Comments on the EDPB Guidelines 02/2026 on Anonymisation

DIGIfor1healthSE · Oct 30, 2026

1. Introduction

DIGIfor1healthSE welcomes the EDPB Guidelines 02/2026 on Anonymisation. DIGIfor1healthSE is a Swedish collaboration project (2024–2027), funded by Vinnova and coordinated by SciLifeLab. Collaborating within DIGIfor1healthSE are two national strategic initiatives and five EU-funded DIGITAL programme projects. The national initiatives are Genomic Medicine Sweden (GMS) and SciLifeLab. The EU DIGITAL projects are Genomic Data Infrastructure (GDI), Genome of Europe (GoE), European Federation for Cancer Images (EUCAIM), Testing and Experimentation Facility for Health AI and Robotics (TEF-Health), and Health Data Sweden (HDS), which is a European Digital Innovation Hub (EDIH).

The guidelines fill a long-standing need. We particularly value the practical assessment of record isolation, linkage and inference, and the attention given to aggregated data, AI models and synthetic data.

Health data raise specific challenges: small patient groups, long care histories and data types that are identifying by their very nature. Our comments aim to make the guidelines easier to apply in the health sector. Section 2 sets out our key recommendations, Section 3 explains them with practical examples, and the Annex addresses highly identifying health data in more depth.

2. Key recommendations

We recommend that the EDPB:

1. **Add worked examples from the health sector**, covering rare conditions, small patient groups and longitudinal care pathways (Section 3.1).
2. **Clarify when a dataset that combines anonymous and personal data falls within the scope of the GDPR**, building on paragraph 36 of the Guidelines (Section 3.1).
3. **Explain how the risk of re-identification should be assessed in research and AI settings**, including the role of the recipient, available reference data and the intended use (Section 3.2).
4. **Describe concrete tests for synthetic data**: membership inference, memorisation, attribute inference and linkage, including the cumulative risk of repeated queries and releases (Section 3.3).
5. **Include an example of federated health analysis under the EHDS**, showing how reviewed aggregate outputs may leave a secure processing environment, without equating architectural safeguards with anonymisation (Section 3.4).
6. **Clarify how EHDS data permits and secure processing environments affect the contextual assessment**, and state that the simplified approach is a more cautious route of assessment, not a separate legal standard (Section 3.4).
7. **Provide practical guidance on differential privacy and a documentation template** for assessment and reassessment over time (Section 3.5).
8. **Address data that are identifying by their very nature** — genomic data, facial images, three-dimensional medical images and retinal images. Such data should be presumed to be personal data for any holder. The presumption should be rebuttable only by a documented assessment

showing that identification is not reasonably likely, also taking into account matching against reference data (Annex).

3. Explanatory remarks and practical examples

3.1 Longitudinal health records, rare conditions and mixed datasets

We propose a worked example based on a longitudinal patient record. It should involve a rare diagnosis, repeated healthcare contacts and a small regional population. Even without direct identifiers, the combination of diagnoses, treatment dates and care pathways may single out individuals or allow linkage to other information.

The example could show how the No Record Isolation, No Linkage and No Inference criteria apply in a healthcare setting. It could also clarify, building on paragraph 36, how the anonymity of individual records relates to that of the dataset as a whole. In particular, it should explain when a dataset containing both anonymous and personal records is subject to the GDPR.

3.2 Re-identification risk in research and AI

We ask the EDPB to explain more fully how the likelihood of identification should be assessed in research and AI settings. The guidance should address how the following factors affect that assessment:

- the identity and capabilities of the recipient;
- the reference datasets available to the recipient or to third parties;
- the technical means available, including AI-based matching;
- the intended use of the data and of any trained model.

It would also help to explain how such assessments should be documented and revisited when new data sources or identification techniques become available.

3.3 Synthetic health data

Synthetic data are increasingly used in health research, but their anonymity cannot be presumed. The guidance should describe tests for membership inference, memorisation, attribute inference and linkage. These tests should include comparing model outputs with and without a given individual's data. The assessment should also cover the cumulative risk of repeated queries and successive releases.

A worked example showing how a synthetic health dataset is tested, documented and reassessed would make the guidance operational for researchers and healthcare organisations.

3.4 Federated analysis and the EHDS

We propose an example of federated health analysis. Personal health data would remain in separate secure processing environments under the responsibility of each entity, while analytical queries run across the participating environments. Intermediate results would be combined through controlled mechanisms. The final outputs would undergo a documented disclosure-risk assessment before release.

The example should address situations where several individually acceptable aggregate results, once combined, reveal information about individuals. This includes small patient groups, overlapping populations, repeated queries and cumulative disclosure risk across environments.

Under Article 73(2) of Regulation (EU) 2025/327, health data access bodies must review download requests so that only non-personal electronic health data leave the secure processing environment. The example should show how output review relates to that obligation. It should keep three questions apart:

1. the legal authorisation to process the data (the data permit);
2. the technical and organisational safeguards of the secure processing environment (applying for instance secure multi-party computation);
3. the separate assessment of whether the released results are anonymous.

The guidelines should also clarify how a data permit and a secure processing environment affect the contextual assessment. In our understanding, the simplified approach is a more cautious route of assessment, not a separate legal standard; the guidelines should say so explicitly.

3.5 Differential privacy, documentation and reassessment

Differential privacy could be described as one possible method. The guidance should explain how to choose its parameters and overall privacy budget, and the trade-off with data utility.

A structured documentation template would support consistent assessment and reassessment over time. It could record the scope of permitted processing, the safeguards applied, the output review procedure and the reasons for concluding that released data are anonymous. It should also record when and on what triggers the assessment will be revisited.

Annex. Health data that are identifying by their very nature

A.1 The issue

Section 2.4.1 of the Guidelines addresses identifiers and other attributes. Paragraph 23 notes that identification may occur through an identifier that is unique in a given context, such as a name, an IP address, a national identification number or an email address.

This point needs to be developed much further for health data. Some health data — genomic data, facial images and certain medical images — carry identifying characteristics within the data themselves. They cannot be separated from the person in the way that a name can be removed from a record. Such data are very difficult to anonymise, and the Guidelines should state when, if ever, anonymity can be established.

A.2 The contextual approach after the SRB judgment

Under Article 4(1) GDPR, a person is identifiable if they can be identified directly or indirectly, including by reference to factors specific to their physical, physiological or genetic identity. Recital 26 excludes anonymous information from the scope of the GDPR and requires account to be taken of all means reasonably likely to be used for identification.

The EDPB Guidelines 01/2025 on pseudonymisation state that pseudonymised data remain personal data where they can be attributed to a person using additional information, even if that information is held by someone else.

In *EDPS v SRB* (C-413/23 P, judgment of 4 September 2025), the Court of Justice confirmed that the assessment is relative. Pseudonymised data are not necessarily personal data for every recipient. What matters is whether that recipient has means reasonably likely to be used to identify the data subject.

The judgment builds on *Breyer* (C-582/14) and *Scania* (C-319/22). It confirms that the same data may be personal data for one actor and anonymous for another.

We accept that this relative assessment applies to all data, including health data. Our concern is how it is applied to data whose identifying characteristics lie in the data themselves. For such data, the absence of a key or a name in the recipient's hands says little. Identification can occur by comparing the data with reference information, and that possibility must be part of the assessment of means reasonably likely to be used.

A.3 Why some health data are identifying by nature

Article 4(14) GDPR defines biometric data as personal data resulting from specific technical processing of physical, physiological or behavioural characteristics, which allow or confirm the unique identification of a person. The definition is technology-neutral. Facial images and fingerprints are examples, not an exhaustive list.

In *Schwarz* (C-291/12), the Court held that fingerprints are personal data because they objectively contain unique information that allows individuals to be identified with precision [1]. Later case-law treats biometric and genetic data as particularly sensitive and requires a strict assessment of necessity and proportionality. Examples are *Ministerstvo na vatreshnite raboti* (C-205/21), *Direktor na Glavna direktsia 'Natsionalna politsia'* (C-118/22) and *Landeshauptstadt Wiesbaden* (C-61/22) [2].

In *Ryneš* (C-212/13), the Court held that an image of a person recorded by a camera is personal data in so far as it makes it possible to identify that person [3]. No name or other conventional identifier is required.

A.4 Genomic data

Under Article 4(13) GDPR, genetic data are personal data on a person's inherited or acquired genetic characteristics, resulting in particular from the analysis of a biological sample. In its guidelines on processing for scientific research [4], the EDPB states that genetic profiles generally constitute personal data. They may be considered anonymous only in exceptional cases, after a careful analysis of the specific genetic configuration and its prevalence in the population.

Research confirms the risk. Individuals have been re-identified from genomic data by linking them to public genealogy databases [5]. Participation in a study has been inferred even from aggregate genomic statistics [6]. **We ask the EDPB to restate its position on genetic profiles in the Guidelines on Anonymisation.**

A.5 Medical images, including retinal images

Medical images raise similar issues. Two examples show why:

- **Three-dimensional head images.** Faces can be reconstructed from MRI and CT scans and matched against photographs with face-recognition software [7]. This is why defacing is standard practice before such images are shared.
- **Retinal images.** The retinal vascular pattern is distinctive enough to be used as a biometric modality. Retinal images may therefore enable identification when compared with reference images, even without any conventional identifier. They also enable inference: deep-learning models can predict age, sex and cardiovascular risk factors from fundus photographs [8]. This is directly relevant to the No Inference criterion.

The fact that a holder is not permitted to identify individuals is relevant under *Breyer* and *SRB*, but it is not sufficient on its own. The assessment must also consider whether reference data exist and whether matching is reasonably likely, given the recipient's access and capabilities and the state of the art. Where specific technical processing allows or confirms unique identification, Article 4(14) GDPR also applies.

We recommend that the Guidelines include a worked example on retinal images. It should show how the inherent identifying characteristics of the images, the availability of reference data and the context of the recipient are weighed in assessing anonymity.

A.6 Direct and indirect identification

The Guidelines should recall the distinction between direct and indirect identification. In *Buivids* (C-345/17), concerning a video of police officers, Advocate General Sharpston observed that the difficulty of identifying data subjects is not a criterion under Directive 95/46. No name or address is needed for the rules to apply [9]. In *Ryneš* and *Buivids*, people could be recognised from the images themselves.

Genomic data and retinal images are different. They usually identify a person only when compared with reference data. That is indirect identification, and the means reasonably likely to be used, as discussed in *Breyer*, *Scania* and *SRB*, are decisive. The key question is therefore whether suitable reference data exist and are reasonably accessible. Biobanks, genealogy services, imaging registries and AI-based matching are making such reference data more available over time.

A.7 Proposed approach

We propose that the Guidelines adopt the following approach for data that are identifying by their very nature:

1. **A rebuttable presumption.** Genomic data, facial images, three-dimensional head images and retinal images should be presumed to be personal data for any holder.
2. **Rebuttal only by a documented assessment.** The presumption can be rebutted only by a documented assessment showing that identification, including by matching against reference data, is not reasonably likely. The assessment should consider the recipient, the available reference data, the state of the art and the safeguards in place.
3. **Case-by-case assessment of reduced data.** Data reduced to a few parameters, such as short gene sequences or partial images, may be anonymous if individuals can no longer be singled out, linked or inferred. Techniques such as generalisation, noise addition, permutation and defacing can contribute. The number of parameters and their identifying power are relevant.
4. **Regular reassessment.** Reference data and matching techniques evolve. The conclusion should be revisited when they change.

For many clinical and research purposes, reducing the identifying characteristics of medical images or genomic data also reduces their scientific and clinical value. The Guidelines should acknowledge this trade-off. In such cases, secure processing environments and other safeguards for personal data, rather than anonymisation, will often be the appropriate route.

Notes

1. Case C-291/12 *Schwarz*, EU:C:2013:670, paragraph 27. Mr Schwarz objected to having his fingerprints taken for his passport under Regulation (EC) No 2252/2004. The Court found that taking and storing fingerprints interferes with the rights under Articles 7 and 8 of the Charter. It held the interference justified and proportionate, as it serves to prevent passport forgery and fraudulent use.
2. Case C-205/21 *Ministerstvo na vatreshnite raboti (Recording of biometric and genetic data by the police)*, EU:C:2023:49; Case C-118/22 *Direktor na Glavna direktsia 'Natsionalna politsia' pri MVR – Sofia*, EU:C:2024:97; Case C-61/22 *Landeshauptstadt Wiesbaden*, EU:C:2024:251. See also the Opinion of Advocate General Pitruzzella in Case C-205/21.
3. Case C-212/13 *Ryneš*, paragraphs 21–22. The case concerned a home security camera covering a public street, which the Court held fell outside the household exemption of Directive 95/46/EC.
4. EDPB Guidelines 1/2026 on the processing of personal data for scientific research purposes.
5. Gymrek M et al., Identifying personal genomes by surname inference, *Science* 2013;339:321–324.
6. Homer N et al., Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays, *PLoS Genetics* 2008;4(8):e1000167.
7. Schwarz CG et al., Identification of anonymized MRI research participants with face-recognition software, *New England Journal of Medicine* 2019;381:1684–1686.
8. Poplin R et al., Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, *Nature Biomedical Engineering* 2018;2:158–164.
9. Opinion of Advocate General Sharpston in Case C-345/17 *Buivids*, paragraph 33.