Individual Contribution to the European Data Protection Board Public Consultation on Guidelines 1/2026 on Processing of Personal Data for Scientific Research Purposes

Submitted by:

Sandra Carvalho, PhD, Habil.

Assistant Professor with Habilitation, School of Psychology, University of Minho, Braga, Portugal, E-mail: sandrarc@psi.uminho.pt

Date: 24/06/2026

Introductory remarks

I welcome the European Data Protection Board's initiative to provide dedicated guidance on the processing of personal data for scientific research purposes. Given the increasing complexity of data-intensive and collaborative research, such guidance has the potential to enhance legal certainty and promote a more consistent interpretation of the GDPR across Member States.

At the same time, it is important that the final Guidelines maintain a proportionate, risk-based and technologically neutral approach that both safeguards the rights and freedoms of data subjects and enables responsible, transparent and reproducible scientific research. Effective data protection and scientific progress should be regarded as complementary objectives that can and should coexist.

From the perspective of a researcher working in cognitive neuroscience, electrophysiology and non-invasive brain stimulation, I would like to highlight three aspects that, in my view, would benefit from further clarification in the final Guidelines.

1. Context-dependent interpretation of neuroscience and physiological data

The final Guidelines would benefit from the inclusion of additional examples involving data commonly used in neuroscience and psychology, including electroencephalography (EEG), magnetic resonance imaging (MRI), autonomic and psychophysiological recordings (e.g., electrodermal activity, heart rate variability and respiratory measures), behavioural measurements, wearable sensor data, datasets acquired in conjunction with

neurostimulation protocols, and emerging digital biomarkers. These modalities are increasingly employed in both basic and translational research and frequently involve healthy populations and non-clinical contexts.

Although such data may, under certain circumstances, reveal information relating to health or other special categories of personal data, the mere theoretical possibility of drawing sensitive inferences should not automatically result in all such datasets being classified as special-category data under Article 9 GDPR. Rather, the assessment should remain context-dependent and should take into account the purpose of processing, the methods actually applied or reasonably envisaged, the nature and reliability of the potential inference, and the intended use of the data.

This issue is becoming increasingly relevant as advances in computational methods and artificial intelligence expand the ability to extract information from complex physiological, behavioural and multimodal signals. A purely theoretical or remote possibility of deriving health-related information should not, in itself, lead to an overly broad interpretation of Article 9. Otherwise, there is a risk that a wide range of datasets routinely used in cognitive neuroscience and psychology would automatically be subject to requirements that may not be proportionate to the actual risks involved.

The increasing use of multimodal datasets combining neurophysiological signals, psychophysiological measurements, behavioural measures, wearable-device data, digital phenotyping and computational modelling further highlights the need for context-dependent assessments. In many cases, the scientific value arises from the integration of multiple data modalities rather than from the interpretation of any individual measure in isolation. Moreover, the same dataset may support different types of analyses depending on the research question, making purpose and context central elements of any assessment under Article 9 GDPR.

Providing illustrative examples involving neurophysiological, psychophysiological and behavioural data would improve legal certainty, facilitate more harmonised interpretation across Member States, and support the implementation of proportionate safeguards consistent with the risk-based approach underpinning the GDPR. Such examples would be particularly valuable in light of the growing use of artificial intelligence and multimodal approaches in neuroscience, mental health and digital medicine, where the boundaries

between physiological, behavioural and health-related information are becoming increasingly complex.

2. Scientific reproducibility and research integrity

Scientific reproducibility and research integrity frequently require the long-term preservation of stable, analysis-locked datasets following the completion and publication of a study. Such datasets play a critical role in enabling independent verification of findings, replication studies, investigation of alleged errors or scientific misconduct, correction of the scientific record, and compliance with disciplinary, regulatory or funding requirements.

The final Guidelines should explicitly recognise that these objectives constitute legitimate scientific interests that deserve careful consideration when applying data subject rights, particularly the right to erasure under Article 17 GDPR. In many cases, deletion of individual records from a locked dataset may affect derived statistics, model parameters or quality-control procedures and may compromise the ability to reproduce published analyses accurately.

This issue becomes particularly important in studies involving multimodal datasets, computational pipelines and machine learning models. In such contexts, individual observations contribute not only to statistical analyses but also to trained models, feature extraction procedures and quality-control frameworks. The removal of individual records following publication may therefore affect model parameters and compromise the reproducibility and interpretability of previously reported findings.

At the same time, the protection of data subjects and the preservation of scientific integrity should not be regarded as conflicting objectives. A proportionate approach may allow both interests to coexist through the implementation of appropriate safeguards, including restricted access, separation of identifiers, pseudonymisation, retention of archival datasets solely for verification and integrity purposes, and exclusion of data from future analyses where technically and scientifically feasible.

The publication of scientific findings should not automatically create a permanent exemption from the right to erasure. Equally, the exercise of this right should not necessarily imply the retraction of published studies or the recomputation of anonymous aggregate results. The stage of the research, the legal basis for processing, the degree of anonymisation

or pseudonymisation, and the potential impact of deletion on reproducibility and scientific integrity should all be considered when assessing such requests.

Furthermore, the growing adoption of open science practices and FAIR principles increasingly encourages data sharing, independent verification and secondary analyses. The long-term preservation of well-curated datasets under appropriate governance frameworks is therefore essential not only for individual studies but also for cumulative scientific progress and the efficient use of research resources. Recognising these considerations in the final Guidelines would provide greater legal certainty for researchers and institutions while promoting trustworthy, transparent and reproducible science. Such clarification would be particularly valuable in increasingly data-intensive and computationally driven research environments, where scientific integrity depends on the ability to preserve, document and verify research outputs over time.

3. Proportionate and technologically neutral safeguards

Research institutions differ considerably in their technical infrastructure, available resources and organisational capacity. While secure processing environments and trusted research environments can provide strong safeguards, particularly when dealing with highly sensitive data, they should not be regarded as the default or universally appropriate solution for all research settings. Contemporary scientific research frequently relies on specialised software, distributed computing resources, international collaborations and rapidly evolving analytical methods. Consequently, the final Guidelines should preserve a risk-based and technologically neutral approach that allows the selection of safeguards to be tailored to the nature of the data, the research purpose, the realistic risks of identification and the available technical and organisational measures.

Alternative safeguards, including pseudonymisation, controlled access procedures, federated analysis, privacy-enhancing technologies, encrypted computation and appropriately monitored local environments, may provide effective protection while maintaining scientific flexibility and facilitating collaboration. The GDPR does not prescribe a single technological model, and the final Guidelines should avoid creating the impression that one particular infrastructure or governance approach is inherently preferable in all circumstances.

This consideration is particularly important for smaller institutions, investigator-led projects and research groups with limited financial and technical resources. Requirements that are proportionate for large national infrastructures or highly specialised centres may be difficult to implement in other settings and may inadvertently increase disparities between institutions and Member States. Furthermore, contemporary research increasingly relies on cloud-based infrastructures, high-performance computing facilities, federated learning approaches, artificial intelligence and large-scale multimodal datasets. The rapid evolution of these technologies makes it particularly important that the Guidelines avoid implicitly favouring specific architectures or infrastructures that may become obsolete or unsuitable as scientific practices evolve.

Technological neutrality is also essential for promoting innovation and supporting open science initiatives. Research communities increasingly adopt FAIR principles, controlled data-sharing mechanisms and privacy-enhancing technologies that allow scientific collaboration while minimising risks to data subjects. A flexible and risk-based framework is more likely to accommodate future methodological developments than prescriptive approaches tied to particular technical solutions. In addition, the proportionality principle should consider not only the sensitivity of the data but also the realistic capabilities of institutions and research teams. Compliance requirements that are excessively complex or resource-intensive may inadvertently disadvantage smaller institutions and early-career investigators, thereby creating barriers to scientific participation and widening existing inequalities within the European Research Area.

The final Guidelines should continue to emphasise proportionality and technological neutrality, ensuring that safeguards are selected on the basis of risk and effectiveness rather than through the implicit endorsement of a particular technical architecture. Such an approach would maintain a high level of protection for data subjects while preserving the diversity, feasibility and innovative capacity of scientific research across Europe.

Concluding remarks

Strong protection of personal data and scientific progress should not be regarded as competing objectives. Rather, both are essential components of responsible, trustworthy and socially beneficial scientific research. The GDPR provides a framework that seeks to balance

fundamental rights with the societal value of scientific advancement, and the final Guidelines represent an important opportunity to reinforce that balance.

Further clarification regarding the context-dependent interpretation of neuroscience and physiological data, the role of reproducibility and research integrity, and the implementation of proportionate and technologically neutral safeguards would enhance legal certainty and facilitate more consistent application of the GDPR across Member States. Such clarification would be particularly valuable in increasingly data-intensive, multidisciplinary and international research environments, where responsible data sharing, transparency and reproducibility are fundamental to scientific progress.

The rapid development of artificial intelligence, digital phenotyping, multimodal datasets and large-scale collaborative infrastructures further highlights the importance of maintaining a flexible and scientifically informed regulatory framework. As research methodologies and analytical capabilities continue to evolve, guidance should remain sufficiently adaptable to accommodate innovation while preserving a high level of protection for individuals.

Continued dialogue between data protection authorities, ethics committees, research institutions and the scientific community will be essential to ensure that regulatory frameworks remain responsive to emerging challenges and technological developments. Such collaboration can help foster a shared understanding of how fundamental rights and scientific progress can be advanced in a mutually reinforcing manner.

By maintaining a risk-based, proportionate and technologically neutral approach, the final Guidelines can support both robust protection of individuals and the continued development of innovative, transparent and trustworthy scientific research across Europe. In doing so, they will contribute not only to legal certainty but also to the sustainability, inclusiveness and global competitiveness of the European Research Area.

Assinado por: **Sandra da Conceição Ribeiro de Carvalho**
Num. de Identificação: 12104622
Data: 2026.06.24 11:39:35 +0100

Sandra Carvalho, PhD, Habil.

Assistant Professor with Habilitation

School of Psychology, University of Minho, Braga, Portugal