Guidelines 02/2026 on Anonymisation

The assessment of genomic data: identification through relatives

DR GEORGIA S. CHATZITHEODOROU

Attorney at Law (Athens Bar)

Submitted in a personal capacity · Switzerland · October 2026

SUMMARY

1. The recipient-specific approach of the draft (paragraphs 6 and 12) is welcome and consistent with the judgment of the Court of Justice in *EDPS v SRB*.[1] Its application to genomic data is, however, left to inference. A patient may be identifiable through the genetic information of relatives that neither the disclosing controller nor the recipient holds: a sequence is matched against a consumer genealogy database, a genetic relative is found, and the patient is identified from the match. The technique is in routine use by genealogists and law enforcement and relies on openly accessible resources, not on any breach of security.[2] The draft's concepts of a chain of means (paragraph 28), third parties holding additional information (paragraph 30), inference (paragraph 72) and linkage (paragraph 80) cover this means of identification, but no example shows how its likelihood is to be assessed for a given recipient.[3]

2. The need is not a theoretical one. Under the European Health Data Space (Regulation (EU) 2025/327), health data access bodies must supply data in anonymised form wherever the purpose of the processing can be achieved with such data, and in pseudonymised form only where it cannot (Article 66(2) and (3)). The Regulation does not set its own test of anonymity; it applies the GDPR's. The final guidelines will therefore be the standard against which access bodies decide, request by request, whether genomic data leave in anonymised or pseudonymised form, from 26 March 2031 for the genetic and genomic categories listed in Article 51.[4] Three additions are proposed.

[1]Judgment of 4 September 2025, EDPS v SRB, C-413/23 P, EU:C:2025:645. The judgment concerns Regulation (EU) 2018/1725, whose Article 3, point (1), corresponds to Article 4, point (1), GDPR. The Court assessed identifiability from the recipient's perspective for the data transmitted to it (paragraphs 77 and 86), but from the controller's perspective, at the time of collection, for the duty to inform (paragraphs 110 to 113).

[2]Y. Erlich, T. Shor, I. Pe'er and S. Carmi, "Identity inference of genomic data using long-range familial searches", Science 362 (2018) 690–694, doi:10.1126/science.aau4832. About 60 per cent of searches for individuals of European descent return a third cousin or closer match, and a database covering 2 per cent of a population provides such a match for nearly everyone in it. The mechanism was already noted in Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP216), p. 10, the authority cited at footnote 243 of Guidelines 1/2026. Law enforcement has used the technique since 2018. Swedish police first used it in 2020 to identify a suspect in the Linköping murders of 2004; the Swedish Authority for Privacy Protection then found that the use lacked a legal basis, and legislation permitting it in defined cases took effect on 1 July 2025.

[3]Participants in the EDPB's stakeholder event of 12 December 2025 named "the identifiability of genetic data for different kinds of recipients" as a point needing guidance: EDPB, Report on stakeholder event on anonymisation and pseudonymisation of 12 December 2025, question 4.

[4]Regulation (EU) 2025/327 on the European Health Data Space, OJ L, 5.3.2025. Under Article 105, Article 51(1), point (f) (human genetic, epigenomic and genomic data) applies from 26 March 2031. Article 2(1), point (a), adopts the GDPR definition of personal data, and Article 2(2), point (b), defines non-personal electronic health data as data that have never related to a data subject or that have been anonymised so that they no longer relate to an identified or identifiable natural person.

I. A PRACTICAL GENOMIC EXAMPLE, FROM ACCESS TO IDENTIFICATION

3. Example 22 (paragraph 94) treats linkage as depending on knowledge that a given individual is in the dataset, and finds that no relevant entity has that knowledge. With genomic data that knowledge is not needed: the sequence is the starting point, relatives' profiles are the matching resource, and the patient is identified at the end, without any relative receiving the data. The draft's general factors (paragraph 29) do not, on their own, resolve how likely that is for a particular recipient, and paragraph 31 rightly cautions against relying on the recipient's lack of motivation: motivation can change and a recipient may be compelled, while a genome does not.

4. Two points of the draft need to be settled within such an example. First, inference. Paragraph 72 recognises that inferences which do not at first appear specific and meaningful may still assist identification. A kinship inference, that a record belongs to a relative of a known person, says little new about the relative, but it is a step towards identifying the person whose record it is; it should be assessed under the chain-of-means and linkage analysis (paragraphs 28 and 80) rather than under paragraphs 69 to 71, which concern inferences as an end point. Second, contractual restrictions. Paragraph 34 distinguishes a prohibition set out in a contract from a prohibition by law. Consumer genealogy services restrict uploads and searches through terms of service that bind only their users, and both the terms and the ownership of these services have changed in recent years.[5] Such terms do not remove the search from the assessment; they are evidence bearing on its likelihood, and a change in the terms, in the ownership of a resource or in the coverage of its database is a trigger for reassessment under paragraphs 35 and 88.

RECOMMENDATION The addition of an example in which the same individual-level sequences go to two or three recipients with different means, such as an academic institute, a commercial entity and a public body able to compel disclosure (paragraph 29) would be useful. For each, it would answer three questions and, in parallel, showcase the necessary evidence. *Access*: can this recipient actually search a matching resource, given the resource's upload and search rules and the recipient's own capabilities? *Kinship*: would a search find a relative, given the size of the resource and how many people of the patient's ancestry have e.g., a third cousin or closer relative in it? *Identification*: could the recipient get from a match to the patient through family trees, obituaries or other records? Under paragraph 19, identification does not require a name; it does require that the patient, and not merely a relative, be distinguished in a way that makes it possible to treat them differently. Taken together, the answers show whether the whole chain of means is reasonably likely to be used; no single answer settles it. The example should also show that a kinship inference is assessed as a step towards the patient (paragraphs 28 and 80), not under paragraphs 69 to 71; that a genealogy service's terms of use may make a search less likely but are not a prohibition by law (paragraph 34); that new terms, a sale of the service or marked growth in its database call for a fresh assessment; how paragraph 36 applies to a cohort in which the kinship question is answered positively for some records and not for others, when it is not known in advance which records those are; and whether, under paragraph 21, the institute's actual or planned sharing of the sequences with a collaborator able to identify the patient makes them personal data in its hands too.

---

[5]GEDmatch changed its terms in May 2019 to require users to opt in to law enforcement matching; it was acquired by Verogen in December 2019 and, with Verogen, by Qiagen in January 2023. 23andMe entered Chapter 11 proceedings in March 2025, and its assets were acquired by the TTAM Research Institute in July 2025.

II. SECURE PROCESSING ENVIRONMENTS UNDER THE EUROPEAN HEALTH DATA SPACE

5. Three features of the Regulation relate directly to the draft. *First*, the access body's choice under Article 66(2) and (3) is an anonymisation assessment made for a defined recipient, the holder of a data permit, and so is precisely the recipient-specific analysis the draft describes. *Second*, that recipient operates under legal and technical restrictions of a kind the draft has not yet addressed: analysis only within a secure processing environment (Article 73(1)), a prohibition on giving access to third parties not named in the data permit and on re-identification (Article 61(2) and (3)), and the information needed to reverse pseudonymisation held by the access body or a trusted third party (Article 66(3)). Paragraphs 32 and 33 allow the general assumption that the law will be followed, and paragraph 90 states that technical, organisational and contractual measures are not by themselves sufficient; how the two apply to an authorised user inside such an environment is not explained. *Third*, Article 61(4) requires that the results or output of secondary use contain only anonymous data, and Article 73(2) permits only non-personal data to be downloaded from the environment; both require an anonymity judgement on results and outputs that the draft does not cover.

6. Contributions to the consultation on Guidelines 1/2026 have argued that genetic data made available in such an environment may be treated as anonymous for the user, because the external linking described in the authority cited at footnote 243 of those guidelines is said to be impossible there.[6] Whether that is so depends on what the user can do inside the environment, including with datasets brought in under Article 67(2)(f),[7] and on what leaves it. The guidelines should name the facts that decide the question, such as permitted inputs and output checks, rather than presume the answer either way.

RECOMMENDATION A section on secure processing environments, or an example set in one, would settle three points. *First*, an authorised user is a restricted recipient, and the assessment should record what restricts them: the data permit, the ban on re-identification, any datasets the user has brought in, and the checks on queries and outputs. *Second*, each result leaving the environment under Articles 61(4) and 73(2) needs its own assessment under paragraphs 17 and 76 to 78; a table of allele frequencies from a small rare-disease cohort, a risk documented since 2008,[8] would illustrate the point. *Third*, a finding of anonymity inside the environment does not carry over if the same data are later released outside it (paragraph 90). If agreed with the Commission and the health data access bodies, this guidance would let every access body read "anonymised format" in Article 66(2) the same way.

III. THE "BY DEFAULT" POSITION OF GENETIC DATA, WHAT A CONTROLLER MUST DOCUMENT, AND THE POSITION OF RELATIVES

7. Guidelines 1/2026 on scientific research describe genetic profiles as personal data "by default", anonymous only "in exceptional circumstances" (paragraph 165); they note that the risks of

---

[6] See, for example, the joint contribution of UMCNL, COREON, NVVP, FMS and Health-RI of 25 June 2026 to the consultation on Guidelines 1/2026, remarks on paragraphs 164 and 165.

[7] Article 67(2), point (f), of Regulation (EU) 2025/327 provides for the applicant to "bring datasets already held by that health data applicant into the secure processing environment"; Article 67(2), point (g), requires the application to describe safeguards "to prevent any re-identification of natural persons in the dataset".

[8] N. Homer and others, "Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays", PLoS Genetics 4(8) (2008) e1000167. In August 2008 the United States National Institutes of Health withdrew aggregate genotype data from open access; since 2018 it allows open access to summary results except for studies of populations from isolated geographic regions or with rare traits (NIH Notice NOT-OD-19-023).

processing genetic data extend to other data subjects related to the research participant, including through consanguineal kinship (paragraph 167); and they ask controllers to consider whether processing may affect persons related to the data subject (paragraph 153).[9] The anonymisation draft applies a single test under Article 4(1) and recital 26 and requires adequate documentation of the assessment (paragraph 41). The two documents should be read together, and controllers need to know what the default requires of them before they conclude that genetic data are anonymous for a recipient. The relatives whose genetic information makes identification possible are neither participants nor recipients, yet the assessment turns on information about them.

RECOMMENDATION The EDPB could state that "personal data by default" is the starting point of the Article 4(1) assessment rather than a standard of its own. A controller may still find genetic data anonymous for a given recipient, but should establish and record, under paragraph 41, the data's granularity; the matching resources open to that recipient and the populations they cover; whether a match could actually lead to identification; available family-tree and demographic information; the statutory and technical limits on the recipient; the remaining uncertainties; and a review date, with the events that would bring it forward. A cross-reference to paragraphs 153 and 167 of Guidelines 1/2026 would ensure that the record also shows how relatives' genetic information could lead to the patient, and what the disclosure means for them.

SUMMARY OF RECOMMENDATIONS

	ADDITION PROPOSED	RELEVANT PROVISIONS
I	An example on individual-level genomic sequences, assessed for two or three recipients, separating access, kinship and identification and stating the evidence for each, and addressing the status of the cohort as a whole.	Draft: paragraphs 19, 21, 28, 29, 31, 34, 35, 36, 69 to 72, 80, 88, 94
II	A section or example on secure processing environments: the authorised user as a restricted recipient, the anonymity of outputs, and no carry-over to other disclosure contexts.	Draft: paragraphs 17, 32, 33, 76 to 78, 90; EHDS Articles 61, 66, 67, 73
III	How the "by default" position operates within the Article 4(1) test, what a controller must document under paragraph 41, and the position of relatives.	Draft: paragraph 41; Guidelines 1/2026, paragraphs 153, 165, 167

Thank you for considering these comments. I would be pleased to provide any further information or clarification that may be helpful.

[9]EDPB, Guidelines 1/2026 on processing of personal data for scientific research purposes, adopted 15 April 2026, version for public consultation, paragraphs 153, 165 and 167.