DE BRAUW
BLACKSTONE
WESTBROEK

Pseudonymisation: big data opportunities in the GDPR

The EU General Data Protection Regulation ("GDPR") introduces pseudonymisation as a tool to help companies meet high data protection compliance standards. Amidst the many provisions and recitals of the GDPR, the benefits of pseudonymisation may not be immediately apparent.

Under certain conditions, applying GDPR-proof pseudonymisation allows companies to reuse personal data for big data analytics without the consent of individuals, even if its purpose is different from that for which personal data was initially collected. The key condition is that big data analytics is not used to support measures or decision regarding a particular individual. Pseudonymisation can also reduce the companies' compliance burden regarding individuals' rights of access to personal data and notification of a personal data breach. Pseudonymisation may also mitigate the risk of administrative fines, which can be as high as EUR 20 million or 4% of an undertaking's worldwide annual turnover.

When is pseudonymisation GDPR-proof?

Pseudonymisation is a technique to replace personal data in a data set with another piece of information, or attribute. Personal data may include an individual's name, date of birth or national identification number, or any other information which links the data set directly to an individual. An attribute replacing personal data can be a random number, or can be derived from the identifying information, for example, a hash value (a code generated based on identifying information) or a cipher (encrypted identifying information).

Application of pseudonymisation techniques does not in itself guarantee de-identification of the personal data. Such techniques vary widely and are not always robust. The GDPR contains a legal definition of pseudonymisation in Article 4(5), which imposes a set of criteria that methods of pseudonymisation must meet in order to create legal effects. Under the GDPR, pseudonymisation must ensure that, after identifying information has been substituted, information in the data set can no longer be attributed to a specific individual without additional information, which must be kept separately. "Separately" does not necessarily mean that additional information must be transferred to a third party outside the company. Separation of data across business functions or data 'siloing' within a single company may also ensure that such additional information is isolated from the rest of the organisation, so that it cannot be combined with pseudonymised information (recital 29).

Re-use of personal data without additional consent for big data analytics

Big data analytics

Big data analytics refers to various techniques of processing large and diverse blocks of information – "big data" – to find patterns, trends or correlations between different pieces of such information. Often, the purpose of big data analytics is different from the purpose for which personal data has been originally collected. It is this departure from the original purpose that poses legal challenges for big data analytics from the data protection perspective.

Flexible regime for big data analytics for scientific and historic research and statistical purposes

One of the central concepts of EU data protection under the GDPR is the principle of purpose limitation. This requires that companies can collect personal data only for specified, explicit and legitimate purposes and may not further process such data in a manner that is incompatible with the original purposes (Article 5(1)(b)). Compatibility of the original purpose with the purposes of further processing is thus the core benchmark to determine whether a company may re-use personal data for another purpose without an additional legal ground, such as an individual's consent or company's legitimate interest. Compatibility is presumed when personal data is re-used for scientific research or statistical purposes, provided that the company applies GDPR-compliant pseudonymisation techniques (or other appropriate safeguards) to protect the rights of individuals (a.o. Articles 5(1)(b) and 89(1)). As such, the GDPR explicitly allows for such re-use when the strict criteria for compliant pseudonymisation are met.

The thin line between scientific research and statistical purposes and profiling

While the GDPR provides for a flexible regime for big data analytics for research or statistical purposes, a much stricter regime applies to big data analytics that qualifies as profiling. The line between scientific and statistical research on one hand, and profiling on the other, is quite thin. Companies should therefore carefully assess whether big data analytics in a particular case constitutes profiling. This evaluation is one of the major risks in relying on the compatibility assumptions for big data analytics.

Both "scientific research" and "statistical purposes" are defined broadly to include any processing of personal data for statistical surveys, as well as applied and fundamental research whether funded publicly or privately (recitals 159 and 162). Examples of big data analytics include analysis of personal data to identify general trends and correlations, a classification of individuals based on their characteristics such as age or gender for statistical purposes, or aggregate overview of customers (market research) without making predictions or conclusions about such customers.

In contrast, whenever the purpose of big data analytics is to inform measures or decisions regarding individuals, such as behavioural advertisements, data brokering or location-based advertising, the rules on profiling and automated decision-making should apply. The distinction between (a) processing personal data for the purposes of scientific research and statistics, and (b) processing personal data for the purposes of profiling, is further explained by European data protection authorities (DPAs) in the 2013 Guidance on purpose limitation and the GDPR-specific 2017 Guidance on profiling). In the latter case, companies may no longer rely on the presumption of compatibility, and must obtain prior explicit consent of

individuals concerned, or rely on another legal ground under the GDPR. Pseudonymisation is one of the GDPR-endorsed techniques to ensure that the results of big data analytics pursuing such purposes are not used to support measures or decisions regarding a particular individual..

Pseudonymisation as a general compliance tool

Outside the context of big data analytics, the GDPR more generally showcases pseudonymisation as an effective tool of data protection by design (Article 25), especially when it comes to implementing the principles of data minimisation and purpose limitation.

In addition, European DPAs flag several other situations, not explicitly mentioned in the GDPR, where a company may enjoy a milder compliance regime and lower sanctions if it applies pseudonymisation. In particular, the fact that a GDPR infringement only concerns pseudonymised data could be interpreted as a mitigating factor for the purposes of determining whether an administrative fine for such violation should be imposed, and at what amount (Guidelines on administrative fines, at 14-15).

Proper implementation of pseudonymisation, combined with other techniques of making personal data unintelligible (such as encryption or salted hashing), may also reduce the likelihood that a personal data breach will result in a risk to individuals and mitigate potential damage. As a result, if a personal data breach affected only unintelligible data and the company suffering the personal data breach has available backups of the relevant data, the company may not have to notify the personal data breach to the data protection authorities (Guidelines on data breach notification, at 18-19). Keeping data in pseudonymised form may also be a factor in determining the impossibility or disproportionate effort of informing individuals about the processing of their personal data, if the data was not obtained directly from those individuals (Guidelines on transparency, at 31).

Beware of the risks

To reap the regulatory flexibilities that accompany pseudonymisation, companies should implement a comprehensive compliance framework throughout the company and carefully assess the case-specific risks of big data analytics on the privacy of individuals. Only GDPR-compliant pseudonymisation can produce legal effects. When relying on the presumption of compatibility of further processing of personal data for big data analytics, companies should be especially mindful of the thin line between big data analytics for the purposes of scientific research and statistics and profiling.