Response to the EDPB Public Consultation on Guidelines 02/2026 on Anonymisation

Luxembourg National Data Service (LNDS) · 30 October 2026

About LNDS

The Luxembourg National Data Service (LNDS) was created by the Luxembourg Government to implement Luxembourg's strategies in research, innovation and digitalisation.

The mission of LNDS is to enable value creation from the secondary use of data, with transparency, control and trust. It focuses on the responsible reuse of public sector data for research and innovation, and supports public and private data partners with legal, ethical and societal impact assessment, secure end-to-end data processing, and support for data access and pseudonymisation. At European level, LNDS contributes to the European data spaces, including the support to the health data access body under the European Health Data Space (EHDS) Regulation. (LNDS – About us)

Anonymisation is central to this work: it determines when public sector data can be made available for secondary use, and under which safeguards. LNDS therefore has a direct interest in guidelines that protect data subjects and that provide more clarity around this controversially interpreted subject.

Introduction

This response is submitted on behalf of LNDS. We welcome Guidelines 02/2026 on Anonymisation (version 1.0) and support their core approach: anonymity is assessed from the perspective of each relevant entity including downstream actors in the processing chain, on the basis of means reasonably likely to be used, and tested against the No Record Isolation, No Linkage and No Inference criteria.

Our main concern is conceptual. The Guidelines define identification as distinguishing a person from others so that they can be treated differently (para 19), but they never say what an *identity* is, nor what "treat differently" means. Late in the text, in section 3.5.3, a decisive additional requirement appears: an isolated or linked record must be connected to the person's identity (paras 98–102, Examples 24 and 25). This requirement determines the outcome of most assessments, yet it is neither introduced nor explained where identification is explained. Several of our comments follow from this gap.

Our further comments concern the data subject as a relevant perspective, the assessment of entities along the processing chain, operational guidance on means reasonably likely to be used, and a number of examples. Paragraph and example numbers refer to version 1.0 of the Guidelines. Annex 1 lists all proposed changes paragraph by paragraph; Annex 2 lists the case law and EDPB documents relied on.

Key recommendations

1. **Provide an explanation for the concepts of identification and identity in the GDPR.** Add both to the Glossary, drawing on Article 4(1) GDPR, which refers to factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of a person. Recognise the two routes that Article 4(1) names: identifiers, which identify through the entity that binds them to a person, and factors specific to a person's identity, which identify where they are, alone or in combination, specific to that person. Clarify that identification does not require knowledge of the civil identity.
2. **Introduce the link to an identity where identification is defined.** Explain in section 2.4, not only in section 3.5.3, that a unique description singles a person out but identifies them only where it can be anchored to an identity of that person, either directly in the data itself or through information held by someone reasonably likely to contribute it (*OC v Commission*).
3. **Include the data subject as a relevant entity.** The data subject can provide additional information that allows the data holder to identify their record (Articles 11(2) and 12(2) GDPR).
4. **Acknowledge data that can be intrinsically personal.** Data containing factors that are, in themselves, specific to a person's identity, such as genetic data, which by definition give "unique information" about the person (Article 4(13) GDPR), in particular genetic data profiles, or biometric data, should be presumed personal and considered anonymous only in exceptional circumstances, e.g., when aggregated, following a thorough analysis (paras 8, 12, 24 and 50). The EDPB already proposes this for genetic data in its Guidelines 1/2026 (paras 165–166).
5. **Recognise persistent identifiers as a means to treat a person differently.** A pseudonym or other identifier that recurs in a context accessible to the data holder, so that the data holder can recognise the same person across datasets or over time and collect further information about them, allows that person to be distinguished and treated differently within the meaning of para 19, as Example 6 already recognises. The data are therefore personal for the data holder, and para 102 should be qualified for linkage through such identifiers even if the data itself is not (yet) identifying an individual.
6. **Clarify "treat differently".** It should cover processing that reaches beyond the given record to the person and is possible only because of an identifier or a factor specific to their identity: obtaining further data about them, linking data about them from other sources or over time, or acting towards them, including through other entities in the processing chain. Merely selecting or querying a record is not enough.
7. **Assess entities along the processing chain in more detail.** Personal data travel upstream, not downstream: data that the recipient can identify are also personal for the transferor (*EDPS v SRB*, paras 84–85; *Gesamtverband Autoteile-Handel / Scania,* C-319/22), while data that are personal for the transferor are not, for that reason alone, personal for the recipient, although their transfer is a transfer of personal data (*EDPS v SRB*, paras 77, 86 and 110–112). Where record-level data or results may return to an entity for which they are personal, they are personal for every entity in between. This also applies where only an identifier is passed on. Joint controllers jointly determine the processing of data that are personal for at least one of them, so the data are personal for all of them.
8. **Improve several examples and correct drafting errors** (Annex 1, which lists all proposed changes by paragraph of the Guidelines).

1. Identification: identity, identifiers and "treat differently"

1.1 The Guidelines need a notion of identity

The word "identity" appears only twice in the Guidelines, both times in passing: para 18 refers to keeping individuals' identities private, and Example 24 concludes that answers cannot be linked to respondents' identity. Neither the term nor the act of identification is defined in the Glossary, although "re-identification" is.

Article 4(1) GDPR offers a starting point. It provides that a person can be identified "by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity" of that person. The GDPR thus names two routes to identification, and recognises that a person has several identities, which are not necessarily the civil identity. The Court of Justice relies on the second route: in *Buivids*, it applied the reference to "factors specific to [a person's] physical … identity" and held that recorded images of persons who can be seen and heard are personal data (C-345/17, paras 30–32; see also *Ryneš*, C-212/13, paras 21–22). The same reasoning applies to the other dimensions of identity named in Article 4(1), such as genetic identity. As Article 4(1) refers to "one or more factors", such factors can also be combined across dimensions.

The two routes work differently:

- **Identifiers** are assigned to a person, such as an examination number, an IP address, a vehicle identification number, a forum nickname or a pseudonym. They identify because an entity binds them to one person: the examination body, the internet service provider, the registration authority, the forum operator or the pseudonymising controller. They can lead to the person's identity through that entity, but they also allow that person to be recognised, tracked and treated differently without any identity being reached.
- **Factors specific to a person's identity** identify because they are part of the person. A genome, a face or fingerprints are specific in themselves and need no guarantor; the person can always prove the match. Other identity factors, such as sex, nationality or profession, are shared by many people and only become specific in combination.

The civil identity illustrates the second point. It consists of a set of factors recorded by a civil registry, such as name, date of birth, address and national identification number. A name connects directly to the person, as the Court recognised in *Nowak* (C-434/16, para 29), where it distinguished direct identification through the candidate's name from indirect identification through an examination number. But a name alone is often not specific enough: in *Rīgas satiksme* (C-13/16, para 30), the first name and surname did not allow the person to be identified with the precision required to bring an action against them, so that the address or identification number was also needed. A name is therefore unique only in some contexts, which para 23 of the Guidelines could make clearer. The national identification number sits on the boundary: it is assigned, like an identifier, but it is the registry's key to the civil identity.

The term “name” covers different things. A first name and surname are part of the civil identity, recorded and guaranteed by a registry. A name chosen within a particular context, such as a nickname in an online forum, is an identifier within that context, without being a factor of the person’s identity.

On this basis, we suggest distinguishing four kinds of information:

Kind of information	Examples	How it identifies
Factors specific to an identity	genome, face, fingerprints	intrinsically, where specific to the person
Civil identity	name, date of birth, address, national identification number	in combination; a name connects directly to the person but is often not specific on its own
Assigned identifiers	examination number, IP address, VIN, nickname, pseudonym	through the entity that binds them to a person; they also allow tracking and treating differently without reaching an identity
Actions, statements and circumstantial information	an opinion, a traffic violation, a grant amount, a relative's occupation	only by being linked to one of the above

Some identity factors can change: faces can be altered, names changed, and phone numbers or email addresses reassigned to other people over time. This affects the likelihood of identification over time, but not the nature of the factor at a given moment.

1.2 Uniqueness is not yet identification

Para 19 defines identification as distinguishing a person from others in a given context, and para 24 suggests that a unique combination of attributes is enough. Example 5 illustrates this with a person on CCTV who is the only one in a black suit walking a dog. The example attributes identification to the wrong element. The description “the person in a black suit who is also walking a dog” contains neither an identifier nor a factor specific to the person’s identity. On its own, it does not allow the person to be found again, linked to other data or treated differently: on another day, the same person may wear a different suit and come without the dog, and another person may wear a suit and walk a dog. Whoever can act on that person in the scene can do so because they observe the person directly; in the CCTV setting, it is the image that may identify them, inasmuch as it makes identification possible (*Ryneš*, paras 21–22; *Buivids*, para 31). The example should make clear that a combination of circumstantial features that singles out a person within an observation does not, for that reason, identify them.

A further illustration: a researcher records red-light violations for a road-safety study at a junction without traffic cameras. “The driver who crossed the red light there on Friday at 10 pm” describes exactly one person in the world. The time and place are location data, but a single observation does not recur and is not, on its own, linked to the person. Neither the researcher nor the driver can link the record to the person unless further information is obtained that connects the observation to them, such as a time-stamped photo, a number plate leading to the registered keeper, or location records held by a mobile network or a car manufacturer. Such information is not reasonably likely to be available to the researcher.

Compare this with a forum user known as "LuckyKid". Other users have no access to the email address or civil identity behind the nickname. But the forum operator ensures that LuckyKid is always the same account, and that a newcomer has to choose another name, such as "LuckyKid2". Other users can therefore recognise LuckyKid over time, profile their contributions and target them with replies. The red-light description is a single observation that does not recur: nothing allows it to refer to the same person beyond that moment. The nickname is an identifier, and that is enough for LuckyKid to be identifiable within the forum (section 1.3).

The Court's reasoning in *OC v Commission* (C-479/22 P) supports this distinction. The published description was specific to one person, but the Court went on to examine how it could be linked to her identity. People in her scientific field could recognise her from their knowledge of her professional background (para 60). Journalists and others could use circumstantial details, such as the host institution and the amount of funding, to search the funding agency's public list of some 70 projects, where those details appeared next to the name of the project manager, and run further internet searches (paras 62–63). The traits specific to her identity that were published, her sex and nationality, were shared by many people. It was the circumstantial information, matched against knowledge held by people and against public records, that made identification possible. This also reflects Recital 26 GDPR, which refers to singling out as a means used to identify a person, not as identification in itself.

The Guidelines reach the same result, but only at the end. In section 3.5.3, singling out requires matching a record with the individual's corresponding attributes (para 98); Example 24 finds survey records anonymous because they cannot be linked to the respondents' identity; and paras 101–102 require linked data to lead to the identifiability of the individual. This is a decisive criterion, and it should be introduced in section 2.4, where identification is defined, and reflected in Examples 1 and 5.

Proposal: state in para 19 that identification requires that a person is distinguished from others *and* that the description can be anchored to a trait specific to their identity, either directly or through additional information, including knowledge held by people, reasonably likely to be obtained.

Proposed wording for para 19: “To identify a natural person means to distinguish that person, not merely a record, from others within a given context, in a way that makes it possible to treat them differently. This can be done using identifiers, such as a name, an identification number, location data or an online identifier, that are unique within that context, or one or more factors specific to the person’s identity. A description that distinguishes a record without allowing it to be connected to the person, directly or through additional information reasonably likely to be obtained, does not in itself identify them (see section 3.5.3).”

1.3 Identifiers and persistent pseudonyms

The distinction between identifiers and identities needs to be drawn more carefully. An identifier not built on an identity, such as a pseudonym, needs additional information to establish who the person is. It may nevertheless be enough to track a person and to treat them differently: to address a pseudonymous user in a forum, to request additional data about the person behind a pseudonym, or to link regular update files received from a hospital.

Paras 101–102 provide that, where data can be linked to another dataset, the linked information should be re-tested, and that it is anonymous if it passes the three criteria. For linkage through non-unique attributes, as in Example 25, this is convincing: such linkage connects a record to a group of people sharing the same attributes, not to one person. Not all linkage relies on a persistent identifier, and for such linkage the re-test in para 102 is the right approach.

Linkage through a unique and persistent identifier is different. Where an identifier allows the data holder to recognise that records in different datasets, or received at different times, belong to the same person, that person is distinguished from others and can be treated differently, which is how para 19 defines identification. Example 6 already draws this conclusion for a pseudonym used to link data across websites. Para 102, as currently worded, would allow such data to be considered anonymous as long as the linked record does not, on its own, allow singling out. This is inconsistent with para 19 and Example 6. It also overlooks that with each linkage the record grows, and para 56 recognises that larger records are more likely to be unique.

This view is supported by the Guidelines themselves and by the GDPR:

- para 19 lists a "reference code" among direct identifiers, so an assigned number can identify;
- Example 6 treats a pseudonym used to link data across websites as an identifier, even though the third party never learns a name;
- Article 4(1) GDPR names identification numbers and online identifiers without requiring them to be meaningful;
- Recital 30 GDPR recognises that identifiers can leave traces which, combined with other information, can be used to create profiles of individuals and identify them.

In our view, what matters is not whether the identifier is random or a real identification number, but its **reach**: whether it is unique, persistent and usable across datasets or over time to accumulate information about the same person. A code that the recipient cannot use to obtain anything further is a dead end. A code that remains a live key across datasets functions like an identification number. Where the pseudonymising entity that holds the key remains part of the processing chain, for example by delivering further data under the same pseudonym, the identifier also remains linked to an identity through that entity; and the information accumulated under it can progressively form a behavioural, economic or social identity in its own right.

A randomly assigned number that occurs only once, within a single dataset, distinguishes records, not persons; this is what the No Record Isolation criterion addresses. Persistence alone does not change this. What matters is whether the identifier recurs in a context accessible to the data holder either directly or indirectly, i.e. through other entities processing data related to the identifier: in further deliveries of data, in other datasets, or through a channel for requesting further information. A respondent number in a one-off survey is a dead end, as in Example 24; the same random number used across the waves of a panel survey allows the researcher to follow the person over time and treat them differently.

What makes an assigned identifier reliable is the entity that binds it to an accountable person: the internet service provider for an IP address (*Breyer*, C-582/14, paras 43–44 and 46–49), the registration authority for a vehicle identification number (*Gesamtverband Autoteile-*

Handel), the examination body for an examination number (*Nowak*), the forum operator for a nickname, or the pseudonymising controller for a pseudonym. The data holder can rely on this binding without having access to what lies behind it.

Shared use of the identifier does not remove identifiability. An IP address identifies the subscriber, not the person at the keyboard; a VIN identifies the keeper, not the driver. The data then relate to that person, who may be held responsible for the use of the connection, account or vehicle, and they make each of the few co-users identifiable with little additional information. The Glossary already recognises that the same information may relate to more than one individual. Identifiability is lost only where an identifier is shared by so many people, for example a public IP address shared by thousands of users, that it no longer meaningfully narrows the population.

Proposal: state that an identifier that recurs in a context accessible to the data holder allows the data holder to distinguish the person and treat them differently within the meaning of para 19, as Example 6 already recognises, so that the data are personal for the data holder even where the linked data does not on its own allow singling out. Para 102 should be qualified accordingly, while Example 25 can remain as an illustration of linkage through non-unique attributes. Example 4 should also note that, if the research institute can request further or longitudinal data on the same patients, the records must remain attributable to the same persons, which brings the No Linkage criterion into play.

Proposed addition to para 102: “Where the linkage relies on an identifier that recurs across datasets or over time in a context accessible to the data holder, directly or through other entities, such as a persistent pseudonym, the individual is distinguished from others within the meaning of para 19, and the linked data should be considered personal for that data holder.”

1.4 What it means to "treat differently"

The ability to treat a person differently is central to the definition in para 19, but it is not explained. Two readings are possible. An effect-based reading requires an action with an external effect on the person, such as a decision, an advertisement delivered or a contact. A processing-based reading covers processing that reaches beyond the given record to the person, and that is possible only because the data contain an identifier or a factor specific to the person’s identity: obtaining further data about the same person, linking data about them from other sources or over time, or acting towards them. Merely selecting or querying a record within the given data is not enough, as this is possible for any record. The decisive point is that such processing is possible on the basis of the data. Where a person can be acted upon only because they are directly observed, it is the observation, not the data, that enables the action.

We recommend the processing-based reading. The Court has held that the retention of data is in itself an interference with private life, that a further interference arises when authorities access the data, and that it does not matter whether the information is sensitive or whether the persons concerned have been inconvenienced in any way (*Digital Rights Ireland*, C-293/12 and C-594/12, paras 29 and 33–35). The Court has thus set a very low threshold for an interference with the rights to privacy and data protection. This low threshold should be taken into account when determining what it means to be able to “treat differently” a person: it would be inconsistent to require a tangible effect on the person for identification, when the

Court does not require any effect for an interference. Article 11 GDPR also shows that the legislator anticipated personal data, which the controller cannot attribute to a person, without treating such data as anonymous. The practical difficulty of fulfilling obligations towards a person is therefore not a reason to consider data anonymous.

This matters for pseudonymised data. If a data holder can use a pseudonym to request additional data about the person behind it, the request is targeted at that individual and leads to the processing of their data by the provider. This holds whether the provider acts as a joint controller, a processor or a third party. A request for information about a group of people that does not separate individuals is different: it does not target any one person.

The precision needed for identification also depends on the treatment intended. In *Rīgas satiksme* (C-13/16, para 30), a name was not precise enough to bring legal action against the person, so that an address or identification number was also needed. To reply to a person in an online forum, a nickname is enough. The "given context" in para 19 should therefore be understood as including the kind of treatment for which the person is to be distinguished, not only the population from which they are distinguished.

Effects may also arise through the processing chain rather than directly. Where the recipient's analysis feeds into decisions taken by another entity about the persons concerned, those persons are affected even if the recipient cannot reach them.

The same question arises in two other places. Para 50 lists as a simple case data used to evaluate, treat or influence a specific individual, without explaining what evaluation means if it has no effect on the data subject. And paras 70–71 require an inference to have an effect on the data subject's rights and interests in order to be meaningful, The Guidelines place this requirement under “relating to” (para 71; see *Nowak*, C-434/16, para 35); it should be made clear that it does not limit identifiability.

Proposal: explain "treat differently" as covering processing that reaches beyond the given record to the person and is possible only because of an identifier or a factor specific to their identity, including affecting them directly, requesting additional data about them and effects produced through other entities in the processing chain.

1.5 Data that is intrinsically personal

Para 8 frames identifiability entirely as a question of likelihood, and para 12 starts from the question "for whom is the data intended to be anonymous?". Both should be qualified. Some data contain factors specific to a person's identity that identify them in themselves, such as genetic or biometric data. The Guidelines mention genetic attributes only once, in para 24, as one of several attribute types that may become identifying in combination. This leaves room for the argument that genomic data processed in a secure processing environment is anonymous for a user who has no reference data to match it against.

The case law of the Court of Justice does not support this argument. The Court has held that fingerprints constitute personal data "as they objectively contain unique information about individuals which allows those individuals to be identified with precision" (*Schwarz*, C-291/12, para 27). Their identifying capacity lies in the data, not in the knowledge of the data holder. Likewise, the Court has applied the reference in the definition of personal data to "factors specific to [a person's] physical identity" to recorded images, which are personal data

inasmuch as they make it possible to identify the person (*Ryneš*, C-212/13, paras 21–22). In *Buivids* (C-345/17, para 32), it was sufficient that the persons recorded could be seen and heard; the Court did not ask whether viewers knew their names. As Advocate General Pitruzzella put it, there is an "intrinsic link between biometric data and the identity of persons" (Opinion in C-70/18, point 28).

The need for a reference sample does not change this. In *Schwarz*, the Court of Justice relied on the judgment of the European Court of Human Rights in *S. and Marper v United Kingdom* (§§ 68 and 84), where the argument that fingerprints are unintelligible "without a comparator fingerprint" was expressly considered and held unable to "alter the fact that fingerprints objectively contain unique information about the individual concerned, allowing his or her identification with precision" (§ 84). For genomic data, the same applies all the more.

This is consistent with the relative approach of *EDPS v SRB*. That case concerned comments and alphanumeric codes, not data that objectively identify. For factors specific to a person's identity, the relative test always yields the same result: means of identification reasonably likely to be used exist, because the data subject can provide a reference and prove the match (Article 11(2) GDPR).

Opinion 05/2014, which these Guidelines update, warned that genetic data profiles remain at risk of identification where the only technique used is the removal of the donor's identity, as public genealogical resources combined with donor metadata can reveal who the donor is (p. 10). This warning has not been carried over. In its Guidelines 1/2026 on processing of personal data for scientific research purposes (version for public consultation, paras 165–166), the EDPB goes further: relying on Opinion 05/2014 and *Schwarz*, it proposes that genetic data profiles be considered personal data by default and anonymous only in exceptional circumstances, and recognises that biometric data are similarly unique to individuals. The two sets of guidelines should be aligned.

Proposal: add to paras 8, 12 and 24 that data containing factors that are, in themselves, specific to a person's identity, such as genetic data within the meaning of Article 4(13) GDPR, which give "unique information" about the person, in particular genetic data profiles, and biometric data, are presumed to be personal data and can be considered anonymous only in exceptional circumstances, following a thorough analysis, as Guidelines 1/2026 already propose for genetic data profiles; and add such data to the simple cases in para 50.

2. The data subject as a relevant perspective

The Guidelines look at re-identification almost entirely from the outside. Yet, the person best placed to provide missing information to establish a record is the person it describes, and the GDPR gives them a legal route to make themselves identifiable to the data holder: under Article 11(2), the rights in Articles 15 to 20 apply where the data subject provides additional information enabling identification, and under Article 12(2) the controller may refuse to act only if it demonstrates that it cannot identify the data subject. A controller should not be able to avoid this by declaring the data anonymous in the first place. This would be contrary to the aim of the GDPR to ensure a high level of protection of natural persons and to strengthen the rights of data subjects, as the Court has emphasised in the context of the right of access (*FT*,

C-307/22, para 47). The Court has also interpreted the concept of controller broadly to ensure the effective and complete protection of data subjects (*Google Spain*, C-131/12, para 34; *Wirtschaftsakademie*, C-210/16, para 28); the same objective should guide the interpretation of personal data, a concept to which the Court has assigned a wide scope (*Nowak*, C-434/16, para 34). An interpretation of identifiability that disregards the data subject's own ability to provide the additional information referred to in Article 11(2) would deprive them of rights the GDPR expressly grants. The objective of a high level of protection weighs against such an interpretation.

What matters is not whether the data subject would recognise the record, but whether they can provide additional information, as Article 11(2) GDPR envisages, that would allow the data holder to identify their record. This is generally possible where the record contains factors specific to the person's identity, or an identifier they can prove is theirs: the data subject can provide proof of those factors, such as a reference sample, their date of birth and postcode, or access to a pseudonymous account, or other information that the data holder does not itself have. It is generally not possible for a record of a single event that contains neither (section 1.2). Unless the data have been transformed, for example by generalisation, perturbation or randomisation, to the point where the information the data subject could provide no longer matches a specific record, the data should be presumed to remain identifiable for the data holder. Circumstantial information, such as answers the data subject recalls, may indicate that a record is theirs but does not prove it, as others may know the same information. The argument therefore applies where the data subject can prove the match through factors specific to their identity, or through an identifier they can show is theirs (see also Article 12(6) GDPR).

We therefore recommend:

- adding the data subject to the list of relevant entities in para 30, alongside partners, friends, colleagues and neighbours, whose relevance the Court confirmed in *OC v Commission* (para 60);
- adding to para 12 that the data subject's perspective must always be considered;
- considering whether section 2.2 could be simplified by starting with the data subject: could the person provide the missing information? Only where this is excluded, for example because the data has been generalised, permuted or randomised, would it be necessary to assess other entities and their place in the processing chain.

3. Perspectives along the processing chain

The contextual approach assesses each entity "based on their respective capabilities" (executive summary). Entities should not be assessed in isolation, but within the processing chain: who may receive the data next, who may act on their behalf, who may provide additional information, and whether the data, or record-level results derived from them, may return to an entity for which they are personal. Where they may, the data are personal for every entity in between, as follows from para 21 and as the condition in Example 4 that "the data will never be returned to the hospital" acknowledges. Para 9, footnote 7 and para 28 already contain elements of this; it should be stated in the executive summary and in section 2.2.

Upstream versus downstream. Personal data downstream in the processing chain have an upstream effect: data that are personal for a downstream entity are also personal for any upstream entity. As the Court held in *EDPS v SRB* (paras 84–85), referring to *Gesamtverband Autoteile-Handel* (paras 46 and 49), data that are in themselves impersonal become personal where the controller puts them at the disposal of persons who have means reasonably likely to enable the data subject to be identified, and they are then personal "both for those persons and, indirectly, for the controller". Conversely, data that are personal for the transferor are not, for that reason alone, personal for the recipient, although the transfer itself remains a transfer of personal data (*EDPS v SRB*, paras 75–77 and 86; on the transfer, paras 110–112). Para 21 states the first point and Example 2 the second; a single statement of this asymmetry would help. It applies equally where data flow from a provider to an entity that requested them: if the requesting entity can identify the persons, for example through a pseudonym, the data are personal for the provider too.

Example 2. The statement that the personal nature of the data should "only" be assessed from the controller's perspective at the time of collection may be misread. It explains why the collecting controller's view governs the information obligation, but it should not suggest that recipients need not be assessed where the data would be personal for them, nor that the assessment is made only once, at collection. We also note that the Court did not decide that the data were anonymous for the recipient in that case. In the same judgment, the Court held that data become personal for the controller where they are put at the disposal of recipients who can identify the data subjects (para 84).

Partial transfers. Para 12 considers the transfer of the data as a whole. An entity may also pass on only some features, such as a pseudonym, which another entity uses to process data with effect on the individual. Where this leads to the processing of the individual's personal data, the perspective of that other entity should also be considered.

Processors and joint controllers (para 15 and Example 3). We support treating the processing entity as a processor. The rationale should also rest on the principle that using a processor does not change the nature of the processing: the controller remains responsible for processing it contracts out. Example 3 should also be more specific: a simple statistical analysis fits the processor scenario. But where the agency, in the course of the same analysis, also uses the data to train its own algorithms, the purposes converge and the agency may become a joint controller with the retailer, as in *Wirtschaftsakademie* (C-210/16) and *Fashion ID* (C-40/17, paras 74 and 78–81), and as EDPB Guidelines 8/2020 on the targeting of social media users illustrate (Example 1, section 5.2.1; see also Guidelines 07/2020, para 60). The perspective of joint controllers is not explicitly discussed so far: since joint controllers jointly determine the purposes and means of processing data that is personal for at least one of them, the data is personal for both, as footnote 7 already indicates with reference to *Wirtschaftsakademie* (C-210/16).

Data minimisation. Where a recipient has no need to re-link records to the persons concerned, the pseudonym should not be passed on, but replaced by a recipient-specific token that is not mapped to it. The Guidelines could recommend this as a practical measure.

Proposals:

- state in the executive summary and in section 2.2 that entities are assessed within the processing chain, including whether data or record-level results may return to an entity for which they are personal (paras 9 and 21, Example 4);
- state the asymmetry between upstream and downstream in one place, with reference to *EDPS v SRB*, para 84 (para 21 and Example 2);
- clarify in Example 2 that "only" does not make the recipient's perspective irrelevant, nor limit the assessment to the time of collection;
- add to para 12 that an entity receiving only an identifier, such as a pseudonym, and using it to process data about the person must also be assessed;
- base the processor rule in para 15 also on the principle that using a processor does not change the nature of the processing; make Example 3 more specific and address the perspective of joint controllers;
- recommend replacing pseudonyms with recipient-specific tokens that are not mapped to them where the recipient has no need to re-link records.

4. Means reasonably likely to be used and relevant entities

4.1 Skills, effort and interest

Example 1 distinguishes the average person from investigative journalists with research skills. With AI tools widely available, this distinction is losing its meaning: any person can now carry out research that previously required specialist skills. Paras 92–93 rightly note that AI reduces the time and cost of re-identification; Example 1 should reflect this. The Guidelines also note that an entity may be unable to run a re-identification technique itself (para 30), while Example 7 shows that such an analysis can be bought in. This should be stated as a general rule: since subcontracting is always possible, lack of skills cannot exclude a means that can easily be obtained from others.

OC v Commission offers two concrete benchmarks. First, the Court held that browsing the descriptions of some 70 funded projects, combined with other internet searches, did not involve disproportionate effort, so that the risk of identification by journalists or other persons unfamiliar with the person's background could not be regarded as insignificant (para 63). Second, the Court considered that a press release about alleged fraud was likely to arouse public interest and prompt such searches. Para 31 rightly rejects lack of motivation as a factor. For the same reason, a low value of the information to a potential attacker should not, on its own, be used to exclude means that require little effort. Conversely, as *OC v Commission* shows, public interest in the information increases the likelihood that such means will be used (para 63).

Proposal: for publication or open release, assume a motivated recipient with access to AI tools, and use the effort accepted in *OC v Commission* as a reference point.

4.2 Unlawful actors

The list of potential re-identifying entities in para 30 is useful as a catalogue, but it does not say when each category must actually be expected, and footnote 25 adds that not all entities need to be considered systematically. Without criteria, controllers may either ignore the list or apply it to every case.

- **Cybercriminals.** The Guidelines link the relevance of unauthorised actors to the quality of security measures (paras 33 and 89, Example 8). However, para 89 suggests that appropriate security controls exclude unauthorised actors from the assessment, while in Example 22 the research institute considers them despite state-of-the-art security, because data may leak over a long period of time. The Guidelines should clarify which approach applies, and state that the period for which the data are held must be taken into account. Considering such actors upfront avoids a reassessment after an incident (para 42).
- **Foreign intelligence agencies** become a relevant factor where data is held in, or transferred to, IT systems of companies subject to the law of third countries that grant such agencies access.
- **Processors acting beyond their instructions**. Such processing is unlawful but common. It is unclear whether it falls under "unethical companies"; the Guidelines should clarify this.
- **Rebutting the presumption of lawful behaviour (para 33).** The factors are sound. Repeated fines imposed on the same companies for comparable infringements show that sanctions do not always stop the conduct concerned. Such cases could illustrate the second and fourth factors. The EHDS illustrates this: although it prohibits re-identification (Article 61(3) of the EHDS Regulation), sanctions fall mainly on the user organisation rather than on the individual researcher, and users in third countries may be beyond effective enforcement, so that the gain from re-identification may outweigh the risk for an individual.

5. Further issues

The following points are not addressed in the Guidelines and are of particular relevance for the reuse of public sector data:

- **Pseudonymisation.** *EDPS v SRB* concerned pseudonymised data, but the Guidelines do not mention pseudonymisation. They should explain how the per-entity perspective relates to Article 4(5) GDPR and to the EDPB's guidance on pseudonymisation.
- **Data spaces and secure processing environments.** The introduction cites data spaces as a reason for the update, but the Data Governance Act, the EHDS and secure processing environments are not discussed. This is where the contextual approach matters most, including the checking of aggregate outputs before they leave the environment. Para 58 states that aggregate data will "usually" satisfy No Record Isolation; guidance on small cell counts and output checking is needed.
- **Irreversible releases.** Paras 35 and 42 rely on periodic reassessment, and para 88 recommends safety margins. Published data, however, cannot be recalled: for releases that cannot be undone, the safety margins should be set with this in mind at the time of release.
- **Burden of proof.** Footnote 36 states that the simplified approach cannot be used by third parties to prove that data is personal. Para 41 already requires documentation demonstrating the effectiveness of the anonymisation; the Guidelines should confirm that, under the accountability principle (Article 5(2) GDPR), it is for the controller to demonstrate anonymity.
- **Positive examples.** Some examples end with the data being anonymous (e.g. Examples 22 and 24), but none shows how a specific anonymisation technique

achieves this. Worked examples of record-level data anonymised with particular techniques would help controllers apply the Guidelines and support responsible reuse.

Conclusion

The Guidelines provide a sound and much-needed framework. Our main request is that they make explicit what is now only implicit: that identification requires a link to a person's identity, that such a link can be held by anyone reasonably likely to contribute it, including the data subject, the people around them and other entities in the processing chain, and that data containing identifiers that recur in a context accessible to the data holder, or factors specific to a person's identity, are as a rule personal. With these clarifications, the Guidelines would better protect individuals while giving controllers, and organisations enabling the responsible reuse of public sector data, the legal certainty they need. We would welcome the opportunity to discuss these comments further.

Annex 1: Overview of proposed changes by paragraph

The table lists all proposed changes in the order of the Guidelines. Substantive proposals are argued in the sections indicated; clarifications concern unclear passages or examples; drafting points concern errors and inconsistencies.

Paragraph	Proposed change	Type	See
Executive summary; para 9	State that entities are assessed within the processing chain, not in isolation, including whether data or record-level results may return to an entity for which they are personal.	Substantive	Section 3
Para 8	Qualify the framing of identifiability as a question of likelihood only: data containing factors that are, in themselves, specific to a person’s identity are presumed personal.	Substantive	Section 1.5
Paras 11–12; Glossary	Use one meaning of “relevant entity”: the entities for whom data should be anonymous (paras 11–12) or the entities that could contribute to identification (Glossary).	Clarification	–
Para 12	Add that the data subject’s perspective must always be considered; that data containing factors specific to a person’s identity are presumed personal whoever holds them; and that entities receiving only an identifier and using it to process data about the person must also be assessed.	Substantive	Sections 1.5, 2 and 3
Example 1	Reflect that the Court in *OC v Commission* examined how the description could be linked to the person’s identity; note that AI tools reduce the difference between the average person and investigative journalists.	Substantive	Sections 1.2 and 4.1

Paragraph	Proposed change	Type	See
Section 2.2	Consider starting the assessment with the data subject: could they provide the missing information? Only where this is excluded would other entities need to be assessed.	Clarification	Section 2
Example 2	Clarify that "only" does not make the recipient's perspective irrelevant, nor limit the assessment to the time of collection; refer to *EDPS v SRB*, para 84.	Clarification	Section 3
Para 15	Base the processor rule also on the principle that using a processor does not change the nature of the processing.	Clarification	Section 3
Example 3	Distinguish a simple statistical analysis from an analysis in which the agency also trains its own algorithms, where converging purposes may make it a joint controller (*Wirtschaftsakademie*; *Fashion ID*); address the perspective of joint controllers.	Substantive	Section 3
Example 4	Add caveats on a retained pseudonymisation key or records that can be mapped back to persons; on requests for further or longitudinal data, which require records to remain attributable to the same persons; and on the condition that the data are never returned.	Substantive	Sections 1.3 and 3
Para 16(a)	Information relates to a person by its content also where it expresses that person's view: in *Nowak*, the examiner's comments also constituted information relating to the examiner (C-434/16, paras 34 and 44).	Clarification	–
Para 16(b)–(c)	Explain that purpose and effect can occur independently, for example effect without purpose through publication, with examples of each; clarify where the vehicle identification number in *Gesamtverband Autoteile-Handel* fits (in our view, effect).	Clarification	–
Para 19; Glossary	Explain identification by reference to the two routes in Article 4(1) GDPR, identifiers and factors specific to a person's identity, and add explanations of identification and identity to the Glossary.	Substantive	Section 1.1
Para 19	Introduce here, not only in section 3.5.3, that identification requires that a description can be anchored to an identity of the person, directly or through information reasonably likely to be obtained.	Substantive	Section 1.2
Para 19	Explain "treat differently" as processing that reaches beyond the given record to the person	Substantive	Section 1.4

Paragraph	Proposed change	Type	See
	and is possible only because of an identifier or a factor specific to their identity; explain that the "given context" includes the treatment intended.		
Para 19	Clarify whether "attributes which relate to that person" refers to relating by content only.	Clarification	–
Para 20	Summarise the three routes to indirect identification plainly: obtaining more information, using methods controlled by others, and having others perform the identification.	Drafting	–
Para 21	State the asymmetry in one place: data that the recipient can identify are personal for the transferor (*EDPS v SRB*, para 84), but data that are personal for the transferor are not, for that reason alone, personal for the recipient.	Substantive	Section 3
Para 23	Make clear that a name is unique only in some contexts (*Rīgas satiksme*, C-13/16, para 30).	Drafting	Section 1.1
Para 24	Distinguish factors that are, in themselves, specific to a person's identity, such as genetic and biometric data, from attributes that become identifying only in combination; retain the warning of Opinion 05/2014 on genetic data profiles and align with Guidelines 1/2026 (paras 165–166).	Substantive	Sections 1.1 and 1.5
Example 5	Explain why the person is identifiable: the attributes are not specific to their identity, and the combination is unique only at the moment of observation.	Clarification	Section 1.2
Para 30; Example 7	State as a general rule, building on Example 7, that lack of skills cannot exclude a means that can easily be subcontracted.	Clarification	Section 4.1
Para 30	Add the data subject to the list of relevant entities.	Substantive	Section 2
Para 30; footnote 25	Give criteria for when each category of entity must be considered.	Clarification	Section 4.2
Para 30	Clarify whether unauthorised actors are excluded by appropriate security controls (para 89) or must be considered over the period for which the data are held (Example 22).	Substantive	Section 4.2
Para 30	Note that foreign intelligence agencies are relevant where data are held in, or transferred to, IT systems of companies subject to the law of third countries that grant them access.	Clarification	Section 4.2

Paragraph	Proposed change	Type	See
Para 30	Clarify whether processing by a processor beyond the controller's instructions falls under "unethical companies".	Clarification	Section 4.2
Para 31	State that a low value of the information to an attacker should not, on its own, exclude means that require little effort, and that public interest increases the likelihood of identification (*OC v Commission*, para 63).	Clarification	Section 4.1
Para 33	Illustrate the factors with repeated fines for the same type of infringement and with the EHDS, where sanctions fall mainly on the user organisation.	Clarification	Section 4.2
Example 8	Rewrite the example and explain how the possibility of hacking affects whether the data are anonymous for the authority.	Clarification	–
Para 34	Note that the joint guidelines on the DMA and the GDPR are cited in their consultation version, which may change.	Drafting	–
Paras 35 and 42	Building on the safety margins in para 88, require that releases which cannot be undone, such as published data, are assessed with this in mind at the time of release.	Substantive	Section 5
Footnote 36	Confirm that, under Article 5(2) GDPR, it is for the controller to demonstrate anonymity.	Substantive	Section 5
Para 50	Add data containing factors specific to a person's identity to the simple cases; explain what evaluating a person means if it has no effect on them.	Substantive	Sections 1.4 and 1.5
Para 58	Provide guidance on small cell counts and on checking aggregate outputs, in particular in secure processing environments.	Substantive	Section 5
Para 69	Add an example: a dataset shows that people with gene XY tend to develop breast cancer; one can infer that Mrs Smith, who has gene XY and is not in the dataset, has a high risk. The inference relates to her; the dataset does not.	Clarification	–
Paras 70–71	Make clear that the requirement of an effect on rights and interests for an inference to be meaningful, which the Guidelines place under "relating to" (para 71), does not limit identifiability.	Clarification	Section 1.4
Para 71(b)	"A lack of reliance" should read "reliance".	Drafting	–
Para 71(c)	"Rebuttal presumption" should read "rebuttable presumption".	Drafting	–

Paragraph	Proposed change	Type	See
Footnote 44	The references to footnote 40 and to "A(D′) in footnote 41" appear to be to footnotes 42 and 43.	Drafting	–
Para 89	"A mean" should read "a means".	Drafting	–
Para 97 (table)	Correct the reference to "paras. 98 to 00".	Drafting	–
Section 3.5.3; Example 24; paras 98–102	The requirement of a link to an identity appears here for the first time; introduce it in section 2.4.	Substantive	Section 1.2
Para 98	Clarify "interactions with the individual" and refer to the data subject's perspective.	Clarification	Section 2
Para 102; Example 25	Qualify para 102 for linkage through an identifier that recurs in a context accessible to the data holder; keep Example 25 as an illustration of linkage through non-unique attributes.	Substantive	Section 1.3
Annex 1 (flowchart)	Correct the cross-references to paras 103–104 (the text ends at para 103) and to paras 100–102 for singling out (paras 98–100); "means reasonably like to be used" should read "likely".	Drafting	–
Throughout	Use "contextual" and "simplified" consistently, rather than also "contextualised" and "simple"; "Example 1O" is written with a letter O.	Drafting	–
Various	Para 20: "reasonable likely" should read "reasonably likely"; para 52: "does not necessarily mean that the information must necessarily be considered" repeats "necessarily"; para 53: "reintroduced sub-section 3.5" should read "reintroduced in sub-section 3.5"; Example 14: "uses this insight to" should read "applies this insight to"; para 85: "included in the record" should probably read "included in the dataset"; para 93: "particular seen" should read "particularly seen"; Example 24: "No Records Isolation" should read "No Record Isolation".	Drafting	–
General	Address pseudonymisation and its relation to Article 4(5) GDPR and to the EDPB's guidance on pseudonymisation.	Substantive	Section 5
General	Address data spaces, the Data Governance Act, the EHDS and secure processing environments.	Substantive	Section 5
General	Recommend replacing pseudonyms with recipient-specific tokens where the recipient has no need to re-link records.	Clarification	Section 3

Paragraph	Proposed change	Type	See
General	Add worked examples showing how specific anonymisation techniques make record-level data anonymous.	Substantive	Section 5

Annex 2: Case law and EDPB documents referred to

Case or document	Point relied on
C-291/12 *Schwarz*, para 27	Fingerprints are personal data as they objectively contain unique information allowing identification with precision.
C-293/12 and C-594/12 *Digital Rights Ireland*, paras 29, 33–35	Retention of data is in itself an interference; access is a further interference; sensitivity and inconvenience are irrelevant.
C-131/12 *Google Spain*, para 34	Broad interpretation of key concepts to ensure the effective and complete protection of data subjects.
C-212/13 *Ryneš*, paras 21–22	Images are personal data inasmuch as they make it possible to identify the person, by reference to factors specific to their physical identity.
C-582/14 *Breyer*, paras 43–44 and 46–49	An IP address identifies through the internet service provider that binds it to a subscriber; means reasonably likely to be used include means available through third parties.
C-13/16 *Rīgas satiksme*, para 30	A name alone may not identify a person with the precision required for a given purpose.
C-434/16 *Nowak*, paras 29, 34–35 and 44	Direct identification through the name, indirect identification through an examination number; information relates to a person by reason of its content, purpose or effect; an examiner's comments also relate to the examiner.
C-210/16 *Wirtschaftsakademie*, para 28	Joint controllership does not require access to the personal data; effective and complete protection of data subjects.
C-40/17 *Fashion ID*, paras 74 and 78–81	Joint controllership where the processing serves converging purposes of the parties concerned.
C-345/17 *Buivids*, paras 30–32	Recorded images of persons who can be seen and heard are personal data.
C-70/18, Opinion of Advocate General Pitruzzella, point 28	The intrinsic link between biometric data and the identity of persons.
C-307/22 *FT*, para 47	The GDPR aims to ensure a high level of protection of natural persons and to strengthen the rights of data subjects.

Case or document	Point relied on
C-319/22 *Gesamtverband Autoteile-Handel / Scania*, paras 45–49	Identifiers such as a VIN are personal data where they can be linked to a person; data can be personal for the transmitting entity.
C-479/22 P *OC v Commission*, paras 60, 62–64	Identification through knowledge held by people in the person's environment and through public records; effort and public interest as factors.
C-413/23 P *EDPS v SRB*, paras 75–77, 82, 84–86, 102–112	Insignificant likelihood standard; perspective of the collecting controller for the information obligation; data personal for the transferring entity.
ECtHR, *S. and Marper v United Kingdom*, § 84 (as relied on in *Schwarz*)	The need for a comparator does not alter the objective identifying capacity of fingerprints.
Article 29 Working Party, Opinion 05/2014 on anonymisation techniques, p. 10	Genetic data profiles remain at risk of identification where only the donor's identity is removed.
EDPB Guidelines 1/2026 on processing of personal data for scientific research purposes (version for public consultation), paras 165–166	Genetic data profiles are personal data by default; biometric data are similarly unique to individuals.
EDPB Guidelines 8/2020 on the targeting of social media users, Example 1 (section 5.2.1)	Joint controllership between a social media provider and a targeter where purposes and means are jointly determined.