Comments on the EDPB Guidelines 02/2026 on Anonymisation (version for public consultation)

Road traffic accident data: the purpose of the release, the release channel and the need for a method

Augusto Cesar Torbay – researcher in data protection law – contribution submitted in an individual and academic capacity

Contribution submitted in a personal capacity. The views expressed are the author's own and do not represent those of any organisation. The text is based exclusively on publicly available sources.

Summary of the proposals

We respectfully invite the EDPB to consider the following proposals, which are developed in the sections below:

1. To include a worked example on data concerning events that occur in public space, such as serious road traffic accidents, whose attributes (place, time, vehicles, outcome) are observed by third parties and frequently reported in the press. None of the 25 examples in the Guidelines addresses this category, which is the one in which a unique record is most likely to allow singling out (§§ 98-99), and which is the natural counterpart of Example 24.
2. To state that the documentation of the anonymisation process (§ 41) should record the purpose for which the anonymisation is carried out (for instance, open release for any re-use, or provision to a researcher for a defined study), on which the presumption regarding the legal basis in § 38 depends, and, where the anonymisation serves the purposes of Article 89 GDPR, that the assessment required by recital 156 was made. We do not propose that utility become an element of the identifiability test.
3. To mention, in § 32, Article 5(5) of Regulation (EU) 2022/868, which prohibits re-users of protected public sector data from re-identifying data subjects, as an example of a statutory prohibition of re-identification, without prejudice to § 33.
4. To make explicit that the controller's own releases (record identifiers, companion files published separately) are the first source of linkage to be examined under §§ 29(c) and 102, and to recommend documenting a measurement of the uniqueness of the combinations actually published before release as open data.
5. To clarify how the "separate treatment" contemplated in § 36 may be organised when the records that make a dataset personal are also the most valuable for the purpose pursued, given that removing or selectively generalising those records may itself allow an inference (§ 67).
6. To correct three cross-references, two of which are repeated in the summary published in October 2026 (section 6).

1. Why road traffic accident data deserve specific attention

Several Member States release road traffic accident data as record-level open data, with the location of each accident, its date and time and, in some cases, the age, sex and injury severity of each person involved. They are also data whose preventive use is required by Union law,

since Directive (EU) 2019/1936 obliged Member States to carry out periodic network-wide road safety assessments on the roads within its scope.[1]

Two features distinguish these data from the situations addressed in the examples of the Guidelines. The first is that the attributes of an accident exist outside the dataset by their very nature: the place, the time, the vehicles involved and the severity of the outcome are observed by witnesses, neighbours and emergency services, and serious accidents are regularly reported by the local press. The Guidelines rightly list neighbours and investigative journalists among the relevant entities (§ 30), and explain that a unique record may still be anonymous where its attributes "cannot be matched to any other data available to the relevant entity" (§ 99, Example 24). Accident data are, in our view, the opposite case: the attributes that make a record unique are largely those that are publicly observable. The mechanism is illustrated by a study, with a small sample and conducted outside the Union on clinical notes rather than accident files, in which victims of motor vehicle accidents were re-identified by matching the accident details with local news reports.[2]

The second feature concerns the persons to whom the records relate. The injury severity recorded for each living person may, in our view, be health data within the meaning of Article 4(15) GDPR. And although the GDPR does not apply to the personal data of deceased persons (recital 27), the record of a fatal accident almost always also relates to living persons: other drivers, passengers, injured persons and, at times, the vehicle owner.

2. What the published files show: one criterion, very different solutions, no method

To test how the criterion of identifiability is being applied in practice, we measured, in the most recent public files of eight Member States and one EEA State, the share of accidents whose combination of published location and time attributes is unique in the file ($k = 1$). The files were downloaded from the official portals in October 2026; a note describing the method, and the script that reproduces the results, are available on request.

State (year)	Location published	Time published	Attributes of persons	Unique accidents	Unique fatal accidents
France (2024)	coordinates and address	full date, hour and minute	year of birth, sex, injury severity	100.0 %	100.0 %
Czechia (Jan.–Jun. 2026, half-year)	coordinates to the metre	full date, hour and minute	exact age, sex, nationality, injury	100.0 %	100.0 %
Belgium (2024)	coordinates	year, month, hour (full date in a second file)	type of road user	99.6 %	99.5 %
Germany (2025)	exact coordinates	year, month, weekday, hour (day of month recoverable from the record identifier)	none	100.0 %	100.0 %
Spain (2024)	road, kilometre point and direction	year, month, weekday, hour	counts only	97.8 %	99.4 %
Finland (2025, State roads)	point and road address	month, year, weekday	none	100.0 %	100.0 %

Norway (2025)	point and address	full date, hour and minute	type of road user	100.0 %	n/a (no injury severity in the open file)
Netherlands (2025)	road network element	year only	none	82.7 %	93.0 %
Italy (2024)	municipality, no coordinates	quarter, weekday, hour	age class, sex, outcome	56.2 %	73.6 %

Note on the bases of calculation: figures are rounded to one decimal place (100.0 % denotes a share of 99.95 % or higher, e.g. 54,388 of 54,402 accidents and 3,225 of 3,226 fatal accidents in France, 272,879 of 273,007 and 2,476 of 2,477 in Germany, and 33,578 of 33,590 in Czechia). Records with missing values in the measured attributes are excluded from the denominator: in Spain the measurement covers the 38,814 accidents with road and kilometre point recorded, out of 101,996 (1,219 fatal out of 1,656); in Czechia, the 33,590 with valid coordinates and time, out of 42,511 (153 fatal out of 174); in Belgium, the 33,830 geolocated accidents, out of 35,916 (421 fatal out of 443); in Italy, the 170,124 with valid hour, out of 173,364 (435 fatal out of 455); in Norway, 3,560 out of 3,593. In the Netherlands and Finland, where the open file includes property-damage-only accidents, the table reports injury accidents (26,901 out of 125,621 in the Netherlands, compared with 58.5 % across all accidents; 1,362 out of 5,176 in Finland, compared with 99.8 % across all accidents). In the last column, each fatal accident is tested for uniqueness against all measured accidents in the file, and not only among fatal accidents.

The measurement concerns the uniqueness of the event, not, in itself, that of a person, and the figures should be read with that caution. Each figure refers to a single file as published and, for Belgium, does not take into account the linkage described below. Where a file publishes attributes of persons alongside exact coordinates and time, a unique event may, if matched with an external source, allow what is known about it from that source to be linked to the record of each person involved, and what was not known, such as age, sex or injury severity, to be learnt from the file. Where the file publishes no attributes of persons and the time is coarse, as in the Netherlands, uniqueness mainly reflects the spatial isolation of the event, and much less can be learnt from it about the persons concerned. Uniqueness is not, moreover, identification: a unique record only allows identification if it can be matched with attributes known to, or obtainable by, the relevant entity (§§ 98-99), and we have not attempted, nor would we attempt, any matching with external sources capable of identifying persons; the only linkage we performed, reported below, was between files released by the same publisher. We do not therefore claim that any of these releases infringes the GDPR.

With these cautions, three observations follow. The first is that, under the same Regulation and the same criterion, the solutions adopted range from publishing the year of birth, sex and injury severity of every person involved, together with the exact coordinates and minute of the accident, to publishing only the year of the accident and no attributes of persons. None of these choices is accompanied, as far as we could find, by a published measurement of the risk. We do not suggest that all States should publish the same files, but the GDPR presupposes an equivalent level of protection and a consistent and homogenous application of its rules (recital 10), which in our view requires at least a common method of assessing identifiability before release.[3]

The second observation is that information which is not among the variables described for a file

can sometimes be recovered from another variable or from another file released by the same publisher. In the German file, the day of the month, which is not among the described variables, can be derived from the record identifier (it matches the published weekday in 99.997 % of the 273,007 records of 2025, and in the same proportion in 2024). In Spain, a separate file on accidents involving animals contains the full date and minute, and 548 of its 549 accidents with victims match one-to-one a record in the main file, which thus recovers the day. In Belgium, the coordinates and the full date are published in two different files, and linking them on common attributes determines the day for 80.3 % of the geolocated accidents. We do not know whether the omission of the day was, in each case, intended as a protective measure. But these are precisely the linkages that § 102 requires to be re-tested, alongside the uniqueness of records and the additional information that § 29(a) and (c) requires to be taken into account, and they show that the controller's own releases are the first place where linkage must be looked for.

The third observation is that practice is moving, albeit under another legal framework. In France, the national statistics authority, when renewing the quality label of the road deaths series in November 2025, asked the national road safety observatory to ensure "le respect du secret statistique pour l'ensemble des données détaillées diffusées en open data". The requirement is one of statistical confidentiality, which is not identical to the GDPR test of identifiability, but it shows that, in at least one Member State, the granularity of these files has been questioned by the authority that oversees official statistics.[4]

Our conclusion is that the difficulty does not lie in the absence of a criterion, which the Guidelines provide with considerable rigour, but in the absence of a method by which the criterion is applied to a given file before it is released. This is the gap that proposals 1 and 4 seek to address.

3. The purpose of anonymisation should be documented

We begin by saying what we do not propose. The utility of data is not, and cannot be, an element of the identifiability test: a dataset that is useless for its purpose may be perfectly anonymous, and a useful dataset does not cease to be personal because it is useful.

What we propose lies on another level, which the Guidelines themselves recognise. Anonymisation is a processing of personal data, which must have a legal basis; that legal basis is presumed to coincide with that of the preceding processing only where the anonymisation is part of the same processing activity and the controller pursues the same purposes with it (§ 38). Whether the presumption applies therefore depends on the purpose of the anonymisation, which may be the open release of the data for any re-use or their provision to a researcher for a defined study, and which will often differ from the purpose for which the data were collected. The Guidelines further acknowledge that whoever anonymises for research or analytical purposes seeks "an anonymous dataset which lets you learn something about the people in question" (§ 18), and the summary published by the EDPB in October 2026 presents the anonymisation of personal data as "a valuable way to unlock its potential while simultaneously protecting individuals' rights".[5] They do not, however, state that the documentation required by § 41 should record that purpose.

For processing covered by Article 89 GDPR, the Regulation itself requires an assessment that depends on the purpose. Article 89(1) subordinates each step of its cascade to the same condition: the measures "may include pseudonymisation provided that those purposes can be

fulfilled in that manner", and "where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner". Recital 156 adds that further processing for those purposes "is to be carried out when the controller has assessed the feasibility to fulfil those purposes by processing data which do not permit or no longer permit the identification of data subjects". The EDPB reads the provision in the same way in its Guidelines 1/2026: anonymise in the first place, "as long as the purposes of processing can be fulfilled using anonymised data", and pseudonymise where they cannot (§§ 155-157).[6] Since recital 156 applies to further processing for those purposes, including the decision to anonymise, a controller that anonymises data for research or statistical purposes has made, or should have made, that assessment; we propose only that the documentation required by § 41 record that it was made, and for which purpose the data, as transformed, were judged adequate. We do not suggest that § 41 should document the processing of pseudonymised data, which is governed by other obligations.

Road safety research shows why that assessment cannot be made in the abstract, since what the data must contain differs from one purpose to another. For the periodic screening of the network, multi-year counts of accidents per road section, which presuppose that each accident is located on a section, are generally sufficient; for explaining why accidents happen, aggregation by section and annual average can reverse the sign of a relationship, as has been shown for speed; and the study of road user profiles requires the attributes of the persons involved.[7] Recording the purpose does not alter the legal standard of anonymity; it makes the decision to anonymise verifiable, which is the accountability function that § 41 already assigns to documentation.

4. The release channel and statutory prohibitions of re-identification

The Guidelines treat the context in which data are released as an objective factor of identifiability, "including whether there are any effective and permanent restrictions on access to the data" (§ 29(b)), recognise that, where appropriate security controls are in place, access constitutes a means reasonably likely to be used only for the authorised users (§ 89), and distinguish a prohibition by law (§ 32) from a contractual prohibition (§ 34). We do not suggest that the channel can replace the transformation of the data, which § 90 expressly rules out, nor that data made available in a secure processing environment become anonymous because of it: the Union acts that provide for such environments treat the data made available there as, in principle, pseudonymised personal data.

Our proposal is narrower. Union law already contains statutory prohibitions of re-identification, such as Article 5(5) of Regulation (EU) 2022/868, which prohibits re-users of protected data made available by public sector bodies under Chapter II of that Regulation from re-identifying any data subject; it does not extend to data released as open data. As a provision of a regulation, it binds those re-users directly. But because it usually reaches them together with the conditions of re-use that they accept, and § 32 gives no example of a prohibition by law, a controller may hesitate between treating it under § 32 and treating it under § 34. We therefore propose that § 32 mention it as an example of a prohibition by law, without prejudice to the assessment required by § 33. The graduation of the degree of transformation according to the access channel, which underlies this proposal, is already found in Regulation (EU) No 557/2013, in recital 15 of Regulation (EU) 2022/868 and in the EDPB's own Guidelines 01/2025 (§ 96).[8]

5. Mixed datasets: when the records that make a dataset personal are the most valuable

Under § 36, a dataset should only be considered anonymous if anonymisation is effective for all the individuals included; otherwise the entire dataset is to be treated as personal data, unless its parts are treated separately. Combined with § 87, this means that a single record that allows a living person to be singled out is enough to make an entire annual file personal data. We do not contest this rule.

The difficulty is a practical one. The records most likely to be unique and reported in the press are those of serious and fatal accidents, which are also the records of greatest value for prevention and which, as noted, almost always relate to living persons. The separate treatment that § 36 contemplates appears to us to be the way out, but it is not obvious how it should be organised. If the most serious accidents disappear from the open file, anyone who knows of an accident from the press may infer its severity from the mere absence of the record, which is an inference drawn from "information about the anonymisation process" (§ 67). If, instead, only the records of those accidents are generalised, the generalisation itself flags them. Generalising every record in the same way avoids both effects, but it is not, strictly speaking, a separation of the parts of a mixed dataset: it makes the open file anonymous as a whole, under the main rule of § 36, at the cost of its value for every purpose, while the detailed data reach researchers only through a controlled channel, as personal data. We invite the EDPB to clarify, with an example, whether the separation of parts that § 36 allows can be organised without creating either of the first two effects, or whether, where it cannot, the uniform generalisation of the released file is the only course available.

6. Editorial remarks

- In the table of § 97, in the column of the simplified approach and the row of the No Record Isolation criterion, the cross-reference reads "paras. 98 to 00"; it should read 98 to 100.
- In Annex 1, the box on singling out refers to "100–102"; singling out is addressed in §§ 98 to 100, while §§ 101 to 103 concern linkage.
- In Annex 1, the reference "(see 73, 103 and 104)" points to a § 104, which does not exist; the document ends at § 103.
- The same two cross-references in Annex 1 are reproduced in the flowchart of the summary published in October 2026 (*Guidelines at a Glance*, pp. 3-4), where "SIMPLIFIED" also appears as "SEMPLIFIED".

Notes

[1] Directive 2008/96/EC of the European Parliament and of the Council of 19 November 2008 on road infrastructure safety management, as amended by Directive (EU) 2019/1936 of 23 October 2019, Article 5 (network-wide road safety assessments "at least every five years"; the Directive applies to the trans-European road network, motorways and primary roads, Article 1(2)) and Annex III, point 3 ("number, location and cause of fatalities by road user group; number and location of serious injuries by road user group").

[2] JANMEY, Victor; ELKIN, Peter L. "Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack", *AMIA Annual Symposium Proceedings*, 2018, pp. 1329-1337 (PMID 30815177). Of 14 patients for whom the attack was attempted, 4 were re-identified through local news reports; the authors recommend omitting the accident location for patients involved in significant accidents.

[3] See also Judgment of 22 June 2021, *Latvijas Republikas Saeima*, C-439/19, EU:C:2021:504, § 83, where the Court, with regard to Article 10 GDPR, relied on the requirement that the level of protection be "equivalent and homogeneous in all the Member States".

[4] Avis du 13 novembre 2025 de l'Autorité de la statistique publique sur le renouvellement de la labellisation des séries statistiques produites par l'Observatoire national interministériel de la sécurité routière (ONISR), *Journal officiel de la République française*, 19 November 2025, texte 112; reproduced in AUTORITÉ DE LA STATISTIQUE PUBLIQUE, *Rapport annuel, mars 2025 – février 2026*, p. 135.

[5] EDPB, *Anonymisation: How to assess the anonymity of data. Guidelines at a Glance*, October 2026, p. 1.

[6] EDPB, *Guidelines 1/2026 on processing of personal data for scientific research purposes*, adopted on 15 April 2026 (version for public consultation), §§ 155-157. Footnote 33 of Guidelines 02/2026 refers to them in connection with § 38.

[7] ANDRÁŠIK, Richard *et al.* "Alternative approaches to network-wide road safety assessment", *European Transport Research Review*, v. 18, 2026, p. 8; IMPRIALOU, Maria-Ioanna M. *et al.* "Re-visiting crash–speed relationships: A new perspective in crash modelling", *Accident Analysis & Prevention*, v. 86, 2016, pp. 173-185, at pp. 182-183; IMPRIALOU, Maria-Ioanna; QUDDUS, Mohammed. "Crash data quality for road safety research: Current state and future directions", *Accident Analysis & Prevention*, v. 130, 2019, pp. 84-90 (cited from the accepted manuscript, p. 3). The concern that excessive anonymisation may impair the accuracy of models and the validity of research is also documented in the health sector: KOGUT-CZARKOWSKA, Magdalena; SHABANI, Mahsa. "Anonymization, accountability, and access: legal dimensions of health data sharing in federated networks. Perspectives from empirical study", *Frontiers in Digital Health*, v. 8, 2026, 1719728, p. 8, doi: 10.3389/fdgth.2026.1719728.

[8] Regulation (EU) 2022/868 of 30 May 2022 (Data Governance Act), Articles 3(1) and 5(3) and (5) and recital 15; on the treatment of the data made available in secure processing environments, see also Regulation (EU) 2025/327 of 11 February 2025 on the European Health Data Space, Articles 66 and 73; Commission Regulation (EU) No 557/2013 of 17 June 2013, Articles 2(2)-(3) and 7; EDPB, *Guidelines 01/2025 on Pseudonymisation*, version for public consultation, 16 January 2025, § 96; and, on the release model as a factor of the assessment, ELLIOT, Mark; MACKEY, Elaine; O'HARA, Kieron. *The Anonymisation Decision-Making Framework: European Practitioners' Guide*, 2nd ed., UKAN Publications, 2020, p. 9, and STALLA-BOURDILLON, Sophie. "Between nuance and caution: How to read the CJEU's judgment in EDPS v SRB", *Privacy and Data Protection*, v. 26, no. 1, 2025, pp. 6-10, at pp. 8-9.