

Guidelines 02/2026 on Anonymisation

Version 1.0

Adopted on 07 July 2026

Version history

Version	Date	Adoption information
version 1.0	07 July 2026	adoption of the guidelines for public consultation

Executive summary

These guidelines clarify the notion of anonymous data and provide a practical framework for determining whether data has been successfully anonymised. This guidance will therefore help to ensure that data is safely anonymised where possible and proportionate, allowing for free and fruitful use of data while preserving the rights and protections of natural persons.

Under the GDPR, data is anonymous if it does not relate to an identified or identifiable natural person. Whether this is the case may vary from one entity to another. Consequently, anonymity should be assessed from each relevant entity's perspective – typically any party for whom the data is intended to be anonymous.

Information can relate to a natural person by reason of its content, purpose or effect. The existence of such a link need not be readily apparent and may require some processing to establish. A natural person is "identified or identifiable" if they can be distinguished from others in a given context using means reasonably likely to be used and in a way that makes it possible to treat them differently. "Means" should be interpreted broadly and may include means that are only accessible through a third party. Whether they are reasonably likely to be used will depend on the relevant entity's perspective and should be assessed in light of all objective factors.

This analysis is then incorporated into a framework for assessing anonymity. The framework can be applied in two different ways: one which considers the differences in capabilities between those who might identify the data subject ("the contextual approach") and one which does not ("the simplified approach"). The contextual approach reflects the full nuances of the legal standard for anonymisation and allows the controller to assess whether data is anonymous for each relevant entity based on their respective capabilities. The simplified approach, on the other hand, can go beyond the legal standard and may lead an anonymising controller to treat data as though it is not anonymous even if it would actually be so for some relevant entities. However, it also offers a more convenient option for controllers who choose to use it, provides greater confidence that data is actually anonymous and can be combined with the contextualised approach to refine the findings.

The framework itself presents three criteria which can be used to test if data is anonymous: No Record Isolation, No Linkage and No Inference. These criteria should be used to assess the effectiveness of possible (re-) identification techniques. Generally speaking, (re-) identification is more likely to be successful against record-level data with high dimensionality and high resolution, but other factors are also important. Techniques should be assessed on their ability to generate accurate results, in particular whether they produce an answer which is sufficiently precise and reliable to allow the data subject to be distinguished and treated differently.

If all three criteria are passed, under either the contextual or the simplified approach, then the given data can be safely considered anonymous. If any criterion fails, further analysis should be done to determine if the data may nevertheless be considered anonymous. In particular, it should be tested whether any isolated records, possibly together with linked data, allow for singling out individuals.

Table of Contents

1 INTRODUCTION 4

2 LEGAL ANALYSIS 5

2.1 Introduction 5

2.2 Identifying the applicable perspective(s) 6

2.3 Whether the information “relates” to a natural person 8

2.4 Whether the natural person is “identified or identifiable” 9

2.4.1 Identifiers and other attributes 10

2.4.2 Means reasonably likely to be used 11

2.5 Additional considerations 14

2.5.1 Datasets containing a mix of anonymous and personal data 14

2.5.2 GDPR compliance and the anonymisation process 15

3 THE TECHNICAL ANALYSIS OF ANONYMISATION 16

3.1 Introduction 16

3.2 Approaches to the assessment 16

3.3 Simple cases and data mapping 17

3.4 The three criteria 18

3.4.1 No Record Isolation 18

3.4.2 No Linkage 19

3.4.3 No Inference 21

3.5 Applying the criteria 25

3.5.1 Assessing the effectiveness of (re-)identification techniques 25

3.5.2 Assessing the criteria under the contextual approach 26

3.5.3 Compiling the results 29

4 GLOSSARY 31

ANNEX 1 : FLOWCHART FOR THE TECHNICAL ANALYSIS OF ANONYMITY 33

The European Data Protection Board

Having regard to Article 70 (1)(e) of the 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, and repealing Directive 95/46/EC, (hereinafter "GDPR"),

Having regard to the EEA Agreement and in particular to Annex XI and Protocol 37 thereof, as amended by the Decision of the EEA joint Committee No 154/2018 of 6 July 2018,

Having regard to Article 12 and Article 22 of its Rules of Procedure,

Has adopted the following Guidelines:

1 Introduction

1 Data is put to many uses across a wide range of sectors and purposes, many of which are hugely beneficial for an equally wide range of stakeholders. In some cases, this requires the use of personal data and, in such cases, the General Data Protection Regulation ensures that data processing is done in a way that benefits humankind while being safe for data subjects. In other cases, it may be better to use anonymous information, *i.e.* information which does not relate to an identified or identifiable natural person[1]. Anonymous data is not subject to regulation under the GDPR[2].This allows for greater freedom of use, but also means that it is extremely important to ensure that the data is properly anonymised so that the rights of individuals are not put at risk.

2 The Article 29 Working Party Opinion 05/2014 on anonymisation techniques was published in 2014 and provided three criteria for assessing the anonymity of data. Since then, however, there have been major changes in the legal, privacy, engineering and technological landscapes. The development of case law from the Court of Justice of the European Union (CJEU), the establishment of EU-wide data spaces, and developments in both AI and technology in general, amongst many other things, have raised a need to update that Opinion and refine the criteria presented there[3]. These guidelines provide that update, together with further guidance on the topic.

3 The guidelines begin with a legal analysis of anonymisation, and its requirements under the GDPR. This analysis is then incorporated into a framework for assessing anonymity, addressing the three criteria beyond their original formulation in the previous Opinion. The guidelines also set out two approaches for this assessment: one which takes into account the differences in capabilities between those who might identify the data subject ("the contextual approach") and one which, for simplicity's sake, does not ("the simplified approach"). When assessing whether a dataset is anonymous, the question becomes: does that dataset pass the three criteria of No Record Isolation, No Linkage and No Inference? If those criteria are

[1] Recital 26 GDPR. This definition acts as a mirror to Article 4(1) GDPR, which defines personal data as information which does relate to an identified or identifiable natural person (i.e. the data subject). That same Recital also says that information which starts as personal data can later become anonymous if it is processed so that "the data subject is not or [is] no longer identifiable".
[2] Article 2(1) GDPR.
[3] For the avoidance of doubt, a controller that has assessed a dataset as anonymous in accordance with Opinion 05/2014 before the publication of these new guidelines is not expected to conduct a new assessment. It is nevertheless a good practice for any entity to periodically reassess the likelihood of re-identification for anonymised data that they process.

passed, under either the contextual or simplified approach, then the given data can be safely considered anonymous. If one of those criteria are violated, further analysis should be done: the given data may be personal or it may nevertheless be considered anonymous. These guidelines end by providing guidance on that further analysis.

4 Much of the content in these guidelines sits in the space between anonymous and personal data – since the guidelines set out criteria for testing whether a dataset falls into one side or the other, they adopt some terminology specific for these guidelines. It is not possible, for example, to know if the person holding the data should be described as a controller or not, since that answer will depend on whether the data being tested is personal or not. Therefore, rather than, for example, constantly referring to "the person who may or may not be a controller", these guidelines use a number of key terms for the sake of readability, all of which are set out in more detail in a glossary at the end of this document.

5 Particularly importantly, the guidelines use "individual" to refer to a natural person who may or may not be a data subject and "given data" to refer to data which is being tested and may or may not be personal data. The guidelines then use the word "record" to refer to all values (including images *etc*.) within the given data that are known to relate to the same individual. They also use the term "entity" to refer to a party whose role depends on the situation at hand. For example, an entity could be a potential controller, a potential processor, or a potential third party with additional information necessary to identify the data subject. Each entity then has a "perspective", which depends on the circumstances of the case.

2 Legal analysis

2.1 Introduction

6 The core test for anonymity asks two questions:

a. Does the information "relate" to a natural person? If so,

b. Is that natural person "identified or identifiable"?

If the answer to either of these questions is "no", then the data should be considered anonymous. Importantly, these answers may vary from one entity to another. This means that information may be anonymous for some entities, but not for others[4].

7 For example, information may unambiguously relate to an individual, but only one organisation may be able to actually identify that individual. In this case, the answer to both questions would be "yes" from that organisation's perspective, and so the information would be considered personal data for that organisation. At the same time, the answer to the second question would – outside of particular circumstances that will be explained later[5] – be "no" from everyone else's perspective, and so the information would be considered anonymous for everyone else. It is, therefore, necessary to consider the different perspectives of the relevant entities when performing this test. This means that anonymisation can be done in a way that the data is anonymous for everyone, or in a way that is only anonymous for particular entities (which may or may not include the anonymising controller).

[4] C-413/23 P *EDPS v SRB*.
[5] See, *e.g.*, paras 20, 21 and 29.

8 In many cases, the second question will receive a qualified response ("The individual might be identifiable, but only if…"). In particular, there may be several ways that an entity could identify an individual in theory but where the likelihood of this happening is insignificant in practice. This question therefore needs to be viewed as one of likelihood: rather than asking if the individual is identified or identifiable in an absolute sense, it should ask how likely it is that the individual will be identified or identifiable by some entity[6].

9 It should also be recognised that these questions assume that the entity in question has, or could have, some kind of access to, or control over, the data. If, taking into account the means that the entity is reasonably likely to use, an entity cannot actually access or process the data, is not meaningfully connected to any other entities (or chain of entities) with such access, and is not likely to receive the data from such other entities[7], then it should not be necessary to assess whether the information is anonymous for that entity.

10 Finally, these guidelines refer to the concept of "anonymisation" broadly, as any processing which successfully produces anonymous data under the GDPR, whether or not the original data is deleted. This includes anonymisation that does not lead to a deletion of the original data and so does not contribute to storage limitation. These guidelines do not, therefore, provide guidance for the interpretation of existing specific legal provisions that may mandate erasure of the original data.

2.2 Identifying the applicable perspective(s)

11 As noted above, whether information is personal may vary from one entity to another. It is therefore important to determine the relevant entities for the assessment of the given data. Typically, whether information is personal for a particular entity will be assessed by reference to their own perspective – although, as set out below, there are some cases where another entity's perspective should be used instead. The goal of anonymisation should be to ensure that the resulting data is anonymous for all of the relevant entities concerned by the assessment.

12 It can sometimes be immediately apparent if a certain entity is relevant or not. The basic question is: for whom is the data intended to be anonymous? For example, if the intention is to anonymise data for an entity's own use, that entity's perspective will be relevant for the assessment. Likewise, if the intention is to anonymise the data for independent use by anyone who receives the data, anonymity should generally be assessed from the perspective of those who will have access to the data after it has been transferred. In some cases, there may only be one relevant perspective (*e.g.*, if the data is kept securely by a single entity and nobody else will ever be able to gain access to it) while, in others, there may be many different relevant perspectives (*e.g.*, if the data could be accessed by many entities).

Example 1: In *OC v Commission*[8], the CJEU considered whether a press release contained personal data. The Court noted that the press release was, by its very nature, intended for

[6] See, e.g., C-413/23 P EDPS v SRB and C-582/14 Breyer v Bundesrepublik Deutschland.
[7] This can, for example, be the case where there is collusion or cooperation between such entities, or where it is likely that the information may be (directly or indirectly) transferred from one entity to another. See, for example, C-582/14 *Breyer v Bundesrepublik Deutschland* and C-319/22 *Gesamtverband Autoteile-Handel eV v Scania CV AB*. See also C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH*, where the CJEU found that a party who determines the purposes and means of the processing of personal data can still be considered a joint controller, and are still subject to EU data protection law, even if they do not have access to that data.
[8] C-479/22 P OC v Commission.

distribution to journalists, including investigative journalists with research skills beyond those of an average person. When assessing if the publication contained personal data, it was therefore necessary to perform the assessment from (at least) those investigative journalists' perspectives.

13 In other cases, several factors may be important for identifying the relevant perspectives. In particular, it is important to examine the specific nature and context of the processing, such as who has access to or control over the data, whether the data is being transmitted from one party to another, the relationship between those parties, and whether a receiver of the data might process it on behalf of another party.

14 The relevant perspective[9] may be affected by specific GDPR obligations – and may affect how a certain obligation is applied in practice, as was seen in *EDPS v SRB*.

Example 2: In *EDPS v SRB*, the CJEU examined the obligation to provide information when personal data is collected directly from the data subjects. This obligation required, among other things, that data subjects be informed about the recipients of any transfers of personal data. The SRB argued that, when deciding if this obligation was applicable, it was important to consider the receiving entity's perspective. Following the SRB's line of reasoning, if the information was anonymous from the receiving entity's perspective, then sharing the data with that entity would not qualify as a transfer of personal data *per se* and this obligation would not apply.

The CJEU, however, rejected this argument. Rather, the Court noted that this obligation had to be fulfilled at the time when the data is collected, and that the obligation formed an important part of the legal relationship between the data subject and the controller. The Court held that, when deciding if this obligation was applicable, the personal nature of the data should only be assessed from the controller's perspective at the time the data was collected. Since, at that time, the data was personal for the controller, any transfer of that information qualified as the transfer of personal data regardless of the receiving entity's perspective. As a result, the controller was obliged to inform the data subjects about the recipients of the data[10].

15 Importantly, the applicable perspective used for a particular entity depends on whether they themselves determine the means and purposes of the processing[11]. If an entity processes information on behalf of another, whether the given data is personal for that entity should be assessed by reference to the controlling entity's perspective. In practice, this means that if an entity processes information on behalf of a controller (*i.e.* someone for whom the information is personal data, and who decides the purposes and means of the processing), that information should also be considered personal data for the processing entity. That entity should, therefore, be considered a processor under the GDPR and be subject to the obligations applicable to processors[12]. This follows from a teleological application of the GDPR, noting that:

- EU data protection law relies on the concept of processor to prevent controllers from avoiding GDPR protections by outsourcing their processing activities to third parties,

[9] See the definition of the term in the Glossary.
[10] C-413/23 P *EDPS v SRB*, paras 102 – 103 & 108 – 111.
[11] For more guidance on the concept of processing data on behalf of another, see EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR, paras 79 – 84.
[12] This includes, in particular, compliance with Articles 27, 28, 29, 30(2), 31, 32, 33(2), 37, 38, 44, 46, 48, and 49(6) GDPR, and any orders made by supervisory authorities under Article 58 GDPR.

- the GDPR further reinforces this by placing certain obligations directly on processors, and
- both contractual agreements (or other applicable legal acts) and GDPR obligations need to work together to ensure that the processor handles the information in an appropriate manner.

Example 3: An e-commerce retailer transmits excerpts from its customer records to a third-party agency for analysis. In this case, the retailer determines the purpose and means of the processing activity, while the agency is a separate entity which processes the records on their behalf. It is therefore necessary to use the retailer's perspective for both parties when determining if the information is personal for them. In this case, the retailer is able to identify the customers from those records and the information should therefore be treated as personal data for both the retailer (as the controller) and the agency (as the processor).

Example 4: A hospital transmits excerpts from patient records, which in and by themselves do not allow identification of the individuals, to an independent institute for research purposes. The data will never be returned to the hospital, and the institute is free to determine means and purposes of processing and is not acting on the instructions from the hospital. Since it is acting independently, whether the information is personal for the research institute should be assessed from its own perspective.

2.3 Whether the information "relates" to a natural person

16 Information relates to a natural person where, by reason of its content, purpose or effect, it is linked to that person. The link should be assessed in light of all the circumstances of the specific case:[13]

a. **Content.** The information is about that person *per se*, *e.g.* it describes one or several of their characteristics or actions. For example, this could be an employee record, a patient's medical record, or a picture or video of a person.

b. **Purpose.** The information is either about that person, or is primarily about something which, or someone else who, is directly or indirectly related to that person due to some relationship or interaction, and such information is used or likely to be used for the purpose of evaluating, treating or influencing the status or behaviour of the individual in a particular way. This can include information about an object (like a phone, car or house owned by the person), an organisational entity (like a company, department, association or club where that person is a member), or an associate or relative of the natural person (like their friend or family member). For example, this could be data on the value of a house used for the purpose of determining the owner's taxes, or the service record of a car that could be used for the purpose of ascertaining the productivity of the responsible mechanic.

c. **Effect.** The information is primarily about an object (like a phone, car or house), an organisation (like a company, department, association or club), or another person, but

[13] C-434/16 *Nowak v Data Protection Commissioner*, para 35 and C-413/23 P *EDPS v SRB*, paras. 55 and 56. See also Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, adopted on 20 June 2007, p. 10, although that Opinion using the term "result" rather than "effect".

its processing is liable to have an effect on the data subject's rights and interests. For such a link to be present, there should be a concrete possibility that the individual can be treated differently from other persons as a result of processing of the information. For example, this could be location data of taxis that is collected for the purpose of optimising itineraries, but which also allows for the monitoring of the taxi drivers' performance and shows when they are taking a break.

17 Information may relate to a person even if it is not readily apparent that it does so, or if some additional processing is necessary to reveal that link. This is particularly true for aggregate data which, at first glance, only contains information about groups of people, not individuals. For example, many techniques exist that can reveal information relating to one or several single individuals from aggregate data. When such a technique could be used, the aggregate data may relate to that individual (and to any other individual for which such a technique could be used).

2.4 Whether the natural person is "identified or identifiable"

18 It is not enough for information to simply relate to a natural person. For this information to be personal data, that natural person must also be identified or identifiable using means reasonably likely to be used. This is especially important where information is anonymised for research or analytical purposes, where the goal is often to have an anonymous dataset which lets you learn something about the people in question (*i.e.* information which is still related to those individuals) but does so in a way that keeps their identities private (*i.e.* they are not identified or identifiable).

19 To identify a natural person means to distinguish them from others within a given context, consequently making it possible to treat them differently from those other people. This can be done using direct identifiers (such as a name, ID number or reference code), or one or more attributes which relate to that person.

20 An entity can identify an individual either directly or indirectly. An individual is "directly" identifiable by a particular entity if that entity can, through its own means, identify the individual based on the information itself. An individual is "indirectly" identifiable from a particular entity's perspective if, through means reasonably likely to be used, that entity could identify the individual by:

- obtaining additional information[14],
- applying methods (such as decryption of encrypted parts of the data) that are controlled by other parties, and/or
- putting the data at the disposal of other entities who themselves have means reasonably likely to be used to identify the individual[15].

Consequently, an individual can still be identified even if some of the necessary information – or some of the means of identification – are spread across different entities and require effort to obtain, provided that these elements can be combined through means reasonable likely to be used by the respective entities[16].

[14] C-604/22 IAB Europe v Gegevensbeschermingsautoriteit, para 39.
[15] C-319/22 Gesamtverband Autoteile-Handel eV v Scania CV AB, paras 45 – 49.
[16] Recital 26 GDPR. See also C-582/14 Breyer v Bundesrepublik Deutschland, para 43 and C-319/22 Gesamtverband Autoteile-Handel eV v Scania CV AB, paras 45 – 49.

21 In this regard, the EDPB emphasises that if (a) it is reasonably likely that an entity can transfer the data to a third party, and (b) it cannot be ruled out that this receiving entity has the means to identify the individual through means reasonably likely to be used, the data should be considered personal for both that transfer and for any subsequent processing of the data by that recipient[17]. Further, this transfer means that the transferred information would also indirectly become personal for the transferring entity, even if that entity would not otherwise have means reasonably likely to be used to identify the individuals in question[18].

22 Importantly, for information to be anonymous, the likelihood of the individual being successfully distinguished from others within the given context does not need to be zero; instead, that likelihood should be insignificant in reality[19]. This likelihood may vary dramatically from one perspective to another and the fact that an individual has been, or could be, identified from one perspective does not necessarily mean that the same is true from another.

2.4.1 Identifiers and other attributes

23 In many cases, identification is done by reference to an identifier that is unique within a given context. In the simplest cases, this could, for example, be a single piece of unique data, such as a name, IP address, social security number or e-mail address.

24 In other cases, identification may rely on one or more attributes which, when put together, are unique for a particular individual, even if each of the individual values would not in itself be unique. The type of attribute(s) involved could be extremely broad and may include, among other things, physical, genetic, psychological, behavioural, economic, or cultural attributes. For example, someone could be identified as "the one with the northern accent" or "the one who lives underground".

Example 5: When looking at live CCTV of people on a street, many people are wearing black suits, and many people are walking dogs. Within the context of that street, "the person in a black suit" is not, therefore, unique and that description does not allow for the identification of a particular individual. The same is true of "the person with a dog". However, only one individual on the street is both wearing a suit and walking a dog. Within the context of that street, the combination of those attributes is therefore unique to that individual and the description "the person in a black suit who is also walking a dog" allows for them to be identified.

25 Importantly, the identifiers or attributes do not necessarily have to be about the individual *per se* but can also be about other individuals, groups or objects which are otherwise linked to them. For example, a user's device may have a hardware identifier – such as a MAC address – that can identify the device and, in turn, the individual using the device. Equally, a website may "fingerprint" its visitors based on their unique combination of web browser, operating system, screen resolution, time zone *etc.*, and use this fingerprint to then identify them as they move from page to page and across different visits.

[17] C-413/23 P *EDPS v SRB*, para 85.
[18] C-413/23 P *EDPS v SRB*, para 84, referencing C-319/22 *Gesamtverband Autoteile-Handel eV v Scania CV AB*, paras 45 & 49.
[19] C-413/23 P *EDPS v SRB*, para 82.

Example 6: Several websites use the same third-party provider for targeted advertising. The websites do not share any names, IP addresses, or emails *etc*. with the third party; instead, they provide information about the pages visited together with a pseudonym which is based on fingerprints from the users' devices. The third party then uses these pseudonyms to link the data about users across the various sites, profile those users, and decide which adverts should be shown to them.

In this situation, both the raw fingerprint data (as a combination of attributes) and the pseudonym which is derived from that data (as an identifier) allow for the identification of the individual. This is true for both the websites and the third party: even though it does not interact with them directly, the third-party can (and, indeed, does) still use the pseudonyms to distinguish the individual and treat them differently from others, and those individuals are therefore identified or identifiable from its perspective.

2.4.2 Means reasonably likely to be used

26 Rather than describing which means are reasonably likely to be used *per se*, the CJEU has defined this concept by reference to things which would stop means from being reasonably likely to be used by a particular entity. In particular, means may not be reasonably likely to be used if the likelihood of identification appears, in reality, to be insignificant because it would be impossible to do; because it would involve disproportionate effort in terms of time, cost and labour; or because of a legal prohibition[20]. This sub-section provides further guidance on this concept and on how to assess whether a means can be considered reasonably likely to be used or not.

27 As a starting point, the "means reasonably likely to be used" test only relates to possible, future identification (*i.e.* is the individual "identifiable"?). If the individual to whom the information relates has already been identified, then the information is not anonymous.

28 The concept of "means" should be understood broadly, covering anything from simply reading a document to making a web search to running incredibly complex analyses and inferences. Further, an entity's means can also include taking advantage of means that are available to another entity; where this is possible, it is important to assess the likelihood that this overall "chain" of means[21] comes together.

Example 7: A controller wants to share certain information with its business partner. They try to anonymise data but, when testing the strength of the anonymisation, quickly realise that the data subjects could be re-identified through a certain re-identification technique. The controller knows that their business partner would not itself be able to perform this re-identification technique. However, they could easily hire a third party to do it. That technique therefore still qualifies as a means reasonably likely to be used by the business partner.

29 Whether an entity is likely to use certain means to identify an individual should be assessed in light of all objective factors that might influence the availability of such means and the likelihood of their actual use, including:

[20] C-413/23 P *EDPS v SRB*, paras 82 – 83.

[21] The chain of means includes any transfer of the given data to (or access by) an entity that is involved in the respective identification process. The likelihood of such transfers (or access) needs to be considered starting from the reception of the given data by an entity for whom it is supposed to be anonymous.

a. the properties of the data itself – including whether the data is aggregated, the uniqueness of the records, as well as the accuracy and precision of the information – and how these affect the data's vulnerability to the different possible means of (re-) identification;
b. the context in which the data are released and/or processed[22], including whether there are any effective and permanent restrictions on access to the data;
c. any additional information that would allow for the identification of individuals, and how likely it is that such information can be obtained;
d. the costs and the amount of time that the entity would need to obtain such additional information (in case it is not already available to them); and
e. the available technology at the time of the processing, as well as reasonably foreseeable technological developments.

Notably, if the law provides means to access additional information, it is objectively clear that such access constitutes a means reasonably likely to be used[23].

30 The GDPR does not lay down any conditions regarding the entities who are able to identify the individual[24]. Relevant entities may, therefore, include anybody directly or indirectly receiving the data, including unauthorised and malicious actors. Depending on the circumstances at hand, these may include (among others)[25]:

- The controller of the data which is intended to be anonymised;
- Any receiving entities;
- Rogue employees;
- The individual's partners, friends, colleagues, neighbours, *etc.* These entities will often have limited access to the given data (especially if appropriate security measures are deployed), but may have access to additional information which can then be used to (re-)identify the individual(s) in the given data;
- Investigative journalists. This is particularly relevant for data relating to individuals who are, or who are expected to soon be, in the public eye;
- Domestic law enforcement or intelligence agencies. These bodies often have extensive legal powers to investigate and gather data, as well as large access to financial means, high computational power and technical experts;
- Foreign intelligence agencies. The same considerations that apply to domestic law enforcement agencies also apply to foreign intelligence agencies. Further, foreign intelligence agencies may not be limited by the same laws which limit the actions of domestic law enforcement agencies;
- Unethical companies. These entities may have access to large datasets and may exploit legal loopholes or grey areas to gain access to data that would otherwise be inaccessible; and
- Cybercriminals. These entities may simply break the law to gain access to data not available from a law-abiding entities' perspective.

Different entities may be relevant in different capacities. For example, an individual's friends and family might be able to supply additional information that could aid re-identification, but

[22] See para 148 of Binding decision 1/2021: "[T]he EDPB stresses that the whole context of the processing needs to be considered, as 'all objective factors' affect 'whether means are reasonably likely to be used to identify the natural person'".
[23] This includes, *e.g.*, the ability for a private entity to obtain additional information through legal channels, as referenced in C-582/14 *Breyer v Bundesrepublik Deutschland*, para 47 & 49.
[24] C-479/22 P *OC v European Commission*, para 56.
[25] Not all of these perspectives would need to be considered systematically in each case but this depends on the circumstances of each specific case and the entities directly or indirectly receiving the data.

unable to access the given data itself or actually run the necessary re-identification technique. These different capacities can, therefore, affect the means that each entity might have on their own, how each entity might slot into a possible chain of means, and whether those means (or chains of means) are actually reasonably likely to be used.

31 The EDPB cautions against using "lack of motivation" as a factor in this analysis. As a starting point, the motivation of individuals can be difficult to assess or demonstrate objectively, and may change over time. Moreover, an entity's actions may not be consistent with their motivations. It may be, for example, that identification occurs due to an accident or negligence. This could also happen if an entity is forced by outside circumstances to identify an individual, or to transfer information which allows for such identification[26]. When assessing the means reasonably likely to be used, the potential value of the re-identified data to the re-identifying entity (which is not limited to monetary value)[27] should be taken into account to the extent that it offsets the costs and effort involved in re-identification. Equally, the risks that re-identification entails for the re-identifying entity should be considered in the context of this assessment.

32 The CJEU has stated that means do not need to be considered if, in reality, the likelihood of their use is insignificant because they are prohibited by law[28]. It should be noted that these cases involved actors who were properly limited by the rule of law, so that the law effectively prevented the use of the means in question. Indeed, we can generally assume that people will follow the law. However, there may also be cases where illegality does not provide any meaningful barrier to identification, because the entities involved will not actually be restricted by this prohibition.

33 The general assumption that people will follow the law can, therefore, be rebutted if there is sufficient evidence of a concrete risk that unlawful means are nevertheless reasonably likely to be used. The assessment of this risk should consider, in particular, evidence that the prohibition in question is not respected by the entities involved. Relevant factors may include, but are not limited to:

- evidence that the legal prohibition is not effectively monitored, enforced and sufficiently dissuasive;
- evidence that the gains from breaking the legal prohibition might outweigh costs, effort and potential risks of doing so;
- evidence of a situation where data relevant for re-identification (*i.e.* both the given and, potentially, the additional data useful for re-identification) is especially vulnerable to illegal access using the prohibited means, due to a lack of protection; and
- evidence that the prohibition has previously been breached or circumvented in comparable situations.

Example 8: A media company wants to share certain information with a local public authority, provided that they can anonymise it before doing so. The media company is the only entity with access to this information, and they are confident that it is not available through any other sources. The media company, however, does not want to delete the original sources which was used to generate that information, and wants to use an anonymisation technique that

[26] See, *e.g.*, the ability for a private entity to obtain additional information through legal channels, as referenced in C-582/14 *Breyer v Bundesrepublik Deutschland*, para 47.
[27] The potential value of personal data lies in the benefit that an entity may gain from re-identification and should be distinguished from the sensitivity of the personal data. For example, information about a person's hobby may be non-sensitive but economically valuable for targeted marketing purposes.
[28] See, e.g., C-582/14 *Breyer v Bundesrepublik Deutschland*, para 46; C-479/22 P *OC v European Commission*, para 51; and C-413/23 P *EDPS v SRB*, para 82.

prevents identification in practice unless the (otherwise-anonymous) data is combined with privileged information that only the media company possesses. In its assessment, the media company notes that that privileged information is stored on their servers, and that, while their network security is starting to be out of date, third-party access to those servers is prohibited by law. Since the public authority is properly limited by the rule of law, the media company concludes that the risk of identification is insignificant from the public authority's perspective and that the information is anonymous for that authority.

However, the media company also recognises that there is a risk of the shared data being unlawfully intercepted by unintended third parties. These third parties include entities located outside the scope of applicable Union and Member State law, not effectively bound by such laws prohibiting unauthorised access. Moreover, there have been several incidents of illegal access affecting this and other media company's IT systems storing personal data.

Since the media company's out-of-date security measures mean that those third parties could also likely hack into the media company's servers, these unintended receivers of the data would be able to access the privileged information and therefore re-identify the individuals through means reasonably likely to be used. The shared information would not, therefore, be considered anonymous from their perspective. The media company therefore decides that it will upgrade its network security before re-testing the anonymity of the data.

34 A prohibition set out in a contract (even if it is legally binding upon its parties) should not be treated as a prohibition by law. While contractual terms may have an effect on the means reasonably likely to be used, they should only be used to complement technical measures and it is important to carefully consider the actual strength of their effect. Among other things, the measures should be reliable, verifiable and enforceable[29], bearing in mind that contractual measures can typically be subject to revision, or may even be disregarded entirely by the parties.

35 Finally, the likelihood of re-identification typically increases over time due to advances in the technology and techniques used for re-identification, as well as the increased availability of additional information. If the likelihood of identification increases to a level that is no longer insignificant, the previously anonymous data should again be considered personal and the controller will be accountable for any processing of that personal data[30]. It is, therefore, good practice for entities processing anonymised data, wherever possible and as appropriate, to periodically reassess the likelihood of re-identification.

2.5 Additional considerations

2.5.1 Datasets containing a mix of anonymous and personal data

36 Some anonymisation techniques might only make the likelihood of identification insignificant for some individuals in a dataset. In such cases, it is important to distinguish between the anonymity of particular records and the anonymity of the dataset *per se*. In particular, the dataset as a whole should only be considered anonymous if the anonymisation is effective for all of the included individuals – this does not necessarily require an equal level of protection for all individuals, but does require the likelihood of re-identification to be insignificant for all of

[29] European Commission and EDPB, Joint Guidelines on the Interplay between the Digital Markets Act and the General Data Protection Regulation (version for public consultation), para 181.
[30] For general conditions on the liability of controllers for infringing behaviour, see C-807/21 *Deutsche Wohnen SE v Staatsanwaltschaft Berlin*, paras 76 & 78.

these individuals. Where this is not the case, and so a dataset contains a mix of personal and anonymous data, the entire dataset should be treated as containing personal data (and therefore within the scope of the GDPR) in any situation where the separate parts of the dataset are not treated separately[31].

2.5.2 GDPR compliance and the anonymisation process

37 The EDPB emphasises that when making information anonymous from some perspectives but not others, it is important to consider the possible implications for GDPR compliance. For example, when information is anonymised for the entity which receives the given data, but not for the anonymising controller, that controller will still have to treat the given data as personal data and comply with all of their GDPR obligations in regard to it[32].

38 The principles of data protection apply whenever personal data is processed, which includes processing operations carried out to obtain anonymous information. This means, amongst other things, that anonymisation must have a legal basis under Article 6 GDPR, and that one of the exemptions under Article 9(2) GDPR should apply if the dataset contains special category personal data. Legal basis and, when applicable, the relevant Article 9(2) exemption should be presumed to coincide with the legal basis and exemption of the processing preceding anonymisation if the anonymisation is part of the same processing activity, and that the controller pursues the same purposes with it[33].

39 Anonymisation (at least when combined with the deletion of the original data) may also be helpful for controllers who wish to rely on legitimate interest for any preceding processing activities. In particular, the immediate anonymisation of the given data, and immediate deletion of the original data, can limit the overall impact of the processing activities for data subjects, and so act as a positive factor in the legitimate interest balancing test[34].

40 Controllers should also comply with their transparency obligations under Articles 5(1)(a) and 12 – 15 GDPR regarding the anonymisation processing. Among other things, controllers should clearly state that personal data will be processed to produce anonymous data which then falls outside of the scope of the GDPR, and should avoid any ambiguous or misleading statements. In particular, controllers should not use descriptions like “anonymous”, “de-identified” or “de-personalised” if individuals are actually still identifiable.

41 Equally, controllers should ensure adequate documentation of the anonymisation processing. This documentation of the anonymisation process (including for the testing of the supposedly anonymous datasets) makes it possible to demonstrate both the GDPR-compliance of the anonymisation process, and the effectiveness of the anonymisation itself. This documentation should then be retained after the completion of the anonymisation process.

42 In some cases, a security incident may lead to a reassessment of anonymity. This is particularly the case if the assessment of anonymity relied on certain information being kept confidential, but the incident means that the individuals can now be identified with means

---

[31] See, by comparison, the fact that a dataset containing a mix of special category and non-special category personal data must be collectively treated as special category personal data whenever the different parts cannot be processed separately: C-252/21 *Meta v Bundeskartellamt*, para 89.
[32] See, e.g., C-413/23 P *EDPS v SRB* and the effects of anonymisation (or lack thereof) on certain aspects of the information obligation as discussed in Example 2.
[33] See also EDPB Guidelines 1/2026 on processing of personal data for scientific research purposes, adopted on 15 April 2026, version for public consultation, section 3.1.
[34] For more information on the legitimate interest balancing test, see EDPB Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR.

reasonably likely to be used. Where this is the case, the entity which suffered the incident should then determine its role with respect to the processing and whether it needs to fulfil any obligations under Article 33 and 34 GDPR[35].

3 The technical analysis of anonymisation

3.1 Introduction

43 It will not always be immediately apparent whether information relates to an identified or identifiable individual. This is especially the case when the anonymising controller has deliberately tried to make it difficult to recognise that the information relates to an identified or identifiable individual. It is therefore necessary to apply a technical analysis that considers the state of the art of re-identification techniques and can be used to draw out whether or not the data is anonymous.

44 Given the broad scope of possible datasets and anonymisation techniques, as well as future developments in (re-)identification techniques, assessments will often require a case-by-case analysis and may require considerations not fully covered by this document. Nevertheless, this section will provide guidance on a framework which can be used to test the anonymity of data.

45 The framework in this section can be used in two ways. The first is a contextualised approach, which is based on the legal analysis set out above and uses information about all relevant entities' means and the likelihood of their use. The second is through a simplified approach, which disregards differences between entities regarding the means they are reasonably likely to use. This simplified approach may lead the anonymising controller to treat the data as though it is not anonymous even if it would actually be so for the relevant entities. However, it can provide greater confidence and can be complemented with the contextualised approach to refine the findings.

46 Having set out these two approaches in detail, this section quickly discusses the simple cases, where the given data is clearly personal and use of the full framework is unnecessary. It then moves on to the main part of the framework itself, which is made up of three broad criteria: No Record Isolation, No Linkage and No Inference. If all three of these criteria are met, the information may be regarded as anonymous; if one or more is violated, further analysis may be required. The section finishes by providing guidance on how the framework can be applied and assessed, and what this further analysis involves.

3.2 Approaches to the assessment

47 The legal standard for anonymisation can produce different results depending on the applicable perspectives. This contextual approach leads to a differentiated assessment, where the respective capabilities of the relevant entities should be individually assessed. In practice, this can be quite complex, especially where it is difficult to account for (or even know about) all the possible contextual elements involved. This contextual approach therefore runs the risk of false positives, where the assessor may incorrectly conclude that data is anonymous because they are unaware of a particular entity's means reasonably likely to be used for

[35] For further guidance on this topic, see EDPB Guidelines 9/2022 on personal data breach notification under GDPR.

identification. However, if done properly, this approach allows for a comprehensive analysis of the dataset and allows controllers to be confident in whether the dataset falls within the scope of the GDPR.

48 For the sake of simplicity, then, some anonymising controllers may decide to use a simplified approach for the assessment. Under this approach, an anonymising controller can simply choose to ignore differences between entities regarding access to the given data, to additional information or to other resources that can contribute to re-identification. It must be emphasised that this simplified approach is not an alternative legal standard[36]. Instead, this approach is a way to voluntarily shift the risk from false positives to false negatives, where the anonymising controller treats anonymous data as personal because they have (in effect) overestimated the likelihood that certain means will be used. In some cases, this approach would provide a more cautious level of protection than is strictly necessary, but in doing so it ensures that the legal standards set by the GDPR will always be met in cases of doubt.

49 In practice, a combination of the two approaches may often be helpful. For example, an anonymising controller might wish to begin with the simplified approach, asking if re-identification is even possible in theory. If this is not possible, the data can be safely considered anonymous. If, however, the simplified approach identifies the existence of certain re-identification methods, the anonymising controller can then shift to the contextual approach and consider whether those methods are reasonably likely to be used from the relevant entities' perspectives. Alternatively, the anonymising controller might perform an initial mapping of the given data and immediately realise that there are a huge range of re-identification methods that are possible in theory, but which can quickly be excluded as impossible to be applied to the given data in practice. That anonymising controller may, therefore, decide to go straight to the contextual approach, which allows these methods to be immediately discounted because, in reality, they are not means reasonably likely to be used by any entities.

Example 9: The controller wishes to anonymise a dataset containing a combination of demographic and medical information. They begin with the simplified analysis and discover that the demographic data alone could not allow for re-identification of the individuals in this case. However, they also discover that re-identification could be possible if the demographic information is combined with certain additional medical information. Under the simplified analysis, they would therefore conclude that the information is not anonymous. However, the controller decides to extend the analysis and pivots to a contextual analysis. They therefore begin identifying the applicable perspectives and testing if the respective entities could access that additional medical information with means reasonably likely to be used.

3.3 Simple cases and data mapping

50 In some cases, it may be immediately apparent that information is not anonymous. This includes, for example, a dataset which includes an individual's name, email address, phone number or social security number; a dataset which contains pseudonyms that can easily be reverse engineered to discover an identifier, or which itself can be used to identify the individuals; a dataset which lacks direct identifiers but which clearly contains attributes that allow for identification of the individuals; or a dataset which is clearly being used for the

[36] This means, in particular, that while the simplified approach may be quite useful for anonymising controllers to prove that the information is anonymous and falls outside of the scope of the GDPR, it cannot be used by third parties to definitively prove that the information is personal data and that the controller's processing of that data is therefore in violation of the GDPR.

purposes of evaluating, treating or influencing the status or behaviour of a specific individual in a particular way.

51 To help catch these, and other simple cases, it is recommended to map out the data, as well as any additional information that may aid (re-)identification that is either held by other relevant entities or is accessible to them through means reasonably likely to be used. If the data is clearly not anonymous at this stage, it is unnecessary to apply the criteria set out below. This should be done as an early step in the assessment and, even if no such simple identification becomes apparent, can also help to streamline the next steps.

3.4 The three criteria

52 This sub-section sets out the three criteria: No Record Isolation, No Linkage and No Inference, with further guidance on their application then contained in sub-section 3.5. If all three of the criteria are met, the information may be regarded as anonymous. On the other hand, violating a criterion does not necessarily mean that the information must necessarily be considered as personal data; rather, if one or more of the criteria is violated, it is then necessary to continue the analysis, as set out in sub-section 3.5.3, to assess the impact of that violation and whether the data could still be considered anonymous.

53 The same criteria are used for both the contextual and simplified approaches. For the sake of explanation, the descriptions of these criteria in sub-sections 3.4.1–3.4.3 use the simplified approach, which allows this sub-section to focus on the methods themselves. The contextual matters will then be reintroduced sub-section 3.5.

3.4.1 No Record Isolation

54 A common method used to distinguish individuals from others in a given group is by singling them out by reference to one or more attributes[37]. By its very nature, it is much easier to do this when a record – *i.e.* all of the values within the given data that relate to one single individual – is unique. As such, record isolation can then lead to the violation of anonymity for the given data. This leads to the No Record Isolation criterion.

55 The No Record Isolation criterion is met if the data does not contain a unique combination of attribute values that relate to a single individual.

56 The larger a record is, and the more attributes that it contains, the higher the likelihood that the record will be unique within the given data.

57 The No Record Isolation criterion can generally be tested by reference to the given data on its own. Since a record constitutes all information which is known to relate to a single individual, the No Record Isolation criterion may involve looking at, for example, multiple lines in a table or database table, depending on how the given data is structured.

[37] The concept of singling out will be explained in detail in section 3.5.3.

Example 10: A research institute holds a dataset containing individual records of patients. Each record contains information about the individual's sex, date of birth, postcode, and the autoimmune disease affecting the patient. The data was verified before inclusion in the dataset and has not been modified.

Sex	Date of birth	Postcode	Disease
Female	28-09-1955	43221	celiac disease
Female	06-01-1955	43210	Crohn's disease
Male	08-07-1959	89127	Vitiligo
Male	10-06-1959	89127	multiple sclerosis
Male	27-09-1959	89127	Diabetes mellitus type 1

In this case, all of the records are unique. The No Record Isolation criterion is violated because of the unique combinations of the attributes.

58 Notably, aggregate data will usually satisfy the No Record Isolation criterion – if the aggregation was done correctly, it should result in information that relates to groups of individuals and, therefore, should not actually contain any individual records which could be isolated.

3.4.2 No Linkage

59 Another common method of identifying an individual is by linking information from one dataset to another, or multiple other, dataset(s). By establishing that a record in Dataset A relates to the same individual as a record in Dataset B (and, if relevant, Datasets C, D and E *etc.*), the linkage increases the amount of information about that individual, which in turn makes it more likely that the individual can then actually be identified. Linkages can then lead to the violation of anonymity for the given data. This leads to the No Linkage criterion.

60 The No Linkage criterion is met if the data does not contain an individual's record which could be linked to another record which (a) also relates (with certainty or high likelihood) to that same individual, and (b) comes from a different dataset.

61 Unlike the No Record Isolation criterion, which was assessed by reference to the given data alone, this criterion requires knowledge and understanding of other datasets that could be linked to the given data with means reasonably likely to be used[38].

62 Linkage may often be possible by reference to a common identifier (e.g., an internal customer number) which is used in both datasets, or by matching records based on certain combinations of attributes.

Example 11: A board game shop keeps a record of every purchase made by every customer. This is clearly personal data, since each purchase is linked to a specific user. The shop wishes

[38] As noted in para. 53 above, this section will explain the criteria following the simplified approach, to focus the explanation on the core parts of the criterion. It will not, therefore, consider whether the relevant entities could access those other datasets with means reasonably likely to be used. Under the contextual approach, however, it would also be necessary to ask whether such links are actually possible with means reasonably likely to be used from the relevant perspective, as explained in sub-sections 2.2 and 2.4 and discussed in sub-sections 3.5.2 and 3.5.3 below. This is true for all three of the criteria, but can be especially helpful to emphasise for the sake of the examples set out in this sub-section.

to anonymise the data and so decides to test if simply erasing the direct identifiers would be enough to do so.

The shop immediately notices that many of the records in the given data are unique, and that the given data therefore fails the No Record Isolation criterion. However, since failing a criterion does not necessarily mean that the given data is personal, the shop continues to test the other criteria before starting its assessment of the results.

To test the No Linkage criterion, the shop checks for other datasets which might contain matching records based on certain combinations of attributes (in this case, the specific combinations of purchased board games). This leads to a very popular website where users create lists of the games which they own and mark how often they play each of those games. The shop finds that several records on the website match those in the given data, where customers have purchased games from the shop and then entered that information onto the website. These matching records mean that the datasets can be linked, and the shop therefore concludes that the given data also violates the No Linkage criterion.

63 In practice, data which satisfies the No Record Isolation criterion will often also satisfy the No Linkage criterion. However, the two criteria should still be assessed separately, as there are ways to link non-unique data.

Example 12: A research institute processes a dataset containing individual records of some patients. In the dataset, the date of birth and postcode attributes have been generalised as follows:

Sex	Date of birth	Postcode	Disease
Male	**-**-1946	54***	sclerosis
Male	**-**-1946	54***	sclerosis
Female	**-**-1951	32***	lung cancer
Female	**-**-1951	32***	lung cancer
Female	**-**-1951	32***	lung cancer

The data satisfies the No Record Isolation criterion since no record is unique in the dataset. However, looking at the first two records, this dataset reveals that all individuals born in 1946 and living in any postcode starting with 54 (and whose record was included in the dataset) have sclerosis.

Therefore, the data could be linked to information in other datasets which contain the same individual's sex, year of birth and generalised postcode. The fact that the datasets contain information about the same individual could, for example, be established through prior knowledge that two datasets contained information about the same individual, from information contained in either dataset which makes this link clear, or though information in a third dataset which acts as a bridging link. If this link can be established, the No Linkage criterion is violated.

64 It may still be possible to achieve linkage on records that are not fully accurate (including where records are intentionally modified, *e.g.*, by adding noise to the data). This can often be done to a high degree of accuracy and with a likelihood of success by, for example, using

robust record linkage attacks[39]. It is therefore typically difficult to determine whether inaccuracies are sufficiently large to prevent linkage, since noisy data may leak information in unexpected ways. In addition, the effectiveness of the attack depends heavily on the availability of additional information.

65 The No Linkage criterion is likely to be met if the information has not been recorded elsewhere and is not correlated to similar information about the same individual recorded in other contexts. This criterion may be met, for example, in the case of answers given to an opinion survey, provided that they do not contain, among other things, demographic data which might be traced back to an individual. Another example of this could be records of patients' particular physiological states, provided that, among other things, similar attribute values are collected for many patients and that the said state is likely to vary over time.

3.4.3 No Inference

66 Generally speaking, "inference" is the process of reaching a conclusion by deduction from evidence and reasoning[40]. Using this process, a relevant entity can effectively obtain information from the given data. Inferences can be drawn from both record-level and aggregate data. Inferences can also "add to" the data in a way which then reveals it as personal, even if the inferred information was not obviously contained in the dataset before the inference was made[41]. Such inferences can violate the anonymity of the given data. This leads to the No Inference criterion.

67 The No Inference criterion is met if no specific and meaningful inference can be drawn from the given data. Inferences may be based on the given data and/or on additional information, including information about the anonymisation process. In particular:

- An inference is specific if the inferred information relates to a single identified or identifiable individual (*i.e.* it is personal data); and
- An inference is both specific and meaningful if the processing of the inferred information is also liable to have an effect on the data subject's rights and interests, relies on the given data and could not be obtained from general knowledge or from data about the population at large[42].

68 This criterion should be assessed by reference to the given data, together with any external information that might be used in conjunction with the given data to infer information – including the context in which the given data was produced, and any external datasets which might be brought together with it.

69 It may often be possible to infer some personal data from given data in some way, even if the given data itself is not personal. In such cases, the inferred personal data would certainly

---

[39] See, *e.g.*, Arvind Narayanan & Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets" (2008) 2008 IEEE Symposium on Security and Privacy 111, https://doi.org/10.1109/SP.2008.33; and Julien Freudiger, Reza Shokri & Jean-Pierre Hubaux, "Evaluating the privacy risk of location-based services" in George Danezis (ed.) *Financial Cryptography and Data Security: 15th International Conference, FC 2011, Gros Islet, St Lucia, February 28 – March 4, 2011, Revised Selected Papers* (Springer 2012), https://doi.org/10.1007/978-3-642-27576-0_3.

[40] See the entries "infer" and "inference" in Catherine Soanes & Angus Stevenson (eds.) *Concise Oxford English Dictionary* (11th edition Revised, OUP 2009).

[41] In this regard, it is important to recall that, as discussed in sub-sections 2.3 and 2.4, information may be considered personal data, even if its link to an individual is not immediately apparent or if the information does not itself contain everything necessary to identify the individual to whom it relates.

[42] Here, "information about the population at large" means any large group of people about whom information is not collected individually, but extrapolated from a sample, with any connection to the individuals left behind. In technical terms, such inferences can be seen as part of the "prior belief".

relate to an identified or identifiable individual, but this does not then mean that the same is true for the given data. For example, when dealing with record-level data, it may be possible to infer information about an entirely new individual who was not included in the original dataset, but who has similar attributes to people that were. In such a case, the inferred information certainly relates to this new individual, but this does not itself mean that the given data does so.

70 It is, therefore, important to distinguish between two situations. In the first situation, personal data is inferred from the given data, but is not actually connected to the original dataset (*i.e.*, the information that was used to generate the given data). This kind of inference would not lead to a violation of the No Inference criterion because the inference is not meaningful: it is obtained from information contained in the given data that is general knowledge or is about the population at large – a result of a generalisation from the original dataset represented by the given data. In the second situation, the inferred personal data is actually connected to the original dataset and is reasonably likely to have an effect on the rights and interests of the data subject. The inference, then, is meaningful because the connection to the original dataset shows that it is not obtained from general knowledge or from generalised data about the population at large, but from information about the data subject of the inferred information represented by the given data and coming from the original dataset. This kind of inference would lead to a violation of the No Inference criterion.

71 To consider that the given data actually relates to an individual, the inference should be specific and meaningful. This means that the inference will result in personal data whose processing is liable to have an effect on the data subject's rights and interests, relies on the given data and could not be obtained from general knowledge or from data about the population at large. This can typically be shown in one of three ways:

a. Ideally, this can be done by tracing the way in which the inference is drawn.

b. If this is not possible, *e.g.* due to the opacity of automated processing that led to the inferred information, a lack of reliance could also be shown by comparing the outcome of inference with two versions of the given data – one which was produced from the original dataset, and one which was produced from a modified version of the original dataset and which excludes the record(s) relating to the individual. If the inference can only be drawn from the first version of the given data, it is clear that the inference relies on the records relating to the individual, and consequently that the given data relates to the individual[43].

c. If neither of these can be done, *e.g.* because the information has been aggregated and the original dataset has been deleted, it is also possible to rely on a rebuttal presumption that an inference did not rely on the given data by demonstrating that it

[43] In more technical terms, the argument presented in this paragraph is as follows: Inferring means making a prediction on a property S having observed the given data A(D) and additional information, where A(D) was produced by running the anonymisation process A to the original data D and the additional information serves as a conditioning factor. Suppose that S is a property about a person named John Smith and that data relating to that person is included in D. We denote by D' a copy of the dataset D, with the only difference that all data relating to John Smith have been removed. Therefore, D' does not contain any data relating to John Smith. It may still be possible to make some inference on John Smith based on A(D'), including on property S. For example, this may be possible if the other records in D' relate to persons that are "similar" to John Smith (e.g. because they live in the same city). However, this inference cannot depend on John Smith's data, as there is no data relating to John Smith in D'. Therefore, if the inference on John Smith works in the same way for A(D) and A(D'), it is not an inference drawn from his data but from data relating to the population at large, for which D' is representative.

could have been drawn from general knowledge or from information about the population at large[44].

Example 13: The unqualified statement "Green is a popular colour" is anonymous data since it does not relate to any identified or identifiable natural person. We could infer from this information that Connor likes the colour green. The inferred statement "Connor likes green" is clearly information related to an identified natural person and should therefore be treated as personal data. However, this inference does not mean that "Green is a popular colour" is also personal data.

While the inferred information "Connor likes green" is specific (since it relates to a single identified individual), it is not meaningful because it can (and, indeed, was) obtained from information about the population at large. This inference does not, therefore, lead to a violation of the No Inference criterion for the given data "Green is a popular colour".

Example 14: An anonymised record-level dataset of a bank's historical loan information shows that individuals with a particular combination of attributes are likely to fail to repay the loan. The bank uses this insight to a new loan applicant, who was never part of the dataset, and infers that this applicant has a high risk not to honour the loan obligations. This, however, is not a meaningful inference in the sense defined above, and hence the dataset could still satisfy the No Inference criterion.

72 Importantly, inferences which, on first review, do not appear to be specific and meaningful may still be helpful for (re-) identification and should be further analysed in conjunction with the given data. It may, for example, be that the given data, together with the inferred information, can lead to further inferences that would actually be specific and meaningful, and therefore violate the No Inference criterion.

73 There is a wide spectrum of processes that can lead to relevant inferences. On one end are conclusions reached by simply thinking logically about the data, on the other are those reached by elaborate analyses or the use of automated processing which follows sophisticated algorithms. Some forms of inference may be easier, or more relevant, for record-level data, while others may be particularly good at extracting information from aggregate data. The following sub-sections will discuss and provide examples for some of these processes. However, any given inference may aid re-identification in more than one way and the below should not be taken as an exhaustive list.

Adding an inferred attribute to a given record

74 It may be possible to add inferred information to record-level data in a way which then leads to a violation of the No Inference criterion. This inferred information does not have to be appended to the record– it is enough that it is possible to attach it to the record.

Example 15: A controller receives precise location data about users of a mobile app, removes all direct identifiers and collates them into 24-hour tracks for each user. The controller would like to assess whether those tracks constitute anonymous data. The tracks show that,

[44] In such a case, "Prior belief" (see footnote 40) could include knowledge coming from an alternative version of the given data which has been produced without the record of the persons that are being identified – A(D') in footnote 41 – but which still applies to them and other persons by way of extrapolation. Inferences from the given data should yield a significantly different "posterior belief" (on which an action could turn) to be considered related to the individual in a way that classifies the given data as personal.

every weekday, a single user is present in a particular residential building from 0:00 to 08:00 and 20:00 to 24:00, and at a particular commercial building from 09:00 to 17:00. This information allows a meaningful and specific inference of the user's home and work addresses, which could then be added to that user's record.

75 Simply because you might be able to add inferred information to a record does not necessarily mean that the No Inference criterion is violated. It will almost always be possible to find inferred information that can be added, either by analysing the data or simply by applying general facts to it. It is, therefore, important to consider the nature of the inferred information and whether it is meaningful and specific, as set out above.

Producing record-level data from aggregate data

76 It may be possible to infer record-level data by analysing and/or combining aggregate data. This is particularly true for outlier data, where this kind of inference is often more likely.

Example 16: A company decides to publish information about the total salaries that it pays to each category of employees. The information states that engineers in the company receive an aggregate of €550,000 a year but does not specify how much is paid to each engineer. However, the company's internal reporting also states that the company employs a total of six engineers, that the Product Engineering team is made up of five engineers (and does not contain any other employees) and that the Product Engineering team has an annual salary expense of €480,000. Although both reports are limited to aggregate data, an entity with all of this information can infer that the sixth engineer is paid €70,000 a year.

77 This form of inference can also be significantly more sophisticated. It may, for example, start from large sets of aggregate statistics accessible in bulk, or from data query systems. Often, this kind of inference will use multiple aggregated datasets, especially if those sets were all produced from the same original data. This is particularly true for advanced de-aggregation techniques, including attacks which can extract data from supposedly-anonymous AI models[45]. Nevertheless, a single dataset which contains a sufficiently large quantity of data may allow for the inference of information relating to a single individual, without the need for other sources.

78 Inferences can also be drawn from the absence of a value.

Example 17: A company organises a survey where each team can "anonymously" provide their views on the line manager. The survey includes questions like: "From 1 to 5, to what extent does the company culture encourage you to make and learn from mistakes?" The individual responses are not shared with the line managers, who are only told the total number of their team members that selected each answer. However, if a line manager is told that all of their team members took part in in the survey and none of the answers gave a score higher than 3, that manager can infer that all team members have a critical view of the company culture. Although it is aggregated, this information still relates to each member of the team individually.

[45] See also EDPB: Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models.

Membership inference

79 It may be possible to infer that an individual's data was included in the original dataset which was then used to produce the given data. This type of membership inference may be the first step for other types of inferences or may itself reveal information about the individual.

Example 18: A supposedly anonymous dataset is generated from information about members of a rowing club. However, by analysing certain attributes contained in the given data, it is possible to infer that Alex was included in the original dataset and so is a member of that club.

Inferred linkage to additional information

80 Linkage with other data, such as that discussed in section 3.4.2, can also be seen as a form of inference. By linking the given data to additional information, it is possible to infer that information contained in the given data can be applied to the additional information and *vice versa*. This kind of inference can also go beyond the type of linkage described in the No Linkage criterion. For example, if information cannot be matched on a 1:1 basis between datasets, but the additional information could be matched to several possible entries in the given data, it may still be possible to draw relevant inferences. This is particularly the case where one attribute value occurs in an overwhelming number of matching records.

81 Further, while the No Linkage criterion referred to record-level data in particular, this kind of inference also applies to aggregate data.

Example 19: The data from Example 12 is aggregated to create a histogram of diagnoses grouped by combination of sex, year of birth and abbreviated postcode. Despite the aggregation, it is still possible to infer an individual's diagnosis by linking this histogram to (a) data about the individual's sex, year of birth, and generalised postcode; and (b) data that the individual's record was included in the dataset.

82 This kind of inference could also be drawn from data that has been aggregated to represent correlations between elements of the original information, rather than that information itself. This would, in particular, be the case for AI models or synthetic data. Specific inferences can be made by querying (or, in the case of AI models, prompting) the given data with additional information to elicit new information about a particular individual. Where such an inference is also meaningful, this would be a violation of the No Inference criterion.

3.5 Applying the criteria

3.5.1 Assessing the effectiveness of (re-)identification techniques

83 The three criteria set out above should be used to evaluate the effectiveness of (re-) identification techniques that can be deployed against the given data. These techniques range from the very simple, like using a basic search in other datasets for matching attributes, to the very complex, like using special-purpose AI agents to deploy a sophisticated combination of probabilistic techniques. Depending on the technique, these could be used intentionally or unintentionally, and may be tailored to exploit potential vulnerabilities that arise from specific contexts or specific flaws in the applied anonymisation methods.

84 The state of the art for such techniques is always evolving and should be considered during this assessment. The EDPB therefore encourages the development, dissemination and application of scientifically grounded expertise on this topic. Such expertise will become increasingly important as the complexity of the (re-) identification methods grows and/or the risk of their successful deployment rises.

85 Although a comprehensive list is not possible, several key factors will typically be relevant when assessing the effectiveness of these techniques:

- Whether the data is aggregated or at record-level;
- The dimensionality of the data. This refers to the amount of information that is available on each individual (*e.g.* the number of columns in a table for each individual record);
- The resolution of the data. This refers to the level of detail in the recorded attributes (*e.g.* a full date of birth has a higher resolution than the year of birth alone);
- The diversity of the data. This refers to the amount of differences between the data (*e.g.* if many individuals share the same attribute values, the data has a low diversity);
- The number of individuals whose data is included in the record; and
- The amount of additional information that is available for combination with the data.

86 As a rule of thumb, (re-) identification techniques are more likely to be effective against record-level data with high dimensionality and high resolution. By contrast, many (re-)identification techniques are less likely to be effective against data which is aggregated through mathematical functions into statistical indicators (*e.g.* averages), although this can depend on the amount of data which is aggregated. These key factors are also typically interdependent. For example, the amount of additional information which is necessary for a (re-)identification technique to be effective may be much lower for record-level data than it would be for aggregate data.

87 The effectiveness of a (re-) identification technique should be assessed on its ability to generate accurate results. This does not necessarily require perfect accuracy or absolute certainty; rather, the result should be at least sufficiently close, and provide at least a sufficiently high confidence in the circumstances at hand, to allow the data subjects to be distinguished and treated differently. Further, a (re-)identification technique can be successful even if it only succeeds against a single individual in a larger dataset[46].

Example 20: An entity uses a dataset with information about all of the employees in a big organisation to infer, with reasonable precision, the salary of the employees. However, the inference is only successful for one employee with a significant confidence and the entity has no way, other than guessing, to tell whose salary was correctly inferred. Since this technique cannot, with sufficiently high confidence, lead to the identification of the individual, it is not sufficiently effective and its use does not violate the No Inference criterion.

3.5.2 Assessing the criteria under the contextual approach

88 When using the contextual approach, it is important to recognise that the capabilities of different entities may vary over time. For example, a data leak could provide certain entities with access to new datasets which were previously unavailable to them. This would, in turn, allow those entities to link information with means which were, at least from that entity's perspective, previously unavailable but are now reasonably likely to be used, leading to a

[46] See paragraph 36.

violation of the No Linkage criterion. To ensure the proper protection of data subjects, the EDPB therefore recommends that analyses carried out under the contextual approach should incorporate, as much as possible, adequate safety margins ensuring that the assessment remains robust as the entities' capabilities evolve.

89 Another important consideration under the contextual approach is the accessibility of data. If data is publicly available, anybody may be able to access it with means reasonably likely to be used. If the data is subject to appropriate security controls (particularly regarding the confidentiality of the data), access to the data will constitute a mean reasonably likely to be used only for the authorised users (and not for unauthorised actors).This is relevant for both the given data itself and for any necessary additional information that might be used as part of linkage or inference.

Example 21: Returning to Example 16, the company's internal report is a necessary part of inferring the sixth engineer's salary. When applying the contextual approach, it is therefore important to test if each relevant entity has access to that report with means reasonably likely to be used. If that internal reporting is public, it should be assumed that everyone can access it and that the No Inference criterion is violated from everyone's perspective. If, however, access to that internal report is limited, and assuming that there is no other way to get this extra information, the No Inference criterion will only be violated from those perspectives who actually have access to it.

90 Limits to access should be assessed holistically, taking into account the capabilities of the different entities and any contextual elements which may have a greater or lesser effect on the means they are likely to use. Technical, organisational and contractual measures can help to restrict access, but are not by themselves sufficient to make data anonymous; while controllers are encouraged to adopt these measures wherever appropriate, they should still examine the (re-)identification capabilities of anyone who might be able to access the data with means reasonably likely to be used.

91 Appropriately implemented encryption may be used to limit access to data (or, at least, limits access to a form of the data which is intelligible) and, in this way, can help to support successful anonymisation. Nevertheless, encryption is, by its very nature, intended to be reversible and differs from anonymisation. It is, therefore, important to assess if entities with access to the data could still decrypt it with means reasonably likely to be used, even if they do not currently have the decryption key (e.g. through cryptanalysis).

92 Most currently documented (re-) identification techniques are becoming increasingly accessible, require limited resources and can be performed within a reasonable timeframe, even on commodity hardware. This is especially the case for the re-identification of record-level data. Some techniques which work on aggregate data may require larger computational resources, but access to such resources is still possible as a service, at relatively accessible costs. Time and cost would not generally be prohibitive for running these existing techniques, although they may pose challenges for (*e.g.*) obtaining the additional information necessary to use that technique successfully. The EDPB also notes that developments in AI, and especially agentic AI, will likely further reduce the time and costs necessary to access and deploy these techniques.

93 The effect of time and cost on the means reasonably likely to be used may, therefore, be particular seen where additional information which is necessary for (re-) identification is spread across multiple places or entities, or is hard to locate or access. However, the EDPB again notes that developments in AI may also reduce the time associated with gathering and

combining information from multiple sources (at least in cases where those sources are publicly available), which may lead to significant increases in the means reasonably likely to be used.

94 The following example is a high-level illustration of how one might assess the No Linkage criterion using the contextual approach.

Example 22: The research institute described in Example 12 decides to use the contextual approach to evaluate the anonymity of the data contained in the table after deleting the original personal data.

The research institute starts by identifying anybody who might, through means reasonably likely to be used, be able to access the table itself. The table has not been publicly released, and the research institute currently employs strong state-of-the-art cybersecurity measures, as well as other technical means to restrict unauthorised access to this dataset and prevent its members from creating any copies. Out of caution, the research institute considers that, especially over a long period of time, it is possible that the data will nevertheless leak. This could include, among other things, leaks due to human error or hacking attempts from external actors. The research institute therefore concludes that the table is accessible, through means reasonably likely to be used, by, at least:

1. members of the research institute who have been granted access to the table,
2. individuals who might accidentally receive the data due to human error,
3. individuals who might hack into the system and steal and distribute the data,
4. government or judicial authorities that might access the data in the course of the fulfilment of their public duties, and
5. other entities who might obtain the given data from someone in one of the above four groups.

However, the research institution concludes that it is extremely unlikely that people who are familiar with the data subjects' personal lives would receive the data in the aftermath of a leak, which already is not very likely to occur. The same conclusion applies to all other people who would know about an individual's participation in activities that lead to the inclusion of data relating to them in the given data. Therefore, the research institution concludes that these people would not qualify as relevant entity.

The research institution then considers whether the means available to each relevant entity are reasonably likely to be used, using objective factors as set out in paragraph 29 above.

It has already been concluded that the data can be linked to other data relating to the same individuals by any entity who knows (a) the individual's sex, year of birth and generalised postcode, and (b) that the individual was included in the given data. While this information may be known by, for example, the individuals' close friends and family members, nobody who knows about the inclusion of an individual in the given data are among the relevant entities. This includes the members of the research institute, since its members had no direct interaction with the patients and were not involved in their treatment, and there is no record of which patients were included in the dataset.

In conclusion, no relevant entity would know that a certain individual's data was included in the dataset. Moreover, it is not possible to derive from the dataset itself that a certain individual's data was included in the dataset. These considerations depend on the specific context of the given case and need to be underpinned by careful reasoning. However, in this

instance, the research institute is able to conclude that no relevant entity is likely to have access to the additional information that can be linked to the given data and that the No Linkage criterion has been met.

3.5.3 Compiling the results

95 If the given data fulfil all three criteria under either the simplified or contextual approaches, they can be presumed anonymous. If, however, one or more criteria is violated, further analysis may be needed to see if the information should be considered as personal data.

96 If the data is found to be personal under the simplified approach, it is also possible to perform a further assessment through the contextual approach. If none of the three criteria are violated from the relevant perspective, then the information can still be considered anonymous from that perspective. In some cases, this finding of anonymity may be limited to the perspective of only one entity, while in others it may extend to all of the relevant entities. As previously stated, the goal of anonymisation should be to ensure that the resulting data is anonymous for all of the relevant entities.

97 This is shown in the following table:

	Criterion is satisfied under either approach	**Criterion is violated under simplified approach**	**Criterion is violated under contextual approach**
No Record Isolation	The data can be considered anonymous if the other two criteria are also passed.	The data may not be anonymous. See paras. 98 to 00 below. To reach a conclusive result, consider continuing to the contextual approach.	The data may not be anonymous from at least one of the relevant perspectives. See paras. 98 to 100 below.
No Linkage	The data can be considered anonymous if the other two criteria are also passed.	The data may not be anonymous. See paras. 101 to 103 below. To reach a conclusive result, consider continuing to the contextual approach.	The data may not be anonymous from at least one of the relevant perspectives. See paras. 101 to 103 below.
No Inference	The data can be considered anonymous if the other two criteria are also passed.	The data should not be considered anonymous under the simplified approach, consider continuing to the contextual approach.	The data should not be considered anonymous from at least one of the relevant perspectives.

98 If the data violates the No Record Isolation criterion, it should be checked if the isolated records allow for singling out, so that individuals are distinguished from others by matching those records with the individual's corresponding attributes. These corresponding attributes could,

for example, become apparent to the relevant entity in the course of their interactions with the individual, or may be available through additional information that allows for the individual to be identified[47]. The larger the isolated record and the more attributes it contains, the easier the singling out of the individuals becomes. If the individual can be singled out in this way, then they should be considered identifiable.

Example 23: Returning to the dataset from Example 10 although the dataset violated the No Record Isolation criterion in several places, this does not necessarily mean that the individuals can be identified by isolating their records. However, in this case it is very likely: The records are unique, and their values are likely to be unique for most individuals in this population.

Further, the information needed to identify the individuals by reference to the demographic attributes may be available through means reasonably likely to be used (*e.g.* through public registers). If this is possible, the given data would be personal.

99 If this singling out cannot be done – and the other two criteria have been successfully met – then the given data can still be considered anonymous, despite failing the No Record Isolation criterion. This could happen if, among other things, the record is unique (and so the No Record Isolation criterion is failed) but the attributes cannot be matched to any other data available to the relevant entity or if the unique record is not intelligible.

Example 24: A cinema decides to run a study to better understand its clients' preferences. The survey is answered by a very large number of customers, and only contains questions related to the movie itself, that were written specifically for this survey and that do not reveal anything that can be reasonably linked back to the respondent. The responses are then collected in a secret ballot, and later turned into a dataset:

Survey number	**Question 1**	**Question 2**	**...**	**Question 30**
1	D	C	…	B
2	C	A	...	A
3	…	…	…	...

Because of the large number of questions, it is very likely that at least some (if not all) records are unique in the dataset, violating the No Record Isolation criterion. However, this does not mean that participants can be identified. In fact, given the circumstances, there is no way to link the answers to their identity or to other information relating to them using means reasonably likely to be used, nor to treat any respondent differently based on their individual answers. If the data passes the other two criteria, it is, therefore, anonymous, despite failing the No Records Isolation criterion.

100 If this singling out can be done, the information should not be considered anonymous. However, if the analysis has progressed under the simplified approach (*i.e.* testing whether an entity can single out an individual on the basis of the given data, while assuming that if means reasonably likely to be used exist then at least someone will be able to use them), this

[47] See, for example, C-479/22 P *OC v European Commission*, paras. 60 & 63, where the CJEU looked at a unique combination of attribute values which could then be compared against the corresponding attributes available online.

analysis could be further extended by using the contextual approach, as described in paragraphs 48 and 49, and following the guidance set out in sub-section 3.5.2 above.

101 If the data violates the No Linkage criterion, it should be checked if the linked data can then directly lead to the identifiability of the individual.

Example 25: Returning to Example 12 the No Linkage criterion would be violated if the information from the given data could be linked with another dataset which contained the same individual's sex, year of birth and generalised postcode.

Since the individual's sex, year of birth and generalised postcode are not unique, it is not possible to distinguish the individuals and to identify the individual. However, if the second dataset contains any extra attributes which relate to the same individual and which would allow for the individual to be identified (or if it contains a direct identifier for that individual) then the linkage would directly lead to the identifiability of the individual. In such cases, the given data would be considered personal data.

102 If the given data can be linked to another dataset, the new (linked) information should then be re-tested against the three criteria. If the linked information then passes the No Record Isolation, No (further) Linkage and No Inference criteria, it can be considered anonymous.

103 If the new (linked) information can directly lead to the identifiability of the individual, the information should not be considered anonymous. However, if the analysis has progressed under the simplified approach (*i.e.* testing whether the linked information can lead to the identifiability of an individual, while assuming that if means reasonably likely to be used exist then at least someone will be able to use them), this analysis could be further extended by using the contextual approach, as described in paragraphs 48 and 49, and following the guidance set out in sub-section 3.5.2 above.

4 Glossary

To make these guidelines easier to read, they use several terms that are not legally defined. These terms are included in this glossary.

Aggregate data: Data that relates to groups of people and does not contain individual records.

Anonymising controller: A controller performing the anonymisation and assessing its outcome.

Applicable perspective: The perspective from which the personal nature of some data in the hands of a specific entity needs to be assessed subject to the concrete circumstances of the case. May be the perspective of the entity itself or that of another entity, e.g., in the case of processing on behalf of such other entity.

Entity: A natural or legal person, public authority, agency or other body. Depending on whether or not the information is personal, they could, among other things, be a potential controller, a potential processor, or a potential third party with additional information necessary to identify the data subject.

Given data: Data whose anonymity is to be assessed.

Individual: A natural person who may or may not be identifiable and to whom information contained in the given data may or may not relate.

Perspective: Each entity has a **perspective**, which will depend on their circumstances and informs the means reasonably likely to be used by that entity. If an entity has or may gain means reasonably likely to be used to identify the natural person to whom the information relates, the information can be said to be personal data from their perspective.

Record: All data values (including images *etc*.) within the given data that are known to relate to the same individual, regardless of whether that individual is identifiable or not. For example, in a dataset containing photos of faces, a record could consist of one, several or all the photos showing the same person. Usually, records relating to different persons are clearly distinguishable. However, in some cases the same piece of information may relate to more than one individual and so may also belong to different records. For the sake of clarity, this word is not being used in the sense of a "database record". For the sake of these guidelines, a record may, therefore, be made up of data which (for example) covers several rows in a table, provided that each of those rows relates to the same individual.

Receiving entity: An entity that receives the given data.

Record-level data: Data structured in a way to easily distinguish individual records.

Relevant entity: Entities are **relevant** for assessing anonymity from the perspective of a particular entity if they could contribute to direct or indirect identification by that entity.

Re-identification: A process in which data presumed to have been successfully anonymised are found to be personal by identifying an individual to whom the data relate.

Annex 1 : Flowchart for the technical analysis of anonymity

This flowchart has been developed as a support tool for organisations intending to anonymise data. In a simplified manner, it presents one way of how the assessment could be done in line with the guidelines.

