ANONOS®

On November 10th ANONOS Inc. filed at the Court of Justice of the European Union a request to intervene in an appeal procedure (in case **C-413/23 P**) in support of the European Data Protection Supervisor (EPDS).

> The case concerns a judgment of the General Court (Case T-557/20), *Single Resolution Board v European Data Protection Supervisor* which concerns the definition and the nature of personal information. When is data anonymized as to fall outside of the scope of protection of the right to protection of personal information, when is data pseudonymized which although still being understood as personal data, allows it to be processed for various commercial and other purposes?

In the request for intervention, ANONOS submits to the Court various observations on the future interpretation of the notions of anonymization and pseudonymization within Regulation 2018/1725 (EDPR) and the GDPR.

Under the relevant legislation in the EU, personal data is defined by the notion of the identifiability of a person.

In the request for intervention, ANONOS argues that identifiability must be understood to result from the *possibility* of combining data in a database with other data that is freely available on the internet or readily accessible in existing software. This makes individuals potentially *identifiable* through *re*-identification.

European legislation, including the EDPR and GDPR, must therefore be interpreted in the sense that *identifiable* data is defined not by a *subjective* concept of whether the data was transferred with a key or other means for re-identifying the data. The central notion must be whether *objectively* the recipient has the possibility of re-identifying data.

The argument is made that this interpretation follows from recital 16 of the EDPR and Recital 26 of the GDPR which contain an *objective* notion of data being universally anonymous by stating that "…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". (Emphasis added)

Therefore, in order to qualify as anonymous information, data objectively must not be capable of being cross-referenced with other data to reveal identity. This high standard is essential because when data does satisfy these requirements, it is treated as being outside the scope of legal protection provided under the EDPR and GDPR. Any exclusion from the scope of protection is acceptable only when there is "safe" protection in place for the data that actually satisfies the stringent requirements of not being cross-referenceable or re-identifiable.

This standard of objective or 'universal anonymisation' is also the standard required by the French Commission Nationale de l'informatique et des Libertés (CNIL), having published its recommendations on anonymisation and pseudonymisation stating that "[a]nonymisation is a treatment which consists in using a set of techniques in such a way as to make it impossible, in practice, to identify the person by any means whatsoever and in an irreversible manner…"[1]

Equally, the European Data Protection Board (EDPB) in its *Final Schrems II Guidance* states that, along with other requirements, the standard of EU GDPR pseudonymization can be met only if "a data exporter transfers personal data processed in such a manner that the personal data can no longer be attributed to a specific data subject, nor be used to single out the data subject in a larger group, without the use of additional information." [2]

ANONOS points out that the standard applied for anonymisation cannot be less than that applicable for pseudonymisation, the former taking data outside of the scope of application of the GDPR.

Second, ANONOS submits in its request for intervention that possibilities allowing for attribution of data to natural persons by use of data from the internet or other readily available sources have quickly developed in the past years. Generative AI tools now offer readily accessible possibilities for an extremely wide range of users to re-combine data sets with widely available data leading to unauthorized re-identification and attribution to identified or identifiable natural persons.

Advanced systems such as large language model-based software using generative AI (including GPT, Bard and many other AI services available on the internet) often involve the combination of very large amounts of internal and external data, trade secrets, and personal data, which introduces a very serious privacy risk of re-identification of data sets without requiring the use of 'additional information' that is kept separately and subject to technical and organizational measures to ensure it cannot be attributed to identified or identifiable natural persons.

Therefore, the persons to whom the data pertains will potentially remain at least *indirectly identifiable* even in a seemingly anonymized data set. In fact, such data then fails to satisfy the requirements of either 'anonymous' or 'pseudonymous' data and, therefore, is *de facto* not entitled to any of the associated statutory benefits under the EDPR and the GDPR. As noted by CNIL, anonymization, which takes data outside of the category of personal data, must

[1] See https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

[2] See paragraphs 79, 85, 86, 87 and 88 of the EDPB Recommendations 01/2020 on Measures that Supplement Transfer Tools to Ensure Compliance with the EU Level of Protection of Personal Data Version 2.0 adopted on 18 June 2021 at https://edpb.europa.eu/system/files/2021-06/edpb_recommendations_202001vo.2.0_supplementarymeasurestransferstools_en.pdf ("EDPB Final Schrems II Guidance").

"make it impossible, in practice, to identify the person by any means whatsoever and in an irreversible manner."[3]

Pseudonymisation, on the other hand, is a category of personal data which can be submitted to certain processing activities such as transfer because of the use of certain privacy enhancing technologies, including cryptographic algorithms, in order to ensure security and reduce the risk of re-identification by an unauthorized party.[4]

Recital 16 of the EDPR and Recital 26 of the GDPR explain that the state of the art is relevant to defining anonymization and pseudonymization and that "account should be taken of all the means reasonably likely to be used" including "taking into consideration the available technology at the time of the processing and technological developments. …"

Previous case law such as the CJEU judgement in *Breyer*[5] could be understood to have followed a more subjective approach to anonymization focused on the likelihood of re-identifying an individual with respect to which data is included in a dataset from the perspective of the intended recipient(s) of the data.

It considered the specific context in which the data is processed and the means that intended recipients might use to re-identify data subjects. This approach is "subjective" or "localized" in that it looked at the data environment and the capabilities of the intended recipients to match purportedly anonymized data with other information thereby leading to unauthorized identification.

Prior to the EDPR and GDPR, this approach was generally considered necessary to enable data innovation, including the lawful sharing and multi-party analysis of data for commercial and societal benefit. However, treating such "locally" protected data as being outside of the scope of protected data is today inconsistent with the EDPR's and GDPR's legislative requirements for taking into account the state of the art and its goal of protecting the fundamental rights of data subjects. This is due to today's realities of:

(i) The ease with which "localized" data can be combined with readily available data that is external to a "location" resulting in unauthorized reidentification of individuals;
(ii) The inability of organizational or contractual measures by themselves to prevent *a priori* the misuse of the data;
(iii) The increasing popularity of data processing activities involving innumerable parties;[6]

[3] See Note 1.

[4] See Note 2.

[5] CJEU judgment of 19 October 2016, *Breyer* (C-582/14, EU:C:2016:779).

[6] Increasingly popular multi-party processing activities like those involved in popular Large Language Models (LLMs) like ChatGPT and Bard increase the risk of unauthorized re-identification of individuals when multiple datasets are combined, even if each dataset by itself appears anonymous (referred to as the "Mosaic Effect"). While encryption, access controls, masking, and tokenization can serve as protective "guardrails,"

(iv) The increasing prevalence of data breaches and cybercrime that exposes data to unintended recipients.

In contrast, the objective approach to anonymization adopted under the GDPR considers all means "reasonably likely to be used" to identify a person, directly or indirectly, by the data controller or any other third party. This involves a broader perspective beyond the capabilities of the intended recipients and considers the ability of third parties to re-identify data subjects.

However, this does not mean that the data is unavailable for innovation, including the lawful sharing and multi-party data analysis for commercial and societal benefit. Recital 4 of the GDPR specifically states:

> The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. This Regulation respects all fundamental rights and observes the freedoms and principles recognised in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications, the protection of personal data, freedom of thought, conscience and religion, freedom of expression and information, freedom to conduct a business, the right to an effective remedy and to a fair trial, and cultural, religious and linguistic diversity.

To allow processing of personal data public or for private uses Article 25 GDPR encourage data controllers to leverage the state of the art by complying with new "data protection by design and by default" obligations. Specifically, controllers are supposed to "implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects." *(emphasis added)*

The inclusion - for the first time in EU law - of a statutory definition for "pseudonymization" shows that parties can leverage state of the art technical capabilities to enable data-driven innovation that balances fundamental rights - while staying within (versus outside) the scope of the protection of personal data as defined by the EDPR and the GDPR. Prior to the statutory redefinition, the term "pseudonymization" was often used to describe the result of the failed anonymization of personal data. In contrast, to be entitled to the specific statutory

they fall short of achieving the necessary protection required to prevent unauthorized reidentification. Combining diverse datasets protected with masking and tokenization alone can allow the correlation of seemingly harmless information, leading to unauthorized reidentification via the Mosaic Effect. While access controls and encryption may prevent unauthorized access, they do not stop authorized entities from exploiting data to reveal identities via the Mosaic Effect.

benefits attributable to “pseudonymization” under the GDPR,[7] parties must now show that (a) “the personal data can no longer be attributed to a specific data subject without the use of additional information,” **and** (b) “such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.” *(emphasis added)*.[8]

See the Appendix for information on the requirements for and benefits of statutory pseudonymization under the GDPR, enabling ample opportunities for lawful data innovation and processing within the scope of the statute.

ANONOS argues that following the entry into force of the GDPR, a reference to *Breyer* (C-582/14, EU:C:2016:779) as precedent for the definition of the scope of personal data is no longer useful. In *Breyer,* the Court had taken a decision based on Article 2(a) of the old data protection Directive 95/46/EC which is no longer in force. The old data protection directive had a Recital 26, which was considerably reformed in the GDPR, which now makes repeated and explicit reference to the state of the art.

For the foregoing reasons, the correct approach, ANONOS submits to the Court, is to include data that is either readily directly or indirectly relinkable to the identity of data subjects within the scope of the protection of Articles 7 and 8 of the Charter under the provisions of the GDPR.

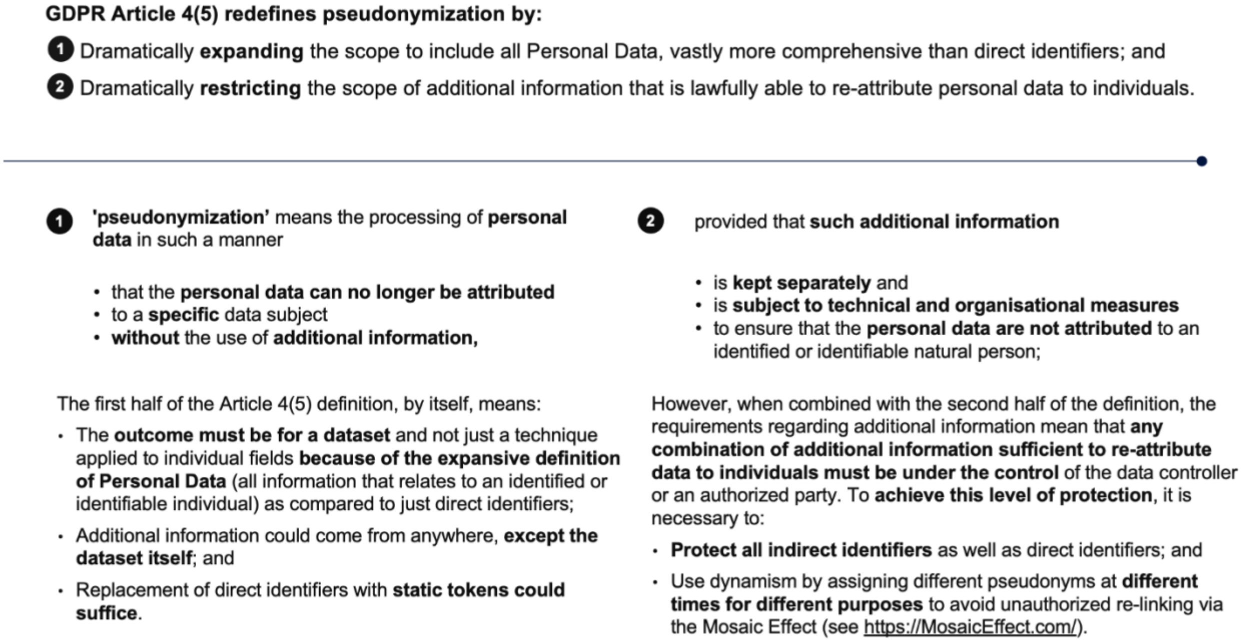
The following explanations **illustrate the background of the technical and legal issues concerning the definition of the scope of personal data** subject to the request for intervention.

[7] In contrast to the term “anonymisation” and derivatives thereof, which appear three (3) times in the GDPR, or the term “encryption” and derivatives thereof, which appears four (4) times in the GDPR, the term “pseudonymisation” and derivatives thereof appear 15 times in the GDPR, many of which, like Article 25 cited above, highlight specific statutory benefits and expanded data use privileges that result from pseudonymising personal data in a manner that satisfies the new heightened statutory requirements under the GDPR and the EDPR. See the Appendix for information on the requirements for and benefits of statutory pseudonymisation under the GDPR, enabling ample opportunities for lawful data innovation and processing within the scope of the statute.

[8] The term “pseudonymisation” is defined in both Article 4(5) of the GDPR and Article 3(6) of the EDPR as follows: “pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”

Requirements For and Benefits of GDPR Statutory Pseudonymization

The following graphic highlights changes to the definition of "pseudonymization" under the GDPR.



GDPR Article 4(5) redefines pseudonymization by:

1. Dramatically **expanding** the scope to include all Personal Data, vastly more comprehensive than direct identifiers; and
2. Dramatically **restricting** the scope of additional information that is lawfully able to re-attribute personal data to individuals.

1. **'pseudonymization'** means the processing of **personal data** in such a manner

- that the **personal data can no longer be attributed**
- to a **specific** data subject
- **without** the use of **additional information,**

The first half of the Article 4(5) definition, by itself, means:

- The **outcome must be for a dataset** and not just a technique applied to individual fields **because of the expansive definition of Personal Data** (all information that relates to an identified or identifiable individual) as compared to just direct identifiers;
- Additional information could come from anywhere, **except the dataset itself**; and
- Replacement of direct identifiers with **static tokens could suffice**.

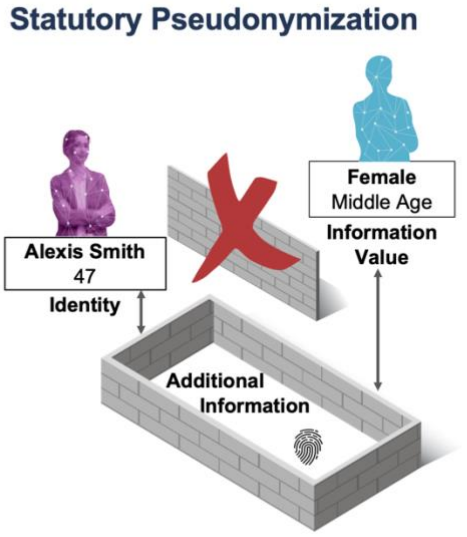
2. provided that **such additional information**

- is **kept separately** and
- is **subject to technical and organisational measures**
- to ensure that the **personal data are not attributed** to an identified or identifiable natural person;

However, when combined with the second half of the definition, the requirements regarding additional information mean that **any combination of additional information sufficient to re-attribute data to individuals must be under the control** of the data controller or an authorized party. To **achieve this level of protection**, it is necessary to:

- **Protect all indirect identifiers** as well as direct identifiers; and
- Use dynamism by assigning different pseudonyms at **different times for different purposes** to avoid unauthorized re-linking via the Mosaic Effect (see https://MosaicEffect.com/).

To satisfy the statutory requirements for pseudonymization under the GDPR, data must not be attributable to a specific subject without additional information that is kept separately and securely by the data controller or designee. The graphic below highlights if it is possible to cross the wall and reconnect "Information Value" (i.e., Female Middle Age) to "Identity" ("Alexis Smith 47") without requiring access to the Additional Information kept separately and securely, then the data is not pseudonymized in accordance with statutory requirements.



Correctly pseudonymized data allows for benefits under GDPR, such as data breach protections, flexibility with data processing for archiving, scientific, or statistical purposes, and supports legitimate interest processing as a legal basis. The following is a summary of these statutory benefits.

- **Support Lawful Data Processing, Repurposing, Sharing and Combining**

    a. Lawful Processing When Consent or Contract Are Not Enough
    [GDPR Article 6.1(f)]

    i. Properly-Pseudonymized data is recognized in the Article 29 Working Party Opinion 06/2014 as **playing "…a role with regard to the evaluation of the potential impact of the processing on the data subject...tipping the balance in favour of the controller" to help support Legitimate Interest processing to protect data in use.**[9]

    ii. The benefits of processing personal data using compliant Legitimate Interests processing as a legal basis under the GDPR include:

    1. Under Article 17(1)(c), if a data controller can show they "have overriding legitimate grounds for processing" supported by technical and organizational measures to satisfy the balancing of interest test, they have greater flexibility in complying with Right to be Forgotten requests.

    2. Under Article 18(1)(d), a data controller has flexibility in complying with requests to restrict the processing of personal data if they can show they have technical and organizational measures in place so that the rights of the data controller properly override those of the data subject because the privacy of the data subject is protected.

    3. Under Article 20(1), data controllers using Legitimate Interests processing are not subject to the right of portability, which applies only to consent-based processing.

    4. Under Article 21(1), a data controller using Legitimate Interests processing may show they have adequate technical and organizational measures in place so that the rights of the data controller properly override those of the data subject because the rights of the data subjects are adequately protected. However, data subjects always have the right under Article 21(3) to *not* receive direct marketing outreach as a result of such processing.

    b. Lawful Repurposing, Sharing and Combining [Article 6.4(e)]

    Pseudonymization is explicitly highlighted in GDPR Article 6(4)(e) as an "appropriate safeguard" that can be used by data controllers "in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected."

- **Separate Processing Benefits from Identity Requirements**
  [Articles 11(2) and 12(2)]

    a. Pseudonymization helps to enable Article 11(2) relaxation of obligations to data subjects under Articles 15 (Right of Access by Data Subject), 16 (Right to Rectification), 17 (Right to Erasure - Right to be Forgotten), 18 (Right to Restriction of Processing), 19 (Notification Obligation Regarding Rectification or Erasure of Personal Data or Restriction of Processing), and 20 (Right to Data Portability) when processing

[9] See Article 29 Working Party 06/2014 at 42 and 67.

does not require identification, if the data controller is not in a position to identify data subjects and the controller has informed data subjects accordingly. Data controllers not in possession of “Additional Information” necessary for re-identification satisfy this requirement.

b. Pseudonymization helps to enable Article 12(2) relaxation of obligations under Articles 15 (Right of Access by Data Subject), 16 (Right to Rectification), 17 (Right to Erasure - Right to be Forgotten), 18 (Right to Restriction of Processing), 19 (Notification Obligation Regarding Rectification or Erasure of Personal Data or Restriction of Processing), and 20 (Right to Data Portability) in addition to the relaxation of obligations under Articles 21 (Right to Object to Automated Decision-Making) and 22 (Automated Individual Decision-Making, Including Profiling) to provide transparent information, communication and modalities for the exercise of the rights of the data subject if the data controller can demonstrate it is not in a position to identify data subjects. Data controllers not in possession of “Additional Information” necessary for re-identification satisfy this requirement.

c. Note that if the “Additional Information” required for relinking to identity (e.g., the Master Index or Keys) is deleted, the data is anonymous under Recital 26 rather than Pseudonymous under Article 4(5).

- **Satisfy Data Protection by Design and by Default Obligations** [Article 25]

a. Article 25(1) requires data controllers - for both primary and secondary processing - to “implement appropriate technical and organisational measures, **such as Pseudonymization**.”

b. Pseudonymization helps data controllers to satisfy their obligations under Article 25(2) to “implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.”

- **Reduce the Risk of Data Breach Liability Obligations and Liability**

a. Article 32 explicitly recognises Pseudonymization and encryption as measures to be considered when“[t]aking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk."

b. Pseudonymization helps to ensure that data breaches are “unlikely to result in a risk to the rights and freedoms of natural persons.” This would mean that an incident would not qualify as a data breach under GDPR and thus would not have to be notified to a supervisory authority under Article 33.

c. Pseudonymization helps to ensure that data breaches are not “likely to result in a high risk to the rights and freedoms of natural persons.” This would mean that an incident would not qualify as a data breach under GDPR and/or (thus) would not have to be communicated to the data subject under Article 34.

- **Enable Benefits of Expanded Lawful Processing**

a. Article 89(1) provides that “[p]rocessing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include Pseudonymization provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.”

b. Article 89(1) Pseudonymization-enabled processing enables greater flexibility under:

i. Article 5(1)(b) with regard to purpose limitation;

ii. Article 5(1)(e) with regard to storage limitation; and

    iii. Article 9(2)(j) with regard to overcoming the general prohibition on processing Article 9(1) special categories.

- **Maximize the Availability of Lawful Profiling and Digital Marketing**

    a. Pseudonymization reduces the risk that profiling "produces legal effects concerning [data subjects] or similarly significantly affects [data subjects]" under Article 22(1) because it can be left up to the data subject whether to choose to participate in opportunities presented to them as a member of a Pseudonymized persona group.

    b. Pseudonymization reduces the risk that profiling "decision[s are made] based solely on automated processing" under Article 22(1) because it can be left up to the data subject whether to choose to participate in opportunities presented to them as a member of a Pseudonymized persona group.

    c. Pseudonymization helps to enable Article 22(2)(b) support for processing "authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests."

    d. Pseudonymization helps to enable Article 22(4) allowance for decisions "based on special categories of personal data referred to in Article 9(1)" premised on Article 9(2)(g) Union or Member State laws by ensuring that "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."

- **Improve Scalability of Data Protection Impact Assessments**

    a. Pseudonymization helps to satisfy Article 35(3)(b) obligations when "processing on a large scale of special categories of data referred to in Article 9(1)."

    b. Pseudonymization helps to satisfy Article 35(8) creation of and adherence to "approved codes of conduct referred to in Article 40 by the relevant controllers or processors shall be taken into due account in assessing the impact of the processing operations performed by such controllers or processors, in particular for the purposes of a data protection impact assessment."

    c. Pseudonymization helps to enable Article 35(10) elimination of separate data protection impact assessment obligations under Articles 35(1)-(7) "[w]here processing pursuant to point (c) or (e) of Article 6(1) has a legal basis in Union law or in the law of the Member State to which the controller is subject, that law regulates the specific processing operation or set of operations in question, and a data protection impact assessment has already been carried out as part of a general impact assessment in the context of the adoption of that legal basis."

For more information, email LearnMore@anonos.com