30 October 2026

EDPB Guidelines on Anonymisation: where unforeseeable is reasonable?

by Peter Craddock

Submission to the public consultation of the EDPB on Guidelines 02/2026

Through its Guidelines 02/2026 on Anonymisation of 7 July 2026, the European Data Protection Board appears to recognise the relative nature of the concept of "personal data", but these Guidelines contain inconsistencies that prevent this recognition from bearing fruit.

Table of contents:

I. Recognition of the relative nature of personal data 1
II. Processing on behalf of a controller 2
III. Information that "relates to" a natural person 3
IV. Information that relates to an "identified" or "identifiable" natural person 4
V. Means reasonably likely to be used 6
V.1. Hypotheticals versus reality 6
V.2. Illegal & unknown sources of data = no problem? 7
V.3. On the role of intent 8
V.4. Contracts not to the rescue, says the EDPB (vs European Commission) 9
VI. Mixed datasets – and the impact of one identified record on the whole 10
VII. Documentation of anonymisation assessments 11
VIII. Contextual and simplified approach 12
IX. Three criteria for anonymisation 13
IX.1. No record isolation (or is it "no singling out"?) 13
IX.2. No linkage 14
IX.3. No inference 15
X. What of evolutions to identification techniques? 17
XI. Conclusions regarding the Anonymisation Guidelines as they currently exist 17

I. Recognition of the relative nature of personal data

The *SRB* case hinged upon the question of whether information can be "personal data" from the perspective of one entity but non-personal data from someone else's perspective. It opposed two views: the "relative" approach (whereby information can be personal data from the perspective of someone who *does* have means of identifying a natural person to whom the information relates but not for someone else), and the "absolute" approach (whereby if information is "personal data" from someone's perspective, it is so also for everyone else – no matter whether those other persons have a way of linking that information to an identifiable natural person).

Paragraphs 7 to 12 of the Anonymisation Guidelines are a recognition by the EDPB that the issue of whether information relates to an identified or identifiable natural person is a relative (or, to use the EDPB's wording, "contextual") one.

For instance, paragraph 7 states that even though certain information may *"unambiguously relate to an individual"*, it may be that *"only one organisation may be able to actually identify that individual"* and in that case *"from that organisation's perspective"* the information would be personal data, but that same information *"would be considered anonymous for everyone else"*.

Paragraph 8 highlights that the key question is therefore a relative one: *"how likely it is that the individual will be identified or identifiable by some entity"*?

Paragraph 9 stresses that if 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"*, there is not even a need to assess *"whether the information is anonymous for that entity"* – a welcome source of reassurance for many recipients of anonymised data.

The EDPB then starts to look at intent as a relevant factor for assessing whose perspective matters: *"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"* (§12). Quoting the *OC v Commission* judgment of the EU Court of Justice (CJEU), which concerned a press release by the Commission's OLAF agency, the EDPB highlights the target audience (*"journalists, including investigative journalists with research skills beyond those of an average person"*) as determining the relevant perspectives.

Yet this idea of intent disappears afterwards and is even disavowed later in the document. *[This is an unfortunate aspect, as recognition of the role of intent would bring legal certainty to the issue of incidental processing. In this respect, I would invite EDPB members to read my piece on "Incidental processing and GDPR: from bystanders to spontaneous notes, do data protection rules apply?"]*

II. Processing on behalf of a controller

Another point that the EDPB tries to explain concerns a lingering post-SRB question: does it matter if the person receiving pseudonymised/anonymised data is a processor? Does that affect their assessment of the data?

In my view, a perspective I have repeated several times since the SRB judgment came out, a processor is acting under the authority of the controller (as per Art. 29 GDPR) while it remains a processor, and there is a statutory logic behind applying the controller's assessment to the processor. In particular, taking the opposite view – that the processor's own perspective must be assessed separately – would introduce inconsistencies:

- if a controller has to view information as personal data, the controller is notably subject to Article 28 GDPR obligations (including the statutory obligation to conclude a data processing agreement with a processor);

- yet if a processor can view that same information as non-personal data, that processor is **not** subject to Article 28 GDPR obligations – and is therefore not required by law to conclude a data processing agreement.

Such a position would mean that the processor is technically free to argue that it has no statutory obligation to sign a data processing agreement under Article 28 GDPR and that it is not even bound by any security obligation or data breach notification obligation under the GDPR. Meanwhile, the controller would remain fully bound by all GDPR obligations and could not legally force the processor to do anything.

In my personal view, this would not meet the statutory logic of Articles 28 and 29 GDPR, and it is preferable to use Article 29 GDPR and the concept of "authority" of the controller over the processor, to say that anything that falls under the authority of the controller must be treated in the same manner. The processor then inherits the legal status of information from the controller's perspective.

Another advantage of the "authority" argument (based on Article 29 GDPR) is that it extends also to employees – but not rogue employees, i.e. individuals who act beyond the instructions of their employer.

The EDPB's reasoning appears to disregard this aspect, and only has a very limited justification in paragraph 15 that in my view is insufficient to withstand a serious challenge.

For this reason, I would recommend that the EDPB adopts a stronger justification for the "processor" assessment section, and relies on Article 29 GDPR and the question of authority to that end.

III. Information that "relates to" a natural person

The Anonymisation Guidelines largely restate in paragraphs 16 & 17 the principles developed by the CJEU in its *Nowak* judgment: information may relate to an individual by reason of its content, purpose or effect.

What the EDPB does, though, is add its interpretation of what "purpose" and "effect" mean in practice:

- On the idea of relating to an individual by reason of the "purpose" of the information: *"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"* [§16, point b].

- On relating by reason of "effect", the EDPB states that *"[f]or 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."* [§16, point c]

These are useful illustrations, though they do raise the question of intent and probability. "Effect", for instance, should not be taken to cover pure hypotheticals, as otherwise *anything* can be said to relate to *anyone* – it is not hard to find far-fetched scenarios that could create a link between a person and

information. The same applies to "purpose", which by definition must be intentional (there is no such thing as an "unintentional purpose" of use of information).

The EDPB then tries to suggest that aggregated data can relate to individuals, but does so in a clumsy manner: *"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)."* [§17]

The issue with this wording by the EDPB is that although one should indeed not exclude all data purely because it is aggregated, this link in this particular paragraph is very tenuous. There should be an **active** attempt to make this about each individual, not just about a group.

Put differently, in §17, the EDPB stretches the concept of "relates to" too far. Just because techniques exist that could reveal information relating to that person does not mean that the aggregated information **does** relate to that person. An additional action is required.

I would recommend amending the wording of §17 to remove the impression that aggregate data relates to individuals *passively*, and instead emphasise that for aggregate information to relate to an individual, **it must be actively coupled with that individual** – i.e. recorded at individual level.

IV. Information that relates to an "identified" or "identifiable" natural person

On the issue of identifiability (which I have examined in depth from the angle of pseudonymisation and in an analysis of the SRB judgment), in paragraph 19 of the Anonymisation Guidelines the EDPB states the following:

> *"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." [§19]*

As I wrote previously: *"Identifying a natural person requires in essence three things: (i) the ability to attribute certain information to a **natural person**, (ii) the ability to **distinguish** that person from any other persons (= i.e. so it is a **specific** natural person, not just any natural person) and (iii) that distinction must be of such a nature as to make it possible to **act upon or in relation to such person**."*

It is encouraging to see that the EDPB shares some of this thinking, yet the wording used in §19 gets overlooked later in the Anonymisation Guidelines, creating a significant inconsistency.

For instance, in §21, it states 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"*. While that might seem at first glance like an application of the Scania and SRB judgments, **the EDPB goes too far by saying that it is reasonably likely that an entity "can" transfer**. *Scania* and *SRB* were about cases of **actual** transfers, not hypothetical ones, and in

Breyer the CJEU explicitly told the referring court to verify whether conditions were in place to obtain additional data.

In other words, **"potential" transmissions are irrelevant** (an issue I have written about previously). "Identifiable" requires assessing concrete situations, not mere hypotheticals, and all case law of the CJEU about these issues has been about actual transfers.

While the EDPB does later recognise this in paragraph 28 (see below), its paragraph 23 appears to **ignore** this – and *Breyer* – when it states that an identifier enabling identification could *"[i]n the simplest cases, […] for example, be a single piece of unique data, such as a name, IP address, social security number or e-mail address"*. After all, *Breyer* teaches us that an IP address **can** be an identifier but is not necessarily so.

Moreover, the association of IP addresses with social security numbers (a truly unique, person-tied government-issued identification number) appears unfortunate, due to significant differences between them in terms of identity.

Before examining these examples, it is important to distinguish between pure hypothetical identifiability (which is easy to show) and plausible identifiability, i.e. foreseeable identification (which is less so). Much of the discussion below concerns whether the EDPB's examples establish mere hypothetical identifiability or whether they genuinely satisfy the threshold of foreseeable identification that the GDPR and case law require.

Similarly, the EDPB tries to use a video-surveillance example (Example 5 in the Anonymisation Guidelines) to illustrate what combination of attributes might allow identification, yet all it seems to do is illustrate **uniqueness** or **singling out**, not actual identification:

> *"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"*

It is unclear how the EDPB interprets this situation as meeting the definition of "identified" that the EDPB itself gave in §19. A review of video-surveillance footage about a "person in a black suit who is also walking a dog" does not after all "*mak[e] it possible to treat them differently from those other people*".

The wording suggests that this conclusion is reached solely on the basis of the first part of the EDPB's own definition of "identified", namely "*distinguish them from others within a given context*", and not on the rest of the definition.

The same observation can be made in relation to paragraph 25, which mentions device identifiers:

> *"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."*

The EDPB explicitly stated (and the GDPR does too) that identification of a **natural person** is required. The link between the device and a natural person who can – according to the EDPB – be distinguished and treated differently must then still be shown, and cannot be presumed to exist.

Example 6 on browser fingerprinting illustrates this issue further:

> *"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."*

Through these examples, the EDPB fails to show how these situations truly identify a natural person.

In other words, although it explicitly mentions the need to be able to treat natural persons differently, through its own illustrations and explanations the EDPB appears to shift the relevant threshold from actual identification to mere distinguishability, treating the ability to distinguish an individual from others within a given context as sufficient for identification.

This is not what the GDPR or CJEU case law say, and it is critical for the EDPB to adapt its Guidelines to include more consistency in this respect (see also Section IX.1 below in relation to "singling out").

V. Means reasonably likely to be used

In paragraphs 26 to 34, the EDPB sets out its vision of what constitute "means reasonably likely to be used" to identify a natural person – and what are *not* reasonable means.

A recurrent issue is that the EDPB raises the threshold without seeming to.

V.1. Hypotheticals versus reality

First, it appears that the EDPB has overlooked a key word the CJEU used in its judgments.

According to the EDPB, *"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"*. While this may look correct at first glance, the CJEU has actually stated that "**practically** impossible" is the standard, not "impossible. Yet "impossible" means no one can ever do something, while "practically impossible" means that the effort to do so would be too great (see also paragraph 49, which also omits the word "practically").

A good illustration of the issue of availability of such means of identification – and a correct application of the test by the EDPB – comes in paragraph 28, which concerns the issue of means of identification that are not available directly to an entity but can be obtained through someone else:

> *"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 comes together"*

This is an important recognition that **pure hypotheticals do not matter**, only reasonable ones do. Recital 26 GDPR explicitly requires (only) consideration of means **reasonably** likely to be used – not all means that are theoretically conceivable. In other words, the GDPR itself incorporates a foreseeability

assessment: the question is not whether identification is possible in some abstract sense, but whether it is realistically foreseeable in the circumstances at hand.

The inconsistencies between paragraph 28 (pure hypotheticals don't matter) and paragraphs 21 and 23 (everything is relevant) create uncertainty as to which position the EDPB wishes to take. My request to the EDPB would be to correct paragraphs 21 and 23 and bring them in line with paragraph 28.

V.2. Illegal & unknown sources of data = no problem?

In paragraph 30, the EDPB tries to explain which "relevant entities" should be taken into account when assessing which means of identification are available indirectly to someone:

> *"The GDPR does not lay down any conditions regarding the entities who are able to identify the individual. 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) […] Cybercriminals. These entities may simply break the law to gain access to data not available from a law-abiding entities' perspective"*

This raises important questions. First, the EDPB admits that a **legal prohibition** means that certain "means" are not reasonable. Yet then it suggests that if someone else breaks the law, the resulting means of identification are reasonable – i.e. using information available through cybercriminals is "reasonable". This must be corrected, as it would otherwise mean that such processing is "reasonable" and bring into question the "lawfulness" principle under the GDPR.

Paragraphs 32 & 33 attempt to deal with this, but the EDPB's argument is that *"evidence that the legal prohibition is not effectively monitored, enforced and sufficiently dissuasive"* could be an example of a factor meaning that (in the EDPB's view) illegal means are reasonable.

In my view, this is fundamentally problematic. "Reasonable" in EU law inherently includes a "lawful" requirement, so just because something is "possible" does not mean that under EU law it counts as "reasonable".

Suggesting that unlawful data sources should be taken into account mixes up two entirely different legal tests:

- When a controller assesses security measures under Article 32 GDPR, it must anticipate risks, including cybersecurity risks and threats from malicious actors. This "ex ante" assessment is intended to be more speculative, because it is about taking steps to prevent threats from occurring.

- On the other hand, when assessing whether information is "personal data" under Article 4(1) GDPR, i.e. whether information actually relates to an identifiable natural person in the first place, the assessment cannot be purely speculative, and hypothetical actions by malicious actors cannot be counted as "means reasonably likely to be used".

If one were required to take unlawful data sources into account, such as potential data leaks from other sources available through the dark web, the mere hypothetical risk of a future data breach would transform truly anonymous data at a given time into personal data **without that leak even occurring**. The **hypothetical unlawful processing of additional data** would then be relevant for assessing

whether **lawful use of information is actually processing of personal data**. Foreseeability – a key component of reasonableness – would be replaced with pure speculation for no good reason other than to assume that information is personal data by default.

In a similar vein, that same list of relevant entities includes **rogue employees** (i.e. employees who **disregard the potential controller's instructions** and decide to add whatever information they can find to identify someone, in a wholly unforeseeable manner) and even **domestic and foreign intelligence agencies**.

It is difficult to see how this is "reasonable", and it does not take foreseeability into account.

Further in the document, the EDPB appears to double-down on this approach – just before contradicting it.

In paragraph 88, the EDPB focusses on the example of **a data leak** as a source of additional data: *"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"*.

Once again, this data is **not lawfully processed**, so it **cannot** count as reasonable means of identification.

Fortunately, the following paragraph then seems to introduce some measure of reasonableness: *"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)"* [§89].

This does reveal, however, a form of contradiction within the EDPB's document:

- Either we assume that illegal processing is "reasonable means",
- Or we assume that any violation of confidentiality is possible and that therefore access to the data is a means reasonably likely to be used **also** for unauthorised users.

The contradiction is thus that (i) illegal processing features heavily among "relevant entities" and even datasets enabling identification, yet at the same time (ii) only authorised actors can be considered as having means reasonably likely to be used.

Legal certainty demands a choice, not flexibility on the part of regulators to be able to decide depending on the case which prevails. I would therefore recommend that the EDPB clarifies that **illegal means are not "reasonable"** and amends paragraphs 30 and following as well as 88 to focus on **lawful** means of (re)identification.

V.3. On the role of intent

In paragraph 31, the EDPB suggests that intent and motivations are irrelevant to the assessment of identifiability, because *"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 paragraph is problematic in several respects.

First, if identification occurs due to an accident or negligence, this raises the question as to how this can (or even should) be anticipated: why would future accidental identification be relevant at the time of assessment?

Later in the document, the EDPB similarly suggests that unintentional identification has to be taken into account: *"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 […]"* [§83]. Once more, this begs the question of how one can anticipate **future and unintentional** identification.

In reality, identification of a natural person must be assessed at **two stages**: before it happens (**identifiable**) and after it happens (**identified**). The first is about **foreseeability** and reasonableness, the second is about an **occurrence**, something that has already happened. In the first case, motivation **is critical**, because accidental identification is by definition unforeseeable.

In other words, paragraph 31 as a whole incorrectly minimises the role that intent plays on identifiability, and should be amended to recognise such role. I refer in this respect to my more extensive analysis on incidental processing.

V.4. Contracts not to the rescue, says the EDPB (vs European Commission)

The EDPB then moves onto contracts, to examine whether a contractual prohibition to (re-)identify should be viewed as a legal prohibition of identification – i.e. whether it is legally relevant to force information to be treated by the recipient as anonymous data by way of contract.

In this respect, it states the following:

> *"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, bearing in mind that contractual measures can typically be subject to revision, or may even be disregarded entirely by the parties." [§34]*

Yet this position of the EDPB appears at odds with the European Commission's own stance on anonymisation under Article 6(11) of the Digital Markets Act.

In a document about proposed measures that Alphabet (Google's parent company) should implement to ensure effective search data sharing with third-party online search engines, the Commission described both technical measures *"to mitigate the risk of re-identification of end users to a residual level"* and contractual measures that *"complement the technical measures and further mitigate the residual risks of re-identification of end users to an insignificant level"*.

"[F]urther mitigate […] to an insignificant level" means that the contractual terms chosen are deemed by the Commission to be sufficient to make the data actually anonymous – based on the CJEU's own case law,

which shows that an "insignificant" likelihood of (re-)identification means that information is not personal data. In other words, information will be viewed as "anonymous data" if such measures are put in place.

The EDPB, on the other hand, by stating that a prohibition in a contract is not a legal prohibition, clearly reserves the right to consider that the combination of contractual and technical measures is insufficient.

While the wording might not be a flat-out contradiction on all fronts (the EDPB also speaks of using contractual terms *"to complement technical measures"*), and while both the Commission and the EDPB focus on verifiability (e.g. through audits, in the Commission's case), the EDPB's highlighting of the risk that contractual measures *"may even be disregarded entirely by the parties"* does not feature in the Commission's own approach under Article 6(11) DMA – which in practice outsources enforcement to Alphabet, while the EDPB would clearly seek to enforce the law itself against those contractual partners.

This leads to a situation where on the one hand contractual measures are described as capable of nullifying the insufficiencies of technical measures of anonymisation (Commission), while on the other hand the EDPB wishes to still be vague about which measures will be acceptable or not.

The EDPB's position is not without its flaws, though, and it all comes down to foreseeability. Is the hypothesis of a contract being changed **on the issue of anonymisation** truly reasonable? And should a given entity truly foresee the hypothesis of a contract being disregarded entirely?

Hypothetically, the strongest encryption is also useless if someone decides at one point in the future, on the spur of the moment, to voluntarily share an encryption key. Yet how realistic and likely is that?

This reinforces the observation that **the Anonymisation Guidelines are missing a key component of foreseeability**.

I would therefore recommend that the EDPB adds **foreseeability** to all relevant sections, in particular paragraph 34.

VI. Mixed datasets – and the impact of one identified record on the whole

In paragraph 36, the EDPB takes a radical stance on the scope of anonymisation for situations where multiple records exist in a dataset:

> *"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 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"*

Footnotes reveal that this reasoning is based on the *Bundeskartellamt* judgment of the CJEU, yet the reasoning itself of the CJEU is unrelated to this issue. In that case, the CJEU was saying that *"where a set of data containing both sensitive data and non-sensitive data is […] collected en bloc without it being possible to separate the data items from each other at the time of collection, the processing of that set of data must be regarded as*

being prohibited, within the meaning of Article 9(1) of the GDPR, if it contains at least one sensitive data item and none of the derogations in Article 9(2) of [the GDPR] applies" (paragraph 89 of *Bundeskartellamt).* In other words, where special categories of personal data are mixed with "normal" personal data with no possibility to separate them, Article 9 of the GDPR applies to the whole of the dataset.

The same reasoning appears further in the document (paragraph 87), where the EDPB states that *"a (re-)identification technique can be successful even if it only succeeds against a single individual in a larger dataset"*.

Yet the reasoning regarding special categories of data is fundamentally different compared to that in relation to anonymisation. Unlike the situation where the GDPR applies to both "normal" personal data and special categories of personal data, just with certain additional obligations, **the GDPR does not apply to non-personal data**. One cannot presume its application to an entire dataset just because one data item within a database might be personal data. While a reidentification technique might be successful against *one* record, it in no way should be assumed to be effective against a whole dataset.

Instead, one should focus on ensuring that **the likelihood of reidentification is as limited as possible as a whole**, and not presume that just because one person can be identified the entire dataset is not anonymous. The more appropriate question is therefore whether the record or records that *are* shown to be personal data materially affect the re-identification risk associated with the remaining records in the dataset.

Moreover, the wording chosen by the EDPB *("not treated separately")* in this respect does not match the wording chosen by the CJEU *("without it being possible to separate the data items from each other at the time of collection")*, making it the EDPB's position inconsistent with case law.

This does not contribute to give the impression that this is an objective assessment of the (in)applicability of the rules.

Ultimately, one should assess whether it is possible to separate the anonymous data from the information that is shown to be personal data – not checking whether they are "treated separately" but whether they are distinguishable from one another. If so, the next step involves determining which parameters and factors make specific records "identifiable", and whether the remainder of the dataset could feature the same characteristics. I recommend that the EDPB adapts paragraphs 36 and 87 along those lines.

VII. Documentation of anonymisation assessments

An important point highlighted by the EDPB is that no matter what is done in terms of anonymisation, *"controllers should ensure adequate documentation of the anonymisation processing*" and *"[t]his documentation should then be retained after the completion of the anonymisation process"* [§41]. This is critical in any situation, whether one agrees with the EDPB's own assessment of what constitutes proper anonymisation or not, notably from an evidentiary perspective.

However, the reason for which the EDPB stresses this point is less aligned with the reasons that should prevail: *"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"* [§41]. While there are GDPR benefits to documenting the process of anonymisation, the real benefit is precisely linked to the **inapplicability** of the GDPR – and it is up to supervisory authorities to demonstrate that information is personal data, not the other way around.

In other words, documentation of the anonymisation assessment is in reality useful not to prove GDPR compliance but to **counter any allegation by a regulator that the GDPR applies**.

While it is not immediately in the EDPB's interest, I would recommend that the EDPB makes this explicit in order to ensure that the Anonymisation Guidelines are as objective as they can be.

VIII. Contextual and simplified approach

The remainder of the Anonymisation Guidelines is dedicated to the actual assessment of anonymisation, and the EDPB makes a distinction between two approaches:

- a "contextual" approach (where an entity examines the (re-)identification capabilities of the entity itself as well as, separately, those of each other "relevant entity" (though hopefully not with the need to consider intelligence agencies, cybercriminals and the like – see above), or

- a "simplified" approach (where all entities are considered together, without any differentiation as to their individual capabilities).

The EDPB describes the contextual approach in paragraph 47 as *"quite complex, especially where it is difficult to account for (or even know about) all the possible contextual elements involved"* – which unfortunately once more ignores the key issue of foreseeability that I have highlighted above. The EDPB states that this contextual approach *"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 identification"* [§47].

Yet this reasoning appears to be incorrect. As mentioned previously, the GDPR only applies if information is personal data, so claiming that false positives might occur *"because [the assessor is] unaware of a particular entity's means reasonably likely to be used for identification"* is a reversal of the burden of proof that is established in the very beginning of the GDPR regarding its own applicability. Moreover, CJEU case law on the topic (in *Breyer*, *Scania*, *OC v Commission*, *SRB*) has always been about **actual, foreseeable** recipients of data and **actual, foreseeable** means of identification at their disposal – not hypotheticals, let alone situations of which someone might be "unaware".

The "contextual approach" to anonymisation described by the EDPB therefore does not seem to have any basis in case law or in law.

Next, the "simplified approach" proposed by the EDPB in paragraph 48 is anything but simple in practice, as it in effect assumes by default that information is personal data, raising the threshold for information *not* to be personal data when the rules regarding GDPR applicability dictate the precise opposite.

Instead, it may have been more interesting for the EDPB to propose an actual simplified approach, for instance along the following lines:

- If the data comes from a given person or entity, is not shared with third parties and the relevant person or entity has removed all means of identification and has documented its reasoning (e.g. de-identification process), that person or entity can then (more easily) take the risk of treating it as non-personal data.

- If the person or entity know that some third parties are likely to have means of identification, but it does not see any way for it to ask their assistance, that person or entity should document its assessment and it could then take the risk of treating it as non-personal data.

- If the person or entity know that some third parties are likely to have means of identification, and it thinks it is possible for it to ask their assistance, the person or entity in question should document its assessment and treat it as personal data.

IX. Three criteria for anonymisation

In paragraph 52, the EDPB sets out three criteria: "No Record Isolation, No Linkage and No Inference" – similar to the three anonymisation criteria that were previously set out by the Article 29 Working Party in its Opinion 05/2014 on anonymisation techniques.

While this may seem like a good start to make an anonymisation assessment more accessible, the rules regarding them mitigate their very usefulness: *"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 […] to assess the impact of that violation and whether the data could still be considered anonymous"* [§52].

In other words, if a data record cannot be isolated and cannot be linked to other data, and provided no inferences are possible based on that record, the information **can be considered** anonymous (meaning this could still be brought into question), yet if any of those criteria are not met, it **can also** be personal data or anonymous data.

These are therefore not black or white criteria, and they do not give legal certainty to the entity making the assessment. They are just guides for the assessment, not parameters.

IX.1. No record isolation (or is it "no singling out"?)

The first criterion – "record isolation" – is a peculiar one. The Article 29 Working Party (WP29) Opinion on anonymisation listed this criterion as "singling out". When the GDPR was adopted, its Recital 26 contained more details on means reasonably likely to be used to identify natural persons than Recital 26 of the GDPR's predecessor, the Data Protection Directive 95/46/EC. Most notably, Recital 26 of the GDPR features "singling out":

> *"To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, **such as singling out**, either by the controller or by another person to identify the natural person directly or indirectly"* (emphasis mine)

In other words, "singling out" is an example of means of identification. But is it sufficient for identification?

There is no definition of "singling out" in the GDPR, but it was defined in the WP29 Opinion as *"the possibility to isolate some or all records which identify an individual in the dataset"*. In effect, the WP29 Opinion showed that singling out is **one possible component** for identification, but that it is not sufficient for identification.

In this context, Recital 26 just includes singling out as an example of means that can contribute to identification, not as a standalone legal test of a technique that automatically means identification. Nothing in the GDPR suggests that singling out is itself equivalent to identification or that it automatically renders a natural person identifiable.

In its new Anonymisation Guidelines, though, the EDPB seems to repeat the myth that singling out *is* identifiability:

> *"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" [§89]*

Yet singling out just means being able to distinguish one record from another, not yet being able to identify the natural person (i) as a specific natural person and (ii) in a way that allows action.

This reveals a possible significant historical rewrite by the EDPB, a possible change to the interpretation of Recital 26 of the GDPR. Where "singling out" was previously *"the possibility to isolate some or all records which identify an individual in the dataset"*, the EDPB now appear to try to elevate "singling out" to something *stronger* than "record isolation", something that establishes identifiability on its own and that is stronger and more individualised than mere record isolation.

This is wholly inconsistent with the EDPB's own definition of identification, though. Just as highlighted earlier in relation to Example 5 (the video-surveillance footage about a *"person in a black suit who is also walking a dog"*), if a video-surveillance feed shows "a person in a black suit who is also walking a dog," the operator has isolated a record. The operator has distinguished them from others in that specific context. But does this actually allow the operator to identify them, to locate them, or to act upon that distinction in the real world? It does not.

In other words, singling out only meets the first part of the EDPB's definition of identification ("*distinguish them from others within a given context"*) but fails to meet the second part *("making it possible to treat them differently from those other people").* It is not identification.

IX.2. No linkage

The "no linkage" criterion is more straightforward in comparison with the WP29 Opinion on anonymisation, as it corresponds to the "linkability" criterion.

Yet the example given is anything but helpful. As Example 11 describes, *"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 assumptions made here are surprising. Shop purchase records are not necessarily personal data, especially given that in this example there is no information about what the "direct identifiers" are that relate to each customer.

The example then lists the idea of a shop owner comparing data from purchases in its own shop with a *"very popular website where users create lists of the games which they own"* and noticing 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"*. In reality, though, one would need much more information (geography, date

of purchase, etc.) to be able to make a useful correlation between a set of shop purchases and a popular website with a list of games owned. Otherwise, a user could be registering a game bought 5 years beforehand or on another continent and be assumed to be a customer who just purchased the game yesterday.

As paragraph 64 admits, the effectiveness of record linkage attacks *"depends heavily on the availability of additional information"* – but that is also the case for assessing linkage in general, and one should not assume that records can be linked together without actual verification of the accuracy and likelihood thereof in practice.

IX.3. No inference

Just as with "no linkage / linkability", the link between the EDPB's "no inference" criterion and the WP29's "inference" criterion is obvious.

Once again, though, the added detail the EDPB brings to the concept merely seems to cause more issues.

The EDPB starts by stating that an inference must be "specific" and "meaningful". "Specific" means that the inferred information *"relates to a single identified or identifiable individual (i.e. it is personal data)"*, while "meaningful" would appear to be met if the processing of the inferred information 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"* [§67].

This points at an important issue regarding macro inferences, i.e. inferences based on general trends: inferences have to be specific about an individual, not stemming from macro-level inferences, to have an impact on the issue of anonymisation.

The EDPB's following paragraphs appear to muddy the waters in this respect, though.

On the one hand, in paragraph 69, the EDPB states that *"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 relate to an identified or identifiable individual, but this does not then mean that the same is true for the given data"*. Yet it is unclear how one can assess whether the data being inferred is "personal data" to start with. One could perfectly infer general characteristics from a dataset without that being personal data, in which case the inferred data **does not** relate to an identified or identifiable individual. Not every inference is personal data.

Paragraph 70 on the other hand is clearer: *"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 other words, the result of a generalisation **is not** a violation of the "no inference" criterion.

Paragraph 71 is equally clearer, bringing the idea of inference back to the "specific and meaningful" criteria it had set out earlier: *"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."*

The EDPB continues on the right track, with examples 13 and 14 illustrating the legal issue in a consistent manner:

> *"**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."*

The conclusion one can draw from these explanations and examples is that general inferences based on statistical analysis **are not** an issue regarding anonymisation or the assessment of what is or is not personal data. Only at the stage of individualisation do they become potentially personal data **in relation to an identified or identifiable natural person**, but **they do not enable identification and are thus insufficient to create personal data in and of themselves**.

Unfortunately, the EDPB does not stop there. Instead, it moves back into the realm of **hypotheticals** with its example 15:

> *"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, 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."*

The problem with this example is that there is no indication that any inference is likely. The EDPB merely states that the information "allows" an inference, but it does not explain how to handle the situation where the information is not used at all in such a way, or where such a use would be **unlawful** (e.g. because the controller has not foreseen any (valid) legal ground). Once more, the EDPB appears to work on the basis of mere hypotheticals.

Still on the topic of inference, the EDPB cautions in paragraphs 76 to 78 against the producing of "record-level data from aggregate data", seeking to explain how this can violate anonymisation. However, the examples given (examples 16 & 17) concern very small datasets – one single team in one example, six people in another. This entire section could be rephrased to emphasise that **the more limited a population size, the likelier identification is possible despite anonymisation**. The issue is that these examples seem to be used to show that anonymisation is flawed, when in reality the very low number of records is the key issue here.

The following example, example 18, seems even more problematic, with key context clearly missing. The sudden statement that "Alex was included in the original dataset" does not include any indication

as to why it should be possible for someone to assume that one record is the same as in another dataset or how "Alex" is even identified to start with.

X. What of evolutions to identification techniques?

The final sections of the Anonymisation Guidelines touch upon the issue of increasingly advanced and accessible techniques for re-identification.

Paragraph 92 in particular raises the question of AI-assisted reidentification:

> *"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"*

While these concerns are perfectly legitimate, so far AI-assisted reidentification stories have focussed on a few scenarios where anonymisation was not effective to start with. Moreover, the key issue of accuracy remains highly relevant: if AI-assisted reidentification creates matches where there are none in reality, is inaccurate identification truly processing of information relating to an identifiable natural person?

It once more comes down to a key question: is reidentifying **one** natural person sufficient to bring into question an entire database? The cost of reidentification of one record may indeed be in theory lessened by way of modern techniques, but it does not mean that reidentification of an entire database is "reasonable", nor does it mean that an entire dataset has to be viewed as "personal data".

XI. Conclusions regarding the Anonymisation Guidelines as they currently exist

In their current state, the Anonymisation Guidelines appear in my view to be more of a source of legal uncertainty than the type of clarification that the EDPB is supposed to bring by way of guidelines. They contain not only contradictions with case law but also contradictions between different paragraphs. What may on the face of it seem like guidelines intended to help understand anonymisation conditions is instead likely to discourage the use of apparently anonymous data, because of the complexity of the anonymisation assessments it demands.

Ultimately, I would recommend that the EDPB introduces throughout its Anonymisation Guidelines a slightly different set of key tenets regarding anonymisation and the concept of "personal data". Namely:

- Information cannot be presumed to be personal data, but must be demonstrated to be so – until then, the GDPR does not apply.

- Identification means not just distinguishing one natural person from others, but also being able to act upon that distinction in relation to that natural person.

- Identification techniques may be increasingly accessible, yet there can be no presumption that identification of one record makes an entire dataset relate to "identifiable" natural persons.

- When assessing means reasonably likely to be used for (re)identification, reasonableness is key – and reasonableness requires lawfulness and foreseeability, not mere hypotheticals or unlawful scenarios.

It should also seriously consider revising its "simplified" approach to anonymisation assessments, and propose more accessible probability-based questions to help guide smaller organisations through a topic that it appears to have made even more complex than it already is.

I hope that these comments can help push the needle in the right direction and mitigate the risk of significant inconsistencies remaining in the final version that the EDPB will adopt.

Thank you for taking these comments into consideration.

Peter Craddock

30 October 2026