VNIVERSITAT DE VALÈNCIA

Contribution to Guidelines 1/2026 on processing of personal data for scientific research purposes

Introduction

By means of these observations, the undersigned wishes to contribute to the public consultation process opened by the European Data Protection Board in relation to Guidelines 1/2026 on the processing of personal data for scientific research purposes, adopted on 15 April 2026. The effort made by the EDPB in addressing an area of extraordinary complexity and strategic relevance for the European Union is welcomed, as is its commitment to providing interpretative criteria that facilitate compliance with the GDPR by controllers operating in scientific research environments.

The observations set out below are limited to four specific sections of the document whose content raises practical or interpretative considerations deemed to be of particular relevance: section 2.1, concerning the indicative factors for determining the scientific nature of the processing; section 2.2, on research data infrastructures; section 3, in particular subsections 3.1 and 3.2, addressing the principles of purpose limitation and storage limitation; and section 8.3, on anonymisation and pseudonymisation. In view of the public consultation deadline, these observations are presented in summary form and without a fully developed alternative drafting proposal.

Observation on Section 2.1

Section 2.1 introduces six key indicative factors for determining whether the processing of personal data is motivated by scientific research purposes, and provides that, where not all of those factors are met, the controller must justify and demonstrate why the activity should nonetheless be regarded as scientific research within the meaning of the GDPR.

The objective of providing guiding criteria in an area in which no universally agreed definition exists at the level of Union law is shared. It is considered, however, that the formulation adopted carries a risk of restrictive application that warrants attention. Taken together, the six factors predominantly reflect a model of institutional academic research, characterised by publication in peer-reviewed scientific journals, the organisational autonomy of the research team, formal doctoral credentials, and contribution to society's general knowledge as a primary objective. This

Avda./ Blasco Ibáñez 13.
VALÈNCIA 46010
Tf. 34 896) 398 32 56
cmicrosoft@uv.es

model, whilst legitimate and prevalent, does not exhaust the diversity of research activities recognised by the GDPR itself, which in Recital 159 expressly refers to technological development and demonstration, fundamental and applied research, and privately funded research.

The application of factors such as the autonomy and independence of the research team, verifiability and transparency through external review, the possession of formal academic qualifications, or an explicit contribution to the growth of general knowledge may give rise to objective difficulties for research conducted in technology centres, start-up companies or industrial laboratories. In such contexts, organisational, competitive, regulatory or budgetary constraints give rise to forms of research that, whilst genuinely scientific in substance, may not fully satisfy all of the factors as formulated.

The practical effect of this approach may be dissuasive: faced with the need to justify the scientific nature of an activity in the absence of one or more of the factors, it is reasonable to expect that certain operators will choose not to frame their processing activities on the basis of the scientific research regime, thereby foregoing the flexibilities legitimately provided for under the GDPR. It is therefore considered appropriate for the Guidelines to state more clearly that the six factors do not constitute a cumulative test or a single model of scientific validity, that none of them should be interpreted as individually determinative, and that the assessment must be carried out in a contextual and proportionate manner, having regard to the specific nature of the activity concerned and the ecosystem in which it is conducted.

Observation on Section 2.2

Section 2.2 addresses the processing of personal data in the context of research data infrastructures, characterising them as repositories or databases intended to make data available for future research projects within a specific area of research. Whilst this approach provides a useful starting point, it is considered insufficient to reflect adequately the nature and function of research infrastructures within the current European ecosystem.

Research data infrastructures are not, in functional terms, simple storage repositories or data transfer mechanisms. They constitute environments of high technical and organisational complexity that perform essential instrumental functions for data-driven research: the curation, enrichment, annotation and cataloguing of datasets; the design and operation of secure processing environments; the implementation of pseudonymisation and

anonymisation mechanisms; the enabling of federated access and interoperability; and the governance of conditions for access and further use. Infrastructures of this nature, regulated at Union level through legal forms such as European Research Infrastructure Consortia and European Digital Infrastructure Consortia, constitute a structural component of the European Research Area and are expressly recognised by the European Health Data Space Regulation as entities authorised to participate in secondary uses of data.

The description contained in section 2.2, and in particular the indication that access to research data will typically be limited to projects in certain fields of research, may not adequately reflect the logic of cross-sectoral and interdisciplinary reuse that characterises a significant proportion of contemporary data-driven research, nor the openness to secondary uses that certain recent European regulatory instruments expressly promote. It is considered appropriate for the Guidelines to expand their treatment of this section so as to acknowledge the functional complexity of these entities, the instrumental character of their activities within the research data value chain, and the need for a compliance framework that is coherent with the broader European regulatory ecosystem governing data.

Observation on Section 3

Section 3 addresses the application of the principles set out in Article 5 GDPR in the context of scientific research, with particular attention to purpose limitation and storage limitation. The explicit recognition of the presumption of compatibility for further processing for scientific research purposes is welcomed, as is the relative flexibility acknowledged with respect to storage.

It is nonetheless considered that the combination of criteria applied in subsections 3.1 and 3.2 may produce a materially restrictive effect on certain forms of data-driven research. In particular, the implicit requirement that future research purposes be reasonably foreseeable in relation to a sufficiently defined scientific area, that the necessary categories of data be capable of being delineated sufficiently in advance, and that storage periods be determined prior to the commencement of processing, may prove difficult to reconcile with the nature of data-intensive research, including correlational methodologies, advanced analytics, machine learning and data-driven science.

In such environments, the purpose of processing cannot always be defined in a linear and static manner from the outset. The scientific value of a dataset may reside precisely in its capacity to generate unanticipated hypotheses, establish interdisciplinary connections or enable reuses that only emerge at later stages of the research process. An excessively rigid application of the principles of purpose limitation and data minimisation in this context may curtail the capacity for scientific discovery, restrict the cross-disciplinary reuse of data and disincentivise the creation of datasets with high future-use potential, even where their processing is subject to robust safeguards.

It is considered appropriate for the Guidelines to reflect more explicitly that, in the field of scientific research, the principles of purpose limitation and data minimisation admit of a dynamic, contextual and proportionate application, adapted to the nature of the research in question and the data governance environment in which it is conducted. In particular, it would be beneficial to clarify that multidisciplinary reuse and the preservation of datasets with high scientific potential are compatible with the GDPR where appropriate safeguards are implemented, alongside periodic necessity reviews and robust access and use controls.

Observation on Section 8.3

An observation is submitted with respect to the approach adopted in section 8.3 on anonymisation and pseudonymisation, on the grounds that, whilst containing useful elements, it maintains an overall conception of anonymisation that is particularly demanding and may hinder operationally viable solutions for data-driven scientific research.

The Guidelines state that personal data processed for scientific research purposes should, in the first instance, be anonymised wherever the purposes of processing can be fulfilled using anonymised data, and further recall that the anonymisation process itself constitutes a processing operation subject to the GDPR. These propositions are understandable within the general architecture of the Regulation; however, their formulation may in practice set an anonymisation threshold that is operationally very difficult to meet in many research environments.

The issue is not to contest the priority assigned to anonymisation where it is genuinely achievable, but to avoid that priority being configured as an idealised standard whose practical consequence is the inviability of functional data reuse models, even in cases where other robust safeguards

could be applied, such as enhanced pseudonymisation, functional separation, secure processing environments, access restrictions, re-identification prohibitions, or complementary organisational and contractual measures.

In this regard, the Guidelines themselves contain valuable elements that merit further development, such as the recognition of secure processing environments, conditions for further access, confidentiality obligations, and the utility of maintaining certain datasets in pseudonymised format where necessary for reproducibility, verification or review of results. It would be desirable for these instruments to appear not merely as ancillary measures, but as components of a layered protection model, equally valid where full anonymisation is not technically achievable or scientifically compatible with the purposes of processing.

Accordingly, it is considered appropriate for the Guidelines to reinforce a more contextual approach to the assessment of anonymisation, having regard not only to the abstract characteristics of the data, but also to the actual processing environment, the technical, organisational and legal barriers to re-identification that are effectively in place, and the scientific functionality of the use case. This approach is consistent with the judgment of the Court of Justice of the European Union in Case C-413/23 P, EDPS v SRB, pursuant to which the determination of whether data constitute personal data — and, by extension, the assessment of their degree of anonymisation — requires consideration of the specific context of processing and the reasonable possibilities of identification available to the controller or to third parties, rather than an abstract assessment detached from the actual processing environment. An approach to anonymisation that does not incorporate this contextual perspective is therefore not fully consistent with the criteria consolidated by the Court, which the present Guidelines should reflect more explicitly.

It is further considered appropriate to expressly acknowledge the legitimacy of maintaining data in pseudonymised format or, where strictly necessary, in identifiable form, where this is indispensable for the purposes of scientific reproducibility, review of results, quality verification or the management of incidental findings.

In light of the observations set out above, the initiative of the European Data Protection Board is once again welcomed, and these contributions are respectfully submitted for the Board’s consideration, with the aim of

supporting a final version of the Guidelines that reflects, in a sufficiently broad and operationally viable manner, the diversity of contexts, actors and methodologies that characterise contemporary data-driven scientific research, and that provides a framework ensuring, in an effective and proportionate manner, both the protection of data subjects' rights and the viability of the European research and innovation ecosystem.

Signed

RICARD JOSEP| MARTINEZ| MARTINEZ

Fecha: 2026.06.25 19:34:28 +02'00'