SND
Swedish National Data Service

Date: 28/02/2025
DNR GU 2025/824

Feedback on the EDPB’s Guidelines 01/2025 on Pseudonymisation

We appreciate the opportunity to submit this feedback on the Guidelines 01/2025 on Pseudonymisation on behalf of the Swedish National Data Service (SND), which is part of Gothenburg University.

SND’s primary function as a research data repository is to support the accessibility, preservation, and reuse of research data and related materials.

We welcome the guidelines and clarifications that confirm our understanding of the use of pseudonymisation. However, there are two matters that we wish to address.

Comments

1. Guidance regarding the distinction between pseudonymous and anonymous data

We would like further guidance and clarification on whether all pseudonymised data always must be considered to be personal data or whether there are situations in which such data could be regarded as anonymous to a third party. We refer in part to the argument in the Judgment of the General Court (Eighth Chamber, Extended Composition) of 26 April 2023, Case T-557/20 and the opinion of Advocate General Spielmann, delivered on 6 February 2025 in Case 2013/23 P. In addition, we wish to present the following arguments:

In our opinion, there are situations in which the risk of identifying an individual is non-existent or insignificant, in accordance with Recital 26 of the GDPR and despite the existence of a code key or similar re-identification mechanism.

We wish to illustrate our position with the following example: Several healthcare providers collect personal data within the framework of a clinical trial or other kind of research study. The personal data is then transmitted to a university (University 1). Before the transmission, each healthcare provider applies robust pseudonymisation, and the recipient, University 1, is unable to identify any individual through legal means (e.g., due to national mandatory laws such as secrecy laws). In the next step, a third party, such as a research institute, conducts a study and collects datasets from multiple universities, including University 1. Each university holds its own dataset containing pseudonymised data originally collected by other entities. Before sharing their respective datasets with the research institute, the universities pseudonymise the data again (applying an additional layer of pseudonymisation). The research institute then processes the data in accordance with their study protocol and seeks to publish e their dataset to make them openly accessible to other scientists. Before publishing the dataset, the research institute removes all code numbers assigned to the original study subjects but still retains the dataset containing the coded study participant numbers. Since there is a “chain” of code keys, “starting” at the research institute, this published dataset could be considered personal data. Nevertheless, we find that there are strong grounds to argue that not only the published dataset, but also the dataset held by the research institute, no longer should be regarded as personal data. In our opinion the data should be considered anonymized (provided that the dataset itself does not contain any means of attribution or re-identification, taken into account all the means reasonably likely to be used, see Recital 26 of the GDPR). According to the reasoning in case T-557/20, even the dataset held by University 1 could be regarded as anonymous for the university.

The healthcare provider in this example could be replaced with any other institution, such as social services offices or municipalities.

In our opinion, further guidance on this matter and a reasonable distinction between pseudonymous and anonymous data are of crucial importance for researchers reusing existing personal data and data repositories working with research data containing personal data. Another aspect is that the interpretation of what constitutes anonymous data appears to vary across different legal systems. Consequently, collaborations with entities in countries outside the European Union (such as the National Institutes of Health in the United States) and research data repositories have significant difficulties when the definitions of personal data differ.

2. Guidance regarding aspects of the risk of re-identification

We have noted that the risk of re-identification is addressed to some extent. However, we would prefer further guidance on how this risk can be both assessed and minimized. The guidelines state that additional information may also exist beyond the immediate control of the pseudonymising controller or processor. In recent years, the rapid development of modern technologies has made it increasingly difficult to determine which data could be re-identifiable using available technology. It would be beneficial to include guidance on the key aspects to consider, whether there are best practices for conducting an adequate assessment, and examples of major errors that should be avoided when assessing re-identification risk.