EFPIA Response to EDPB Consultation on Draft Guidelines on Pseudonymisation

Author: EFPIA ● **Date:** 12 March 2025 ● **Version:** **Final**

On 16 January 2025, the European Data Protection Board ("EDPB") released draft guidelines for consultation on the concept of pseudonymisation ("guidelines"). For the pharmaceutical sector, pseudonymisation is a core data management practice, not only because of its intrinsic privacy enhancing qualities but also because of regulatory obligations, for example, under the Clinical Trials Regulation ("CTR") and Good Clinical Practice ("GCP"). For this reason, EFPIA members took a keen interest in the draft guidelines and are happy to contribute to the consultation, especially as the guidelines in their current form appear to present some specific challenges for the sector and scientific research in general.

General observations

At the outset, EFPIA notes that the guidelines are often quite complex and difficult to understand. As a result, it is impossible to address every argument or sentence that is not entirely clear. EFPIA's comments therefore focus on points of principle for purposes of this response, and EFPIA reserves its views on more specific elements of the guidelines.

Also, a discussion of the concept of pseudonymisation necessarily touches on the boundaries between that concept and the related concept of anonymisation. As the EDPB are aware, pending before the CJEU is a case that focuses upon these related concepts, and the position taken in the guidelines (§22-23) is already the subject of a challenge before the CJEU in the *SRB v EDPS* case (T-557/20). Furthermore, the CJEU has already ruled in the Breyer case (C 582/14) that certain indirect identifiers qualify as personal data only if a party has "legal means" to obtain the necessary additional information to reidentify data subjects. The logical consequence of this ruling is that, in the absence of such legal means, the dataset should be considered anonymous. The current guidelines, however, fail to acknowledge this important nuance and instead assume that any potential for re-identification, even if legally or practically improbable, is sufficient to classify data as personal.

If the recent opinion of the Advocate General in *SRB v EDPS* were to be confirmed, it would have an important impact on the appropriate designations given to data, in particular in the scenarios where coded data are shared with third parties. We thus believe that it would be appropriate for the Board to suspend its work on the guidelines until the CJEU has rendered a decision in this case.

No mandatory or prohibited pseudonymisation techniques

The guidelines appear to fluctuate between characterising pseudonymisation as an optional measure for controllers to implement with recommended techniques (§25) and a mandatory measure that controllers must implement, for example, to meet their privacy-by-design obligation (§30). In the latter case, the recommendations could be read as thinly-veiled technical mandates. This is especially apparent in Section 3 of the guidelines that discusses technical measures and safeguards for pseudonymisation. For example, §116 states that *person pseudonyms* are "admissible if and only if […]", suggesting that some techniques are not acceptable.

It should be clear that controllers can decide on the pseudonymisation techniques that they wish to apply, taking into account the context of their processing operations and recognising that some techniques may be stronger than others. The guidelines should neither exclude nor impose particular techniques.

Similarly, while the examples in the annex to the guidelines are useful, they should only be examples and not prescriptive. These examples may represent best-practices, but they should not constitute mandatory requirements.

Overall, the recommended practices in the guidelines once again appear to increase the compliance burden on controllers and to add much red tape and cost (*e.g.*, the extensive use of trust centres in the examples). It is not always clear that these burdens and cost are commensurate to the added value the relevant techniques provide. An overly detailed assessment like the one set out in the "summary procedure for pseudonymisation" (Section 3.4), for example, almost inevitably results in a time consuming and potentially expensive case-by-case assessment, complicating current pseudonymisation practices that are highly standardised. This risks increasing the cost of scientific research, not only for the pharmaceutical sector, but also for other research bodies such as university hospitals and other public and private research bodies. Such burdens are precisely what policy-makers try to avoid or mitigate.

The standard of pseudonymisation

Pseudonymisation is a protective measure only and does not equate to anonymisation per se (although it could have that result for third party recipients under certain conditions – see *SRB* case). The fact that some level of identification remains possible does not mean that the pseudonymisation applied is not appropriate. Ultimately, pseudonymised data remains personal.

In this respect, EFPIA is of the opinion that the standard for pseudonymisation proposed by the Board is too strict. For example, the guidelines indicate that pseudonymisation requires:

- "measures to ensure that the personal data are not attributed to an identified *or identifiable* natural person" (§5);
- "It requires additional technical and organisational measures to ensure that the personal data are not attributed to an identified *or identifiable* natural person" (§9);
- that recipients of pseudonymised data cannot "*single out* the data subject in other contexts on the basis of what they learned from handling the pseudonymized data (§47, 50 and 64). (own emphasis)

This is overly restrictive. The objective of pseudonymization is not to prevent the singling out of an individual. It does not prevent an individual from being *identifiable* (*i.e.*, by means of additional information and in contrast to *identified* or directly identifiable individuals). The guidelines appear to suggest that data has to be almost anonymous (*i.e.*, no singling out or other possibility to indirectly identify the data subject) in order to be properly pseudonymised. This is unnecessary because recipients of pseudonymised data relating to an *identifiable* individual, or that allow for mere *singling out* of an individual, still process pseudonymised data (if not anonymous data – see *SRB* case).

The guidelines themselves explain that attribution in relation to an identifiable individual means "to link the data *to other information* with reference to which the natural person could be *identified*. Such a link could be established on the basis of one or several identifiers or identifying attributes" (§17). This explanation concedes that *other information* is required, so why is this data not still pseudonymous in the absence of the other information?

Controllers should make sure that data subjects cannot be *directly identified* by parties in the pseudonymisation domain (*i.e.*, parties that can be expected to have access to the pseudonymous data and for whom the data must remain pseudonymous – see §35), because that would undo the pseudonymisation. The scenario in §101 is a good example of such attribution to a directly identifiable individual. Similarly, §131 indicates that controllers should verify "which attributes contained in the

personal data can be used alone or in combination to attribute some of the data to data subjects, directly *or indirectly*, within the pseudonymisation domain, *considering information that can be accessed with reasonable effort from within it*." This latter addition is essential because that is what makes the data directly identifiable and destroys the pseudonymization. It is important that the guidelines clarify this point on additional information, and it would be clearer if the guidelines referred to "directly identifiable" instead of "identifiable". In any event, the prohibition on mere *singling out* should be revised.

Also, even if an individual would be directly identifiable in a data set because of additional information available with reasonable effort to an entity in the pseudonymisation domain, this does not mean that the entire data set is affected. The other data in the data set remain pseudonymised.

Impact on the pharmaceutical sector

The importance of our observations above become clear when applied to the pharmaceutical sector and clinical trials. Clinical trial data must be pseudonymised by the hospital/investigator, pursuant to the CTR, before it is shared with the sponsor of the trial (often a pharmaceutical company). However, this data is very detailed, containing precise measurements, dates of interventions, genetic markers, etc. Such detailed data is necessary in order to identify risks and benefits of treatments depending on the demographic characteristics of clinical trial participants. Moreover, the collection and retention of this detailed data is required by clinical trial rules and necessary to obtain an authorization for a new medicine or designation. While this mandatory pseudonymisation may still allow for the singling out of individuals in such a dataset, this data is still not directly identifiable and should qualify as pseudonymised.

In addition, the regulatory framework is such that sponsors only receive pseudonymised data and are required only to share pseudonymised data with regulatory authorities, for example, in the context of a marketing authorization application. These authorities are thus necessarily part of the pseudonymisation domain. However, these same authorities, domestic *and foreign*, have the authority to access non-coded data at EU hospitals if necessary for investigation purposes. So, they necessarily can access *with reasonable effort* additional information that renders the trial participant directly identifiable. According to the Board's standard, this data would not be properly pseudonymised, yet clinical trial rules qualify it pseudonymised and impose this form of pseudonymisation.

Finally, the proposed pseudonymisation standard also carries risks. For example, the expectation that the pseudonymisation technique used should assign "widely differing pseudonyms to persons with similar identifying attributes" (§29) would mean that trial participants would require widely differing pseudonyms given that they share similar identifying attributes (*i.e.*, they participate in the same trial, have the same condition, meet the same trial eligibility requirements). However, applying widely differing pseudonyms in this context is uncommon and, in contrast to what the guidelines suggest, may lead to more mistakes and erroneous attribution of trial data to widely differing codes. This creates risks for the patients involved and for the integrity of trial data set.

Conclusion

EFPIA is of the opinion that the guidelines are too complex, premature in light of pending CJEU procedures and recommend an unnecessarily high standard. The guidelines should better take into account the regulatory frameworks in place to avoid the risk of conflicts with those frameworks. In addition, EFPIA calls on the Board to make a more robust assessment of cost/burden vs. added value of the pseudonymisation techniques it recommends. It is not because data protection is a fundamental right that such considerations are unnecessary or unimportant. The EU Charter's right to data protection is not an island; it interconnects with various other rights. An unnecessary increase of burden and cost for

controllers risks having an impact on health research performed in the EU with negative consequences for innovation and public health.