RIGA GRADUATE SCHOOL OF LAW

Analysis of the Right To Be Forgotten under the GDPR in the age of Surveillance Capitalism

BACHELOR THESIS

AUTHOR: Raimonds Dolfs Jēkabsons
LL.B 2022/2023 year student
student number B020062

SUPERVISOR: Ēriks Kristiāns Selga

LL.M.

DECLARATION OF HONOUR:

I declare that this thesis is my own work, and that all references to, or quotations from, the work of others are fully and correctly cited.

(Signed) ……..

RIGA, 2023

Table of Contents

ABSTRACT

The definition of personal data is evolving in the modern age. With the emergence of new technology, new commercial practices and the increase in the value of data, companies are looking for ways to extract as much value as possible from the data of their users and gain an edge on their competition. Among these practices there are various legal concerns such as the right to be forgotten under the GDPR, how well it can be ensured and whether it can be ensured. Because of competition, companies may engage in practices that may not be legal in terms of data collection in order to benefit and increase their market dominance.

Overall, the right to be forgotten is not adequately ensured under the GDPR in terms of copied information due to a lack of clear enforcement terms and definitions. Profiling is well regulated and defined, however, in real practice most companies do not admit that their work revolves around profiling or benefitting from an ecosystem built on profiling, which means that in reality profiling is still a big issue. Harmful data extraction is regulated, as well as there is a case brought before Germany's competition authority regarding abuse of market position by a dominant social network. This case can bring attention to harmful data extraction and increase the quality of its regulation, while it is currently not defined under the GDPR. Overall, the GDPR suffers from a lack of definitions and enforcement terms, which could be fixed by computer scientists and legislators collaborating more closely.

SUMMARY

While the contours of personal data are expanding, personal data is defined in Article 4 (1) of the GDPR. The definition can be assessed by paying attention to 4 key components, namely "natural person," "any information," "relating to" and "identified or identifiable." While the definition of personal data may be defined adequately, the scope of its definition widening every day is a cause for concern due to new technology that is able to combine various types of information from various sources to make advanced calculations that may end up in producing information that constitutes personal data.

Big data is a process that consists of accumulating data, processing data and then analyzing the data. Big data is built on ubiquitous computing, which is the accumulation of data through various sensors and ubiquitous devices, such as smartphones, smart watches, computers among other things. Big data relies on the correlation of data among individuals to discover new information and predict behaviors, which allows companies to improve their products and services. Companies are investing more and more resources into sensors as data increases in value while sensors are relatively cheap to finance. Big data transcends privacy norms and possibly the right to be forgotten because companies copy the information they obtain and sell it further to other data processors, while making the data difficult to trace. It can additionally be difficult to forget certain persons if they have already been identified, at which point it is difficult to untangle the link between the personal data an the identified person.

Surveillance capitalism is a similar phenomenon to big data. The surveillance capitalism theory shares the same attributes of big data, such as accumulating data to predict user behavior, however, surveillance capitalism is more predatory towards the user. While collection of data under big data might be conducted only to the extent which is required to provide a certain service, surveillance capitalism defines the idea that data is being collected beyond the extent which is necessary to provide a certain service. Companies can afford to impose such conditions on its users due to dominating market shares where users do not have a choice but to agree to those extreme conditions or otherwise be barred from using that service.

The protection of the right to be forgotten, in the context of copied information and profiling, is not good enough. One of the flaws of the GDPR is the lack of practical advice on enforcement and clear definitions, which ultimately is the reason that the right to be forgotten is not effectively enforceable. In the case of profiling, the issues with enforcement do not necessarily lie in the GDPR, rather in companies and their lack of awareness of their profiling practices.

Harmful data extraction is regulated under the GDPR, as evidenced in the Bundeskartellamt's case against Facebook in Germany. It deals with harmful data extraction well despite not having the proper definition enclosed in the regulation. The GDPR must add the definition of harmful data extraction in order to regulate this practice and bring more awareness to it.

Overall, the GDPR's approach to data protection is working well. The GDPR's biggest issue, in terms of the right to be forgotten and harmful data extraction, is the lack of clear definitions and practical advice on enforcement. By widening the scope of application by including more definitions, the GDPR will bring more awareness to companies and individuals alike in terms of their rights and obligations.

1. INTRODUCTION

1.1 REVIEW OF THE PROBLEM

1.1.1 MASS COLLECTION OF DATA, THE RIGHT TO BE FORGOTTEN AND SURVEILLLANCE CAPITALISM

Since their inception, information technologies and the internet have come a long way and grown in importance over time to reach a status where these technologies have now become an integral part in our day to day lives. The development of the internet and technology has led to a large amount of user data being generated, which is then further collected and processed by companies and organisations. The type of data that is being collected depends on the organization that is collecting it, ranging anywhere from personal information about the person such as name, email, personal address and behavioral data to more sensitive types of data, such as financial information and even medical records. While useful to curate and personalize the content that the user is exposed to, it is the cause of many concerns regarding the protection of user data and also their privacy. The mass collection of data results in information being spread widely across the web, raising questions of whether rights such as the right to be forgotten can be ensured in the current technological world due to the information being difficult to track and erase.

The collection of data regarding people has evolved to include a tendency that some scholars have named "informational capitalism" por "surveillance capitalism", where the personal data that is collected by an organization is turned into a commodity.[1] Due to the rapid growth of the IT industry the amount of personal data and, generally speaking, any data about people being generated and collected is increasing exponentially, highlighting the importance of data governance legislation more and more each day. With the circulation of data across the world wide web it is often difficult to trace the data and find out where it ends up. While data governance regulations such as the General Data Protection Regulation (hereinafter referred to as "GDPR") offer the user various rights that protect their data, there is consensus among scholars that current data governance laws are providing inadequate protection against harmful data extraction.[2]

1.1.2 IMPORTANCE OF THE PROBLEMS RELATING TO ERASURE AND EXTRACTION OF DATA

Due to the amount of data circulating through the cyberspace in combination with unclear personal data extraction terms it is often quite difficult if not impossible for the user to know which data they have willingly or unwillingly given consent to be extracted and processed, as well as know where the data ends up and whether it is really erased when they decide to stop giving consent to companies that extract their data. This raises questions of the efficiency and scope of data governance laws and whether they even afford adequate protection of personal data and whether they can ensure rights such as the right to erasure or right to be forgotten.

It is also important to be aware of the phenomena called "surveillance capitalism" that focuses on turning personal data of users into a commodity. This practice that businesses nowadays engage in oftentimes leads to harmful extraction of data, of which the users are

[1] Salome Viljoen, "A Relational Theory of Data Governance," *Yale Law Journal Forthcoming* (November 11, 2020): p. 577, accessed March 19, 2023. Available: http://dx.doi.org/10.2139/ssrn.3727562.
[2] *Ibid.*

mostly unaware of as it lacks transparency and clarity. Due to this potentially harmful practice it is important to investigate whether data is being collected in a way that is not detrimental to the data subject under any circumstance and whether the owner of the data can exercise all of their rights that are afforded to them under current data governance laws.

The definition of personal data is expanding through various processes of converting real life information about individuals into data and turning it into a commodity. Google and Amazon are such companies that collect real life information through their products (Google Street View, various home assistants like Amazon Echo Dot) and convert the data into a commodity or use it to predict and influence future behaviors of their users. This type of competition between companies nowadays has become fierce, therefore companies have an incentive to extract as much data as possible from their customers. This has led to companies collecting and processing personal data, as well as data that is not necessarily personal data but it has the potential to be considered personal data if it can be used, using new technologies, to identify someone indirectly. This poses various potential privacy violations as individuals can be identified in more and more ways, meaning that the right to be forgotten under the GDPR also becomes difficult to be enforced if not impossible in some cases. Therefore, it is important to identify the potential issues as regards to the effective enforcement of the right to be forgotten in an era where the contours of personal data are expanding every day and where identification of individuals is quite difficult to reverse in some cases. It is also important to be aware of harmful data extraction techniques which result from the fierce competition among companies for personal data and whether it is regulated under the GDPR.

1.2 RESEARCH AIM AND METHODS

1.2.1 RESEARCH AIM

While there is a consensus that current legal regimes regulating data governance are failing to protect people from the harmful extraction of data in an age of informational or surveillance capitalism, a consensus is yet to be reached regarding what exactly constitutes personal data and how to identify it, or rather – the definition of personal data and what is viewed as personal data is unclear, meaning that clarity regarding this topic is becoming increasingly more important in order to develop proper legislation that addresses all of the current issues related to the protection of personal data. The author will investigate the definition of personal data and the possible issues surrounding its identification in the GDPR as well as scholarly writings to obtain clarity as regards to the scope of personal data and its identification as well as to what extent the GDPR ensures one of its core principles – the right to be forgotten. The author will also investigate harmful data extraction tendencies and whether they are regulated under the GDPR.

The research question posed by the author is the following: To what extent does the GDPR ensure the right to be forgotten in regard to the evolving contours of personal data and regulate harmful extraction of data in terms of surveillance capitalism? As the question states, the author will focus on the scope of what ultimately can be considered personal data and to what extent the right to be forgotten under the GDPR can be ensured taking into account the expanding definition of personal data and to what degree the GDPR regulates harmful data extraction.

1.2.2 METHODOLOGY

The author will utilize the analytical approach research method and doctrinal legal analysis in order to analyze the expanding definition of personal data, to what extent the right to be forgotten can be ensured in respect to the evolving definition of personal data, as well as to what degree protection against harmful data extraction tendencies is ensured under the GDPR. The author will focus on interpreting provisions under the GDPR, as well as identifying gaps in the provisions in pursuit to provide clarity to what ultimately constitutes personal data and what is its definition. The author will also analyze to what degree one of the main principles of the GDPR – the right to be forgotten – is ensured under the regulation with respect to the evolving contours of the definition of personal data, as well as to what extent harmful data extraction practices in the age of a phenomena called "surveillance capitalism" are regulated under the GDPR. As there will be an exclusive focus on the GDPR throughout the paper, the author will primarily focus on the jurisdiction of the European Union.

1.2.3 LITERATURE REVIEW

Upon conducting research on scholars' opinions and exploring scholarly literature, there seems to be a broad consensus that current data governance laws are not up to date to the evolving contours of personal data in a world of surveillance capitalism.[3] While there is no concrete consensus as regards to an exhaustive list of what constitutes personal data, there does seem to be a consensus of tendencies and methods of companies regarding the extraction and production of data, which is that the data that companies extract are processed and analyzed at a population level, or rather – how the data they collect from one individual relates to other individuals.[4] Some scholars even take the view that activities relating to surveillance capitalism and data collection reach beyond the borders of technology and extend into the real world via a practice called "datafication", which is a process of converting real world information about individuals into personal data and processing it for monetary gain.[5] Regarding interpretations of the GDPR in terms of the expanding definition of personal data, "any information" under Article 4 of the GDPR is said to be divided into "objective" and "subjective" types of information, with objective information like facts about an individual's appearance going under the "objective" information type and things like employment evaluations going under the "subjective" information type.[6] It is also mentioned that there is no limitation when it comes to the format of the data, meaning that formats of video, audio, photographic *et cetera* can all contain personal data.[7]

While there is consensus on things such as data governance laws failing to protect individuals from the violation of their rights, the fact that the definition of personal data is expanding and that collection of personal data reaches beyond technological borders and into real life, there are gaps in the scholarly literature. While current data governance laws are being criticized, it is not clear in the literature what and how the legislation can be improved. For example, the GDPR is a regulation of proactive nature and its rules are designed to encompass

[3] *Ibid.*
[4] O. J. Gstrein & A. Beaulieu, "How to protect privacy in a datafied society? A presentation of multiple legal and conceptual approaches," *Philosophy & technology*, *35 (1), 3* (2022). Available: https://doi.org/10.1007/s13347-022-00497-4. Accessed March 27, 2023.
[5] *Ibid.*
[6] Richie Koch, "What is considered personal data under the EU GDPR?" Available: https://gdpr.eu/eu-gdpr-personal-data/?cn-reloaded=1. Accessed March 27, 2023.
[7] *Ibid.*

various technological tendencies and practices as it does not limit itself to any specific technology, which is a great approach in a field of law so volatile from a technological standpoint as data protection law, yet no tangible improvements have been proposed. It is also unclear from scholarly literature how the GDPR ensures the right to be forgotten in regards to the expanding definition of personal data and whether it is reasonably possible to ensure it, as well as possible future improvements to make sure that the right to be forgotten can be effectively ensured. There was also no scholarly literature particularly on the topic of harmful data extraction, which is becoming more prevalent in the age of surveillance and whether this tendency is being regulated under the GDPR.

2. PERSONAL DATA – AN EXPANDING DEFINITION

2.1 DEFINITION OF PERSONAL DATA

2.1.1 DETAILED LOOK AT ARTICLE 4 (1) UNDER THE GDPR

Personal data is evolving both in scope and definition. The term "personal data" is defined in the GDPR under Article 4 (1). It reads as follows:

> '[P]ersonal 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.[8]

While the definition of personal data under Article 4 (1) of the GDPR seems to be rather superficial there are a few key words in the first line of the provision that, upon further explanation, could yield more clarity as to what is the scope of personal data under the GDPR.

"Natural person" is an unambiguous term. By using the term "natural person" the GDPR excludes data about companies, which are known as "legal persons," from the definition of personal data, however, the person must be alive, as data about the deceased are not considered personal data under the GDPR.[9]

"Any information" is one of the key words that conveys a broad definition and is quite inclusive.[10] It encompasses both "objective" data (for example, the height or name of an individual or other factual information) and "subjective" data (things such as employee evaluations, surveys etc.) and it does not adhere to any specific format, be it video, audio, numerical, graphical and other means of data.[11] To illustrate, an employee welfare evaluation survey in a workplace could be considered personal data, but only if it reveals information such as the mental state of the employee in relation to their workplace and how the company treats them, as well as behavior of the superior authorities. As long as the information relates to any individual, even if it is incorrect factually, is still considered personal data, however it stops

[8] General Data Protection Regulation, *OJ L* 119, 4.5.2016, p. 1-88. Available on: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed March 29, 2023.
[9] Koch, *Supra* note 6.
[10] *Ibid.*
[11] *Ibid.*

being personal data at the point where it cannot be attributed to any person, such as referring to a non-existent address when asked where one lives.[12]

"Relating to" is another key component in the personal data definition.[13] The essence of this component is that any data that can be attributed to an individual is data relating to that same individual, meaning that it is personal data. If the processing of the individual's data will have an impact on them or if the processing is done in order to learn something about the individual, it is personal data, such as, for example, information relating to the amount of water usage per month in order to define the person's water bill.[14] Information that could have an unintentional impact on an individual when processed is also considered personal data. For example, Uber tracks all of its drivers in order to find the closest driver to the potential customer, however, this could also be used to track whether the drivers are abiding by rules of the road and how productive they are.[15]

"Identified or identifiable" deals with direct and indirect identification of an individual according to available data.[16] Direct identification is a straightforward concept, it is identification where data like names and locations are available and help to identify someone. A person may still be directly identified even without their name being known, however other information would be required such as the location and physical attributes of the person.[17] Indirect identification is more complex, involving more factors. A person is generally considered as identified indirectly when the data processor has insufficient information to identify that person directly, so the processor resorts to using other information at their disposal or information they can reasonably access from third parties (the police identifying an individual through their license plate is one example of indirect identification).[18]

In sum, the definition of personal data under the GDPR is defined by analyzing and breaking down 4 key terms, namely – "natural person," "any information," "relating to" and "identified or identifiable." "Natural person" means that the data has to be linked to a regular person and not a legal person such as a company, "any information" means any type of content of informative nature that leads to an identification of an individual, "relating to" means that the data has to relate to an individual in order to be considered personal data and "identified or identifiable" means that if the data leads to direct or indirect identification of an individual it is considered personal data.

2.1.2 CONTEMPORARY ISSUES REGARDING THE EXPANDING DEFINITION OF PERSONAL DATA

While the definition of personal data under the GDPR and scholarly interpretations of the GDPR provides more clarity as to what constitutes personal data, issues regarding the definition still exist and there are also potential issues that might exist in the future.

> Methods of identification that are not present today could be developed in the future, which means that data stored for long durations must be continuously reviewed to make

[12] *Ibid.*
[13] *Ibid.*
[14] *Ibid.*
[15] *Ibid.*
[16] *Ibid.*
[17] *Ibid.*
[18] *Ibid.*

> sure it cannot be combined with new technology that would allow for indirect identification.[19]

As described in the quote above, in an age where technology evolves at an exponential rate and pace it is imperative that data is reviewed on a regular basis in order to make sure individuals cannot be identified indirectly using new types of technology. As discussed in the section afore regarding direct and indirect identification, indirect identification is done when the processor has insufficient information to identify someone directly, but has ability to reasonably access information regarding the individual from third parties in order to identify them. Couple this with the fact that any data that can lead to direct or indirect identification of an individual is considered personal data and that technology is constantly evolving it is, theoretically, difficult to predict what is and what will be personal data tomorrow, because there is a chance certain data that is not considered personal data today could be used to identify someone indirectly via new technology tomorrow, therefore labeling it as personal data.

Another issue, despite the clarifications in scholarly interpretations of the GDPR, still persists and it is the issue of differentiating and figuring out what is and what isn't personal data.

> The delineation between personal data and non-personal data is of paramount importance to determine the GDPR's scope of application. This exercise is, however, fraught with difficulty, also when it comes to de-personalized data—that is to say data that once was personal data but has been manipulated with the goal of turning it into anonymous data.[20]

There are issues with the scope of application regarding the GDPR, in large part relating to the ambiguous term of "personal data," as evidenced in the quote above. Another part of the issue seems to be personal data that has been manipulated to not look like personal data and rather look like anonymous data. There are certain advantages for a data processor to do this, namely the fact that if the data is completely anonymous and is in no way related to any individual and prevents the processor from identifying the individual, then it is no longer personal data, meaning that the GDPR is no longer applicable to the processor. Not being governed by the GDPR and processing *de facto* personal data (under the guise of anonymous data) would let the processor purchase the data from a data collector, process it as anonymous data and avoid potential fines and punishments stipulated under the GDPR.

Because the world of technology is advancing at an unparalleled pace, it is quite difficult to predict which data will be personal data in the future that is simultaneously not considered personal data in the present. A large part of the problems that exist nowadays regarding personal data are linked also to the risk of re-identification, which can be achieved in many ways that will be further discussed in this paper. Manipulating personal data to look like anonymous data is another problematic tendency that the data governance regime has to deal with as it can effectively allow the selling and processing of personal data under the guise of anonymous data, avoiding any types of regulation.

2.2 EXPANDING SCOPE OF PERSONAL DATA

[19] *Ibid.*
[20] Michèle Finck, Frank Pallas, "They who must not be identified—distinguishing personal from non-personal data under the GDPR," *International Data Privacy Law*, Volume 10, Issue 1 (2020): pp. 11–36. Available: https://doi.org/10.1093/idpl/ipz026. Accessed March 30, 2023.

2.2.1 PERSONAL VS. NON-PERSONAL DATA: RECITAL 26

While it is not yet crystal clear what will constitute personal data under the GDPR there are 2 known tests to aid in the endeavour of analyzing the scope of personal data. One test can be found in Recital 26 of the GDPR and the other test was developed by the now defunct Article 29 Working Party (hereinafter referred to as A29WP).

Personal and non-personal data are two existing types of data. The definition of personal data has been covered in the previous section. Non-personal data essentially is either any data that does not relate to an identifiable individual or data that once qualified as personal data, meaning that identification of any individual is made impossible. Recital 26 defines a legal test, or rather, methods that help to identify and differentiate personal data from non-personal data.[21]

> 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. 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.[22]

The approach taken by Recital 26 is a risk-based approach to identifying what is personal data and what is not, where if the risk of the data leading to identification of an individual is high, then the data is to be considered personal data, however, if the risk of identification is relatively low or even negligible, then it will not constitute personal data under Recital 26 of the GDPR.[23] To clarify the components of the test: firstly, the data has to relate to a natural person, as stated before in the analysis of the definition of personal data under Article 4 (1) of the GDPR. Secondly, there must be a possibility that the natural person can be identified with the data in question, whether directly or indirectly. Indirect identification is where it becomes more complex as there are various degrees of indirect identification. One of the ways of identifying a person indirectly is by having access to additional information, such as the example of the police identifying a person from searching up the license plate of their car. However, there is an even more ingenious way of identifying someone, such as performing advanced methods of identification with the available data, for example, singling someone out or finding out ways how to interlink data with the individual.[24] If it is reasonably likely that indirect identification is possible, it will constitute personal data under Recital 26. The term "reasonably" might also not be the most clear term, nonetheless the implications of this term are whether the costs, time invested and technological progress to identify someone are "worth it" for the data processor, meaning that the value the processor seeks to obtain from the data exceeds the transaction costs they are incurring.[25]

2.2.2 PERSONAL VS. NON-PERSONAL DATA: ARTICLE 29 WORKING PARTY

The Article 29 Working Party was an advisory body for European data protection authorities that dealt with personal data and privacy related issues that were primarily linked to the Data

[21] *Ibid.*
[22] GDPR, *Supra* note 8.
[23] Finck, *Supra* note 20.
[24] *Ibid.*
[25] *Ibid.*

Protection Directive.[26] The body was renamed to the European Data Protection Board after the GDPR came into force, meaning that A29WP was the main legal body at the European level for data protection related cases before the GDPR. Since A29WP was the predecessor of the current data protection advisory board, its opinion on the material issue of what is considered personal data can be compared to the current approach taken by the GDPR and how the approach has evolved to what it has become now.

Compared to the risk-based approach of Recital 26 of the GDPR, the A29WP is more strict in its nature.[27] In its Opinion 05/2014 on Anonymization Techniques, the A29WP does point out that there is a risk factor regarding anonymization, however it is described as an inherent part of anonymization and describes that there is bound to be a residual risk of re-identification regardless of the anonymization technique.[28] The opinion also mentions that anonymization must be carried out in such a way that no individual can be "reasonably likely" identified after the anonymization.[29] It further goes on to state that the only acceptable way of carrying out anonymization is that it is irreversible.[30] Anonymization being irreversible in principle means that the data that has gone through the process of anonymization has absolutely zero chance of being used to re-identify someone, it is completely anonymous data. With this wording the opinion of the A29WP takes a more strict position in what is to be considered personal data than the more loose risk-based approach taken by Recital 26 of the GDPR.[31] The opinion of the A29WP does not prescribe any risk regarding the possible re-identification of an individual after the anonymization of personal data, therefore advocating for a more strict regime in regards to the anonymization of personal data. The components of the test are largely the same as the risk-based approach of Recital 26, the difference being how those components are treated. The opinion is effectively stating that personal data, that has gone through the process of anonymization, must essentially be in the same state as it being erased, leaving no trail of possible re-identification.[32] Therefore, data that contains any risk, be it reasonable or unreasonable, of it being used to re-identify someone, is considered personal data according to A29WP.

To conclude, the two most popular legal tests for determining whether certain data counts as personal data are the Recital 26 test in the GDPR and the test developed by Article 29 Working Party. The Recital 26 test employs a risk-based approach, where the risk of whether an individual can be identified, directly or indirectly, using the data in question is assessed. If there is a reasonable risk that identification, in combination with various technologies and additional information, is possible, then under the Recital 26 test that particular data is to be considered personal data. However, the approach taken by the A29WP is not as lenient. A29WP does not tolerate any amount of risk in its developed approach, therefore the approach mentions the requirement of anonymization as mandatory, as well as adding that it must be irreversible as to minimize the risk of re-identification. The components of the tests are very similar, but

[26] European Data Protection Board. Available on: https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party_en.
[27] Finck, *Supra* note 20
[28] Article 29 Data Protection Working Party. *Opinion 05/2014 on Anonymisation Techniques*. Available on: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. Accessed April 4, 2023.
[29] *Ibid.*
[30] *Ibid.*
[31] Finck, *Supra* note 20.
[32] Article 29 Working Party, *Supra* note 28.

the risk tolerance is much lower for the A29WP approach, stating that the data must essentially be in the same state as erased after anonymization.

3. BIG DATA AND ITS IMPACT ON DATA PROTECTION

3.1 BIG DATA

3.1.1 DEFINITION

Big data is a phenomenon that describes various processes in which businesses and even the government combine data that is available to them with statistics and other components to obtain new, interlinked information by use of correlation and data analysis.[33] By use of data mining, the process includes extracting previously unknown, implicit and potentially useful information.[34] It relies on correlation instead of causation by applying various algorithms to obtain new information that can be quite unpredictable at times, showing that the potential of big data is endless.[35]

Big data consists of three components – accumulation of data, the processing of data and the storage and analysis of the data.[36] The accumulation of data occurs across a broad range of methods nowadays, including the collection of data not only online but also from mobile phones with location tracking capabilities, smartphone applications that share information with different parties, interactions with smart environments, monitoring systems in the physical environment and the human body, which is used for biometric authentication and genetic testing.[37] With all of the data that has been generated and is available in the digital space, accompanied with the various methods of accumulating data and analyzing it, it is quite obvious that privacy concerns will be a given and that the field requires comprehensive regulation in order to address these issues. Big data can also be used for good, such as monitoring behavior on social media regarding posts about health and tracking behavior of users and their responses to certain health related content, which can be used to make healthy suggestions to users.[38] It can also be used to track biometric data, however, biometric data is a very unique personal identifier which requires its own regulation.[39]

3.1.2 UBIQUITOUS COMPUTING

Ubiquitous computing is the large accumulation of data from ubiquitous (or rather, always present) devices that is obtained from various elements like sensors and location trackers in order to provide an optimized experience for the user in terms of daily tasks like providing

[33] Ira S. Rubinstein, "Big Data: The End of Privacy or a New Beginning?" *International Data Privacy Law, Volume 3, Issue 2* (2013): pp. 74–87. Available on: https://doi.org/10.1093/idpl/ips036. Accessed April 9, 2023.
[34] *Ibid.*
[35] *Ibid.*
[36] *Ibid.*
[37] *Ibid.*
[38] Annika Richterich, "Big Data-Driven Health Surveillance." In *The Big Data Agenda: Data Ethics and Critical Data Studies*, (University of Westminster Press, 2018), pp. 71-90. Available on: https://doi.org/10.2307/j.ctv5vddsw.7 Accessed May 11, 2023.
[39] Fiona Q. Nguyen, "The Standard for Biometric Data Protection*," Journal of Law & Cyber Warfare 7, no. 1* (2018): pp. 61–84. Available on: https://www.jstor.org/stable/26777963. Accessed May 10, 2023.

optimized driving routes or healthcare advice.[40] It collects data from various devices like smartphones and wearable devices, processes and analyzes the data with advanced machine learning techniques discover people's emotions, traits and behaviors.[41] Ubiquitous computing is one part of Big Data and people undoubtedly benefit from it with all of the customized experiences that are automatically adjusted to each individual, however there are deeper issues regarding privacy and data protection that come from ubiquitous computing.[42]

With the broad aggregation of data that is not necessarily collected for a pre-determined purpose, ubiquitous computing often contradicts privacy norms.[43] In reality, when data are collected, whether we are aware of it or not, in most cases it is available to multiple parties, not just the initial collector of the data.[44] Because most of us are using smartphones and wearable devices that constantly collect our data daily, there are a lot of traces of personal data and data that is not necessarily personal data upon collection but becomes personal data after ubiquitous computing and smart machine learning.[45] This data is available to more than one party, especially since we often use multiple apps or applications, however, what we are not always aware of is the fact that this data gets copied countless of times and sent or sold to other data processors or collectors.[46] This alone can create various privacy issues, as this data is often used for profiling and marketing and becomes more and more difficult to truly erase when the data subject wishes to do so because traces of these data have already travelled to many different places across the digital world.[47] Because it enables profiling through smart machine learning techniques, organizations and businesses can build data profiles on many users and many new users without collecting too much information.[48] There are also questions with how these data are collected that concern privacy, where businesses can come up with sneaky ways to collect seemingly harmless information about the person that could in reality violate either their privacy or someone else's through means of profiling.[49] Seemingly meaningless data can be used to build meaningful data profiles on individuals, which makes profiling a prominent issue.[50] The ethics of these practices have been questioned, however, companies seem to have little regard to any ethical issues.[51]

---

[40] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis, "Forgetting Personal Data and Revoking Consent under the GDPR: Challenges and Proposed Solutions," *Journal of Cybersecurity 4, no. 1* (January 1, 2018). Available on: https://doi.org/10.1093/cybsec/tyy001. Accessed April 8, 2023.
[41] *Ibid.*
[42] *Ibid.*
[43] *Ibid.*
[44] *Ibid.*
[45] *Ibid.*
[46] *Ibid.*
[47] *Ibid.*
[48] *Ibid.*
[49] Christoph Busch, "Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law," *The University of Chicago Law Review 86, no. 2* (2019): pp. 309–332. Available on: https://www.jstor.org/stable/26590557. Accessed May 9, 2023.
[50] Shraddha Kulhari, "Data Protection, Privacy and Identity: A Complex Triad," *Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain Technology to Secure Privacy and Identity*, 1st ed. Nomos Verlagsgesellschaft mbH (2018): p. 27. Available on: http://www.jstor.org/stable/j.ctv941qz6.7. Accessed May 11, 2023.
[51] Linnet Taylor and Lina Dencik, "Constructing Commercial Data Ethics," in *Technology and Regulation* (Tilburg: Open press TiU, 2020), pp. 1-8. Available on: https://www.jstor.org/stable/community.34023115?searchText=personal+data+GDPR&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dpersonal%2Bdata%2BGDPR%26efqs%3DeyJjdHkiOlsiWTI5dWRISnBZblYwWldSZlltOXZhM009Il19&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3A0dec42937d7cc80ed84820d10692c9ce&seq=14. Accessed May 11, 2023.

3.2 IMPACT REGARDING PRIVACY AND DATA PROTECTION

3.2.1 RE-IDENTIFICATION, PRIVACY NORMS AND THE RIGHT TO BE FORGOTTEN

Big data is of great concern for data protection and privacy as a whole. There are several risks related to privacy in the world of Big data in the modern day and age.

> Big data poses big privacy risks. The harvesting of large sets of personal data and the use of state of the art analytics implicate growing privacy concerns. Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world. As more information regarding individuals health, financials, location, electricity use, and online activity percolates, concerns arise regarding profiling, tracking, discrimination, exclusion, government surveillance, and loss of control.[52]

As stated in the quote, there are various concerns regarding privacy and data protection because of the massive collection of personal data relating to various things of an individual, such as their health, financials and other data for the use of profiling, tracking, discrimination, surveillance and more. While there are a lot of potential issues relating to big data and privacy as well as data protection, for the sake of this paper, because it deals with the right to be forgotten, focus will be on the re-identification of individuals and the erasure of data.

While data that are collected might not always lead to the identification of a specific individual, once the data has been linked to the person's identity it becomes entangled and the question whether this data can truly be separated once again from the identified individual and fully deleted arises.[53] According to some researchers, data that has been de-identified, such as information from anonymous surveys and so on, can actually be used in combination with open databases that anyone has reasonable access to in order to re-identify an individual. Researchers Narayanan and Shmatikov used a Netflix dataset and online information to de-anonymize movie viewing records and movie recommendations and re-identify individuals.[54] The researchers developed an algorithm and demonstrated that an adversary, that knows very little about a specific individual, can successfully identify that individual from a crowd of 500,000 subscribers just by referencing the Netflix dataset with a publicly available Movie Database.[55] What's more is that in order to beat the algorithm that the pair constructed, the data has to be altered in such a way that its utility essentially becomes negligible, meaning that data that is being collected nowadays and that has been linked to an identifiable individual is very difficult to truly un-link from the already identified individual and erase it.[56] There are a multitude of data that the adversary could uncover this way, such as political preferences and other potentially sensitive information.[57]

Because even the smallest amount of personal data can spread throughout the internet world like wildfire the European Commission proposed the right to be forgotten when drafting

---

[52] Omer Tene, Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property* 11, no. 5 (April 2013): p. 251. Accessed April 10, 2023.
[53] *Ibid.*
[54] Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets," *The University of Texas at Austin* (2008): p. 14. Available on: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf. Accessed April 10, 2023.
[55] *Ibid.*
[56] *Ibid.*
[57] *Ibid.*

the GDPR. Data is quite difficult to untangle and separate from an individual once they are identified, therefore there has to be a right like the right to be forgotten to counter the adverse effects of big data and restore some privacy norms. Privacy norms have been a debated topic, with the right to be forgotten's central privacy norm being the ability and right to forget one's past information that is no longer valid or applicable to the individual because of changed opinions and/or values, therefore allowing the person to wipe away the information that is no longer valid.

In sum, Big Data is a process of accumulating large amounts of data in order to extract value from it by conducting various types of analyses. It relies on correlation because of the data that they collect relates to other people, allowing companies to discover new information and improve their products. The 3 main components of Big data are the collection, analysis and processing of the data. The collection of data is becoming more prevalent as companies are investing more in technology that collects data, such as smart sensors and other things. Ubiquitous computing is a large part of Big Data, as it uses ubiquitous devices in order to collect data, such as smartphones, smartwatches, cars among other things. Ubiquitous computing may contradict privacy norms, as the information the companies collect gets copied many times over the internet, making it very difficult to trace the data or predict where it eventually ends up. Data that might also not be personal data upon collection may become personal data, if it gets copied countless times across the internet and is used in combination with other information or new technology to identify a specific individual. Big Data creates worrisome privacy threats as well as some issues with the implementation of the right to be forgotten. Namely, once an individual has been linked with a piece of data, it is quite difficult to un-link this person from the data and forget that this individual was ever linked with that specific data. The essence of the right to be forgotten is that individuals may choose to "forget" their past selves, or in other words, delete information that is no longer relevant due to changes in beliefs or other things. It was shown in a research done by Narayanan and Shmatikov that re-identification of individuals is very possible after personal data has been anonymized, however, it is not necessarily abundantly likely.

4. SURVEILLANCE CAPITALISM – A NEW FORM OF INFORMATIONAL CAPITALISM

4.1 SURVEILLANCE CAPITALISM

4.1.1 DEFINITION

Big data is a phenomenon that has been around for some time and its impact on the modern world is becoming more and more noticeable every day. Big data is mostly described as the accumulation of personal data and then the processing of that data using intelligent computer systems to create algorithms and applying them to discover new information that could uncover certain patterns about people's behavior, which would enable profiling and other important discoveries about how to link certain interests to people that one knows relatively little about. However, upon reviewing academic literature, some scholars believe that this definition is slightly inaccurate, or rather – incomplete or explained using the wrong approach. Shoshana Zuboff, a information systems scholar, has coined the term "surveillance capitalism," which is, in her words, a system which "aims to predict and modify human behavior as a means to

produce revenue and market control."[58] As the quote states, surveillance capitalism focuses on the accumulation of data, but what makes it different than big data in a way is that there is a monetization dimension to it. Big data, in most cases, is defined as a byproduct of technological advancement, it is defined as a technological process of accumulating large datasets and analyzing them, however Zuboff argues that this definition does not show the full picture of what is happening nowadays.[59] Surveillance capitalism illustrates a more sinister image about what is going on in the world of data, defining it as being more of a social phenomena than a technological one.[60]

4.1.2 A NEW FORM OF CAPITALISM

While in the previous century it was labour or land that defined the way capitalism was monetized, surveillance capitalism is a new form of capitalism that focuses on turning the private experience of individuals into a commodity that is traded on the financial market.[61] Surveillance capitalism is a rapidly evolving phenomena that cannot be attributed to any single company or entity, it is practiced by many companies nowadays that strive to capitalize on the information that they obtain from individuals in ways that challenge the people's autonomy, privacy and democratic norms.[62] The market for data is increasing every day, which provides for an incentive to accumulate as much data as possible in order to influence and even change people's behavior, which often forces companies to compete with each other, resulting in a brawl over profit and influence that often may transcend legal norms. Google is a prominent figure in the surveillance capitalism field, even deemed as its inventor by Zuboff.[63] According to Zuboff, Google has been accumulating data in inconspicuous and innovative ways that no one else is doing or has the capacity to do it.[64] When Google was working on its Street View service for Google Maps, they used cars to drive around cities and neighbourhoods and photographing houses, while also collecting information from Wi-Fi networks in the area.[65] Besides various potential privacy violations regarding the collection of information from local Wi-Fi networks, Google was photographing people's houses without their permission.[66] Actions have been brought before the court regarding this practice by Google, however, the cases usually end in a settlement where Google merely pays for any damages the individual has incurred.[67] This would typically sound like an expensive affair for Google, however the settlement fees they pay are a small price to pay for the return they get from the value of the data they obtain from the photographies of those houses. Google attempts to accumulate data in different ways as well, essentially disregarding any legal norms at first until they meet resistance from individuals that claim their privacy has been violated, then they proceed to pay the individuals in what ultimately is only a fraction of what they gain from every case.[68]

---

[58] Shoshana Zuboff, "Big other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology 30* (2015): pp. 75-89. Available on: https://doi.org/10.1057/jit.2015.5. Accessed April 13, 2023.
[59] *Ibid.*
[60] *Ibid.*
[61] Shoshana Zuboff, "Surveillance Capitalism and the Challenge of Collective Action," *New Labor Forum,* vol. 28, no. 1 (2019): pp. 10-29. Available on: https://doi.org/10.1177/1095796018819461. Accessed April 15, 2023.
[62] *Ibid.*
[63] *Ibid.*
[64] *Ibid.*
[65] *Ibid.*
[66] *Ibid.*
[67] *Ibid.*
[68] Zuboff, *Supra* note 58.

4.1.3 IMPACT ON EVOLVING CONTOURS OF PERSONAL DATA

In her various works that attempt to shed some light on the new phenomena that she has termed "surveillance capitalism," Zuboff focuses on defining its history, its origins and how it shapes the present and will continue to shape the future. Surveillance capitalism is the answer why Google has always been an innovator and one step ahead of its competition – they were the first company to handle large amounts of data that they later realized can be turned into a commodity in the form of predicting the future behavior of people.[69] The information regarding future behavior of people is a very powerful commodity that companies can use to improve their products, advertisement and marketing strategies.[70] Companies like Google and Amazon employ home assistants that help personalize the experience of their users while these companies actually benefit from this as it is a sneaky way to getting access to the private experience of individuals. The private experience of individuals is a commodity that is growing in value every day while there are more and more ways companies are accessing our private lives nowadays – through ubiquitous computing and various sensors that are cheap to finance in comparison to the massive return companies get.

While Zuboff focuses more on the actual functionality of surveillance capitalism, there is another important conclusion that can be drawn – that surveillance capitalism is the main driver behind the evolving definition of personal data. As discussed before in this paper in section 2.1.2 regarding contemporary issues of the definition of personal data, the 2 main issues that were mentioned were that, firstly, data that is not necessarily personal data can potentially be combined with new technologies (such as sensors and home assistants) to become personal data under the definition of personal data under the GDPR and secondly, companies or data processors are attempting to manipulate personal data into appearing as anonymized data in order to avoid compliance with GDPR. In this particular case the former is the more relevant point that is directly linked with surveillance capitalism. Data that has the potential to be considered personal data when combined with new technology to identify or re-identify someone has to be reviewed constantly so that it does not lead to indirect identification of a specific individual. Since Google showed that enormous surplus can be created from personal data with relatively little consequences, many other companies and sectors have engaged in the phenomena that is called surveillance capitalism.[71] This economic creation, just as any form of capitalism, attempts to take something that is outside the market and transform it into a commodity that can be traded on the market, in this case personal data.[72] Since the sector of this phenomena is growing, so is the competition and new technologies and ways to identify people that was not possible before. Because of this fierce competition, companies are developing technologies with as many sensors to monitor behavior as possible in order to obtain as much personal data as possible. "Datafication," as defined in the paper previously, is the transformation of real life information into data, which is made possible via technology such as smart home assistants. Because of this large accumulation of data and the struggle of companies to extract as much data as possible in many ways, the definition of personal data is expanding every day and every time new technology is invented.

---

[69] Zuboff, *Supra* note 61.
[70] Liam Welch, "Grave New World: Mass Surveillance and Labour Rights," *Socialist Lawyer, no. 83* (2019): pp. 36–41. Available on: https://doi.org/10.13169/socialistlawyer.83.0036. Accessed May 8, 2023.
[71] Zuboff, *Supra* note 61.
[72] *Ibid.*

4.2 NECESSITY OF IDENTIFICATION

4.2.1 IS IT NECESSARY FOR COMPANIES TO IDENTIFY INDIVIDUALS?

Companies engaging in surveillance capitalism collect an enormous amount of data. This data is then used for purposes such as profiling, where companies build data profiles of users to then apply to other users in the form of curated recommendations even if the user only recently signed up to the service and has not provided much information. The more specific data can be attributed or related to previously collected data, the more valuable it is as it makes it easier to extract value from customers and also provide value in the form of curated content.[73] However, there are important questions that lie in the whole process of data extraction and further analysis by companies – is it important for the companies to be able to identify individuals in order to create automated and curated content? Is it enough to identify a single individual and then based on the information they provided make recommendations for others with similar information provided? Is it completely unnecessary to identify anyone?

The ability for companies to identify individuals may pose privacy threats and the violation of the GDPR in some cases (namely the right to be forgotten), therefore it is important to investigate whether identification of individuals is inevitable under the current surveillance capitalism regime. The practice of surveillance capitalistic companies is the following:

> [I]f the records that belong to a given person can be connected during a series of visits and between different websites or applications, a far more sophisticated user profile can be generated. […] [W]ebsites use trackers that are able to track the entire activity during the browser session. It is also a reason that Facebook and Google are highly motivated for users to use their Facebook or Google accounts for identification with other service providers. Third, the correlation of many of these detailed user profiles makes it possible to make statistical predictions about user behavior, and thereby make sophisticated assumptions that are not even necessarily limited to what the users themselves are conscious of. Because of this intra- and inter-personal connection, new data is more valuable the more it can be correlated with already-held data. As a consequence, a network effect occurs that amplifies the centralization of data.[74]

Companies use trackers on their websites that track the entire browsing session – meaning that if the user, for example, visits any Google service like Gmail, a tracker may be attached to that person's browsing session, which results in Google obtaining data about that user's behavior even outside of Google services.[75] This results in the accumulation of a large amount of data that Google can further use to build data profiles.

However, the important question is whether companies need to identify a specific individual in order to extract value from the collected data. While the data that companies collect oftentimes includes personal identifiers that are attached to the data, companies do not necessarily always need the personal identifiers in order to keep the value of the data.[76] In fact, in the cases where data that have personal identifiers attached to them is being processed, companies use tools to de-identify the data and not lose its value.[77] It was argued previously in this paper that de-identification is a very difficult task to the extent where re-identification of

---

[73] *Ibid.*
[74] Marvin Landwehr, Alan Borning and Volker Wulf, "Problems With Surveillance Capitalism and Possible Alternatives for IT Infrastructure," *Information, Communication & Society, 26:1* (2023): pp. 70-85. Available: https://doi.org/10.1080/1369118X.2021.2014548. Accessed April 20, 2023.
[75] *Ibid.*
[76] *Ibid.*
[77] *Ibid.*

individuals becomes impossible or negligible, however, some scholars argue that the ability to re-identify individuals after de-identification has taken place is not nearly as common as it has been made to believe.[78] In order to enable third parties to use the data that the original company collected, the company performs de-identification, however, it is not always completely effective.[79] In order to de-identify data companies have to remove any possible identifiers that could enable someone to identify a specific individual, potentially with malicious intent. The issue seems to be that it is not always a simple task to remove identifiers. Previously in the paper the concepts of direct and indirect identification were discussed. It is apparent that removing the identifiers that allow for direct identification is a straightforward task, however, the identifiers that could enable indirect identification or quasi-identification are much more difficult to fully remove, as they are linked with unique identities and in the overall scheme can be correlated to specific individuals in combination with additional reasonably accessible information.[80] While in some cases after the process of de-identification the data might lose its utility, it is still possible to preserve a high grade of utility of the data after de-identification.[81] In conclusion, companies do not necessarily need to know the identity of a specific individual because data can retain its utility even after it has been identified, however, in some cases some data may not be de-identified as it would lose its utility, meaning that situations exist where identification is an unintentional byproduct of data analysis and collection by companies.

5. THE RIGHT TO BE FORGOTTEN UNDER THE GDPR

5.1 APPLICATION OF ARTICLE 17 OF THE GDPR

5.1.1 PERSONAL RIGHTS

Article 17 of the GDPR lays down the terms for the right to be forgotten. Article 17 offers various rights for protection of personal data and also includes cases where those rights are overridden in case of a legitimate interest. The right to erasure or right to be forgotten is ensured in these cases:

> the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
>
> the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;
>
> the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);
>
> the personal data have been unlawfully processed;
>
> the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;

[78] Ann Cavoukian and Daniel Castro, *Big Data and Innovation, Setting the Record Straight: De-identification does work* (Ontario: Information and Privacy Commissioner, 2014).
[79] *Ibid.*
[80] *Ibid.*
[81] *Ibid.*

> the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).[82]

While the focus is mostly on the data subject and the protection of their data by giving them more control over it, there are some cases where that protection is overridden:

> for exercising the right of freedom of expression and information;
>
> for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
>
> for reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3);
>
> for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or
>
> for the establishment, exercise or defence of legal claims.[83]

The rights under Article 17 are clear: the data subject has the right to request erasure of their data without undue delay (undue delay is said to be a period of about a month) when the data is no longer necessary for the purpose it was originally collected for, when the data subject withdraws their consent, when the data subject objects to processing, when the data have been unlawfully processed, when the data controller is bound by national or union level legislation that requires that the data is deleted and if the data have been collected in relation to the offer of information society services. However, in a nutshell, the right is overridden if there is a legitimate public interest that justifies it. The Covid-19 pandemic is one example of such a case. It is obvious that when there are cases of privacy concerns in the name of public health, a balance will have to be struck between the right to privacy and interest of public health, which in itself contains principles of privacy and also the principles of the wider public and the European Union. As there are competing human rights, the principle of proportionality is the key factor in determining that the right to be forgotten is overridden when a health crisis is considered.[84]

5.1.2 THE RIGHT TO BE FORGOTTEN – A NOVEL RIGHT

The concept of the right to be forgotten is not exactly new – it has been a right since the Data Protection Directive (DPD), which was the predecessor to the GDPR. While the GDPR does not bring any new substantive changes to the right to be forgotten, it is important symbolically as it defines the right and conditions under which it can be used.[85] The right to be forgotten is a new right with a broader scope of application compared to the one under the DPD while also

[82] GDPR, *Supra* note 8.
[83] *Ibid.*
[84] Mónica Correia, Guilhermina Rego & Rui Nunes, "The Right to Be Forgotten and COVID-19: Privacy versus Public Interest," *Acta bioethica 27* (2021): pp. 59-67. Available: http://dx.doi.org/10.4067/S1726-569X2021000100059. Accessed April 6, 2023.
[85] W. Gregory Voss, "European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting," *The Business Lawyer 72, no. 1* (2016): pp. 221–34. Available on: https://www.jstor.org/stable/26419118. Accessed May 5, 2023.

being applicable retroactively.[86] There are, however, technical challenges in enforcing the right, as well as the scope of application not being cristal clear in every situation.[87] The right to be forgotten is considered a human right and an extension of the right to privacy, despite it being questioned and facing resistance from various businesses and free speech advocators due to it conflicting with other rights and interests.[88]

5.2 CHALLENGES AS REGARDS TO THE ERASURE OF PERSONAL DATA

5.2.1 DATA BACKUPS

The right to be forgotten is a legal concept that is quite problematic in terms of implementation in the digital world. The GDPR is proactive in terms of legal nature and does not provide technical instructions on how to implement all of the provision in order to achieve compliance, meaning that certain technology companies and platforms cannot gain an unfair advantage over others, however, this is still problematic in nature.[89]

Another important problem with the implementation of the right to be forgotten are data backups. With there being more and more reliance on ICT services, companies must keep backups in case of any accidents so that the data does not get deleted and disrupt the service.[90] When a user invokes the right to be forgotten and asks for the data controller to erase the data, the controller has to erase the data from the backups as well.

> Apparently, according to the GDPR this deleting action must be performed in the backups as well, opening thus the door to potential data abuses, deliberate exploitations or even accidental mistakes. Propagating the required erasure mechanisms to backups, empower users and financial institutions to manipulate data integrity according to their needs, like hiding transactions from audit controls when deemed necessary. [...] Therefore, once a user requests the deletion of his data, non-automated, and –contrary to the legal framework within the institution operates– actions have to be performed, leading to additional costs and possible legal deadlocks. Such issues may become more evident in financial institutions where records must always follow the information reliability, integrity and transparency principles.[91]

As stated in the quote, data backups are a complicated topic in respect to the right to be forgotten. From one side, the erasure from data backups is necessary, as it is the only way to truly ensure the right to be forgotten. However, there are a couple of problems with this, namely the fact that it paves the way for various forms of data exploitation and even mistakes.[92]

6. ANALYSIS OF THE RIGHT TO BE FORGOTTEN IN RESPECT TO THE EVOLVING DEFINITION OF PERSONAL DATA

6.1 MAIN ISSUES REGARDING THE RIGHT TO BE FORGOTTEN

[86] Politou, *Supra* note 40.
[87] Ashley Nicole Vavra, "The Right to Be Forgotten: An Archival Perspective," *The American Archivist 81, no. 1* (2018): pp. 100–111. Available on: https://www.jstor.org/stable/48618003. Accessed May 5, 2023.
[88] Politou, *Supra* note 40.
[89] *Ibid.*
[90] *Ibid.*
[91] *Ibid.*
[92] *Ibid.*

6.1.1 ANALYSIS METHODOLOGY

As discussed above, there are a bundle of issues regarding the implementation of the right to be forgotten. Since the contours of personal data are also evolving, it is becoming quite an important topic that needs to be addressed – to what extent the right to be forgotten can be ensured in respect to the evolving definition of personal data and whether it is possible to effectively implement it in an age of surveillance capitalism. In order to investigate to what extent the right to be forgotten can be ensured in respect to the evolving contours of personal data, it is necessary to recap the main reasons why the contours of personal data are evolving and what is/are the main drivers behind it. The essence of the right to be forgotten is that personal data have to be erased in such a way that re-identification of individuals using that specific data is made irreversible, thus – impossible. It is necessary to carry out these measures because an individual may not be considered as "forgotten" if the information that the individual requested to be deleted can be used to identify them (one of the main components of personal data is its relation to the data subject and it is important to dissociate the data from the data subject in order to truly "forget" the data that the data subject wishes to be forgotten).

6.1.2 BIG DATA AND COPIED INFORMATION

Big Data, as previously established in 3.1.1., is the accumulation and subsequent processing of the data to uncover new, potentially useful information and insights about the population, which involves making smart machine calculations to obtain new information. This new information could be used in combination with older, previously accessible information in order to identify certain individuals, if needed. The problem in this scenario that is linked to the right to be forgotten is the fact that this information is sold or copied[93] over to many other individuals over the cyberpace, which not only increases the chance of identification, but also makes it extremely difficult to truly erase as the original data controller is no longer the sole owner and possessor of the personal data and tracking where the data ends up is a difficult task. Therefore, it is important to assess how protected the individual is under the GDPR in terms of the right to be forgotten against this practice.

6.1.3 UBIQUITOUS COMPUTING AND PROFILING

Ubiquitous computing is used to accumulate large amounts of data through various sensors, which are easy to finance because of the value companies obtain from the extracted data. These sensors are everywhere – GPS tracking, smartphone cameras, gyroscope sensors in our everyday devices that are constantly collecting data on us. These data, all of which we might not be aware of that are collected, are used to build data profiles using inferred data where the quality of data does not necessarily matter.[94] It is important to investigate how protected we are from the fact that profiles are being built around us and for us using our data. The data companies use, such as behavioral data, that have the potential to identify certain individuals, are used to build profiles and the data that a person wishes to be forgotten may be used to build a certain profile, meaning that the data has not been truly erased.

---

[93] Copying is the process of duplicating certain pieces of data for the purpose of redistributing it or selling it to 3rd parties.

[94] Bart Custers, "Profiling as inferred data: amplifier effects and positive feedback loops," In *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen,* ed. Emre Bayamlioglu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt (Amsterdam: Amsterdam University Press, 2018), pp. 112-115. Available on: https://doi.org/10.2307/j.ctvhrd092.23. Accessed May 11, 2023.

6.2 DOCTRINAL ANALYSIS OF THE MAIN ISSUES REGARDING THE RIGHT TO BE FORGOTTEN

6.2.1 ANALYSIS OF COPYING INFORMATION

Information that is collected, whether with or without consent, gets copied or sold many times over sometimes that eventually creates a web of data and makes it less clear where specific data ends up in the cyberspace.[95] This is one of the biggest problems with the right to be forgotten as it directly obstructs the main process that must be carried out in order to ensure the right – and that is "erasure" of personal data. Article 17 (2) of the GDPR vaguely mentions copying information:

> Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.[96]

On paper, the paragraph acknowledges the existence of potential copies of data and attempts to regulate it by imposing an additional obligation on data controllers of informing the other processors that hold the same data that the personal data in question must be erased pursuant to Article 17 (1) of the GDPR. Although the requirement of erasing copied information is defined, there is a lack of clarity in the article regarding some terms. The term "erasure" is nowhere to be defined in the article and neither is it throughout the GDPR, the wording implies that the process of erasure is simple and straightforward, just as destroying a regular file in real life.[97] In addition, the term "reasonable," despite being tackled earlier in the paper, still does not offer full clarity as to what are the "reasonable" steps that the data controller has to take, with the current position being that economic incentives and legal obligations have to be balanced in order to ensure the steps taken are "reasonable." While the whole approach of the GDPR is agnostic in terms of technological phenomena which in turn makes the GDPR more flexible and proactive in its nature in terms of technological advancement which is very necessary in the field of data protection, the absence of clear definitions and the presence of slightly ambiguous wording are decreasing the chances of effective implementation of the GDPR and the right to be forgotten. While technology is advancing, GDPR seems to be having a difficult time catching up to the latest technology and all of the complexities surrounding it.[98] Another issue regarding implementation of this article is knowing where the data has travelled, or rather – who are the 3rd parties or other data controllers that are processing this information.[99] The lack of clear definitions of how to properly and efficiently obtain information from all of the third parties and how to effectively erase it has also shown to raise doubts about its enforcement.[100]

Article 17 (2) does indeed define and acknowledge the practice of copying information, however its tech-agnostic wording, intended to encapsulate various technological practices,

---

[95] Politou, *Supra* note 40.
[96] GDPR, *Supra* note 8.
[97] Eduard Fosch Villaronga, Peter Kieseberg, Tiffany Li, "Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten," *Computer Law & Security Review* 34, 2 (2018): pp. 304-313. Available: https://www.sciencedirect.com/science/article/abs/pii/S0267364917302091. Accessed May 9, 2023.
[98] *Ibid.*
[99] Politou, *Supra* note 40.
[100] *Ibid.*

widens the scope of the statute at the expense of clear practical terms that facilitate enforcement, which ultimately backfires on the legislators that had good intentions in the first place. Because of the lack of clear terms that hinder enforcement and potentially, liability of the third parties that refuse or ignore the data controllers' requests to delete copies of the data, the GDPR is not competent enough to fully ensure the "right to be forgotten." It is difficult to determine exactly to what extent the right to be forgotten is ensured under the GDPR in terms of copying information taking into account all of the technical complexities, however it is clear that there are fundamental issues in terms of enforcement, leading to the conclusion that the erasure of copied information is underregulated, thus inadequately ensuring the right to be forgotten.

6.2.2 ANALYSIS OF PROFILING

To reiterate, profiling is the process of tracking user behaviors online and running that data through machine learning techniques in order to create detailed and accurate profiles on individuals that allow for precise predictions of future behavior in terms of interests, preferences and traits of certain individuals.[101] Because machine have been equipped with feedback loops that allow them to learn and teach themselves, they are able to constantly improve on their own with little to no human involvement needed.[102] The technicalities surrounding profiling are as follows:

> [A]lgorithms that use a so-called "knowledge-base" for calibration, i.e., the algorithm takes the knowledge-base with pre-calculated results as reference data and extracts the common artifacts. It then uses these "learned" rules on new data, which have to be very close to the training data in terms of data structure and statistical properties. Furthermore, the resulting categorizations are again fed into the knowledge base in order to get even better training data for the next run, thus iteratively extending the knowledge base.[103]

"Knowledge base" in this case is meant as certain pieces of personal data that are used to "train" the machine algorithms to apply new rules or criteria that allows the machines to relate the data from one individual of the population to others. This process of relating individuals to other individuals is the foundation of profiling and its main objective is to predict future behavior of individuals by observing past behavior. Profiling is linked to the right to be forgotten because personal data is used to train algorithms, meaning that for the right to be forgotten to be effectively ensured, personal data that are used to train algorithms are deleted from the machines that use feedback loops to train themselves in order to make sure any residual traces of personal data are removed. Research suggests that deleting the data from the machines does not always cause results of the machine learning to change drastically, however, it is also important to consider that the research was conducted by deleting random points of data and not specific points of data that could, in real practice, drastically change the results.[104]

The GDPR covers profiling extensively, as it is found in multiple articles in various different contexts.[105] It is defined in Article 4 (4), Article 13 (1) (f) states that the data subject should be informed if their personal data are used for profiling purposes, Article 14 (2) (g)

---

[101] Karen Yeung, "Five fears about mass predictive personalization in an age of surveillance capitalism," *International Data Privacy Law, Volume 8, Issue 3* (2018): pp. 258–269. Available: https://doi.org/10.1093/idpl/ipy020. Accessed May 10, 2023.
[102] *Ibid.*
[103] Villaronga, *Supra* note 97.
[104] *Ibid.*
[105] Klaus Wiedemann, "Profiling and (automated) decision-making under the GDPR: A two-step approach," *Computer Law & Security Review, Volume 45* (2022). Available: https://doi.org/10.1016/j.clsr.2022.105662. Accessed May 10, 2023.

refers to the same thing as Article 13 (1) (f), but in a context of data being collected from a third party and not directly from the data subject, Article 15 (1) (h) defines the right of the data subject to obtain information whether profiling is performed on their personal data, Article 21 (1) and (2) defines a right to object to profiling, Article 35 (3) (a) requires a data protection impact assessment to be carried out if profiling is taking place.[106] Because of such extensive coverage, on paper, profiling seems to be governed and defined well enough[107] and it does not suffer from the same lack of clarity as copied information, therefore from a legal aspect the regulation displays that it is well aware of the practice and its severity. The reality, however, is different in real practice. While the GDPR adequately deals with profiling as a practice, the problem at hand is that most businesses, largely being unaware, engage in practices that amount to profiling, contribute or benefit from an ecosystem that is built on profiling while not considering themselves as entities that engage in the practice.[108] Understanding when profiling takes place is a key part of the GDPR being enforced in terms of profiling as data subjects become aware of their rights, which unfortunately is not the case most of the time, which is why profiling continues to happen.[109] There is a disconnect between the GDPR and real practice, which is the main problem, as well as the lack of awareness of companies.

The right to be forgotten is defined in Article 17 of the GDPR, where no definition or presence of profiling is present.[110] Instead, Article 17 (1) (c) refers to Article 21 (1) and (2) which is the right to object, which includes profiling in its definition.[111] Since profiling is well defined in the statute, GDPR adequately offers protection against profiling in terms of the right to be forgotten, as well as any right under the GDPR in general. The main issue with protection against profiling is the lack of awareness of the companies, or potentially intentional lack of awareness, to the engagement in the practice which is why profiling remains an issue that may not always be enforced.

7. HARMFUL DATA EXTRACTION REGULATION UNDER THE GDPR

7.1 DOCTRINAL ANALYSIS REGARDING THE RIGHT TO BE FORGOTTEN IN TERMS OF HARMFUL DATA EXTRACTION

7.1.1 THEORY OF HARMFUL DATA EXTRACTION

The theory of harmful data extraction was first discovered in a German case concerning Facebook.[112] The issue of harmful data extraction arises from Facebook's third party tracking of users across different websites and applications.[113] The Bundeskartellamt, Germany's

[106] GDPR, *Supra* note 8.
[107] Wiedemann, *Supra* note 105.
[108] Chiara Rustici, "GDPR Profiling and Business Practice," *Computer Law Review International, Volume 19 issue 2* (2018): p.34. Available on: https://doi.org/10.9785/cri-2018-190203. Accessed May 10, 2023.
[109] *Ibid.*
[110] GDPR, *Supra* note 8.
[111] *Ibid.*
[112] Bundeskartellamt, *Facebook*, B6-22/16, 2019. Available on: https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4. Accessed May 10, 2023.
[113] Viktoria Robertson, "The Theory of Harm in the Bundeskartellamt's Facebook Decision," *Competition Policy International* (2019): pp. 2-3. Available on: https://www.competitionpolicyinternational.com/wp-content/uploads/2019/03/EU-News-Column-March-2019-Full-1.pdf. Accessed May 10, 2023.

competition authority, held that Facebook had access to third party data, namely from other immensely popular services owned by the social media giant in Instagram and WhatsApp.[114]

> By combining extensive third-party data sets with the data it gathers through its own website and applications, Facebook is able to turn multi-source data into comprehensive user profiles. Users do not freely agree to this practice, as theirs is an all-or-nothing choice: Either access Facebook's popular social networking services and accept its exploitative data practices, or be shut out from that dominant social network.[115]

Even if users don't use Facebook or have never used it, Facebook can and has still built user profiles through its own data and the data available from Instagram and WhatsApp.[116] In the eyes of the Bundeskartellamt, being a user of Instagram and WhatsApp does not constitute consent of processing data for Facebook, as Facebook ultimately has a dominant market position in terms of social media and users have little to no choice in terms of alternatives.[117] The link regarding GDPR in this case is the question of consent, which according to the Bundeskartellamt, is linked to the users not having full comprehension of the extent of the processing of their data.[118]

7.1.2 ANALYSIS OF HARMFUL DATA EXTRACTION REGULATION

While working closely with the European Data protection authorities, the Bundeskartellamt came to a conclusion that the GDPR and its values have been violated in this case to the detriment of the users of the platform.[119] Recital 43 is particularly relevant:

> […] Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance.[120]

Consent is presumed not to be freely given in cases where it is not necessary to carry out a service. This effectively forbids Facebook to use data obtained from Instagram and WhatsApp and other third party applications on the basis of the user not giving consent, as it is presumed to not be freely given in the case where it is not needed, and in the case of users of Instagram and WhatsApp, these users do not need to be giving consent to Facebook in particular to use their data and build profiles for them to be able to use Instagram and WhatsApp respectively.

In its judgment, the Bundeskartellamt named violations of Article 6 (1) (a), (b) and (f) mainly.[121] Article 6 (1) (a) deals with effective consent, and, in the Bundeskartellamt's opinion, there was no effective consent ensured as users of Facebook concluded the terms and conditions contract for the sole purpose of concluding the contract because Facebook has a dominant

---

[114] *Ibid.*
[115] *Ibid.*
[116] *Ibid.*
[117] *Ibid.*
[118] Viktoria Robertson, "Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the era of Big Data," *Common Market Law Review 57* (2020): p. 181. Available on: https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2020006.pdf. Accessed May 11, 2023.
[119] Bundeskartellamt, Bundeskartellamt prohibits Facebook from combining user data from different sources (2019). Available on: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html. Accessed May 11, 2023.
[120] GDPR, *Supra* note 8.
[121] Bundeskartellamt, *Supra* note 112.

market position and no alternatives.[122] Article 6 (1) (b) deals with the necessity to process data in order to carry out a service or contract, however, the Bundeskartellamt held that Facebook does not need to carry out the processing of third party data in order to fulfill its obligations to its users.[123] Article 6 (1) (f) deals with legitimate interest of processing data, and it is unjustified in this case.[124] Facebook's legitimate interests that were brought forward did not outweigh other interests, such as consequences of the affected users, the data type and the way it was processed, reasonable expectations of users and positions of Facebook and its users.[125] Facebook, because of its market position, was able to impose data processing conditions that were far-reaching, obtaining more data than was necessary to operate, which cannot be justified without proper user consent.[126]

Harmful data extraction practices have long been under everyone's radar, but have now surfaced in the wake of the age of surveillance capitalism with the help of the Facebook case in Germany. This case could be a landmark case which could have the potential to uncover the wrongdoings of many other technology giants, namely Google, as well as bring more awareness to their illegal practices and surveillance capitalism as a whole. Harmful data extraction, as evidenced in the Facebook case, seems to be regulated well under the GDPR, however, it is not acknowledged or defined as a practice under the regulation, which makes room for improvement of the legislation in terms of harmful data extraction. The essence of harmful data extraction is regulated under the GDPR, however, it has gone unnoticed until now most likely because of the lack of awareness of the true practice and intentions behind companies' data processing conditions, which will now come to light. Spreading awareness of the rights of individuals and the practices of companies and including clearer terms are some of the ways improvements could be made to the GDPR to ensure that harmful data extraction is regulated in a more comprehensive way and make sure that future violations of this nature do not take place.

8. CONCLUSIONS

The definition of personal data is enshrined in Article 4 (1) of the GDPR that can be broken down into 4 components – "natural person," "any information," "relating to," "identified or identifiable." "Natural person" relates to persons that are not legal persons and are alive, "any information" relates to information of any kind that is not limited to any specific format, "relating to" deals with the data being able to be related to a person and "identified or identifiable" deals with identification of individuals as a result of information being able to be related to certain persons. The issues with the definition of personal data still persist, mainly the issues of new technology enabling data that is not personal data to become personal data via indirect identification, as well as data being manipulated into being anonymous data hwne in reality it is not.

The fact that the contours of personal data are evolving can be largely attributed to surveillance capitalism. The definition of personal data is expanding due to practices that are fundamentally rooted in surveillance capitalistic intentions. Profiling and copying information are the maine practices discussed in the paper in terms of the right to be forgotten and to what

[122] *Ibid.*
[123] *Ibid.*
[124] *Ibid.*
[125] *Ibid.*
[126] *Ibid.*

extent it is ensured. Profiling is the process of accumulating large amounts of data and running it through machine learning mechanisms with feedback loops that train themselves to build ddata profiles of users, as well as individuals that are not users to the specific service but are users of other third party services that the data controller has acces to the information therein. Copying information also lets other companies in the data ecosystem, to benefit from profiling and other data analysis practices. Because of the competition between companies to obtain more data as it becomes a crucial part of a new form of capitalism, it drives technology forward in terms of obtaining data and learning new information, thus turning information that was not previously personal data into personal data that can be used to effectively identify someone, therefore expanding the definition of personal data.

Copied information is inadequately regulated under the GDPR. The GDPR does define the practice of copying information, but ultimately it suffers and falls short due to ambiguous and unclear wording. Because of the lack of clear practical enforcement terms regarding technological complexities, the GDPR's ability to ensure the right to be forgotteb in terms of copied information is suboptimal, meaning that the right to be forgotten cannot effectively be ensured under the GDPR in terms of copied information.

Profiling, on the other hand, is well defined throughout the GDPR in various contexts and articles. On paper, it does not suffer from a lack of clarity as does the practice of copying information, however, profiling is met with an entirely different issue that may or may not be out of the GDPR's reach to regulate – which is the lacj of awareness, or intentional lack of awareness of companies that engage in profiling or benefit from an ecosystem that is built on profiling. With the current state of the GDPR, profiling is sufficiently regulated in order to effectively ensure the right to be forgotten, however one of the shortcomings of the GDPR in this case may be the identification of profiling, which, evidently, has not been fruitful.

Harmful data extraction practices have existed for longer than we have been aware of them, however, the Bundeskartellamt's Facebook case could be a landmark case to bring the issues surrounding harmful data extraction practices to light and expose the wrongdoings of other tech giants that abuse their dominant market position. Harmful data extraction is considered a violation of the GDPR after extensive analysis, meaning that the GDPR does well to regulate the practice, however, it lacks the definition of harmful data extraction and its inclusive practices that could only improve the protection against this phenomenon and also bring more awareness to it.

To answer the research question – to what extent does the GDPR ensure the right to be forgotten in regard to evolving contours of personal data and regulate harmful extraction of data in terms of surveillance capitalism? Starting with the right to be forgotten, the GDPR does not ensure the right to be forgotten effectively in terms of copied information and while it does extensively regulate profiling and from the legal aspect contains everything needed for effective enforcement, ultimately the lack of awareness of companies is what prevents the right to be forgotten to be ensured in practice. Harmful data extraction is regulated under the GDPR as it violates some of its core principles, however, harmful data extraction is not defined under the GDPR, while having such a definition could only bring more awareness and more strict regulation that would ultimately govern harmful data protection effectively. Therefore, as it stands now, the GDPR does regulate harmful data extraction, however, with clearer definitions, it could improve its regulation immensely.

Overall, the author concludes that the GDPR's framework and current legal approach are working well – a proactive approach in such a volatile field as data protection law is

imperative to ensure that the GDPR is keeping up with the technology. Although one of its shortcomings is the difficulty to keep up with the technology, the author concludes that another very important issue is the lack of clarity in terms of definitions and enforcement terms. The GDPR does not require a massive overhaul of its fundamental operation, however it will benefit immensely if the legislators and computer scientists that know the ins and outs of the technical complexities have a closer collaboration to make the GDPR more intuitive and inclusive. The common denominator of the inefficiencies with copied information, profiling and harmful data extraction in the GDPR is the lack of clarity and lack of additional, but at the same time paramount, information. There are a few questions left unanswered to perhaps serve as future points of research. Are there other surveillance capitalistic practices we are currently unaware of that are illegal under the GDPR? If so, how well regulated and defined are they? Will adding more definitions and enforcement terms to the GDPR make it inclusive and exhaustive? These are important questions that could raise awareness and provide clarity for individuals regarding their rights and the exploitative behavior of companies, if answered.

9. BIBLIOGRAPHY

PRIMARY SOURCES

STATUTES

1. General Data Protection Regulation, OJ L 119, 4.5.2016, p. 1-88. Available on: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed March 29, 2023.

CASES

1. Bundeskartellamt, *Facebook*, B6-22/16, 2019. Available on: https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4. Accessed May 10, 2023.

SECONDARY SOURCES

BOOKS

1. Cavoukian, Ann and Daniel Castro. *Big Data and Innovation, Setting the Record Straight: De-identification does work*. Ontario: Information and Privacy Commissioner, 2014.
2. Custers, Bart. "Profiling as inferred data: amplifier effects and positive feedback loops." In *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen,* edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt (Amsterdam: Amsterdam University Press, 2018), pp. 112-115. Available on: https://doi.org/10.2307/j.ctvhrd092.23. Accessed May 11, 2023.
3. Kulhari, Shraddha. "Data Protection, Privacy and Identity: A Complex Triad." In *Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain Technology to Secure Privacy and Identity*, 1st ed. Nomos Verlagsgesellschaft mbH (2018): p. 27. Available on: http://www.jstor.org/stable/j.ctv941qz6.7. Accessed May 11, 2023.
4. Richterich, Annika. "Big Data-Driven Health Surveillance." In *The Big Data Agenda: Data Ethics and Critical Data Studies*, (University of Westminster Press, 2018), pp. 71-90. Available on: https://doi.org/10.2307/j.ctv5vddsw.7. Accessed May 11, 2023.
5. Taylor, Linnet and Lina Dencik. "Constructing Commercial Data Ethics." In *Technology and Regulation* (Tilburg: Open press TiU, 2020), pp. 1-8. Available on: https://www.jstor.org/stable/community.34023115?searchText=personal+data+GDPR&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dpersonal%2Bdata%2BGDPR%26efqs%3DeyJjdHkiOlsiWTI5dWRISnBZblYwWldSZlltOXZhM009Il19&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3A0dec42937d7cc80ed84820d10692c9ce&seq=14. Accessed May 11, 2023.

JOURNALS

1. Busch, Christoph. "Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law." *The University of Chicago Law Review 86, no. 2* (2019): pp. 309–332. Available on: https://www.jstor.org/stable/26590557. Accessed May 9, 2023.
2. Correia, Mónica, Rego, Guilhermina & Rui Nunes. "The Right to Be Forgotten and COVID-19: Privacy versus Public Interest." Acta bioethica 27 (2021): pp. 59-67. Available: http://dx.doi.org/10.4067/S1726-569X2021000100059. Accessed April 6, 2023.
3. Finck, Michèle and Frank Pallas. "They who must not be identified—distinguishing personal from non-personal data under the GDPR." *International Data Privacy Law*, Volume 10, Issue 1 (2020): pp. 11–36. Available: https://doi.org/10.1093/idpl/ipz026. Accessed March 30, 2023.
4. Gstrein, O. J. & A. Beaulieu. "How to protect privacy in a datafied society? A presentation of multiple legal and conceptual approaches." *Philosophy & technology, 35 (1), 3* (2022). Available: https://doi.org/10.1007/s13347-022-00497-4. Accessed March 27, 2023.
5. Landwehr, Marvin, Alan Borning and Volker Wulf. "Problems With Surveillance Capitalism and Possible Alternatives for IT Infrastructure." *Information, Communication & Society, 26:1* (2023): pp. 70-85. Available: https://doi.org/10.1080/1369118X.2021.2014548. Accessed April 20, 2023.
6. Nguyen, Fiona Q. "The Standard for Biometric Data Protection." *Journal of Law & Cyber Warfare 7, no. 1* (2018): pp. 61–84. Available on: https://www.jstor.org/stable/26777963. Accessed May 10, 2023.
7. Politou, Eugenia, Efthimios Alepis and Constantinos Patsakis. "Forgetting Personal Data and Revoking Consent under the GDPR: Challenges and Proposed Solutions." Journal of Cybersecurity 4, no. 1 (January 1, 2018). Available on: https://doi.org/10.1093/cybsec/tyy001. Accessed April 8, 2023.
8. Robertson, Viktoria. "Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the era of Big Data." *Common Market Law Review 57* (2020): p. 181. Available on: https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2020006.pdf. Accessed May 11, 2023.
9. Robertson, Viktoria. "The Theory of Harm in the Bundeskartellamt's Facebook Decision." *Competition Policy International* (2019): pp. 2-3. Available on: https://www.competitionpolicyinternational.com/wp-content/uploads/2019/03/EU-News-Column-March-2019-Full-1.pdf. Accessed May 10, 2023.
10. Rustici, Chiara. "GDPR Profiling and Business Practice." *Computer Law Review International, Volume 19 issue 2* (2018): p.34. Available on: https://doi.org/10.9785/cri-2018-190203. Accessed May 10, 2023.
11. Tene, Omer and Jules Polonetsky. "Big Data for All: Privacy and User Control in the Age of Analytics." *Northwestern Journal of Technology and Intellectual Property* 11, no. 5 (April 2013): p. 251. Accessed April 10, 2023.
12. Vavra, Ashley Nicole. "The Right to Be Forgotten: An Archival Perspective." *The American Archivist 81, no. 1* (2018): pp. 100–111. Available on: https://www.jstor.org/stable/48618003. Accessed May 5, 2023.
13. Viljoen, Salome. "A Relational Theory of Data Governance." *Yale Law Journal, Forthcoming* (November 11, 2020). Available: http://dx.doi.org/10.2139/ssrn.3727562. Accessed March 19, 2023.

14. Villaronga, Eduard Fosch, Peter Kieseberg and Tiffany Li. "Humans Forget, Machines Remember: Artificial Intelligence and the Right To Be Forgotten." *Computer Law & Security Review* 34, 2 (2018): pp. 304-313. Available: https://www.sciencedirect.com/science/article/abs/pii/S0267364917302091. Accessed May 9, 2023.
15. Voss, W. Gregory. "European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting." *The Business Lawyer 72, no. 1* (2016): pp. 221–34. Available on: https://www.jstor.org/stable/26419118. Accessed May 5, 2023.
16. Welch, Liam. "Grave New World: Mass Surveillance and Labour Rights." *Socialist Lawyer, no. 83* (2019): pp. 36–41. Available on: https://doi.org/10.13169/socialistlawyer.83.0036. Accessed May 8, 2023.
17. Wiedemann, Klaus. "Profiling and (automated) decision-making under the GDPR: A two-step approach." *Computer Law & Security Review, Volume 45* (2022). Available: https://doi.org/10.1016/j.clsr.2022.105662. Accessed May 10, 2023.
18. Yeung, Karen. "Five fears about mass predictive personalization in an age of surveillance capitalism." *International Data Privacy Law, Volume 8, Issue 3* (2018): pp. 258–269. Available: https://doi.org/10.1093/idpl/ipy020. Accessed May 10, 2023.
19. Zuboff, Shoshana. "Big other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology 30* (2015): pp. 75-89. Available on: https://doi.org/10.1057/jit.2015.5. Accessed April 13, 2023.
20. Zuboff, Shoshana. "Surveillance Capitalism and the Challenge of Collective Action." *New Labor Forum,* vol. 28, no. 1 (2019): pp. 10-29. Available on: https://doi.org/10.1177/1095796018819461. Accessed April 15, 2023.

WEBSITES

1. Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets," *The University of Texas at Austin* (2008): p. 14. Available on: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf. Accessed April 10, 2023.
2. Bundeskartellamt, Bundeskartellamt prohibits Facebook from combining user data from different sources (2019). Available on: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html. Accessed May 11, 2023.
3. European Data Protection Board. Available on: https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party_en.
4. Richie Koch. "What is considered personal data under the EU GDPR?" Available: **https://gdpr.eu/eu-gdpr-personal-data/?cn-reloaded=1**. Accessed March 27, 2023.

LEGAL OPINIONS

1. Article 29 Data Protection Working Party. *Opinion 05/2014 on Anonymisation Techniques*. Available on: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. Accessed April 4, 2023.