THE DEFINITION OF PERSONAL DATA

Working Paper 02/2017

IAB Europe
GDPR Implementation Working Group

Version 1.1
18 July 2018

About IAB Europe

IAB Europe is the voice of digital business and the leading European-level industry association for the interactive advertising ecosystem. Its mission is to promote the development of this innovative sector by shaping the regulatory environment, investing in research and education, and developing and facilitating the uptake of business standards.

About the GDPR Implementation Group

IAB Europe's GDPR Implementation Working Group brings together leading experts from across the digital advertising industry to discuss the European Union's new data protection law, share best practices, and agree on common interpretations and industry positioning on the most important issues for the digital advertising sector. The GDPR Implementation Working Group is a member-driven forum for discussion and thought leadership, its important contribution to the digital advertising industry's GDPR compliance efforts is only possible thanks to the work and leadership of its many participating members.

Acknowledgements

This working paper on the definition of personal data has been prepared by the members of the IAB Europe GDPR Implementation Group under the leadership of Ghita Harris-Newton, Chief Privacy Officer & Deputy General Counsel at *Quantcast*.

Contacts

Townsend Feehan (feehan@iabeurope.eu)
CEO, IAB Europe

Matthias Matthiesen (matthiesen@iabeurope.eu)
Director, Privacy & Public Policy, IAB Europe

Chris Hartsuiker (hartsuiker@iabeurope.eu)
Manager, Privacy & Public Policy, IAB Europe

Contents

Executive Summary

- The definition of personal data under the GDPR is very broad and intentionally all-encompassing. Pseudonymous data is defined as a sub-category of personal data, and still triggers full application of the GDPR.
- Cookies and other device and online identifiers (IP addresses, IDFA, AAID, etc.) are explicitly called out as examples of personal data under the GDPR.
- Due to this broad definition, it is highly likely that any data being processed in the online advertising ecosystem falls within the definition of personal data. As the definition is extremely broad, it is prudent to err on the side of caution and assume data is personal.
- Where data might appear to fall outside of the scope of personal data, a careful analysis should be carried out to substantiate this on a case-by-case basis. Depending on the circumstances, the same piece of data (i.e. an IP address) may be personal, pseudonymous, or anonymous data. This depends on the circumstances in which an IP address is obtained, for which purposes it is used, and who receives the IP address.

Overview

On 27 April 2016, the European Union adopted the General Data Protection Regulation ("GDPR").[1] The GDPR became directly applicable law in the European Union ("EU") and European Economic Area ("EEA") on 25 May 2018, replacing previous national data protection laws.

The GDPR does not only apply to companies based in the EU but also to companies all over the globe offering goods and services to people based in the territory of the Union, or monitor the behaviour of individuals located within it. Data protection law regulates the processing of personal data, defined broadly as any information that relates to an identified or identifiable natural person, which may include, amongst others, online and device identifiers that can be used to single out a natural person, for example for digital advertising purposes.

The GDPR grants data protection authorities the power to levy significant administrative fines against businesses found in breach of the law. Depending on the severity of the infringement, fines can reach up to € 20,000,000 or 4 per cent of a company's annual global turnover – whichever is higher.

This document has been prepared by members of the IAB Europe GDPR Implementation Group to provide guidance to companies across the globe on understanding what the definition of personal data means for them.

Personal Data Under the GDPR

The definition of "personal data" is fundamental to data protection law because the GDPR only applies to personal data. Data that is not personal data falls outside the scope of the GDPR. While the digital advertising industry, and other businesses that use similar technologies, have often interpreted unique online identifiers such as cookie IDs and mobile device advertising IDs to be outside the scope of data protection law where they were not coupled with personally identifying details (such as name or email address), these online identifiers are likely to fall within the scope of personal data under the GDPR in many circumstances. **Therefore, it is critical that companies involved in digital advertising understand how the definition of personal data in the GDPR applies to them.**

This paper examines the scope of personal data under the GDPR, including the concepts of anonymous data (which is *not* personal data and *not* regulated under the GDPR) and pseudonymous data (which *is* personal data and *is* regulated under the GDPR).

Personal Data

The definition of personal data in the GDPR expands upon the text of the definition contained in the Data Protection Directive (Directive 95/46/EC, "DPD") by explicitly referencing additional examples

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available at http://eur-lex.europa.eu/eli/reg/2016/679/oj/.

of identifiers, such as online identifiers, and factors that can be used to identify a person. Article 4(1) of the GDPR states:

> *"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular* ***by reference to an identifier such as a name, an identification number, location data, an online identifier*** *or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."*

Recitals 26 and 30 provide additional insight into the definition of personal data. Recital 26 introduces the concept of making a person identifiable by "singling out" that person, directly or indirectly. It indicates that one must consider all means reasonably likely to be used to identify the person, taking into account all objective factors when making such a determination. Recital 26 states:

> *"...To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used,* ***such as singling out****, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."*

Recital 30 indicates that identification may occur by associating online identifiers, such as cookie IDs and IP addresses, with other information to create profiles. Recital 30 states:

> *"Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as* ***internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags****. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers,* ***may be used to create profiles of the natural persons and identify them****."*

These provisions expand the scope of personal data significantly from common interpretations under the DPD. While the digital advertising industry has often interpreted unique online identifiers such as cookie IDs to be outside the scope of data protection law where they were not coupled with personally identifying details (such as name or email address), these online identifiers are likely to fall within the scope of personal data under the GDPR in many circumstances. As described in Recital 26, one will need to look at all means reasonably likely to be used in the circumstances to identify the underlying natural person to determine if the data is personal data; however, as a general matter under the GDPR, data such as online identifiers should be treated as personal data

unless a valid argument can be made that the data subject is not (directly or indirectly) identifiable and cannot be singled out.

The determination whether a piece of data is personal data will be context specific. For example, an IP address that corresponds to a public "hot spot" such as a coffee shop, and is used by hundreds of customers every day, by itself is unlikely to comprise personal data. However, if the company links that common IP address with other information that would allow it to single out one individual, then the IP address is likely to be personal data.

Similarly, a truncated IP address would not be personal data where the holder of that truncated IP address has no reasonable means to identify the individual. However, if the holder of the truncated IP address can, using reasonable means at its disposal, collect additional information that would allow it to single out the individual, then even that truncated IP address is likely to be personal data.

Companies should remember that personal data encompasses more data than what is typically considered personally identifiable information (or PII) in some jurisdictions outside of the EU. In instances where it is unclear whether data is personal data, treating it as personal data would be the prudent course of action, particularly given the potential for high fines under the GDPR.

Anonymous Data

Like the DPD before it, the GDPR does not apply to anonymous data. Recital 26 explains that anonymous information does not relate to an identified or identifiable person. Recital 26 states:

> *"The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."*

The Article 29 Working Party, in its prior Opinion 05/2014 on Anonymisation Techniques, referred to anonymisation as "a technique applied to personal data in order to achieve **irreversible de-identification**." That opinion sets out various anonymisation techniques and highlights that "case studies and research publications have shown how difficult it is to create a truly anonymous dataset whilst retaining as much of the underlying information as required for the task."

Where a company holds data that is truly anonymous, the GDPR does not apply to that data. For example, a piece of general location information that does not identify an individual is anonymous data that is not subject to GDPR. If a company holds the name of a large city (e.g., Brussels), does not associate any other identifying information, and is not reasonably likely to obtain or use

additional information that could associate the location with an individual, then the data is anonymous.

Aggregated data that does not relate to one user, but relates to an entire group of users, is anonymous data as long as the individuals whose data is in the pool cannot be identified.

The analysis of whether a particular piece of information, or group of information, is anonymous is context specific and not always clear. Where a company is unsure whether the data it holds is personal data or anonymous data, treating the data as personal data is a prudent course of action.

Pseudonymous Data

The GDPR introduces the concept of **pseudonymous data as a subset of personal data** that cannot be attributed to a specific data subject without additional information. Article 4(5) states:

> *"'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."*

Pseudonymisation was not addressed in the DPD and many in the advertising industry have considered pseudonymous data to be outside the scope of personal data in the DPD and thus outside the scope of the DPD. **Under the GDPR, pseudonymisation does not render a data set anonymous (and therefore out of the GDPR's scope). Recital 26 clarifies that pseudonymous data is in scope of the GDPR:**

> *"…Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person…"*

However, as described in Article 11, pseudonymisation may exclude the data from certain GDPR obligations that specifically require identification, such as subject access and the right to rectification, erasure and data portability (Articles 15-20). **Online identifiers, such as cookie IDs that are associated with online browsing history, are often going to be personal data under the GDPR, although a context specific analysis always applies.**

Pseudonymisation is recognized as a safeguard that reduces the risks to data subjects and helps controllers and processors meet their data protection obligations. Recital 28 recognizes this benefit of pseudonymisation, stating:

> *"The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations. The explicit introduction of 'pseudonymisation' in this Regulation is not intended to preclude any other measures of data protection."*

The GDPR explicitly recognizes pseudonymisation as a safeguard that can contribute to permissible processing for a secondary use. Article 6(4) says:

> *"Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: … (e)* ***the existence of appropriate safeguards, which may include encryption or pseudonymisation.****"*

Article 89(1) recognizes pseudonymisation as a safeguard for processing for archiving in the public interest, scientific or historical research purposes or statistical purposes. Article 89(1) states:

> *"Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner."*

Importantly, the GDPR recognizes that pseudonymisation of personal data is possible by a controller where that controller holds additional information that could be used to attribute that data to an individual data subject, as long as the controller has taken technical and organisational measures to keep that information separate. Recital 29 says:

> *"In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and* ***that additional information for attributing the personal data to a specific data subject is kept separately****. The controller*

processing the personal data should indicate the authorised persons within the same controller."

Article 25(1) of the GDPR recognizes that technical and organizational measures such as pseudonymisation should be designed "both at the time of the determination of the means of processing and at the time of the processing itself." This design supports two equally relevant concepts of pseudonymisation.

First is the collection of data in a way that allows a controller to hold data that cannot be attributed to a specific data subject without the use of additional information. In other words, the data is pseudonymous at its collection, use and storage. For example, some ad tech companies never collect information to directly identify the end user; rather, they only collect a randomised cookie ID and associated URLs visited, which allow a browser to be recognised but the end user cannot be directly identified. This data is pseudonymous in the hands of that ad tech company because that company does not have nor has reasonable access to additional information that would allow it to directly identify the data subject.

The second concept of pseudonymisation is as a process that companies can apply to personal data, for example using encryption, hashing or tokenization techniques, to ensure the data is not linked to an identified or identifiable natural person. For example, a company may collect full name, mailing address, account number and URLs visited. If it holds that information in its subscriber database, it could create a separate database of data that has been pseudonymised by removing the name and mailing address information and hashing the account number. If the company puts appropriate technical and organisational measures in place to keep the databases separate and prevent re-attribution of the pseudonymised data, then the second database is a pseudonymous database that could, for example, be used for research purposes in a privacy-friendly way.

An IP address is an example of data that could be anonymous data, pseudonymous personal data, or non-pseudonymous personal data, depending on the specific circumstances. Referenced earlier in this paper is an example of a common IP address at a "hot spot" that is anonymous data when held without any other information because it does not identify or make an individual identifiable. Also, referenced earlier in this paper is an example of a truncated IP address that alone is not personal data, but becomes personal data if the holder of that truncated IP address can reasonably associate the truncated IP address with additional information to allow the holder to identify the individual. If the only additional data held is the missing octet, then the data would be pseudonymous personal data; however, if the additional data held is the missing octet plus information such as a name and address associated with the IP address, then that combined data would be non-pseudonymous personal data. **Companies are urged to engage in a context specific analysis of the data they hold to determine whether it is personal data.**

Special Categories of Personal Data

The GDPR, like the DPD, recognizes certain special categories of personal data that cannot be processed unless stringent requirements (contained in Article 9(2)) are met, such as explicit consent

by the data subject. Article 9(1) outlines the special categories of sensitive data as: “personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership;” genetic data or biometric data; and data concerning health or a natural person's sex life or sexual orientation.

Data relating to criminal convictions and offences are also subject to restrictions. Article 10, like the DPD, restricts its processing to the control of official authorities or instances where national law may provide derogations.

Companies wishing to process special categories of personal data or data relating to criminal offences should be sure to comply with the more stringent processing requirements.

Conclusion

The GDPR expands on the definition of personal data contained in the DPD and thus expands the scope of EU data protection law. **Under the GDPR, online identifiers and information associated with those online identifiers will often constitute personal data. Where the information collected is pseudonymous, it will be considered personal data, and the pseudonymisation will act as a safeguard, bringing benefit to the data subject and excluding the data from certain GDPR obligations.** The types of pseudonymous data commonly used by companies in the online advertising industry, such as device advertising identifiers and cookie ids, will (depending on the specific situation of the company processing the data) generally fall into the category of personal data and thus be subject to the requirements of the GDPR. Companies in the digital advertising space should carefully examine their data processing activities to ensure that if they process personal data, that processing complies with the GDPR.

About the IAB Europe GDPR Implementation Working Group

IAB Europe's GDPR Implementation Working Group brings together leading experts from across the digital advertising industry to discuss the European Union's new data protection law, share best practices, and agree on common interpretations and industry positioning on the most important issues for the digital advertising sector.

The GDPR Implementation Working Group is a member-driven forum for discussion and thought leadership, its important contribution to the digital advertising industry's GDPR compliance efforts is only possible thanks to the work and leadership of its many participating members.

For more information please contact:

Townsend Feehan (feehan@iabeurope.eu)
CEO
IAB Europe

Matthias Matthiesen (matthiesen@iabeurope.eu)
Director, Privacy & Public Policy
IAB Europe

Chris Hartsuiker (hartsuiker@iabeurope.eu)
Manager, Privacy & Public Policy
IAB Europe