23.06.2026

Statement on the Public Consultation: Guidelines 1/2026 on processing of personal data for scientific research purposes

The University of Helsinki appreciates the opportunity to provide its opinion on this Guideline.

General comments on the guidelines

Overall, the guideline includes several good and welcomed clarifications and explanations. Our main concern is the guideline not taking into account that Article 11 of the General Data Protection Regulation (GDPR) reflects a strong data minimisation perspective by establishing that controllers are not obliged to maintain, acquire, or process additional information solely for the purpose of identifying data subjects when such identification is not necessary for the processing purpose. This provision reinforces the principle laid down in Article 5(1)(c) GDPR, according to which personal data must be adequate, relevant, and limited to what is necessary for the purposes of processing, thereby discouraging excessive data collection and promoting anonymisation or pseudonymisation where possible. The guidelines are partly in conflict with the aforementioned principles.

The principle of freedom of research, enshrined in Article 13 of the Charter of Fundamental Rights of the European Union, protects scientific research as a fundamental value, allowing researchers a degree of autonomy in accessing and using data. The GDPR acknowledges this through specific provisions, particularly Article 89, which permit certain flexibilities and derogations from data subject rights for scientific research purposes. The guidelines seem to have partially forgotten this fundamental right and fails to balance it with data protection.

Scope of concept of “processing of personal data for scientific research”

The University of Helsinki considers the conceptualisation of scientific research to be well executed and welcome the clarification that research can also have commercial objectives.

What we consider to be missing from the guidelines is verification of the research results. An integral part of research quality assurance is the verification of research results by another researcher outside the original research group. For this purpose, the processing of original research data (that might include personal data) is often needed. It is quite common nowadays that the publishers of research articles require that the data is made available to other researchers (for verification purposes). This should be noted in the guidelines, along with the clarification that this is also part of the processing for the research project in question and not regarded as the use of data for additional purposes.

Data Protection Principles

Presumption of purpose compatibility

It is good that EDPB has emphasized that a compatibility test is not needed when the data is processed for scientific research purposes. It is also good specification that the same legal basis can be used. However, the guidance does not take into account that the legal basis can also change. For example, if the data is collected based on legal obligation or contractual relationship, it should be possible to further process the data for scientific research based on for example the legal basis “research carried out in public interest” (in Finland this legal basis is commonly applied).

Storage limitation

University of Helsinki appreciates the clarification that if it is not always possible to specify the purposes of a future research project at the time of data collection, it is enough to specify potential research in a certain area or research.

Lawfulness

The role of consent as a basis for processing has been emphasized and has received significant coverage compared to other bases for processing. In comparison, the possibility to use ‘public interest’ as a legal basis has only been mentioned shortly. In Finland, the legal basis ‘public interest’ as a basis for processing is used widely, as it is defined in the § 4.1. of the Finnish Data Protection Act.

We would like to point out that University of Helsinki considers the use of consent to be a problematic basis for data processing in the context of research. This is due, among other things, to the fact that a large portion of research is conducted by public organizations, i.e., universities, universities of applied science, and state-owned research institutes, that also act as public authorities. Consent is an incompatible option in cases where the data subject is in a weaker position related to the data controller. This is why the processing should primarily be based on another legal basis. Another reason is that if the research subject were to withdraw their consent, the researchers would have no choice or opportunity to assess whether they still have a legitimate argument to process the data for research purposes (while respecting the data subject’s rights).

When the research is carried out on the legal basis of public interest, it allows the researchers to assess, on a case-by-case basis, whether the purpose of the research would be endangered if a data subject’s data were removed from the dataset after its collection. In many cases, the analysis of the data may already be underway, and some results may even have been published. If, at this stage, a data subject withdraws their consent and seeks the erasure of their data, this could also affect the ability to verify the research results, as the dataset would no longer correspond to its original form.

Can guidelines about the broad consent be applied as a guide for informing when the legal basis is public interest?

University of Helsinki considers that what is written, for example, about the broad consent, can be applied also to cases where the legal basis is public interest when it comes to informing the data subjects. This could be clarified further in the guidelines. For example, when it comes to the statement ”a controller can obtain a broad consent from data subjects when collecting and storing personal data for processing in future research project within certain areas of scientific research” the same logic could be applied when one is collecting data from data subject and informing them that the legal basis for processing is research carried out in public interests. It would also then be

enough to inform that the data will be processed in future research project within certain areas of scientific research.

Consent to participate vs. legal basis for processing

It is good that the guidelines distinguish between the consent to participate and consent as a legal basis. However, the guidelines do not mention the possibility of obtaining consent to participate and use another legal basis than consent as a basis for processing (for example, research carried out in the public interest). This is common practice in Finland, as the Finnish Data Protection Act permits public interest to be used as a legal basis. This aspect should be included in the guidelines.

Public interest or exercise of official authority as a legal basis

University of Helsinki appreciates the clarification that the use of legal basis “public interest or exercise of official authority” is not limited only to public entities conducting scientific research.

Obigations to inform

General

The chapter about informing appears, taken as a whole, to approach informing obligations in a way that gives limited weight to the characteristics of scientific research and gives the impression that information should be provided purely for the sake of providing information. In research contexts, data is generally pseudonymized whenever possible and processed in a form that does not allow the controller to identify the research participants. Research usually focuses on broader phenomena rather than on individuals. In addition, research does not generally involve decision-making that affects data subjects directly. If the risks to data subjects can be minimized to a reasonable level, and the data is processed in a form that does not allow the research subjects to be identified, what is the purpose and value to use limited resources to inform the research subjects about every study in which their data has been used? For example, just in the University of Helsinki there are thousands of research projects, this would cause a significant burden for research in a big picture.

Compared on the data minimization principle and Article 11, the requirements regarding informing seem somewhat contradictory, as the guidelines urge the collection and retention of data that would not be otherwise necessary for the conduct of the research, but solely so that the research participants can be informed about how their (mostly otherwise pseudonymized and even quite anonym) data is processed as part of the overall research dataset.

It is clear that one of the fundamental principles of the GDPR is transparency in data processing. However, the question arises: is being informed such core value that should be pursued merely for the sake of formality? Especially in the case of register-based studies, where data is processed in a pseudonymized form (meaning that very few people would even be able to identify the research subjects) providing information seems like a somewhat meaningless formal step that would result (if it were even possible to implement in accordance with the EDPB guideline) in a significant burden and possible flood of notifications.

Contact details also change quite often so in longlasting studies e.g., age cohort studies, it would be quite impossible to keep contact details up to date. At least, that would cause a significant burden to the research group. A better solution would be to give data subjects a link to a webpage where the data subjects can read how the research projects evolve, if interested. It would be justified to store identification and contact information only if it is necessary for the research itself, for example, because there is a known need to contact the data subjects to collect additional information or samples.

About further processing when the data is collected from the data subjects themselves (article 13 (3) GDPR)

This section is broadly aligned with GDPR principles of transparency and purpose limitation, but it could be improved to reflect better the realities of academic research and long-term data reuse. The requirement for transparency should also not constitute an obstacle to apply Article 89 and Article 5 (1. b)) of the GDPR, especially if the data is originally collected by an organization other than the one who is carrying out the research (for example, customer data that will be used as registry data in scientific research).

In particular, the requirement to inform data subjects of further processing “in advance” should more explicitly recognize the operational and methodological constraints of largescale, longitudinal, and secondary-use research that is typical in universities where re-contacting participants may be impractical. The guidance should therefore more clearly acknowledge that appropriate safeguards under Article 89(1) GDPR, such as ethics approvals, data minimization, and technical and organizational security measures can, in most cases, provide an equivalent level of protection when direct re-contact is not feasible. Also, it could be stated clearly that it is enough to inform about the further use of the data in other research projects when the data is collected for the research from the data subjects. Furthermore, the expectation not to delete contact details if future research use is anticipated should be compared against the principle of data minimization and storage limitation.

A more balanced approach would allow controllers to rely on layered transparency (e.g. public websites) as a primary mechanism in cases of secondary research use, without creating a de facto obligation to retain identifiable data solely for potential future contact. Clarifying these points would help ensure the guideline supports both data protection and the societal value of scientific research.

About provison of information when data is collected from other sources than the data subjects (article 14)

This section would also benefit from clearer alignment with the practical realities of academic research. The guidance should clarify that the obligation to provide information must be proportionate to the context of the research, and that it may be fulfilled through layered transparency mechanisms, such as publicly accessible project descriptions, particularly where individual notification would involve disproportionate effort or risk undermining the scientific research or the validity of the research.

The recommendation to enable data subjects to exercise their rights via pseudonyms should be carefully considered, as in many university research settings re-identification is intentionally restricted under Article 89(1) safeguards. The guidance should therefore clarify that alternative rights-enabling mechanisms (such as institutional contact points or mediated access procedures) are acceptable where direct identification is neither possible nor appropriate.

The section about exceptions contains important and generally welcomed recognition of the exemptions under Article 14(5) GDPR, particularly in relation to disproportionate effort and impairment of research objectives. However, the overall framing could be improved to better reflect the central role these exemptions play in enabling large-scale, retrospective, and register-based research commonly conducted by universities. In particular, the statement that the exceptions “must be interpreted restrictively and should not be relied on routinely” risks creating legal uncertainty and unnecessarily discouraging legitimate reliance on Article 14(5)(b) in contexts where informing data subjects individually is genuinely impracticable or would undermine scientifically valid methodologies. The guidance should more clearly acknowledge that, in many

research settings such as longitudinal cohort studies, national registries, and secondary use of administrative data reliance on the “disproportionate effort” and “impairment of objectives” grounds are not exceptional but an integral and well-justified aspect of GDPR compliant research, provided that appropriate safeguards under Article 89(1) are in place.

The guidance includes an example: “A high number of data subjects to inform, for example large registries with patient data, population data or data on pupils accumulated nation-wide, while many contact details are outdated due to the age of the dataset. This exemption cannot be relied on if the controller can lawfully obtain a high number of contact details easily, for example through the use public registries.” As stated above, this example is partially in conflict with the data minimisation principle and the Article 11 because usually this would require storing the data in a form that the data subject can be identified and findable from the data set, usually this is not relevant for research purposes. E.g., to be able to specify wich one of the “Sam Smiths” is in the research data it would require the collection of the personal identification numbers to be able to find the right contact details. This information is not relevant in many (or most) of the research fields. And as mentioned, it is quite common that researchers are not interested about the individual as a person, rather solely the data and larger phenomenon. It is thus not sensible to store the data in a format that could help to identify the persons on the datasets, this could also be a risk towards the loss of anonymity and pose an unnecessary security risk to the data subjects.

Overall, the guidance would benefit from a more balanced articulation that equally emphasises both transparency and the necessity of proportionate, feasible implementation in order to safeguard the societal value of research. This is a good mention in the guidelines and should be highlighted: “When assessing the effort involved to provide the information to data subjects against the impact and effects on the data subject if they are not provided with the information the controller should consider the safeguards that it has adopted pursuant to Article 89(1) of the GDPR, as well as measures to provide information indirectly.” If the research groups really were to inform individuals about all the research projects their data is involved, it would cause a flood of informing which cannot be the goal when the data can be processed for scientific research without causing too much risk to data subjects.

Attribution of responsibility

Joint controllers

About scope of joint controllership

When assessing controllership, greater value should be taken to the diverse affiliations under which research is conducted at universities. A researcher may not necessarily be employed by the university, yet may rely on the university’s authority, reputation, and infrastructure in carrying out the research, while also contributing to the university’s mission of promoting scientific research. At the same time, an individual researcher or research group may retain an independent ability to define the research questions and research protocol.

Such circumstances arise, for example, in relation to doctoral researchers and grant-funded researchers, whose activities appear externally as part of the university’s research activity and who present themselves to data subjects as representatives of the university. This is relevant to the data subject’s reasonable expectations: a data subject may reasonably understand that they are participating in research in which the university is involved as a controller. This may also strengthen the data subject’s willingness to participate in the research.

The university may also influence the researcher’s research plan, for example where the researcher operates as part of a university research group or where the funding obtained by the researcher has been awarded to the university.

Even where the researcher independently determines the research question and research design, access to university's support services may be conditional on the use of university-provided devices, systems, document templates, and other procedures, as well as compliance with the university's instructions on the processing of personal data and information security.

The position of students, too, is often less independent than the formal setting might suggest.

A master's thesis meets the criteria for scientific research as defined in the guidelines. At the same time, the processing carried out by a master's student is connected to their objective to complete a degree within a university degree program, which places the student in a relationship of dependence with the university and under the supervision of their academic supervisor. The nature of the degree program may also require the processing of personal data as part of the master's thesis. This is particularly the case in medicine, educational sciences, and many fields within the humanities.

In such situations, controllership should not be assessed solely on the basis of the student's formal independence. Rather, the assessment should take into account the actual influence exercised over the purposes and means of the processing of personal data, as well as the institutional framework within which the processing takes place.

A bachelor's thesis is an academic exercise, and at the University of Helsinki it has not been regarded as meeting the threshold of scientific research from the perspective of the legal basis for processing. Nevertheless, the same considerations concerning controllership and the relationship of dependence are also relevant in that context.

The University of Helsinki thus considers that the university and non-employed researchers, such as grant-funded researchers, doctoral candidates, and undergraduate students, should be regarded as joint controllers, because even though a researcher or student has the autonomy to choose their research topic, this research is conducted in in close connection with the university and the university as an organization is supporting and enabling the research.

Allocation of responsibilities

The guideline states that the joint controllers must lay down their mutual responsibilities in an agreement. However, this exceeds the requirement of the GDPR Article 26 which requires an arrangement in transparent manner. The term arrangement allows the more freedom for the joint controllers to assess what is the most suitable manner to organize their mutual responsibilities.

Occasionally a written agreement may be appropriate but in certain situations other arrangements are more functional. For example, in regard of joint controllership with students and non-employed researchers, the University of Helsinki has implemented a clause in the data protection notice in which the allocation of responsibilities is written down. A written agreement would be too rigid in situation where the responsibilities are divided based on each party's *de facto* abilities, and when there is little to no room for negotiating the division of responsibilities.

Appropriate safeguards

In the guideline, it is stated that "personal data used for scientific research should, wherever possible, be anonymised, provided that the research purposes can still be achieved with anonymised data. Under the GDPR, anonymised data refers to information that does not, or no longer, relate to an identified or identifiable natural person. Once data is truly anonymised, its further processing falls outside the scope of the GDPR. However, the anonymisation process itself constitutes processing under the GDPR and must comply with its requirements, including data protection principles and lawfulness."

The guidance appropriately emphasises that anonymisation should be pursued where research purposes can be achieved without identifiable data and correctly clarifies that truly anonymised data falls outside the scope of the GDPR. However, the wording that the further processing falls outside the scope of the GDPR may be overly simplistic and lead to misunderstandings, and next sentence might not clarify what is meant because the language used is to some extent “legal jargon”. It would be useful to emphasize that, regardless of the GDPR’s applicability, scientific research remains subject to research ethics standards and possible sector-specific regulation. For example, a researcher cannot collect data for research purposes with informing something and then anonymise the data and use it for whatever purpose.

Another example is when a university requests a dataset which contains patient health data from a hospital. The hospital removes all identifiers and ensures that the information cannot be linked to identifiable individuals by the research group. The dataset provided to the university could thus interpreted to fall outside the scope of the GDPR if the research group cannot identify the individuals from the data. It is worth pointing out that the actual research might still be subject to sector-specific legislation, rules regarding ethical review, etc.