Med Law consult

Comments on draft Guidelines 01/2025 on Pseudonymisation

By Evert-Ben van Veen, LL.M. , March 14 2025

Introduction

Two motives have led to this commentary.

The first is my ongoing involvement in observational health research. The second is my concern in general about the 'rule of law' and the separation of powers when it concerns powerful unelected single issue supervisors. Both motives come together when the EDPB issues draft guidelines on what have traditionally been considered the workhorses of health research, namely pseudonymised data, originally called coded data.

In this introduction I start with the first motive and then discuss the second.

About the first: There is a data-chain in observational health research: from the original data sources to the secure research databases where data are analysed. Privacy enhancing techniques from the data sources to that research database have a long tradition in this respect.[1] However, in that data chain participants must be uniquely distinguished to find possible statistical patterns between exposure and health. Direct identifiers of each participant will be removed and replaced by a unique distinguishing number. Researchers have a long track record of keeping these data safe. Even if the research data in the research database would in theory be re-identifiable, there are no instances known to me where that has actually happened in real world scenarios of inside or outside adversaries.

As what was originally called coded data, are now often pseudonymised data, further discussion on pseudonymisation as defined in the GDPR is welcome. But the EDPB should in that case correctly reflect the present case law or even wait till the CJEU has reached its

[1] For an example see: Veen, Evert-Ben van. 'Obstacles to European Research Projects with Data and Tissue: Solutions and Further Challenges'. *European Journal of Cancer*, Cancer Control in Europe: State of the Art in 2008, 44, no. 10 (1 July 2008): 1438–50. https://doi.org/10.1016/j.ejca.2008.03.011.

verdict in the SRB-EDPS case. So, there comes the second motive. The draft guidelines consider all pseudonymised data also personal data. As such this draft does not sufficiently reflects present case law which has a moderate view on when data can be considered anonymous. The draft is also untimely as the decision of the CJEU will shed a new light on the status of pseudonymised data, certainly if the Court would follow the Opinion of the Advocate General.[2]

In the following I will briefly expand on the issues raised in this introduction and their relevance. I will conclude with a few comments that will probably not be appreciated but hopefully will also lead to some introspection at the EDPB.

The draft Guidelines do not reflect the case law on whether data can be considered anonymous

In their current form, the draft guidelines fail to address recent case law and even appear to tacitly endorse the WP29 opinion on anonymisation. I have mentioned the data chain already. The possibility of pseudonymised data being personal data in the hands of one party, such as the sender of the data, and anonymous data for another, such as the recipient of the data, is not only ignored, but also seemingly dismissed.

The rebuttal of the EDPB could be that the Guidelines only discuss pseudonymisation in the sense of the GDPR and not pseudonymised data. In the GDPR pseudonymisation is defined as a specific privacy enhancing processing of personal data and the draft Guidelines only discuss how you can organise that pseudonymisation properly. Such a rebuttal would not be fair. As the draft guidelines never raise the point that, as a result of these transformations, the data might not be personal data anymore. I will come back to that in the concluding remarks.

The importance of the question whether pseudonymised can also be considered anonymous data by the recipient

There are two reasons.

The first one that it currently matters for observational research from a regulatory point of view. Many member states still rely on the 'consent or anonymise' approach for much observational health research. The EDPB also seems to hold this view in example 5 of the draft guidelines where the study takes place 'with data subjects signing up for the study'.

[2] ECLI:EU:C:2025:59

Though a prospective study into unknown effects of labour conditions is certainly feasible and can be interesting, much of this research is retrospective. Unexpected incidences of disease lead to looking back to persons who were in similar circumstances and might have been exposed to a yet unknown variable, trying to find a pattern for a specific variable. That is amongst other things how the link between mesothelioma and working with asbestos products was discovered. Outside labour conditions the link between (lung) cancer and smoking is a prime earlier example. Such retrospective research cannot be based on consent for a variety of reasons, ranging from the practical issues to reaching out to a large number of persons concerned on their current addresses (which might be unknown), the costs of doing so especially as usually reminders must be sent to bias created when certain groups do not respond also after a reminder. It should be noted that the statistical outcomes of this research must be robust as they are often unwelcome and likely to be challenged by the operators in the field which can also be governmental agencies and not only industry. To consider this research as research with pseudonymised anonymous data, is of extreme importance to those member states which currently rely on the 'consent or anonymise' approach.

The second reason is the following and is also important for those member states which have a more public health research friendly approach in their legislation. It must always be explained by the data sources what is done with the data when they are released for research. Research institutions as well must be transparent. Apart from the conditions in the GDPR from article 12 to article 14, or article 11 when applicable, it is then important how the message can be conveyed. That only anonymised data are released and that the research institution only analyses anonymous data is obviously a much easier and still honest message than that the data are pseudonymised and that researchers cannot reasonably re-identify you but that in theory they might. With a clear statement that pseudonymised data can under circumstances also be anonymous data, as these are two different concepts under the GDPR, much confusion in the field and discussions with lawyers can be avoided. Of course, in those cases researchers must still justify that the data they analyse as recipients in the secure processing environments of their research institution, are indeed anonymous to them. That justification can be based on the present case law and not on the outdated Opinion 05/2014 of the then article 29 Working Group. Researchers can rely on the many safeguards in place in this respect, ranging from the technical safeguards, various transformations in the data though they still must retain their usability for often quite nuanced research, and the professional and legal conditions which forbid re-identification.

Singling out is not re-identification and does not as such make data personal data

As shown above, researchers need to uniquely distinguish between the persons in the data they analyse. The variables must be connected to the persons to whom they apply and not to others. However, this will never have direct effects for each of them. Persons with the same variables will be grouped together and the outcome is statistics about such groups. Those will be discussed in the literature and other fora, such as the political. In the end, these results can lead to policy decisions which may affect certain groups, such as higher taxes on addictive substances, or to a ban of certain products. This is to show that the seemingly wider interpretation of personal data in the IAB case[3] does not apply to health research with data in the research environment of the research institution. In the IAB case there were direct consequences for the data subjects, being possible 'personalised' advertisements when opening a website.

When doing -omics research, there can be incidental findings. In that case that result will be transmitted to the health care provider who sent the data. It is then to the health care provider given his professional treatment relation with the patient, whether and how this finding will be communicated to the patient. Communication will in principle not take place if the patient has indicated a whish not to be informed of incidental findings ('the right not to know'). Also in this case, observational research as described here, will not directly lead to consequences for the data subject. Insofar as there are, the route is indirect, mediated by the health care provider and wanted by the patient.

Concluding remarks

The GDPR has a specific definition of pseudonymisation. At the advent of the GDPR I warned already that this definition does not correspond with how the concept of 'pseudonymised data' is used in health research and that another term for those coded data, as a more generic term, might be more appropriate.[4] But the term is there and then the question is how it can be used diligently. The EDPB ignores that the result of pseudonymised data can result in anonymous data. It issues the draft Guidelines while not referring to earlier case law and while it obviously knows that new case law is arriving. This might be seen as influencing the debate but worse would be that the EDPB simply denies the case law. If you

[3] ECLI:EU:C:2024:214

[4] Veen, Evert-Ben van. 'Observational Health Research in Europe: Understanding the General Data Protection Regulation and Underlying Debate'. *European Journal of Cancer* 104 (November 2018): 70–80. https://doi.org/10.1016/j.ejca.2018.09.032.

apologise that I don't give references, I remember how the EDPB/article 29 WP immediately embraced the 'Google Spain' verdict of the CJEU[5] but was silent for a very long time about the Breyer judgement[6] and still is in the present draft Guidelines. The EDPB might not have liked Breyer and its consequences as the EDPB probably prefers the GDPR as the 'law of everything'.[7], [8] I wonder whether with the draft Guidelines the EDPB acts more as an interest group promoting its single issue of interest than as a supervisory authority giving nuanced guidance to the law as it is interpreted by the highest authority in this respect.

[5] ECLI:EU:C:2014:317

[6] ECLI: EU: C: 2016: 779

[7] Nadezhda Purtova (2018) The law of everything. Broad concept of personal data and future of EU data protection law, Law, Innovation and Technology, 10:1, 40-81, DOI: 10.1080/17579961.2018.1452176

[8] From a different perspective Groos, Daniel, and Evert-Ben van Veen. 'Anonymised Data and the Rule of Law'. *European Data Protection Law Review* 6, no. 4 (2020): 498–508. https://doi.org/10.21552/edpl/2020/4/6.