Criteo observations on the Draft European Data Protection Board's Guidelines 01/2025 on Pseudonymisation

March 2024

Criteo welcomes the issuance of the EDPB draft guidelines 01/2025 on Pseudonymisation adopted on January 16th, 2025 (**DG**).

Pseudonymisation is a key driver to **enable data led innovation and the development of Artificial Intelligence (AI) in a responsible manner** by enabling the safe reuse and secondary use of data. Since 2018, the EU has put in place a data economy and adopted several key legislations to enable the circulation of data and the development of data driven products and services (Data Governance Act, Data Act, Artificial Intelligence Act, Data Spaces). These regulatory developments have reshaped the role that data plays in our societies and economies and highlighted the need for data to be shared and reused to create value as long as data is used in compliance with the GDPR.

The interpretation of the GDPR is an important factor for the competitiveness of European businesses and innovations. The Draghi report has rightfully highlighted the economic and technological risks the EU could face, emphasizing the need to redirect the EU's political choices to support growth. Reinforcing the culture of risk assessment and mitigation fostered by the GDPR is a key element to navigating these choices effectively.

Criteo welcomes the explicit recognition in the DG that **pseudonymisation techniques present numerous benefits** such as:

- Reducing the risks (including the severity of the risk) to the data subject by preventing the attribution of personal data to individuals during the processing and in the event of unauthorised use;
- Enabling controllers to rely on the legitimate interest legal basis;
- Contributing to the compatibility of further processing;
- Enabling an essentially equivalent level of protection in data transfers;
- Contributing to data protection by design and by default;
- Being a safeguard enabling to meet the requirements of data protection law and demonstrating compliance with the data protection principles;
- Contributing to ensuring a level of security appropriate to the risk.

Criteo believes however that there are missed opportunities in the DG:

1. The DG are very long and **can be complex to understand**. They introduce new concepts that go beyond the wording of the GDPR. These new terms add legal uncertainty to an already complex concept — for example, 'pseudonymization transformation', 'pseudonymization domain', 'pseudonymization secrets', 'pseudonymization proxy'. Pseudonymisation should be presented as a tool promoting innovation by the reuse of data. Therefore, the DG should be easy to read and understand, as they are relevant for stakeholders of all sizes and nature, including SMEs and start-ups, the public sector, the research community or academia. They should avoid long developments and include tables, toolboxes, charts, decision trees or simple lists.

2. The DG should take more into consideration **the potentially unlimited use cases that can benefit from pseudonymisation**. Most of the examples provided in the DG relate to the health sector whereas the objectives of pseudonymisation can be very diverse and benefit potentially all private and public use cases, such as for example, public transportation, e-commerce, banking and finance, mobility, etc. A wider variety of examples would be especially relevant for SMEs. In some instances, the pseudonymisation will be set up so that it is impossible to revert to the original data and raise the question as to whether the data would not even become anonymous as per the case law of the CJEU [see Breyer case or IAB Europe case].

3. The DG make no reference whatsoever to **artificial intelligence** and the huge promise of pseudonymisation for the development of AI and model training (the latest EDPB opinion on AI models and GDPR explicitly mentions pseudonymisation in paragraphs 51 & 101). As the bar for personal data to be considered anonymous remains very high and even impossible to reach in some use-cases, the only economically viable way for businesses to train their AI models is to use pseudonymised data. The DG should clearly acknowledge this.

4. The DG rely on **a very high bar** for measures that limit access to the additional information enabling data reidentification to be deemed efficient. These measures are based on worst-case scenarios, considering not only risks stemming from lawful activities but also from illegal activities. By doing so, they put an onus on the controller to assume how third parties could behave and to know the type of data they could already hold. Unfortunately, this standard is impossible to meet for most controllers and may discourage the reliance on pseudonymisation. This may in turn reduce the effective protection of data in practice by reducing the risk of unlawful access. The DG **fail to consider the notion of good faith** and the reference to "standard market practices":

    - Paragraph 11 provides that "*the effect of pseudonymisation will have to be measured against the capabilities of persons or parties acting without authorisation*".
    - Paragraph 21 recognizes that "*additional information may also exist beyond the immediate control of the pseudonymising controller or processor*" and that such controller or processor should take this into account – which is a contradictory statement unless the controller assumes that additional information beyond its immediate control always exists.
    - Paragraph 38 is very far reaching in mentioning that "*the controller may define the pseudonymisation domain to encompass, [...] a range of or all external entities that may attempt to gain access to the data without authorisation*."

5. The DG do not make any reference to the notions of **re-identification risk and residual risk.** These are extremely important concepts, particularly when carrying out data protection impact assessments or evaluating impacts on data subjects in the case of a data breach. Such risk assessment methodology is the cornerstone of data security under the GDPR and should also be applicable in the context of pseudonymization (rather than a risk zero approach to re-identification). In the same vein, because pseudonymization enables to reduce the risks to the rights and

freedoms of individuals, pseudonymized data should be recognized as a separate specific category of data. The EDPB could acknowledge this through the creation of a category of "**low risk personal data**". This would mirror the category of "**high risk personal data**" that that EDPB created in its previous guidelines.

6. We welcome the **different use-cases,** which highlight simple and realistic pseudonymization processes that businesses can put in place in the health sector. More examples in various sectors are needed, including in the online advertising sector. We also note that some of the use cases or close variants are already used in the online advertising industry (e.g. use-cases 1 and 8). The **online advertising sector relies heavily on pseudonymisation techniques** to ensure that insights or key information about an ad placement is shared with the relevant advertising actors of the supply chain. Not only does it enable to protect personal data and reduce the risk to the rights and freedoms of individuals, but also it ensures that competitors do not have access to proprietary data. As a seminal example, personalised advertising requires the processing of user timelines (i.e. sequences of actions shown or made to a specific user, such as product views or sales events) which are personal data. Such timelines are the cornerstone of the online advertising industry as they materialise the user purchase intent and needs. The cookie ID used to build such user timelines is stored within a matching table with restricted and controlled access and replaced by a random pseudonym in databases used to train AI advertising models. Without the matching table and other sources of data, it is impossible to re-identify a specific user only using user timelines.

7. The DG should remain **neutral and not assume that certain practices**, industries or use cases cannot rely on pseudonymised data. Such an approach limits innovation and discriminates against an entire industry in the application of the GDPR. Paragraph 48 provides that pseudonymisation can help prevent that data is sent and processed for some incompatible purposes and explicitly cites personalised advertisement as an incompatible purpose. This is an over interpretation of article 6(4) as the test of compatibility or incompatibility must take into account the specific context of the data processing activity.

8. The DG do not make any reference to **existing international standards** – failing to establish bridges with existing internationally wide known and used concepts, such as:

    - **ISO/IEC 20889:2018 – Privacy enhancing data de-identification terminology and classification of techniques:** this standard provides a description of privacy-enhancing data de-identification techniques designed to support the development of de-identification measures in accordance with the privacy principles in ISO/IEC 29100. It specifies terminology, a classification of de-identification techniques according to their characteristics, and their applicability for reducing the risk of re-identification. It is relevant for all types and sizes of public and private organizations that are controllers or processors, implementing data de-identification processes for privacy enhancing purposes.

    - **ISO/IEC 27559:2022 – Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework**: this standard provides a framework for identifying and mitigating re-identification risks and risks associated with the lifecycle of de-identified data. It is applicable to all types

and sizes of public and private organizations that are controllers or processors acting, implementing data de-identification processes for privacy enhancing purposes.

9. It is important that controllers can **test the robustness of their pseudonymization techniques and use cases**. It is therefore necessary that DPAs work on codes of conduct (article 40.2 of the GDPR expressly mentions pseudonymization as a topic to be specified by a code of conduct), certifications or regulatory sandboxes. This would provide legal certainty to controllers to innovate. It is not possible that the DPAs only assess the robustness of the pseudonymization technique at the enforcement stage.

10. On the **rights of data subjects on pseudonymized data**, it is key to ensure that the person claiming rights is effectively the person to whom the pseudonymized data relates to avoid providing data to an unauthorized person. The DG should provide concrete examples as to how controllers can perform such verification effectively, particularly when they receive pseudonymized data from another controller and do not have any possibility to access additional information. Some DPAs have been refusing reliance on a national ID. Controllers need alternative methods to control the risk of impersonation.

11. The DG fail to explain **the role of pseudonymization in enforcement cases** and how it should be considered as a mitigating factor contributing to reducing the risk for data subjects. This could be covered under article 83 (g) "*the categories of personal data affected by the infringement*" as pseudonymized data cannot be treated as "data in clear" and should be recognized as a separate category of data (see point 5 above). It could also fall under article 83 (k) "any *other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement*". This recognition as a mitigation factor in enforcement cases is paramount as it serves to acknowledge the efforts and the investment of the controller to reduce the risk and better protect the data in practice. It would also incentivize market players to implement such techniques. By creating this virtuous circle, this brings more effective protection on the ground, which is the objective of the GDPR. As the DG show, these techniques are complex – it is therefore paramount to promote investment so that they become accessible to more market players, including SMEs and start-ups that are key innovation drivers in data and AI.