persons from the dataset provided under this Regulation, and where they do so they should be subject to administrative fines and enforcement measures laid down in this Regulation or possible criminal penalties, where national law so provides. Moreover, a health data applicant should be able to request a response to a health data request in an anonymised statistical format. In such cases, the health data user will only process non-personal data, and the health data access body will remain sole controller for any personal data necessary to provide the response to the health data request.

Given the sensitivity of electronic health data, it is necessary to reduce risks for the privacy

(…)

(92) Certain categories of electronic health data can remain particularly sensitive even when they are in anonymised format and thus non-personal, as already specifically provided for in Regulation (EU) 2022/868. **Even where state-of-the-art anonymisation techniques are used, there remains a residual risk that the capacity to re-identify could be or become available, beyond the means reasonably likely to be used**. Such residual risk is present in relation to rare diseases, that is to say a life-threatening or chronically debilitating condition affecting not more than 5 in 10 000 persons in the Union, where the limited numbers of cases reduce the possibility of fully aggregating the published data in order to preserve the privacy of natural persons while also maintaining an appropriate level of granularity in order to remain meaningful. Such residual risk can affect different categories of health data and can lead to the re-identification of the data subjects using means that are beyond those reasonably likely to be used. Such risk depends on the level of granularity, on the description of the characteristics of data subjects, on the number of people affected, for instance in cases of data included in electronic health records, disease registries, biobanks and person-generated data, where the range of identification characteristics is broader, and on the possible combination with other information, for example in very small geographical areas, or through the technological evolution of methods which had not been available at the moment of anonymisation. Such re-identification of natural persons would present a major concern and would be likely to put the acceptance of the rules on secondary use provided for in this Regulation at risk. Furthermore, aggregation techniques are less tested for non-personal data containing for example trade secrets, as is the case in the reporting on clinical trials and clinical investigations, and enforcement of breaches of trade secrets outside the Union is more difficult in the absence of a sufficient international protection standard. Therefore, for those categories of health data, there remains a risk of re-identification after the anonymisation or aggregation, which cannot be reasonably mitigated initially. This falls within the criteria indicated in Article 5(13) of Regulation (EU) 2022/868. Those types of health data would thus fall within the empowerment set out in Article 5(13) of that Regulation for transfer to third countries. The special conditions provided for under the empowerment set out in Article 5(13) of Regulation (EU) 2022/868 will be detailed in the context of the delegated act adopted under that empowerment and need to be

> proportional to the risk of re-identification and to take into account the specificities of different data categories or of different anonymisation or aggregation techniques.

This future regulation imposes certain obligations on users with data access authorisation:

▪ The results or output of secondary use shall contain only anonymous data (Article 61(4).

▪ In practice, Article 66 of the Regulation imposes the principle of anonymisation by default by forcing the data access applicant to justify access to pseudonymised data:

> 3. Where the health data user has sufficiently demonstrated that the purpose of processing cannot be achieved with anonymised data in accordance with Article 68(1), point (c), health data access bodies shall provide access to electronic health data in pseudonymised format. The information necessary to reverse the pseudonymisation shall be available only to the health data access body or an entity that acts as a trusted third party in accordance with national law.

This justification requires the following elements to be included in the application (article 67):

> (e) a description explaining whether the electronic health data need to be made available in a pseudonymised or anonymised format; in the case of a pseudonymised format, a justification as to why the processing cannot be carried out using anonymised data;
>
> (…)

In the issuance of data permits, Article 68 applies the same criteria. The health data access bodies shall assess whether all the following criteria are fulfilled:

> (c) the processing complies with Article 6(1) of Regulation (EU) 2016/679 and, in the case of pseudonymised data, there is sufficient justification that the purpose cannot be achieved with anonymised data;

On the other, hand there is a limitation for Health Data Access Bodies that «shall provide electronic health data in an anonymised format, where the purpose of processing by the health data user can be achieved with such data, taking into account the information provided by the health data user» (article 66 (2)). Under Article 68 (3) these Bodies must issue a data permit deciding to provide a response in an anonymised statistical format where the requirements for issuing a data permit are not met, on condition that providing that response would mitigate the risks and, if the purpose of the health data access application can be fulfilled in this manner, that the health data applicant agrees to receiving a response in an anonymised statistical format.

3. Interplay with Artificial Intelligence Act (AIA)

However, there is another essential question from the point of view of secondary use of electronic anonymised health data, when it comes to the processing of massive amounts of data, the so-called big data, or the use of artificial intelligence. The inspiring principles of the AI Regulation call for a risk-based approach. For this, there are essential values such as accountability and the exclusion of bias. And when it specifically addresses data governance in high-risk systems (Article 10), it emphasises that data governance requires a delicate balance between the principle of data minimisation and the variety, variability, and diversity of data.

The anonymization techniques recommended by WP216 necessarily involve the reduction of variables, the exclusion of whole groups of subjects, or the use of techniques that may have certain collateral effects. There is nothing to prevent a properly anonymised data set from losing its scientific value or, even more painfully, from failing to support the risk analysis that is essential for the use of artificial intelligence. In other words, the anonymised dataset is fully acceptable from a GDPR perspective but creates risks in areas such as the exclusion of bias or explainability. On the other hand, given the practical impossibility of obtaining the consent required in most jurisdictions, large-scale processing of pseudonymised data is currently impossible. **Thus, unless the EHDSR is seen as legitimising the processing of large volumes of pseudonymised data without consent, the system would be unworkable**. In any case, the system would be unworkable for areas of scientific research different than scientific research related to health or care sectors considering those will not have any regulation as EHDSR that would legitimise the processing of pseudonymised data without consent.

4. Secure processing environments, robust anonymization and software intermediation as combined techniques that ensure properly anonymisation and an adequate risk management

Secure processing environments in European Data Spaces implement layered technical and organizational measures to robustly protect anonymized data from re-identification risks. When combining closed environments, traceability, encryption, and multi-party computation tools, these systems create defense-in-depth barriers against reidentification attempts.

4.1 Closed environments with strict access controls

European Health Data Spaces mandate secure processing environments that:

- Restrict access to pre-authorized personnel listed in data permits.
- Separate data storage from processing zones to prevent direct downloads of raw data.
- Implement granular role-based permissions (e.g., read-only access for analysts).

These closed systems reduce attack surfaces compared to open architectures.

4.2 Comprehensive traceability mechanisms

Mandatory logging features enable retroactive detection of reidentification attempts through forensic analysis:

- Recording all data accesses with timestamps and user IDs.
- Tracking query patterns to detect anomalous behaviour.
- Maintaining immutable audit trails for 10+ years post-processing.

4.3 Encryption protocols

Combined cryptographic protections:

Technique	Protection Scope	Implementation Example
Homomorphic Encryption	Protects data during computation	EHDS-compliant systems process encrypted health records without decryption
AES-256 Transport Encryption	Secures data in motion	Used in EU data space implementations
Multi-Layer Key Management	Prevents single-point failures	Separate keys for storage, processing, and access revocation

4.5 Multi-party computation (MPC) safeguards

Advanced MPC frameworks:

- Split data processing across multiple trusted nodes.
- Require consensus from ≥3 parties to reconstruct partial datasets.
- Integrate TEEs (Trusted Execution Environments) with hardware-enforced isolation.

Recent implementations show these techniques reduce or prevent linkage attacks.

4.6 Traceability

Traceability of user actions serves as a critical defence mechanism against re-identification risks in data environments through layered monitoring and accountability measures. Here's how it operates:

a) Real-Time Anomaly Detection

Traceability systems continuously log user interactions with data assets, enabling:

- Identification of suspicious patterns like bulk data exports or unauthorized access attempts.
- Immediate alerts for atypical actions (e.g., 3 AM database queries by marketing staff).
- Correlation of access events with contextual factors (user role, device location, data sensitivity).

b) Forensic Audit Capabilities

Immutable activity logs create permanent records that:

- Allow retrospective analysis of re-identification attempts.
- Establish chain-of-custody evidence for legal proceedings.
- Support GDPR's "demonstrable compliance" requirements through timestamped proof of data handling.

c) Behavioural Profiling

Advanced UEBA (User and Entity Behaviour Analytics) systems:

- Create baseline activity patterns for each user/role.
- Detect deviations like sudden access to anonymized health records by logistics personnel.
- Assign risk scores to prioritize investigations.

d) Deterrent Effect

The knowledge of comprehensive monitoring:

- Reduces insider threat attempts.
- Enforces least-privilege access policies through visible accountability.
- Prevents credential sharing through individual action tracing.

e) Integrations with Security Controls

Traceability amplifies other protections by:

- Providing access logs for encryption key rotation schedules.
- Feeding MPC (Multi-Party Computation) systems with user authentication metadata.

- Triggering automated responses like session termination when high-risk thresholds are breached.

In the European Health Data Space framework, traceability systems could reduce successful re-identification attempts when combined with encryption and access controls, primarily by enabling rapid containment of suspicious activities. A well-implemented traceability reduces residual re-identification risks below compliance thresholds when paired with additional technical safeguards.

4.7 Intermediation Tools

Additional software layers enforce strict access policies:

1. **Conclave/Sharemind MPC**
    - Optimizes MPC for big data via hybrid Spark/MPC execution plans.
    - Restricts raw data exposure by preprocessing locally before secure computation.
2. **Firebase Firestore Rules**
    - Enforces field-level access controls (e.g., published == true).
    - Blocks unauthorized queries at the API layer.
3. **Homomorphic Middleware**
    - Tools like **Microsoft SEAL** or **OpenFHE** manage encrypted data pipelines.
    - Integrate with databases to process encrypted columns directly.
4. **Environment Design for Zero-Exposure**

A robust implementation combines:

1. **Closed processing zones**: Data storage isolated from query interfaces.
2. **Query sandboxing**: Tools like **Conclave** auto-reject requests attempting raw data extraction.
3. **Immutable audit logs**: Track all query attempts to detect/prevent rule circumvention.
5. **Limitations**

- **Performance overhead**: HE increases computation time by 30-100x vs cleartext.
- **Output utility trade-offs**: DP noise reduces dataset accuracy by 8-12%.
- **Implementation complexity**: Requires cryptographic expertise to avoid misconfigurations.

When properly implemented with MPC, HE, DP, and intermediation tools can create environments where users query data **without direct access to datasets**. The combination of cryptographic processing, output perturbation, and strict access controls prevents downloading, copying, or viewing raw data while maintaining usability for approved analytical tasks.

4.8 Combined Effectiveness

The integration of these measures creates overlapping protections:

1. **Pre-attack prevention**: Closed environments and encryption block external recon attempts.
2. **During-processing security**: MPC and TEEs prevent internal data exposure during analysis.
3. **Post-processing auditability**: Trace logs enable rapid containment of breaches.

5. Allegations regarding Guidelines 1/2025

5.1 Submissions regarding paragraph 22

Regarding paragraph 22, we think that it is the most important paragraph of the Opinion because it lays down the key rule to determine whether any subject is processing pseudonymised data or anonymised data. This key rule is worded on the sentence "If pseudonymised data and additional information could be combined having regard to the means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal." However, the opinion doesn't explain how the EDPB considers that a subject (data controller, processor, recipient or third party) needs "means reasonably likely" or more than these means to combine the pseudonymised data and additional information in order to reidentify the data subjects.

Through the document, the opinion states different scenarios where say that they are pseudonymised data but it doesn't explain the argument that justifies it. We think that is necessary to know the argument applied to each scenario in order to understand the reasoning of the EDPB and, therefore, to have legal certainty.

> 22. Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person, and is therefore personal. **This statement also holds true if pseudonymised data and additional information are not in the hands of the same person.** If pseudonymised data and additional information could be combined having regard to the means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal. **Even if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data becomes anonymous only if the conditions for anonymity are met**.

The sentences underlined in bold lead inescapably to an objectification of the concept of pseudonymisation that was already present in Opinion 5/2014, and it isn't aligned with the risk-based approach of the GDPR. In practice, the effort made for anonymisation in a Secure Processing Environment (SPE) will be irrelevant since under the slightest risk condition anonymisation will be considered an impossible objective. The possibilities of re-identification of a real user are relevant in the concept of anonymisation. An SPE, in fact guarantees:

> **1. That there is no third-party user outside the SPE with the capacity to re-identify.**
>
> **2. That technological measures are adopted which, together with robust anonymisation, guarantee the impossibility of re-identification.**

3. That additional measures are adopted such as user training and the signing of binding non-re-identification commitments by users.

Therefore, it seems necessary to adopt clearly and directly the concept of "de facto anonymization"[1]. The de facto anonymization methodology is compatible with a methodology based on a risk-based approach. This involves applying anonymization methodologies in successive phases following the steps recommended for example by the Spanish Data Protection Agency combined with the design of secure processing environments provided of security and traceability measures. In addition to anonymization techniques such as synthetic data, the software intermediation - such as differential privacy, secure multi-party computation or homomorphic encryption - could be considered as processes equivalent to anonymization. These are technologies with a high cost in terms of processing capabilities[2]. But at least they offer a framework of seemingly reliable solutions.

---

[1] «De facto anonymization (sometimes also referred to as relative anonymization) describes de-identification operations by which so many identifiers are removed and further techniques (see 6.1) to reduce personal reference (e.g. randomization or generalization) are applied that re-identification with reasonable efforts in accordance with the current state of the art (see 3.3.4) is no longer possible and the personal reference is eliminated»
BDI (2021). *Anonymization of personal data. A cross-sector practical guide for industrial companies*. Available at https://issuu.com/bdi-berlin/docs/202103_handbook_bdi_anonymization-of-personal-data

[2] « **Anonymisation: Secure Multiparty Computing**
Secure Multiparty Computation150 or SMPC. This is a cryptographic protocol that, by means of Additive Secret Sharing, allows a secret data to be segmented into different parts, so that, when the information is shared, the original data cannot be revealed by any of the sources. In the protocol, the desired result is obtained without the need to reveal any sensitive data, and the result obtained does not suffer any type of deviation.
This strategy is useful in certain scenarios and requires technological assistance to implement it.
Anonymisation: Differential privacy
Differential privacy guarantees, by incorporating random noise to the original information, that in the result of the analysis process of the data to which this technique has been applied, there is no loss in the utility of the results obtained. It is based on the Law of Large Numbers, a statistical principle that states that when the sample size grows, the average values derived from it approach the real mean value of the information. Thus, the addition of random noise to all the data compensates for these effects and produces an 'essentially equivalent' value.
One example of use can be found in the US Census Bureau, which applies differential privacy to ensure the accuracy of its statistics and prevent personal information from being disclosed even through the statistics and thus increase citizens' confidence in the security of the data they provide.
Anonymisation: Anonymisation-oriented documents
Recital 9 of the DGA, in the case of re-use of data, states the need to develop data processing in which anonymisation is built into the concept of the data and in which data formats allow for efficient anonymisation 'by design': *'In order to facilitate the protection of personal data and confidential data and to speed up the process of making such data available for re-use under this Regulation, Member States should encourage public sector bodies to create and make available data in accordance with the principle of 'open by design and by default' referred to in Article 5(2) of Directive (EU) 2019/1024 and to promote the creation and the procurement of data in formats and structures that facilitate anonymisation in that regard.*'
Other techniques for safeguarding data protection
Without aiming to be exhaustive, there are other techniques used to safeguard data protection when sharing data. For example, homomorphic encryption, the recovery of private information, or the federated learning techniques in machine learning. The following is a brief overview of each of these techniques.
Homomorphic encryption is a privacy-by-default technique that is suitable for cases where a controller outsources a part of an activity to a processor and wants to technically ensure that the processor will not access the data.
In a traditional scheme, the data controller transmits the information to the processor in encrypted form, to protect confidentiality during transit. Once the processor has received it, it is decrypted and processed. However, this scheme presents both legal and technical risks, so ideally, to minimise the risks, the processor should not have the possibility to decrypt the information, and all processing should be carried out on the data encrypted by the data controller. This would prevent a disloyal processor or a third party from accessing the data and using it for different purposes. One way to achieve this protection is through the so-called homomorphic encryption.

The technical deployment could be completed by a legal apparatus based on three levels of interaction:

1.-The data holder provides the data by means of a data sharing and/or data transfer agreement or responsive declaration.

A set of formal and material guarantees are required from the entity:

▪ Ensuring that the data processing for making the data set available is legitimate and has been authorised by the data controller or by law.

▪ Ensuring compliance with any applicable ethical requirements.

▪ Declare and catalogue restrictions on use in accordance with national law.

▪ Implement and accredit a reliable anonymization process.

In this regard, the platform from which data is processed should integrate a risk analysis. If there is a risk that the data are not correctly anonymised or that the infrastructure itself is used for this process, the legal support to the data holder necessarily implies the signing of a data processor contract.

2.-A legal-technical governance frameworks for data access must be established by the platform.

The achievement of this objective implies that, irrespective of the data access rights referred to in the EHDSR, the user with access to data must be subject to certain requirements:

▪ A registration process on the platform, which implies an awareness of the rules of the platform.

▪The acceptance of a clear set of terms and conditions. This legal act does not correspond, strictu sensu, to the applicant for access to the data or to the user of the data, if they are different subjects, but to the legal representative of the entity in which they carry out their activities.

▪ The explicit acceptance by the user(s) of the information system of the system-specific security obligations and of the commitments not to re-identify. This should be done by taking a positive action such as ticking a box and accepting by receiving a confirmed message. The process must be digitally evidenced.

De facto anonymization therefore consists of a technical deployment that includes a robust anonymization process complemented by a secure processing environment. The latter prevents both the improper downloading or manipulation of data and any access

---

Homomorphic encryption therefore makes it possible to perform operations on encrypted data and obtain results, also encrypted, equivalent to the operations performed directly on the original information.
On the other hand, Private Information Retrieval (PIR) is a cryptographic technique that allows the user to retrieve an entry from a database without revealing to the data custodian the item that has been retrieved and unlink the information that could be inferred regarding who is performing the access. (…)
Lastly, we can also highlight federated learning techniques, both horizontal and vertical, for artificial intelligence applications based on Machine Learning. Federated learning techniques are a category of PET (Privacy-Enhancing Technology) that allow the development of machine learning systems without the need to communicate personal data between participants. These techniques can be both horizontal and vertical and are key in new scenarios for the improvement and development of society, such as Data Spaces».
Agencia Española de Protección de Datos (2023). *Approach to Data Spaces from GDPR Perspective*, pages 51-53. Available at https://www.aepd.es/documento/approach-to-data-spaces-from-gdpr-perspective.pdf

by third parties. It thus excludes the risk of re-identification by third parties and the risk associated with the use of malicious software since no processing tools other than those previously authorised by the platform are supported.

This method should be acceptable, adequate, and legally admissible. If this matter is not adequately resolved, we are faced with a particular situation. In most national laws there are no exceptions to consent for the processing of pseudonymised data.

The risks associated with the objectification of the concept of pseudonymisation are as follows:

▪ By default, all data sets created in the field of healthcare should be considered pseudonymised only.

▪ During the transition to the EHDSR the processing of pseudonymised data will not be feasible. If consent is to be used, it is necessary to take into account the position of the European Data Protection Board in the Guidelines 05/2020 on consent under Regulation 2016/679 and its consequences, i.e. the impossibility to collect large volumes of data on every occasion and for every data access request[3].

▪ Only an interpretation of the EHDSR that implies that this regulation authorises the pseudonymised processing of data for secondary purposes without consent and takes precedence over national law would solve the above problem.

▪ The anonymisation objectives resulting from the interaction of WP216 and Guidelines 1/2025 can lead to the generation of low-quality data sets for research, particularly in the development of high-risk AI systems. This should objectively lead notified bodies to refuse the CE mark due to non-compliance with the data governance requirements of Article 10 of the AIA or due to the presence of systemic risks to explainability, bias exclusion or reproducibility. In the case of medical devices, this can imply a severe risk to the integrity and health of the patient.

▪ In practice, the design of the EHDSR will be altered. Instead of the user, it will be the data holder and the Health Data Access Bodies that declare a dataset pseudonymised (and not anonymised). In this case, the only possible justification for the data access applicant when justifying the pseudonymised use of the data can be none other than the absence of anonymised data. This undoubtedly denatures the objectives of the EHDSR.

5.2 Submissions regarding paragraph 42

> 42. If a controller or processor wants to use pseudonymisation to reduce confidentiality risks from some or all unauthorised third parties, they will include those third parties in the pseudonymisation domain and assess the means they are reasonably likely to use for attribution. Relevant third parties include not only cyber-crime actors, but also employees or maintenance service providers acting in their own interests rather than on instructions from the controller. Taking due account of contextual elements and the circumstances at hand, it is recommended to consider both actions in good faith, and those executed with criminal intent.

Regarding paragraph 42, the opinion says, "It is recommended to consider both actions in good faith, and those executed with criminal intents". Considering that

[3] See paragraphs 143 to 160. Available at https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en

Recital 26 of the GDPR stablishes the rule of the "means reasonably likely" to be used to reidentify data subjects, this rule should be applied on this sentence. Therefore, the data controllers should consider only criminal intents that have been done during, for example, the last year and inside the sector or region of the data controller. We consider that the data controller does not have to take into account the criminal intents done in, for example, Argentina, whether the data controller is processing personal data in a region of Italy when there is not any connection between these two regions. Moreover, the scope of criminal intents to be considered should be limited to the criminal intents done inside a professional sector (public bodies, healthcare services providers, hospitals, universities, energy supply companies, etc.)

5.3 Submissions regarding paragraph 43

> 43. For instance, pseudonymisation may be performed prior to transmission of the data to a processor or third party that ensures only a level of security that would not be appropriate for the processing of the original data, but is appropriate for the risk connected with the processing of data that cannot be attributed to data subjects. In this case, all means available to unauthorised parties that might access the pseudonymised data while the (authorised) recipient of that data processes them need to be considered.
>
> 51. For external processing, i.e. processing under instruction by a processor or transmission to an independent controller, more extensive measures and risk assessment may be necessary to prevent attribution to data subjects. In particular, all intended recipients of the pseudonymised data need to demonstrably assure that the pseudonymised data are not disclosed to unauthorised recipients beyond the defined domain. For processors, additional tools under Art. 28 GDPR such as audits are available to support this assurance.

Regarding paragraph 51, the opinion states that "In particular, all intended recipients of the pseudonymised data need to demonstrably assure that the pseudonymised data are not disclosed to unauthorised recipients beyond the defined domain". In this statement the opinion requires a probatio diabolica, i.e. it requires that the data controller should be able to prove that something has not happened and, consequently, impossible to prove because, while evidence proves the existence of something, lack of evidence does not disprove it.

5.4 Submissions regarding Example 5: Secondary use for research

Regarding example 5, the common practice is that the medical centres or hospitals send pseudonymised data to the data centres or they send personal data to trust centres that will pseudonymise data, but it is uncommon that hospitals send personal data to data centres before been pseudonymised by them or by trust centres. It is a significant difference to be considered on the assess of re-identifiability of data subjects according to Recital 26 of the GDPR.

Moreover, in relation to section "Processing of pseudonymised data" of this example, we think that is necessary to know the reasons that justify why the Opinion considers that the data received by the research groups will be pseudonymised data whether there will be technical and organisational safeguards from access to any additional information that would allow the reidentification. In particular, we think that is necessary to know the subjacent reason why research groups will still only need "means reasonably likely" to reidentify the data subjects.

Something else should be added. One of the clear effects of the EHDSR is that it is increasingly common for hospitals and healthcare systems to opt for the anonymisation methodologies described in section 4.1.