ROBERT KOCH INSTITUT

Comments on the EDPB Guidelines 02/2026 on Anonymisation

Reference: Guidelines 02/2026 on Anonymisation, version 1.0, adopted on 7 July 2026 for public consultation

We welcome the Guidelines and consider their reasoning consistent and well founded from a data protection perspective. We comment from the perspective of a data-sharing unit at a public research institution that is mandated to make sensitive personal data available for research and as open data.

From this perspective, the Guidelines show clearly how difficult it has become to achieve anonymity, and that the way data are shared, and what is called anonymous, needs to change. Anonymisation requires ever higher effort while leaving ever less detail in the released data. We would welcome if the Guidelines stated this consequence explicitly, pointed controllers to alternative routes for detailed data, and gave more practical guidance on how to meet the high standard they set.

On the other hand, the scope of the data affected by this Guideline is soon expected to experience a significant increase through outputs from Secure Processing Environments. These outputs will in most cases need to meet these standards for anonymous data, which the controllers will have to ensure during output checking. One example is the upcoming European Health Data Space (EHDS).

Effort and level of detail

Assessing linkage and inference scenarios in line with the Guidelines, for each relevant entity from acquaintances of data subjects to malicious actors and foreign intelligence services (para. 30), requires highly specialised staff supported by standardised, validated tools. The simplified approach, in turn, seems workable only for releases with very little detail. In consequence, data holders not obliged to share their data are likely to be pushed toward not sharing at all, at least not as open data, which runs counter to policy objectives that promote the re-use of data, such as open data policies.

In our assessment, the Guidelines imply that record-level data with a high level of detail, as typically provided in Scientific Use Files (pseudonymised microdata shared under contractual safeguards), can only rarely be considered anonymous. For research requiring such detail, other routes should be prioritised, such as access via secure processing environments or the provision of personal data on an appropriate legal basis with safeguards. We would welcome if the Guidelines stated this consequence explicitly.

It would also be helpful if the Guidelines provided criteria or examples for when a contextual assessment is likely to succeed and is therefore worth the effort, for

example: aggregated data or data of low dimensionality, a low value of re-identification for potential attackers, a well-defined and controlled group of recipients, few previous releases from the same source data, and information that is not recorded elsewhere (cf. para. 65).

Linkage, inference and cumulative releases

Linkage and inference criteria are becoming increasingly difficult to assess, and the Guidelines lack a guardrail as to how far the assessment has to go. In case of uncertainty, data must be considered personal, which rules out releasing them as anonymous data. Technical capabilities are growing rapidly, especially through AI, and the state of the art is hardly foreseeable. The spectrum of potential attackers is widening (hackers, unauthorised and malicious actors, autonomous AI agents), and available datasets can hardly be surveyed in a decentralised landscape. Under the contextual approach, assessors therefore have to invest considerable effort to gather information on potential downstream recipients or reduce the level of detail drastically to be on the safe side.

Applied consistently, the Guidelines require every data release to take earlier releases into account, since they could be linked once published (cf. para. 77). This suggests tracking cumulative risk per dataset, a kind of “anonymity budget” (comparable to the “privacy budget” in differential privacy) that evaluates a further data release in the context of earlier ones. Even result sets released via secure processing environments, including through EHDS-mediated data sharing starting soon, may have to be counted in such a budget. Once the budget is exhausted, no further anonymised releases from the same source data would be possible, and further access would have to be provided through secure processing environments.

We would appreciate if the Guidelines addressed this difficulty, including practical guidance on how technological development is to be taken into account and how releases can be recorded in a standardized way for linkage risk assessments. It also remains open who should keep a record of releases across projects and data holders. All of this requires governance structures that typically receive no additional funding or staff.

Tools and infrastructure

Meeting the standard set by the Guidelines requires a supporting toolset that is currently lacking: secure processing environments, standardised and validated tools, for example state-controlled AI tools, that can assess the current, concrete risk profile of a release, and tools that log earlier releases and assess their combined risk. It might help to couple the introduction of the Guidelines with the establishment of such infrastructure. At the least, we suggest including an appeal to Member States and the scientific community to develop these tools.