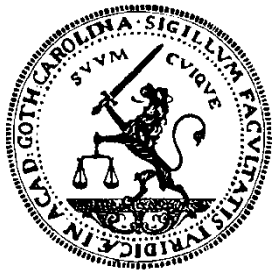

FACULTY OF LAW
Lund University

Hannes Westermann

Change of Purpose

The effects of the Purpose Limitation Principle in the General Data Protection Regulation on Big Data Profiling

JURM02 Graduate Thesis

Graduate Thesis, Master of Laws program
30 higher education credits

Supervisor: Vilhelm Persson

Semester of graduation: Period 1 Spring Semester 2018

Contents

Summary

Over the past few years, many companies have started to adopt Big Data technologies. Big Data is a method and technology that allows the collection and analysis of huge amounts of all kinds of data, mainly in digital form. Big Data can be used, for example, to create profiles of online shopping users to target ads. I call this Big Data Profiling. Facebook and Google, for example, are able to estimate attributes, such as gender, age and interests, from data provided by their users. This can be worrisome for many users who feel that their privacy is infringed when the Big Data Profiling companies, for example, are able to send advertisements to the users that are scarily relevant to them.

Big Data Profiling relies on a vast amount of collected data. Often, at the time of collection, it is not clear how exactly this data will be used and analyzed. The new possibilities with Big Data Profiling have led to companies collecting as much data as possible, and then later figuring out how to extract value from this data. This model can be described as "collect-before select", since the data is first collected, and then "mined" for correlations that can be used to profile users.

In this thesis I analyze whether this form of collection and usage of Personal Data is legal under the General Data Protection Regulation (GDPR), which enters into force in the European Union on 25 May 2018. While many of the provisions of the GDPR already existed in the Data Protection Directive (DPD) since 1995, they have been reinforced and extended in the GDPR.

One of the main principles of the GDPR is that of Purpose Limitation. While the principle already exists under the DPD in a very similar fashion, it is likely that it will be enforced more under the GDPR, since the GDPR is directly applicable in member states instead of having to be implemented. The enforcement mechanisms, such as sanctions, have also been significantly strengthened.

The Purpose Limitation Principle requires the data controller (such as companies processing Personal Data, like Facebook and Google) to have a specified purpose for and during the collection of Personal Data. Further, the Personal Data cannot be processed beyond this purpose after it has been collected. This seems to run contrary to Big Data Profiling, which regularly looks for purposes only after the Personal Data has been collected. However, I have identified three potential ways the "collect before select" model could still be possible under the GDPR.

The first possibility is the anonymization of Personal Data. If data can be efficiently anonymized, it will fall outside of the scope of the GDPR because it will not contain Personal Data after the anonymization. The controller is then free to analyze the data for any purpose, including creating

models that could be used to profile other users. However, I found that Big Data methods can often reidentify Personal Data that has been previously anonymized. In such cases even purportedly anonymized data may still fall under the scope of the GDPR. If on the other hand enough Personal Data is removed to make reidentification impossible, the value of the data for large parts of the business world is likely destroyed.

The second possibility is collecting Personal Data for a specified purpose that is defined so widely that it covers all potential future use cases. If a controller can collect Personal Data for a vague purpose, such as "marketing", the controller will have a lot of flexibility in using the data while still being covered by the initial purpose. I found that the GDPR requires data controllers (such as companies) to have a purpose for the data collection that is specific enough so that the data subject is able to determine exactly which kinds of processing the controller will undertake. Having a non-existent or too vague purpose is not sufficient under the GDPR. Companies that collect data with no, or an only vaguely defined, purpose and then try to find a specific purpose for the collected data later will therefore have to stop this practice.

The third possibility can be used if the controller wants to re-use Personal Data for further purposes, after the controller has collected the Personal Data initially in compliance with the GDPR for a specified purpose. In this case, the GDPR offers certain possibilities of further processing this data outside of the initial purpose. The GDPR allows this for example if the data subject has given consent to the new purpose. However, I found that Big Data Profiling companies often come up with new purposes later by "letting the data speak", which means by analyzing the data itself to find new purposes. Before performing an analysis, often the company might not even know how the processing will be done later. In that case, it is impossible to request the data subject's specific consent, which is required under the GDPR. Even without the data subject's consent, there are however other possibilities of further processing data under the GDPR, such as determining whether the new processing is compatible with the initial purpose. My thesis examines some of those possibilities for a change of purpose under Big Data Profiling.

My conclusion is that the GDPR likely means a drastic impact and limitation on Big Data Profiling as we know it. Personal Data cannot be collected without a purpose or with a vague purpose. Even Personal Data that was collected for a specific purpose cannot be re-used for another purpose except for in very few circumstances. Time will tell how the courts interpret the GDPR and decide different situations, how the companies will adapt to them and if the legislator will react to this reality.

Sammanfattning

Över de senaste åren har många företag börjat använda sig av ”Big Data”. Detta är en metod och teknik som tillåter insamling och analys av enorma datamängder i digital form. Big Data kan användas till exempel för att skapa detaljerade profiler av användare på internet för att anpassa reklamen de ser. Jag kallar detta för Big Data Profiling. Facebook och Google, till exempel, kan räkna ut egenskaper som kön, ålder och intressen av användare. Detta kan väcka oro hos användare som känner att deras privatsfär inkräktas om företag skickar överraskande relevanta erbjudanden till dem.

Big Data Profiling använder sig av enorma mänger av personuppgifter. När dessa uppgifter samlas in är det ofta inte klart exakt hur de kommer att användas. De nya möjligheterna har lett företag till att samla in så mycket personuppgifter som möjligt, och först i efterhand komma fram till hur man kan utvinna värde ur den. Denna modellen kan beskrivas som ”samling-innan-utval”, då datan först samlas och sedan analyseras för korrelationer som kan skapa profiler om användare.

I denna uppsats analyserar jag huruvida denna form av insamling och användning av personuppgifter är laglig under Dataskyddsförordningen (GDPR) som träder i kraft i den Europeiska Unionen den 25 maj 2018. Dataskyddsdirektivet (DPD), som varit i kraft sedan 1995, har till stor del samma innehåll, men bestämmelserna har utvecklats och utökats i GDPR.

En av de viktigaste principerna i GDPR är ändamålsbegränsningen. Principen finns redan i DPD, men det är troligt att den kommer att tillämpas till en större grad under GDPR, eftersom GDPR är en förordning och därför är direkt tillämplig i alla medlemsstater. Mekanismerna tillgängliga för upprätthållande av lagen, som sanktioner, har även utökats starkt.

Principen om ändamålsbegränsning innebär att den personuppgiftsansvarige (till exempel företag som behandlar personuppgifter, som Google och Facebook) ska ha ett särskilt syfte med insamlingen. Personuppgifter får inte heller behandlas för något som inte ryms inom detta syfte efter insamlingen. Denna begränsning är inte förenligt med Big Data Profiling, som ofta försöker hitta ett syfte för personuppgifter efter att den samlats in. Dock verkar det finnas tre möjligheter som ”samling-innan-utval” potentiellt ändå kan vara tillåtet under GDPR.

Den första möjligheten är anonymisering av personuppgifter. Om data kan anonymiseras på ett effektivt sätt, faller den utanför tillämpningsområdet för GDPR, då den inte längre innehåller personuppgifter efter anonymiseringen. Den personuppgiftsansvarige kan då analysera datan för vilket syfte som helst, till exempel för att skapa en modell som kan skapa profiler av andra användare. Dock kan individer ofta identifieras med hjälp av Big Data trots att data anonymiserats. Då är datan fortfarande personuppgifter som faller

under GDPR. Om tillräckligt med data tas bort för att en identifiering ska vara omöjlig går värdet av datan nog förlorat för en stor del av den kommersiella användningen.

Den andra möjligheten är att samla in personuppgifter för ett syfte som är så brett att det täcker alla potentiella användningar av personuppgifterna. Om den personuppgiftsansvarige kan samla in personuppgifter för ett väldigt vagt syfte, som ”marknadsföring”, kommer den personuppgiftsansvarige att ha mycket flexibilitet i hur personuppgifterna ska bearbetas och ändå fortfarande falla under det angivna syftet. Dock fann jag att GDPR kräver att den personuppgiftsansvarige har ett syfte som är specifikt nog att den registrerade (som t.ex. kunden) kan fastställa exakt hur personuppgifterna kommer att bearbetas. Att samla personuppgifter utan syfte eller med ett vagt sådant är inte acceptabelt under GDPR. Företag som samlar personuppgifter och sedan avgör hur de ska användas kommer därför att behöva upphöra med denna verksamhet.

Den tredje möjligheten till att hitta ett syfte för personuppgifter efter insamling kan utnyttjas om den personuppgiftsansvarige vill återanvända personuppgifter för ett nytt syfte efter att den redan samlat in dessa i enlighet med GDPR. I detta fall erbjuder GDPR vissa möjligheter att använda personuppgifter till ett annat syfte än de samlats in för. Till exempel tillåts återanvändning om den registrerade ger sitt samtycke för en behandling till ett nytt syfte. Dock fann jag att Big Data Profiling-företag ofta kommer fram till nya syften genom att ”låta datan tala”, vilket betyder att man vill analysera personuppgifterna själv för att hitta nya syften. Innan denna analys har gjorts vet företagen i dessa fall inte hur personuppgifterna kommer att bearbetas. I detta fall är det omöjligt att samla in en registrerads specifika samtycke, vilket krävs under GDPR. Det finns även vissa andra möjligheter att bearbeta personuppgifter till ett nytt syfte, till exempel att bearbeta dem på ett sätt som är kompatibelt med det initiala syftet. Min uppsats analyserar dessa möjligheter.

Min slutsats är att GDPR innebär en stor inskränkning av Big Data Profiling. Personuppgifter kan inte samlas in utan ett syfte, eller med ett vagt sådant. Personuppgifter som samlats in för ett särskilt syfte kan endast under väldigt få omständigheter återanvändas för ett nytt syfte. Tiden kommer att utvisa hur domstolarna kommer att tillämpa GDPR, hur företagen kommer att anpassa sig och om lagstiftaren kommer att reagera på denna verklighet.

Preface

The 25 May 2018 the General Data Protection Regulation enters into force. While this marks a new beginning for data protection in Europe, it also marks an end for my legal studies at the University of Lund. While I am sad that this wonderful time is coming to an end, I am also excited about what the future may bring. And I know that I will carry the wonderful memories, the friendships with the fantastical people that I have met and everything that I have learnt with me for the rest of my life.

I would like to say a huge thank you to Vilhelm Persson, my supervisor for this thesis. Without your poignant questions, insightful comments and constructive criticism, this essay would not be what it is today.

I would also like to thank my father Dirk Westermann, who through helpful discussions, long phone calls and valuable feedback helped me clarify my thoughts. A big thank you also goes to him and the rest of my family, Maria, Anneli and Helena, for being there for me over my entire studies. You guys are the best.

Lastly, I would like to thank everyone who made my time in Lund and Montréal to the amazing journey it has been. I will miss the cozy fikas, walking around and having fun in the occasionally sunny Lund, and the interesting conversations with all of you. Wherever our paths may take us I look forward to our visits and you can be sure that I will always have a couch for you, no matter where I am.

I hope you enjoy the reading!

Hannes Westermann
Lund, the 23 May 2018

Abbreviations

AI	Artificial Intelligence
BDP	Big Data Profiling
DPD	Data Protection Directive
ECJ	European Court of Justice
EU	European Union
GDPR	General Data Protection Regulation
ICO	Information Commissioner's Office
ML	Machine Learning
WP29	Article 29 Data Protection Working Party

1 Introduction

1.1 Background

In April 2018, it was revealed that the political consultancy firm Cambridge Analytica had systematically harvested the Personal Data of 87 million Facebook users.[1] Supposedly, Cambridge Analytica was able to use this and other data to generate sophisticated profiles of voters in the United States, including their personality traits. This data could then be used to precisely target political advertising to sway the opinions of individuals.[2] As far as we know, Cambridge Analytica used its technologies in the Trump presidential campaign and the Brexit-campaign in the United Kingdom.[3]

Cambridge Analytica is part of what I refer to as the Big Data Profiling industry. Here, millions of data points of different users are mined for correlations. One such correlation could be, for example, that users who like certain webpages are likely to fall into a certain age range. Using many such correlations, the Big Data Profiling company is able to create a sophisticated profile of attributes and interests of the individual and use this to display relevant ads on webpages to the person in question, to recommend items that the person might like to buy or people that they might know and want to connect with.

Big Data Profiling relies on two emerging technologies: (1) Big Data and (2) Artificial Intelligence. (1) Big Data strives to collect huge amounts of data and then perform analytics on this data to gain valuable insight into the real world.[4] Unlike previous data collection efforts, where data is collected to solve a specific problem or assess a specific metric, Big Data often goes in the opposite direction. First, huge quantities of data are collected. Only after this will data scientists perform analyses on the data, and thus arrive at conclusions that can, for example, improve business.[5] (2) Even more recently, Artificial Intelligence has become a trend in data analysis. It

[1] Issie Lapowsky, 'Facebook Exposed 87 Million Users to Cambridge Analytica' [2018] *WIRED* <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/> accessed 2 May 2018.
[2] Sasha Issenberg, 'Cruz-Connected Data Miner Aims to Get Inside U.S. Voters' Heads' *Bloomberg.com* (12 November 2015) <https://www.bloomberg.com/news/features/2015-11-12/is-the-republican-party-s-killer-data-app-for-real-> accessed 15 May 2018.
[3] Carole Cadwalladr, '"I Made Steve Bannon's Psychological Warfare Tool": Meet the Data War Whistleblower' *The Guardian* (18 March 2018) <http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> accessed 2 May 2018.
[4] Tal Z Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2016) 47 Seton Hall L. Rev. 995, 5–6.
[5] Istvan Borocz, 'Clash of Interests-Is Behaviour-Based Price Discrimination in Line with the GDPR' (2015) 153 Studia Iuridica Auctoritate Universitatis Pecs Publicata 37, 8.

autonomously creates sophisticated analysis and prediction models from huge amounts of data with minimal input from human engineers.[6]

Big Data Profiling promises to generate tremendously valuable and actionable insights and allows businesses to tailor the experience of the customer, for example as written above by displaying ads that are specifically geared towards the individual customer's interest.[7]

This potential value have prompted actors, especially on the internet, to collect a huge amount of data about large groups of users.[8] Every webpage visited, every link clicked and every search engine query is recorded by the companies in question.[9] This data, in combination with artificial intelligence, allows the companies to build sophisticated profiles of the users, and get to know large amounts of facts about them.[10] To many users, this seems worrying, and they see the collection of use of data about them as a breach of their personal integrity and privacy. There are also fears of Big Data Profiling being used to filter information displayed to the user on the internet, for example by showing only views and opinions that exactly correspond to the user's own views, thereby segregating different opinions from each other and preventing important discourse in society between people of different views.[11]

The European Union has taken notice, and in May 2018 the General Data Protection Regulation (GDPR) will become active in all European Union Member states. The GDPR is the successor of the EU's Data Protection Directive 95/46/EC (DPD) and was "designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organizations across the region approach data privacy".[12] While the DPD already contained many of the principles of the GDPR (including the Principle of Purpose Limitation), it suffered under varying implementation and enforcement levels in the different member states.[13] The

---

[6] 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (Information Commissioner's Office) 8 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 20 February 2018; Zarsky (n 4) 6.

[7] 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 16.

[8] Logan Kugler, 'The War over the Value of Personal Data' (2018) 61 Communications of the ACM 17, 1–2.

[9] 'Google Chrome Privacy Notice' (*Google Chrome*, 6 March 2013) <https://www.google.com/chrome/privacy/> accessed 2 May 2018.

[10] 'Browser Tracking' (*me and my shadow*, 16 February 2017) <https://myshadow.org/> accessed 2 May 2018.

[11] Lokke Moerel and Corien Prins, 'Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2784123 24 <https://papers.ssrn.com/abstract=2784123> accessed 27 March 2018.

[12] 'EU GDPR Information Portal' (*EU GDPR Portal*) <http://eugdpr.org/eugdpr.org.html> accessed 2 May 2018.

[13] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) 2016 (OJ L119/1) recital 9-11.

GDPR, on the other hand, is a regulation and thus directly applicable uniformly in all member states. It reinforces existing data protection mechanisms and introduces new ones, such as increased fines for violations.[14] The regulation places heavy restrictions on corporations collecting Personal Data. Personal Data can only be collected to fulfill a certain purpose (Purpose Limitation). Further, this data cannot usually be used for another purpose after collection.[15] My thesis examines if and to which extent these rules cause problems for Big Data Profiling companies, which rely on collecting millions of data points and performing exploratory analyses later in the process.[16]

1.2 Purpose & Problem

In this thesis, I will specifically investigate how the *Purpose Limitation Principle* of the GDPR will impact the practice of Big Data Profiling. For this reason, I will analyze the following questions:

- Does the data typically used in Big Data Profiling fall under the GDPR?
- Can the Personal Data be anonymized to enable unhindered processing?
- To which extent does the purpose of collection have to be specified at the time of collection?
- Can the purpose of the processing be altered after collection?

1.3 Delimitations

In this essay, I will specifically determine how the aforementioned concept relates to Big Data Profiling. I will not deal with the numerous other requirements imposed by the GDPR on data controllers and processors.

For example, I will not go in-depth on the requirement for a legal basis for the processing of Personal Data, such as consent, under article 6 of the GDPR. While important and interesting, the requirement for a legal basis is separate and independent from the requirement for Purpose Limitation. I have, however, described the requirement for a legal basis to the extent necessary to understand the Purpose Limitation Principle.

I will also not write about the specific provisions relating to automated decision making, which can be found in Article 22 of the GDPR. Article 22 gives the data subject the right to avoid being subject to certain automated decisions. These situations are those where the decisions produce legal or

[14] Paul Voigt and Axel von dem Bussche, *The EU General Data Protection Regulation (GDPR)* (Springer International Publishing 2017) 2 <http://link.springer.com/10.1007/978-3-319-57959-7> accessed 20 February 2018.
[15] General Data Protection Regulation (n 13) article 5.1(b).
[16] Zarsky (n 4) 1006.

similar effects on the individual. While these provisions are interesting, I believe them to not yet be as broadly applicable as the Purpose Limitation Principle, as least today. Profiling affects almost everyone using the internet, but only few people are today affected by automated decisions producing legal effects. Advertising, for example, likely does not produce effects significant enough to fall under the prohibition.[17] The provisions have also been found to only apply to certain conditions, as well as being potentially easy to circumvent.[18]

I will neither analyze the different duties between controllers and processors, nor what happens when data leaves the controller's sphere, such as in a sale, or when a controller purchases data about an individual from an external source. While these situations can have some connection to Big Data Profiling, they carry their own different sets of issues that would be outside of the scope of this essay to analyze.

1.4 Method, Material & State of Research

To write this thesis, I have used a number of sources from different fields. To determine the definition of Big Data Profiling and the way it is used in practice, I have looked to legal doctrine in the area of data protection and several online articles. Since there is considerable hype surrounding Big Data and some entities might exaggerate or understate their capabilities, I have tried to use a critical perspective on these sources. The views on what constitutes Big Data are also divergent. Through comparing different sources, I have attempted to gain an understanding and convey this area.

In order to determine the meaning of the legislation, I have - besides the legislation itself - used several sources throughout the legal doctrine. Due to the novelty of the GDPR, many of the sources I have used are about the DPD, to the extent that they still apply. I found especially the opinions of the Article 29 Working Party helpful. This working party was instituted under article 29 of the DPD and has an advisory role on the implementation of the provisions of the DPD.[19] Since the regulation is not yet implemented, there are no court cases, and the number of relevant court cases concerning the DPD is also limited.

There has also been a large amount of interest and research in specifically the application of the GDPR to Big Data analysis. For a critical perspective, see Zarsky, who criticizes the GDPR as being incompatible with the age of

[17] Voigt and von dem Bussche (n 14) 182.
[18] Zarsky (n 4) 1015–1018.
[19] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data 1995 (OJ L281/31) article 29-30.

Big Data.[20] Mayor-Schonberger and Padova are more optimistic and see the GDPR as enabling Big Data.[21] The Information Commissioner's Office (ICO) of the UK has also posted a notable discussion paper on the treatment of Big Data under the GDPR.[22] Nikolaus Forgó et al analyse the application of the Purpose Limitation Principle to Big Data in general.[23]

I aim to distinguish my thesis from these previous works by providing a focused and in-depth analysis of specifically the effect of the Purpose Limitation Principle on Big Data Profiling. I will accomplish this by first determining the defining features of Big Data Profiling. I will then provide an overview of the GDPR, and then focus on explaining the relevant provisions more in-depth. Finally, I will use the presented material to determine the way these provisions should be applied to Big Data Profiling.

The method I will use is thus the legal dogmatic method (rättsdogmatisk metod) which focuses on interpreting legal sources in order to solve a legal problem.[24] First, the legal sources, such as legislation, case law and legal doctrine are analyzed and weighed against each other to determine the legal rule. This is then interpreted and applied to a specific legal issue in order to find how the rule should be applied in a certain case.[25] The legal dogmatic method is used not only by legal scholars, but also by judges and lawyers in order to solve practical cases.[26] Since my aim with this thesis is to determine how the principle of Purpose Limitation affects Big Data Profiling, I believe this is the most well-suited method for my thesis.

1.5 Outline

In the second chapter, I will explain what Big Data Profiling is and how it relates to the concepts of Big Data and Artificial Intelligence. This is the factual situation that I will later apply the legislation to.

In the third chapter, I will briefly present the General Data Protection Regulation. I will focus on the aspects that will be important for the analysis. This will put my later analysis in context and explain how the principle of Purpose Limitation relates to other principles in the GDPR.

---

[20] Zarsky (n 4).
[21] Viktor Mayer-Schonberger and Yann Padova, 'Regime Change: Enabling Big Data through Europe's New Data Protection Regulation' (2015) 17 Colum. Sci. & Tech. L. Rev. 315.
[22] 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6).
[23] Nikolaus Forgó, Stefanie Hänold and Benjamin Schütze, 'The Principle of Purpose Limitation and Big Data', *New Technology, Big Data and the Law* (Springer, Singapore 2017) <https://link.springer.com/chapter/10.1007/978-981-10-5038-1_2> accessed 15 May 2018.
[24] Fredric Korling, *Juridisk metodlära* (Mauro Zamboni ed, Studentlitteratur AB 2013) 21.
[25] ibid 28–29.
[26] ibid 42.

In the fourth chapter, I will present the legislation concerning the definition of Personal Data. I will then assess whether the data typically used in Big Data Profiling is likely to fall under this definition. I will also analyze whether anonymization of data is feasible as a tool for further use and how Big Data affects this. This will set the stage for which data is included in my analysis of Purpose Limitation, and whether anonymization can be used as a tool for unlimited reuse of data.

The fifth chapter will describe and analyze the core principle that is subject to this thesis, namely the Purpose Limitation Principle.

I will start with the principle's history, in order to convey its significance and set the stage for the analysis over the next two sections.

Then, I will describe the first part of the Purpose Limitation Principle, namely Purpose Specification (5.1). It requires controllers to have a specified, explicit and legitimate purpose for the processing of Personal Data. This section will explore whether it is possible to specify a purpose that is sufficiently broad that the controller can process the data in different and new ways, while still remaining within the same purpose.

In the next section, I will analyze the change of purpose of Personal Data processing at a later stage (5.2), namely once the Personal Data has already been collected under a specified purpose. This chapter will deal with whether it is possible for the data controller to use Personal Data it has already collected for another purpose than initially planned. The GDPR offers some avenues of reuse of collected Personal Data and I will analyze whether this can be used in a Big Data Profiling context.

Finally, I will summarize my findings and briefly explain which effect I think the GDPR will have on companies using Big Data Profiling.

2 Big Data Profiling

This section will provide an overview over Big Data Profiling. First, it will define Profiling (2.1), and then elaborate how it uses Big Data and Machine Learning to create stunningly accurate profiles (2.2). It will explain the Big Data Profiling process (2.3) and some salient examples of Big Data Profiling (2.4), as well as highlight some risks (2.6). It will then determine the defining features of Big Data Profiling that will later be used in the analysis (2.7). This chapter will set the factual basis that the GDPR will be applied to in later chapters.

2.1 Profiling definition

Profiling, as defined by the Article 4(4) GDPR, means:

"any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;"[27]

Companies use profiling to observe actions of a user over time to build a profile of this user, containing information such as interests, knowledge, background, skills, goals and likely future behavior.[28] They can then use this profile to target advertisements to the user or tailor his experience in some other way.[29] This phenomenon is also known as behavioral advertising.[30]

2.2 Big Data & Machine Learning

It appears that two core technologies have developed over the past few years that have helped turning profiling into a billion-dollar business, and which I will further analyze in the following sections: (2.2.1) Big Data techniques, that are able to collect and analyze vast amounts of data, thus making the profiles more accurate and comprehensive.[31] (2.2.2) Artificial Intelligence, that is able to automate the discovery of correlations in the data, which

[27] General Data Protection Regulation (n 13) article 4(4).
[28] Omar Hasan and others, 'A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case' (IEEE 2013) 2 <http://ieeexplore.ieee.org/document/6597115/> accessed 9 May 2018.
[29] 'Opinion 2/2010 on Online Behavioural Advertising' (Article 29 Data Protection Working Party 2010) WP171 4.
[30] ibid.
[31] Hasan and others (n 28) 1.

increases the accuracy of the profiles and allows for a surprising level of detail in inferred data.[32]

2.2.1 Big Data

The most popular definition of Big Data stems from the Gartner IT glossary:

"high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."[33]

This definition looks at the amount, quality and efficiency of the data processing. Another approach would be to define Big Data by differentiating it from normal data processing. A useful distinction seems to be the fact that Big Data has enough data-points to enable a new perspective on the reality that it tries to describe.[34] Instead of analyzing *why* something is happening, Big Data limits itself to analyzing *that* it is happening. Big Data can therefore be used even in situations where the analyst does not understand the mechanisms governing a specific occurrence.[35] These shifts allows the data to "speak for itself" – no hypothesis of the connection between factors is required. For example, Amazon is able to determine the links between user interactions and likeliness of purchasing another product simply based upon the sheer amount of available statistical data, without having to conduct surveys or manually grouping products together.[36]

Instead of data being collected for, and tied to, a singular purpose, all data is under the new model likely to contain some hidden value.[37] As many have said, "data is the new oil".[38] This has led to companies capturing as much data as possible, in the hope of deriving future value and insight.[39] Instead of determining a problem first, and then looking for data points to solve it, Big Data enables firms to do the opposite – collecting data first, without any

[32] Josh Sutton, 'Demystifying the Role of Artificial Intelligence in Marketing and Advertising' (*eMarketer*, 22 April 2016) <https://www.emarketer.com/Article/Demystifying-Role-of-Artificial-Intelligence-Marketing-Advertising/1013864> accessed 11 May 2018.
[33] 'What Is Big Data?' (*Gartner IT Glossary*) <https://www.gartner.com/it-glossary/big-data> accessed 21 February 2018.
[34] Mayer-Schonberger and Padova (n 21) 5.
[35] Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight* (New and expanded edition, John Murray 2017) 18–19.
[36] ibid 20.
[37] ibid 21.
[38] Perry Rotella, 'Is Data The New Oil?' [2012] *Forbes* <https://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/> accessed 21 February 2018.
[39] Maria Helen Murphy, 'Algorithmic Surveillance: The Collection Conundrum' (2017) 31 International Review of Law, Computers & Technology 225.

specific purpose, and then finding one or several uses for the data.[40] This can be referred to as "collect before select", since the data is first collected, and then a purpose is selected.[41] Alternatively, data that is collected for one purpose can be repurposed and used to serve a different purpose.[42]

2.2.2 Machine Learning

The quantities of data collected in a Big Data context are enormous. Using traditional tools for processing of data, such as queries for specific data points, would struggle to make full use of the data and require a lot of time and manpower.[43] However, a new set of tools that improve and partially automate the analysis of data, known as "Machine Learning", has come into the spotlight.[44] Machine Learning is the act of a computer, creating a model or algorithm from data that it can later use to predict unknown cases.[45] Recently, "Deep Learning" has emerged as an even more sophisticated model of Machine Learning, being able to detect structures in unstructured data such as sound-, text- and image-files. Since Big Data often consists of a vast amount of unstructured data sets, Deep Learning is therefore especially well-suited for the analysis of Big Data.[46]

Machine Learning can therefore be seen as a tool for unlocking the value of collected Big Data.[47] Instead of manually sifting through the data to detect correlations and patters, engineers can set a few parameters and then let the computer create a model for the correlation between different factors.[48]

Big Data and Machine Learning seem like important tools in Profiling. The following example may illustrate how it changes the dynamic in a Profiling operation: Under the traditional model of data processing, a company specialist would decide that a user reading about certain shoes is more likely to be interested in certain additional other shoes, and then start tracking such information coming from the user's behavior of reading certain web pages. In a Big Data Profiling operation however, the data about which websites

[40] Mayer-Schonberger and Padova (n 21) 319; Borocz (n 5) 43.
[41] Borocz (n 5) 8.
[42] Mayer-Schonberger and Padova (n 21) 319.
[43] Zarsky (n 4) 1001; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 6.
[44] 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 7.
[45] Bernard Marr, 'What Is The Difference Between Artificial Intelligence And Machine Learning?' [2016] *Forbes* <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/> accessed 22 February 2018.
[46] Maryam M Najafabadi and others, 'Deep Learning Applications and Challenges in Big Data Analytics' (2015) 2 Journal of Big Data 1.
[47] 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 7.
[48] Merle Temme, 'Algorithms and Transparency in View of the New General Data Protection Regulation' (2017) 3 European Data Protection Law Review 473, 3; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 7.

users have visited has already been collected. The specialist would then let algorithms, such as Machine Learning, figure out the correlations in the collected data. The algorithms would realize that users that are interested in one kind of shoes are generally also interested in certain other kinds of shoes. User that are reading about these shoes would then receive ads for other kinds of shoes. The same data could also be used to infer other attributes about the user.

2.3 The process of Profiling

The typical Big Data Profiling system contains three stages: (1) Data collection, (2) data mining and (3) decision making.[49] The steps all have different data protection implications.

(1) During the data *collection* stage, the company collects as much data as possible from a variety of sources about the individual. This can be data that is provided by the individual himself (such as by uploading an image), data that is recorded by the company (such as the location of the user if he has GPS enabled on his phone) or data obtained from a third party (such as personal information bought from a third party).[50]

A common way of collecting data on the internet is done via "tracking", e.g. with the help of so-called "cookies". Cookies are pieces of information that a website places in a user's web browser when he accesses it. The next time the user visits the same or another website containing a tracking code, the website will recognize the cookie and therefore know that it is the same user. This can enable advertisers to track the same user on different websites and record which kinds of websites he visits.[51] A survey conducted of 144 million websites showed that 75% of the websites included some sort of tracking. Many of the websites do not only feature one tracker, but several: The Daily Mail, for example, was found to send over 19,000 cookies to users that visited their webpage.[52]

Another common way of tracking users is via "loyalty cards". A loyalty card is a customer card that gives discounts or other rewards if a customer registers the card during a purchase. They are commonly used, for example,

[49] Dimitra Kamarinou, Christopher Millard and Jatinder Singh, 'Machine Learning with Personal Data' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2865811 8 <https://papers.ssrn.com/abstract=2865811> accessed 1 March 2018; 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (Article 29 Data Protection Working Party) WP251rev.01 7 <http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826> accessed 28 February 2018.
[50] Frederik Zuiderveen Borgesius and Joost Poort, 'Online Price Discrimination and EU Data Privacy Law' (2017) 40 Journal of Consumer Policy 347.
[51] Jessica Davies, 'Know Your Cookies: A Guide to Internet Ad Trackers' (*Digiday*, 1 November 2017) <https://digiday.com/media/know-cookies-guide-internet-ad-trackers/> accessed 22 February 2018.
[52] ibid.

by grocery stores and airlines. When a customer buys a product using a loyalty card, the purchase is recorded in that user's profile.[53]

(2) The second step of profiling is "*data mining".* During this step, algorithms are used to find correlations in the collected data. For example, an algorithm might identify the following correlation: "A user who looks at Product A is likely to also look at Product B." or "A male user who likes the page of Britney Spears is more likely to be homosexual." While these conclusions are just correlations in the individual case, combining them allows to create a comprehensive profile of a user. I will refer to the combination of these correlations as "models". Researchers at the university of Cambridge were able to extract personality traits by correlating "likes" on Facebook, i.e. the clicking of a "like" button on Facebook-pages that a user agrees to or enjoys, with quizzes that the users had answered on other Facebook pages. This system was able to predict, for example, the political party affiliation of a user with 85% accuracy, based on an average of 68 likes that the user had clicked on Facebook. The system was also able to predict users' skin color with an accuracy of 95%.[54]

(3) The final step of the profiling process is *decision making*. Here, the generated model or algorithm is applied to infer attributes of the user. These attributes, for example the users' interests or likely behavior[55], can then be used to show the users specifically tailored ads[56], or suggest people to connect with that the user is likely to know.[57] It can also be used to show content the user might be interested in, such as movies.[58]

2.4 Examples of Big Data Profiling

There are many examples of companies that use Big Data Profiling in their businesses, for example the following noteworthy ones:

One example is the discount store retailer "Target" in the United States. Target assigns an ID to each of their customers, for example by correlating

---

[53] Paul Michael and Wise Bread, '8 Ways Retailers Are Tracking Your Every Move' *Time.com* (23 September 2016) <http://time.com/money/4506297/how-retailers-track-you/> accessed 2 May 2018.
[54] Mikael Krogerus and Hannes Grassegger, 'The Data That Turned the World Upside Down' [2017] *Vice Motherboard* <https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win> accessed 14 March 2018.
[55] 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 49) 7; Kamarinou, Millard and Singh (n 49) 11.
[56] 'Opinion 2/2010 on Online Behavioural Advertising' (n 29) 4.
[57] Kashmir Hill, 'How Facebook Figures Out Everyone You've Ever Met' (*Gizmodo*, 11 July 2017) <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691> accessed 22 February 2018.
[58] Libby Plummer, 'This Is How Netflix's Top-Secret Recommendation System Works' [2017] *WIRED UK* <http://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like> accessed 13 May 2018.

their use of credit cards or loyalty cards. Each purchase is then tracked and added to the customer's dataset in Target's database. This data then allows Target to predict future purchase patterns, based on previous purchases, and allows Target to send out ads for products that the user is likely to want. In one case, Target sent out ads for diapers and other baby related products to a pregnant girl that lived with her parents, before she even had informed her parents of her pregnancy. The only thing needed to infer the pregnancy of the woman were her purchasing patterns.[59]

Another example is Google: Google places trackers on many web sites, which allows it to create a profile about users and target ads.[60] It also uses information collected by many of its own services to make ads more relevant. This includes searches, location information, websites visited and apps used.[61] A survey showed that 46.4% of 144 million surveyed websites contained trackers by Google.[62] During the first quarter of 2018, the company earned over 26 billion USD from advertisements.[63]

Facebook runs a similar tracking network and even creates "shadow" profiles for people that do not have a Facebook-account. [64] Each user in the US and Canada brought Facebook an average of 26.76 USD of revenue during the fourth quarter of 2017.[65]

There are also businesses, such as BlueKai, whose only activity is to profile users and sell this data. They are known as data brokers.[66]

Recently, the press has highlighted the use of Big Data Profiling in political elections. Allegedly, the corporation "Cambridge Analytica" was able to leverage profiling insights to affect the U.S. elections in 2016.[67] Cambridge Analytica obtained detailed user data of about 87 million user accounts from

---

[59] Kashmir Hill, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did' [2012] *Forbes* <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> accessed 22 February 2018.
[60] Olivia Solon, 'Google's Ad Tracking Is as Creepy as Facebook's. Here's How to Disable It' *the Guardian* (San Fransisco, 21 October 2016) <http://www.theguardian.com/technology/2016/oct/21/how-to-disable-google-ad-tracking-gmail-youtube-browser-history> accessed 22 February 2018.
[61] 'How Google Uses Your Data for Ads' (*Google Ads*) <http://privacy.google.com/intl/en-GB_ALL/how-ads-work.html> accessed 14 March 2018.
[62] Rahul Chadha, 'Ad Trackers Are on More than 75% of Websites' (*eMarketer*, 8 January 2018) <https://www.emarketer.com/content/ad-trackers-are-on-more-than-75-of-web-pages> accessed 22 February 2018.
[63] Jillian D'Onfro, 'Alphabet Earnings Q1 2018' (*CNBC*, 23 April 2018) <https://www.cnbc.com/2018/04/23/alphabet-earnings-q1-2018.html> accessed 3 May 2018.
[64] Hill (n 57).
[65] Anita Balakrishnan, 'Facebook Earnings Q4 2017: ARPU' (*CNBC*, 31 January 2018) <https://www.cnbc.com/2018/01/31/facebook-earnings-q4-2017-arpu.html> accessed 3 May 2018.
[66] Alexandra Suich, 'Getting to Know You' [2014] *The Economist* <https://www.economist.com/news/special-report/21615871-everything-people-do-online-avidly-followed-advertisers-and-third-party> accessed 22 February 2018.
[67] Cadwalladr (n 3).

Facebook.[68] This information was then used to create detailed profiles of users, enabling Cambridge Analytica to target ads designed to appeal to certain personality types at people that appeared being likely to change their minds about who to vote for in the presidential elections.[69]

Profiles could also be used to assess credit worthiness[70] or to detect and provide medical and psychological help to users that are at risk of self-harm and suicide.[71]

2.5 Other uses of Big Data

It is also possible to use Big Data in a more general way. For example, by analyzing Personal Data of individuals, companies are able to identify general market trends or aggregate interests of entire groups of people.[72] This kind of analysis can be seen more as statistics. The results are not applied back to individual users but can be used to influence strategic decisions of the company. While this is an important use of Big Data, this essay is focused on Big Data Profiling, where the results of the analysis are used to reach decisions regarding individual users. This distinction will become relevant in chapter 5.2.4.

2.6 Risks of Big Data Profiling

There are many risks that can come with Big Data Profiling, some of which I will describe here.

Perhaps the biggest risk is the infringement of privacy that Big Data Profiling can enable. The right to privacy is protected in Article 12 of the Universal Declaration on Human Rights.[73] It is commonly defined as the right to control one's Personal Data. An individual thus has the right to choose when, how and to whom he wants to reveal information about

---

[68] Lapowsky (n 1).
[69] Carole Cadwalladr and Emma Graham-Harrison, 'How Cambridge Analytica Turned Facebook "Likes" into a Lucrative Political Tool' *The Guardian* (17 March 2018) <http://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm> accessed 2 May 2018.
[70] Mikella Hurley and Julius Adebayo, 'CREDIT SCORING IN THE ERA OF BIG DATA' (2016) 18 Big Data 70.
[71] Diana Kwon, 'Can Facebook's Machine-Learning Algorithms Accurately Predict Suicide?' [2017] *Scientific American* <https://www.scientificamerican.com/article/can-facebooks-machine-learning-algorithms-accurately-predict-suicide/> accessed 15 March 2018.
[72] Bernard Marr, 'How To Use Analytics To Identify Trends In Your Market' [2016] *Forbes* <https://www.forbes.com/sites/bernardmarr/2016/08/16/how-to-use-analytics-to-identify-trends-in-your-market/> accessed 2 May 2018.
[73] Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A(III) (UDHR).

himself.[74] For example, an individual might want to reveal health information to a physician, but not to their employer.[75] Big Data Profiling, however, is able to predict attributes of individuals by creating profiles based on their behavior. This challenges the individuals' right to self-control of information, since the companies engaging in Big Data Profiling can deduce information without the consent or even knowledge of the individual.[76] An example of this is Target using loyalty cards to predict the pregnancy status of a girl and sending advertising for pregnancy products to her parents' address. The girl had not revealed her pregnancy to Target or her parents. The Big Data Profiling used by Target took this choice from her by predicting her pregnancy, and then revealing it to her parents through the ads. This was a violation of her right to control her information.

Another risk of Big Data Profiling is that the result might be discriminatory. An algorithm tries to generate a model based upon previous data. However, this data might contain some biases, leading to the algorithm inheriting these biases. Due to the opacity of algorithms, this might be hard to detect.[77] For example, Google was found to show ads for more high-paying jobs to male than to female users.[78]

Big Data Profiling can also create so called "filter bubbles". Algorithms designed to engage users to certain activities are more likely to show the users news and information that they already agree with and hide information that they dislike. The result is the creation of a filer bubble, meaning that users only receive and see information and news that correspond with their own worldview, for example on their Facebook feed. This could result in compromising people's ability to form an independent and comprehensive view on politics, news and social aspects, preventing their independent opinion-making and thus be a threat to democracy.[79]

[74] Bart Custers, 'Predicting Data That People Refuse to Disclose' 2012 Privacy Observatory Magazine 2.
[75] ibid 1.
[76] ibid 2.
[77] Temme (n 48) 2, 5; Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11 Northwestern Journal of Technology and Intellectual Property 38, 254.
[78] Amit Datta, Michael Carl Tschantz and Anupam Datta, 'Automated Experiments on Ad Privacy Settings' (2015) 2015 Proceedings on Privacy Enhancing Technologies 92, 14.
[79] Jacob Weisberg, 'Bubble Trouble' [2011] *Slate* <http://www.slate.com/articles/news_and_politics/the_big_idea/2011/06/bubble_trouble.html> accessed 4 May 2018; Tene and Polonetsky (n 77) 252; Martin Degeling and Thomas Herrmann, 'Your Interests According to Google - A Profile-Centered Analysis for Obfuscation of Online Tracking Profiles' [2016] arXiv:1601.06371 [cs] 2 <http://arxiv.org/abs/1601.06371> accessed 17 May 2018.

2.7 Defining features

Based on the results above, I will try to determine the important features of Big Data Profiling. These are the features that I will use during my analysis to determine how the GDPR affects Big Data Profiling.

The first defining feature is the *latent value of data*. Instead of all data having a purpose at the time of collection, in Big Data Profiling it is common to collect as much data about an individual as possible, and then mine this data to identify correlations. One piece of data is not bound to a specific purpose either, but is likely to be used in many different models, aimed to infer different attributes and likely behaviors of the user.

The second defining feature is the *volume of the data*. Data storage and collection is expensive. Previously, only the data that seemed relevant was kept by a company. However, due to the belief that all data is potentially valuable in Big Data Profiling, the volume of the worldwide data collection and storage has exploded. Almost every mouse click is tracked, collected and stored in massive data centers.

The third defining feature for Big Data Profiling is the *sophistication of the analysis.* Previously, in order to analyze and create statistics from data, data scientists had to specifically create queries to look for a specific correlation or statistic. With Big Data and Machine Learning, this process has been partially automated and become a lot more efficient. Instead of building the models for correlations themselves, the data scientists can give the computer a few parameters and set it to work to find these correlations instead.

These new computer tools allow the market players to easily build sophisticated profiles of people and use them, for example, to target ads for products or political advertisements.

2.8 Conclusion

Big Data Profiling carries tremendous potential for targeted advertising. It has also resulted in an important shift in the way data is collected and analyzed at companies. Instead of collecting specific data points to solve a specific predefined problem, the companies can let computers extract correlations and models from vast amounts of purposelessly collected data. The recent developments in Machine Learning further increase these capabilities.

However, from the individual perspective of a consumer and person, Big Data Profiling can be negative. It gives Big Data Profiling companies the possibility to closely get to know the individuals in question and trying to manipulate them, for example by directing tailormade commercials to them

in order to get their money or their vote. Using Big Data Profiling, this can be done more efficiently and targeted than ever before. Moreover, the individual is often not even aware of how its Personal Data is used and who has access to it. This is where data protection regulation, which I will discuss in the next chapter, comes in.

3 The General Data Protection Regulation

In this section, I will describe the General Data Protection Regulation (GDPR).[80] First, I will briefly describe the history of the legislation (3.1). Then, I will explain some provisions that are related to and may have impact on Big Data Profiling. I will start by explaining the material (3.2) and geographical scope of the legislation (3.3), to determine if the large Big Data Profiling companies, such as Google and Facebook, will fall under the GDPR. Next, I will explain the three main roles in the GDPR (3.4). Then I will give an overview of the general principles the GDPR mandates for all processing (3.5) and describe how a legal basis is necessary for the processing of data (3.6). I will explain how these requirements are enforced (3.8), and finally at which stage of Big Data Profiling the different requirements become active (3.9). This will help contextualize the analysis in the following chapters within the system of the GDPR.

3.1 History and purpose

In 1995, the data protection levels of different states were very varied, making it hard for corporations to operate over the border and hard for individuals to evaluate how their data might be used.[81] In order to facilitate the transfer of information between different member states and cross-border activities, the European Community (now European Union, EU) introduced a directive[82] to harmonize the legislation across different member states. However, the harmonization efforts failed as member states implementations differed greatly,[83] resulting in obstacles for cross-border data processing.[84]

Since the directive of 1995, globalization and technological developments have led to a substantial increase of the processing of Personal Data.[85] In response to these developments, in 2016 the European Union adopted the General Data Protection Regulation, which is set to enter into force the 25 May 2018.[86] Unlike the directive of 1995 that needed to be implemented by each member state, the GDPR is a regulation and thus applies directly in all

[80] General Data Protection Regulation (n 13).
[81] Voigt and von dem Bussche (n 14) 2.
[82] Data Protection Directive (n 19).
[83] Voigt and von dem Bussche (n 14) 2.
[84] General Data Protection Regulation (n 13) recital 9.
[85] ibid recital 6; Ernst, 'DS-GVO Art. 1 Gegenstand Und Ziele' in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 10; General Data Protection Regulation (n 13) recital 5.
[86] General Data Protection Regulation (n 13) article 99.2.

member states.[87] The regulation not only increases fines and enforcement mechanisms compared to the directive, but also introduces new data protection rules.[88]

3.2 Material Scope

According to Article 2 of the GDPR, the regulation applies:

"to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system."[89]

This scope is interpreted in a very broad manner.[90] It depends upon two important definitions: Personal Data and Processing.

Personal Data

Personal Data means information that is relating to an identified or identifiable natural person.[91] I will describe this in-depth in chapter 4.

Processing

The definition of processing can be found in Article 4(2). It reads:

"'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;"[92]

This definition is designed to make the legislation independent from technological change.[93] It covers as good as all uses of Personal Data, including collection and storage.[94] Even short-term uses of small amounts of

[87] Voigt and von dem Bussche (n 14) 2; Monika Wendleby and Dag Wetterberg, *Dataskyddsförordningen GDPR: förstå och tillämpa i praktiken* (Första upplagan, Sanoma Utbildning 2018) 12.
[88] Voigt and von dem Bussche (n 14) 2; Wendleby and Wetterberg (n 87) 27–28.
[89] General Data Protection Regulation (n 13) article 2.1.
[90] Voigt and von dem Bussche (n 14) 9.
[91] General Data Protection Regulation (n 13) article 4(1).
[92] ibid article 4(2).
[93] ibid recital 15; Ernst, 'DS-GVO Art. 2 Sachlicher Anwendungsbereich' in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 5.
[94] Ernst, 'DS-GVO Art. 2 Sachlicher Anwendungsbereich' (n 93) Rn. 4; Wendleby and Wetterberg (n 87) 44.

Personal Data, such as storing it in a cache, are covered.[95] Also manual processing can be covered in certain cases, mostly if the data is part of an organized collection.[96]

Due to this very broad definition, I believe that all three previously described stages of Big Data Profiling (data collection, data mining and decision making) will fall under the GDPR's definition of processing. They are all operations performed on Personal Data. During data collection, Personal Data is collected from different sources. During data mining, the Personal Data is analyzed in order to build models that are able to predict attributes of individuals. Even this creation, before it is applied to any individual, is therefore processing. Finally, the decision making generates new Personal Data about individuals. It is also processing, as it uses personal information of the individuals to create new Personal Data.

3.3 Geographical scope

According to Article 3 of the GDPR, the regulation applies to processing of Personal Data if the controller or processor (see below) is established within the European Union.[97] However, even if the controller is outside of the European Union, processing can fall under the regulation if the company offers goods or services to people in the EU, or monitors people in the EU.[98] This includes profiling of individuals.[99] Most large corporations, such as Google etc., which work with Big Data Profiling and have users in the European Union, will therefore fall under the regulation.

3.4 Roles

The three main roles in the data protection context are the data controller, the data processor and the data subject.

The data controller is the entity that initiates the processing of the Personal Data. He alone or together with others determines the means and purpose of

[95] Voigt and von dem Bussche (n 14) 10.
[96] ibid 10–11; Ernst, 'DS-GVO Art. 2 Sachlicher Anwendungsbereich' (n 93) Rn. 8.
[97] General Data Protection Regulation (n 13) article 3.1; Ernst, 'DS-GVO Art. 3 Räumlicher Anwendungsbereich' in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 3-4.
[98] General Data Protection Regulation (n 13) article 3.2; Ernst, 'DS-GVO Art. 3 Räumlicher Anwendungsbereich' (n 97) Rn. 15-20.
[99] General Data Protection Regulation (n 13) recital 24; Voigt and von dem Bussche (n 14) 27; Ernst, 'DS-GVO Art. 3 Räumlicher Anwendungsbereich' (n 97) Rn. 19-20.

the processing.[100] It is the entity mainly tasked with conforming to the obligations under the GDPR.[101]

In some cases, the controller might want to delegate some of the processing to another entity, for example by contracting another company for services within the processing of process Personal Data. This entity is known as the processor. [102] While the main burden of complying falls on the controller, the processor also has responsibilities, such as having guarantees for safeguards and being bound by a contract to the instructions of the controller.[103]

The main beneficiaries of the GDPR are the data subjects, that is, identified or identifiable natural persons.[104] Data subjects have a large number of rights under the GDPR,[105] including the right to be informed about the processing activities by the controller,[106] the right to rectify or delete data,[107] and the right to not be subject to automated decision making in some circumstances.[108]

Google, Facebook and other companies that process Personal Data for their own and other companies' benefit can be found both in the roles of controllers and processors, depending on if they process the data for the use in their own business, or process data as a service for other companies. Their customers and other people involved whose Personal Data is processed qualify as data subjects under the GDPR.

3.5 General Principles

While the GDPR contains a lot of specific rules, Article 5 lists the general principles that apply to the processing of Personal Data. Many of the same principles, including the Purpose Limitation Principle, can be found in Article 6 DPD. These principles are enforceable and backed by the heavy sanctions found in Article 83.[109] However, they are also likely to be used in

[100] General Data Protection Regulation (n 13) article 4(7).
[101] ibid article 5(2), article 24.
[102] ibid article 4(8).
[103] ibid article 28.
[104] ibid article 4(1).
[105] ibid articles 12-23.
[106] ibid articles 12-15; Wendleby and Wetterberg (n 87) 92.
[107] General Data Protection Regulation (n 13) articles 16-17; Wendleby and Wetterberg (n 87) 96.
[108] General Data Protection Regulation (n 13) article 22; Wendleby and Wetterberg (n 87) 102–103.
[109] Eike Michael Frenzel, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 1-2.

the future by the courts to interpret the rest of the provisions in the GDPR.[110]

The basic principles of Article 5 GDPR are, and requires that Personal Data must be:

- *Lawfulness, fairness and transparency:* Processed lawfully, fairly and in a transparent manner in relation to the data subject.
- *Purpose Limitation:* Collected for a specific purpose and not further processed beyond that purpose. I will analyze this principle in depth in chapter 5.
- *Data minimisation:* Adequate, limited and necessary to what is necessary in relation to the purposes for which they are processed.
- *Accuracy:* Kept accurate and up-to-date, giving the data subject the possibility to correct if inaccurate.
- *Storage limitation:* Only kept for as long as necessary for the purpose the data is processed.
- *Integrity and confidentiality:* protected against unauthorized or unlawful access and accidental damage.[111]

As can be seen, the purpose for which the data was collected plays an integral role throughout the entire lifecycle of the data processing operation. According to Article 5.1(b) GDPR, the purpose for which the Personal Data is collected has to be specified at the time of collection, and this purpose later limits what the Personal Data can be used for and how long it is allowed to be kept and stored.

As written above, Big Data Profiling is characterized by purposeless and broad data collection, before at a later stage different possible uses of he collected data are determined. It is obvious that there is a field of conflict between this kind of data processing and the GDPR's principle of Purpose Limitation, which requires a clearly defined initial purpose before collecting Personal Data.

3.6 Legal basis of processing

For the processing of Personal Data to be lawful at all, it has to be justified by one of the legal bases in Article 6. These are:

- The consent of the data subject
- The necessity for performance of a contract with regards to the data subject.
- The necessity for compliance with legal obligations on the controller
- The necessity in order to protect the vital interests of the data subject

[110] Voigt and von dem Bussche (n 14) 87.
[111] General Data Protection Regulation (n 13) article 5.1(a)-(f).

- The necessity for a task in the public interest
- The necessity for legitimate interests pursued by the controller.[112]

I believe that consent according to Article 6.1(a), which conditions are further specified in Article 7 GDPR, is likely to be the basis for much processing in a Big Data Profiling context.[113] This means that companies will have to ask users to agree that the company will collect, analyze and use their Personal Data to, for example, provide them with individually tailored ads or other specific user experience.

Apart from consent, legitimate interest according to Article 6.1(f) could also play an important role. In some cases, as mentioned in recital 47 GDPR, direct marketing is considered a legitimate interest of the controller.[114]

Purpose Limitation according to Article 5.1(b) GDPR as general principle for the processing of Personal Data is a cumulative requirement with the legal basis according to Article 6 GDPR.[115] This means that even if processing is based on a valid legal basis, it also has to be compliant with the Principle of Purpose Limitation for processing to be legal.[116] While the identification of a legal basis is a very important part of data collection, this thesis is mainly focused on the principle of Purpose Limitation.

3.7 Profiling

The GDPR contains a definition of profiling in Article 4(4) as:
"any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person."

However, while this definition is mentioned in some provisions, such as Article 21 and 22, there is no specific legal effect tied to this definition in the GDPR. It is therefore seen mostly as symbolic.[117] Of course, the regular data protection principles apply to profiling.[118]

[112] ibid article 6.1(a)-(f).
[113] For more about consent, see chapter 5.2.2.
[114] Voigt and von dem Bussche (n 14) 103; General Data Protection Regulation (n 13) recital 47; Wendleby and Wetterberg (n 87) 63.
[115] 'Opinion 03/2013 on Purpose Limitation' (Article 29 Data Protection Working Party) WP 203 27 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> accessed 28 February 2018.
[116] See chapter 5.2.6.
[117] Peter Schantz, 'Die Datenschutz-Grundverordnung – Beginn Einer Neuen Zeitrechnung Im Datenschutzrecht' [2016] Neue Juristische Wochenschrift 1841, 1844.
[118] General Data Protection Regulation (n 13) recital 72.

3.8 Enforcement

One of the large changes of the GDPR compared to the DPD is the increase of fines and enforcement mechanisms.[119] This was seen as necessary to ensure the efficiency of the regulation.[120]

The position of the National Supervisory Authorities has been strengthened in the GDPR by giving them the main tasks of monitoring and enforcing the GDPR. [121] In case of a breach of the regulation, the authorities can impose administrative fines of up to 20 million euro or 4% of the worldwide annual turnover of the offending entity.[122] Entities in breach of the GDPR are also liable to compensate any person who suffered damage under the breach.[123]

3.9 Relevant provisions for reuse of data

The GDPR does not only provide regulations for the collection and first use of Personal Data, but also for later re-use of previously collected Personal Data. Due to the nature of Big Data Profiling that to a large extent uses previously collected data to later create user profiles, this aspect plays an important role for Big Data Profiling companies.

There are different phases during the processing of Personal Data, with different requirements on the controller that apply to Big Data Profiling:

Before or latest with the start of the data collection:
The data controller has to determine whether the data he will collect will be considered Personal Data according to Article 4(1) of the GDPR. If the data is considered Personal Data, it falls under the GDPR, and the controller must comply with the GDPR throughout the use of the data. This issue will be covered in chapter 4.

Furthermore, at the time of collection, the controller has to have specified a purpose for the collection, according to Article 5.1(b) of the GDPR. This requirement will be analyzed in chapter 5.1. Additionally, the data controller has to identify a legal basis for processing according to article 6 of the GDPR. This could be, for example, the consent of the user (Article 6.1(a)), or that the processing is necessary to fulfil a legitimate interest of the controller (Article 6.1(f)).

[119] Voigt and von dem Bussche (n 14) 201; Eike Michael Frenzel, 'DS-GVO Art. 83 Allgemeine Bedingungen Für Die Verhängung von Geldbußen' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 1.
[120] General Data Protection Regulation (n 13) rec. 11.
[121] ibid article 57.
[122] ibid article 83.
[123] ibid article 82.

After collection:
After the data is collected, the controller is typically bound to the purpose specified at collection and normally may not use the data for other purposes (Article 5.1(b)). However, there are some instances where the controller can process the collected data for another purpose than the original one, as mentioned in Article 5.1(b) and Article 6(4) of the GDPR. I will analyze these possibilities in chapter 5.3. Also, if the new processing is not covered by the legal basis identified at collection, a new legal basis has to be found for the re-use. For example, if the initial processing was performed under user consent, the controller could identify a legitimate interest to cover the new processing of the data.

After use of the data:
Typically, after the Personal Data has been used to fulfil its purpose, it has to be deleted according to the principle of Storage Limitation according to article 5.1(e) GDPR. However, if the controller manages to anonymize the data, he will be able to continue processing outside of the bounds of the GDPR. Despite not being linked to an individual, the data can be used to find general correlations and build models. Due to the proximity with Personal Data, this possibility will be analyzed in chapter 4.

3.10 Conclusion

The General Data Protection Regulation is the replacement of the Data Protection Directive, which has been in force since 1995. While many of the core principles of the directive, such as the concept of Personal Data and the Purpose Limitation Principle, remain intact, the enforcement mechanisms and fines for non-compliance have been greatly increased. The GDPR is also a regulation, which makes it directly applicable in all European Union member states. The directive, on the other hand, had to be implemented by each member states, leading to varying levels of protection.

The GDPR applies to any entity processing Personal Data. In a Big Data Profiling context, processing includes the collection and analysis of Personal Data and the application of the created results to individuals. The entity mainly tasked with complying with the GDPR is the “controller” of Personal Data, which is the entity that initiates and decides the purpose of processing. The controller has to comply with several general principles in the GDPR. One such general principle is the Purpose Limitation Principle, which regulates how Personal Data can be collected and re-used after collection. The Purpose Limitation Principle will be the main focus of this thesis. The controller also has to find a legal basis for processing to comply with the GDPR. This can be, for example, the consent of the user or that the processing is necessary for legitimate interests by the controller.

These provisions apply throughout the entire use cycle of Personal Data for Big Data Profiling companies. However, if data is not considered Personal

Data, it falls entirely outside the scope of the GDPR and is not limited by the regulation. The next chapter will analyze which data is considered Personal Data, and if it is possible to easily anonymize data to continue processing without the limitations imposed by the GDPR.

4 Personal and Anonymous Data in Big Data Profiling

In order to determine how Big Data Profiling (BDP) activities will be impacted by the GDPR, it is first necessary to investigate whether the data used in BDP falls under the definition of Personal Data. This is the first question in this chapter. I will first describe the regulatory approach of Personal Data and the views in the literature of how the provisions should be applied to Big Data. Then I will examine the relevant BDP activities. Finally, I will apply the regulation to the results of my examination and determine which BDP data falls under the regulation.

The second question in this chapter is whether it is possible for the controller to anonymize data, thereby removing it from the scope of the GDPR. If this were the case, the controller would be able to continue using the data for data mining after the initial purpose was completed, unencumbered by Data Protection rules. Thereby, some of the value of the data would be preserved.

4.1 Personal Data

In order for the activity to fall under the GDPR, the processed data has to be regarded as "Personal Data". Personal Data is defined in Article 4(1) GDPR as:

"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;"

This means that the definition of Personal Data in Article 4(1) GDPR contains four important requirements:

- Any information
- Relating to
- Identified or identifiable
- Natural person

These elements will have to be examined and applied to the Big Data Profiling context. I will follow the structure and use material from the

"Opinion on the Concept of Personal Data"[124] of the Article 29 Working Party, which has an advisory role for the application of the Data Protection Directive, according to Article 29-30 of the Data Protection Directive. Since the provision in the DPD defining Personal Data is very similar (Article 2(1) DPD), this Opinion should still be relevant.

4.1.1 Any information

The definition of personal information covers "any information". As pointed out in the WP29 Opinion, this is meant to make the definition very broad.[125] As for the *nature* of the data, it does not matter whether the information is objective (for example, the age of a data subject) or subjective (for example, assessments about the data subject based on objective data).[126] It also does not matter whether the information is correct or not.[127]

The data used in Big Data Profiling is typically (1) provided data (which the user explicitly provides to the service, such as by liking a certain post or entering his phone number) and also (2) observed data (which is observed by the controller, such as when a user visits a website containing tracking code).[128] From this data, the controller can then (3) infer other attributes of, and therefore data about, the user, [129] such as interests and expected behavior.[130] For example, a social network might infer that a user is interested in hockey from him liking several hockey players' webpages. Due to the broad definition set out in the guidelines, all the data that a controller holds about a user, whether provided, observed or inferred, is captures by the requirement of "any information". The accuracy of the data concerning a person does not matter, therefore even wrongly inferred data is Personal Data.

Regarding the *content* of the data, the guidelines set out another very broad explanation, including everything relating to the individual's private and professional life, as well as data being created from the individual's behavior or relations. The position of the data subject, whether consumer,

---

[124] 'Opinion 4/2007 on the Concept of Personal Data' (Article 29 Data Protection Working Party) WP 136 <http://collections.internetmemory.org/haeu/content/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 20 March 2018.
[125] ibid 6.
[126] ibid; Wendleby and Wetterberg (n 87) 37.
[127] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 6.
[128] Martin Abrams, 'The Origins of Personal Data and Its Implications for Governance' [2014] SSRN Electronic Journal 6–7 <http://www.ssrn.com/abstract=2510927> accessed 9 May 2018; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 13–14.
[129] Abrams (n 128) 9; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 14.
[130] 'Opinion 2/2010 on Online Behavioural Advertising' (n 29) 7.

employee, employer etc. does not matter. The definition thus extends further than the private sphere to include even the commercial and public sphere.[131]

As we have seen, the data used in a Big Data Profiling operation usually originates from some kind of behavior or action of a user. This can be pages visited, goods purchased or a physical location visited by the user.[132] This should clearly satisfy the requirements on content for data to be any information.

The *format* of the data is also very loosely defined. The data can be stored in any format, including digital and on paper. It also does not need to be stored in a structured database, even stray personal information contained in an email would be considered Personal Data.[133] The guideline mentions the example of a drawing of a child revealing information about the parents parenting methods as well as the child's feelings about them. It could thus hold personal information about the parents and the child.[134]

Big Data is as good as always stored electronically. The format can be different, such as in an electronic file containing a picture, a database with rows off account information or sound-files recorded by the user's device when asking a query. All of these should fall under the Working Party's broad format requirements.

The conclusion must be that all data that is used in a Big Data Profiling context falls under the definition of "any information" in the GDPR.

4.1.2 Relating To

The second requirement in the GDPR is the requirement of a relationship between the data and an individual. The Working Party sees this as the data being "about" an individual.[135] This is likely easy to establish in many cases, such as patient records and images of a person. In other cases, however, it is not as easy to establish the link. The Working Party mentions the example of properties of objects that can be utilized to infer properties of people, through ownership, influence or vicinity. For example, the price of a house is telling on the net worth of an individual. The service record of a car can contain information both about the driver and the mechanic repairing the car.[136] German literature surrounding the GDPR has the concept of "Data about things" (Sachdaten), which includes all data about things, and not people. An example is the maximum speed of a person's car. If the data is detailed enough, it can be used to infer legal, economical and social

[131] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 6–7.
[132] See chapter 2.3.
[133] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 7–8.
[134] ibid 8.
[135] ibid 9.
[136] ibid 9–10.

positions on the car owner. In these cases, even pure data about things can qualify as Personal Data.[137]

To make this determination easier, the Working Party has specified three categories that can be used to assess whether data relates to an individual. (1) The first is that the content itself relates to the individual. This is the most straight-forward case and deals with situations where the data is clearly about the individual, such as medical analysis results.[138] (2) The second alternative category is that the purpose of the data relates to an individual. This is the case if the data is likely to be used for the purpose of evaluating or treating the individual in a certain way, for example, if phone log data is collected with the purpose of evaluating individuals answering phones.[139] (3) The third category concerns the result of the data. If, as a result of the data, an individual is likely to be treated in a different way, the data also relates to an individual. This is the case for example for taxi location data collected for the purpose of making taxi routings more efficient. The same data can also be used to estimate a certain driver's efficiency and working ethics.[140]

In a Big Data Profiling context, a lot of the data will be considered falling under the first category, i.e. relating to an individual. Using the analysis provided by the Working Party, it can be determined that a greater amount of data than one might initially think can relate to a person. In the most straight-forward case, data is collected about the individual. Website visits and user location are both directly related to a user. Even if data is not "about" an individual, but it will be used to profile the individual, it is Personal Data. For example, if data about how long a user's computer is turned on is used to evaluate the user's sleeping patterns, this is Personal Data. Data about things might also be related to people in some instances. An example of this might be for example if a corporation provides webpages with data sheets of expensive and new cell phones, in order to determine the socioeconomic status of users accessing the web pages.

4.1.3 Identified or Identifiable

The definition of Personal Data covers any data that can be traced back, directly or indirectly, to a person. The GDPR differentiates between "identified" and "identifiable". In order to fall outside of the legislation in this respect, the GDPR suggest anonymizing data.[141] Pseudonymization

[137] Hans-Herrman Schild, 'DS-GVO Artikel 4 Begriffsbestimmungen' in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2018) Rn. 22-24.
[138] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 10.
[139] ibid.
[140] ibid 11.
[141] General Data Protection Regulation (n 13) recital 26.

separates the link between Personal Data and the identity of the data subject in order to make compliance with the regulation easier.[142]

4.1.3.1 Identified

An identified person is one that is distinguished from other people within a group.[143] This is done using a so-called "identifier", a piece of information closely linked to that person's identity, which makes it possible to distinguish a person from other individuals.[144] The most common way to *directly* identify a person is by the person's name or social security number.[145] However, it is also possible to hold other identifiers which *indirectly* identify the person, such as a phone number, physical appearance, car registration number or social insurance number. It can also be an attribute or job[146] – for example, the information that somebody is the prime minister of Sweden is able to distinguish that person from other people.

It should be noted that an identifier does not have to be linked to a name. As long as the identifiers are enough to distinguish and single out the person, the person is considered identified.[147] Factors specific to the "physical, physiological, genetic, mental, economic, cultural or social identity" of a natural person can therefore be enough to identify that person.[148] If an online service places a cookie in a user's browser, collects data about this user and stores it linked to the cookie, the person can be identified.[149] This is the case even if the online service has no idea of that person's name or real-life identity. Since the controller is able to profile the individuals social or economic identity and use this to affect the decisions relating to this user, all data linked to this user is considered Personal Data.[150] In a case in 2017, the European Court of Justice ruled on whether a website containing the hobbies of people and their names was to be seen as containing Personal Data. The court found that it did and seems to indicate that even the hobbies or working conditions themselves can sometimes be enough to identify an individual.[151]

Compared to the DPD, the GDPR adds that a user can be identified with an online identifier, location data or the genetic identity.[152] This seems to lower

[142] ibid recital 26-29.
[143] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 13.
[144] ibid; Schild (n 137) Rn. 17.
[145] Schild (n 137) Rn. 16.
[146] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 13; Schild (n 137) Rn. 17.
[147] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 14; Schild (n 137) Rn. 17; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 8.
[148] General Data Protection Regulation (n 13) art. 4.1.
[149] ibid recital 30; 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 14.
[150] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 14.
[151] *Criminal proceedings against Bodil Lindqvist (Reference for a preliminary ruling from the Göta hovrätt (Sweden))* [2003] European Court of Justice Case C-101/01 reference 27.
[152] Data Protection Directive (n 19) article 2(a); General Data Protection Regulation (n 13) article 4(1).

the threshold for what is considered being an identified individual.[153] However, as seen above, online identifiers in the form of cookies were already included under the DPD if they allowed the profiling of the individuals identity.

4.1.3.2 Identifiable

In some cases, the dataset itself might not contain enough data to identify the individual. However, through combination of the dataset with other datasets the person can be identified. The *possibility* of identifying that person is enough to make the data personal.[154] If, for example, the controller has access to two datasets, the first containing a number of clicks by a certain person linked to unique but anonymous identifiers, and the other containing the link between the identifiers and the name of the person, even the first dataset would be considered Personal Data. This is the case even if the second dataset is not held by the same controller.[155]

A typical situation is where the data is anonymous by itself, but possible to identify using another publicly available dataset. This second dataset is known as auxiliary data.[156] Perhaps some of the data points overlap between the supposedly anonymous dataset and one containing identifiers. Even tiny overlaps are often enough – a study showed that 87% of US-citizens are identifiable using only the three values zip-code, gender and date of birth.[157] Rubinstein and Hartzog describe the difficulty of completely anonymizing data.[158] They mention the example of Netflix releasing an "anonymous" set of movie ratings by users for a competition. By correlating this data with profiles from the Internet Movie Database, researchers were able to identify some users in the dataset.[159]

Narayanan and Shmatikov go even further than this assessment. They claim that it is possible to link *any* information that is relatively static over time and fine-grained enough to an identity by using additional outside data.[160]

Based upon this research, it seems extremely hard to guarantee that a dataset is anonymous. Given a corresponding dataset that the controller has access

---

[153] Detlev Gabel and Tim Hickman, 'Chapter 5: Key Definitions – Unlocking the EU General Data Protection Regulation' (*White & Case*, 13 September 2017) 5 <//www.whitecase.com/publications/article/chapter-5-key-definitions-unlocking-eu-general-data-protection-regulation> accessed 14 May 2018.
[154] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 12; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 9; Wendleby and Wetterberg (n 87) 40.
[155] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 13; Wendleby and Wetterberg (n 87) 40.
[156] Ira S Rubinstein and Woodrow Hartzog, 'Anonymization and Risk' 91 WASHINGTON LAW REVIEW 59, 712–713.
[157] Latanya Sweeney, 'Simple Demographics Often Identify People Uniquely' [2000] Carnegie Mellon University, Data Privacy Working Paper 3 34.
[158] Rubinstein and Hartzog (n 156).
[159] ibid 713.
[160] Arvind Narayanan and Vitaly Shmatikov, 'Myths and Fallacies of "Personally Identifiable Information"' (2010) 53 Communications of the ACM 24, 26.

to, any information would seem to be identifiable. Some of the authors therefore suggest stepping away from a black-and-white distinction between personal and anonymous data towards a risk-based approach where the process used to safeguard the data, or the computations performed with the data, are in focus rather than the data itself. [161]

The GDPR acknowledges the difficulty of guaranteeing anonymity and uses a more nuanced approach in recital 26. In order to determine whether a person is identifiable in a dataset, all the "means reasonably likely to be used" by the controller or another person to identify the person have to be taken into consideration. This can be determined with regards to the cost and time required to perform such an identification, with the state of technology at the time of processing in mind as well as future developments in processing.[162] The Working Party suggests other factors to consider, such as the purpose of the processing, the interests at stake for individuals and the risk for breaches at the corporation.[163]

The ECJ has performed this analysis in a case relating to the DPD concerning the identifiability of IP-addresses.[164] Instead of applying an objective standard, it used a subjective one. The court considered whether an additional dataset required for identification is "a means likely reasonably to be used to identify the data subject." If the identification is illegal or requires a "disproportionate effort in terms of time, cost and man-power", the risk of identification is insignificant.[165] Since, in the case, the controller had the legal means to access the data from the internet service provider that linked an IP-address to an individual, the data was considered personal.[166]

According to the ECJ, the theoretical possibility of identifying a person in a dataset is not enough to make data personal. If the effort associated with the identification is so large that no one would likely attempt an identification, the data is likely not personal.[167] Data that can only be obtained by a court order should therefore not be seen as auxiliary information, while data that is freely accessible online should.[168] According to the literature, if the data is freely accessible or available for purchase on the market, this should be considered being "means reasonably likely to be used".[169]

---

[161] Narayanan and Shmatikov (n 160); Rubinstein and Hartzog (n 156) 729.
[162] General Data Protection Regulation (n 13) recital 26; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 10.
[163] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 15.
[164] *Patrick Breyer v Bundesrepublik Deutschland (request for a preliminary ruling from the Bundesgerichtshof — Germany)* [2016] European Court of Justice (Second Court) Case C-582/14.
[165] *Case C-582/14* (n 164) rec 45-46.
[166] ibid rec 47-49.
[167] Schild (n 137) Rn. 18.
[168] Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 11.
[169] Voigt and von dem Bussche (n 14) 12; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 11.

The legislator thus seems to have recognized that a pure identifiable contra anonymous approach is difficult to achieve and moved slightly in the direction of a risk-based approach, as suggested by some authors.[170] However, the analysis still focuses on the dataset itself, without taking into account the access controls implemented by the controller. As long as the dataset in itself is identifiable with means reasonably likely to be used, it will therefore be personal information and fall under the GDPR. Note however that even a partial anonymization can be helpful, as this can be seen as a safeguard that protects the user.[171]

4.1.3.3 Anonymization & Pseudonomization

The GDPR recitals 26-29 suggest two approaches for controllers to meet their obligations with regards to personal information.[172] (1) The first is anonymization. Anonymization is the process of removing or altering certain aspects of the data to make it impossible to link it to a certain person. Anonymization can therefore be a very important tool for controllers to be able to continue using data for statistical purposes etc., since the anonymized data does not fall under the GDPR.[173] As we have seen in the preceding section, simply removing the identifiers is often not enough to create an anonymous dataset[174] because it could still be referred to an individual by combining a dataset with other datasets. There are two main strategies for anonymizing data: Randomization and generalization.[175]

Randomization subtly alters some of the attributes of the dataset to make it harder to rely on the information about individuals, thereby making it harder to identify people in the dataset while still retaining enough of the information to make analysis of the data meaningful.[176] An example of this could be to alter the height of an individual in a dataset by a random but small amount. This makes this property harder to use for identification but still allows an overall use of the data for certain analysis.[177]

Generalization, on the other hand, aggregates the data of individuals into entire groups of categories, thereby making it harder to single out a particular individual.[178] An example of this could be generalizing the location of an individual to the country, instead of the city.[179]

---

[170] Rubinstein and Hartzog (n 156) 713.
[171] See under chapter 7.4 and 7.5.
[172] General Data Protection Regulation (n 13) recital 26-29.
[173] ibid recital 26.
[174] 'Opinion 05/2014 on Anonymisation Techniques' (Article 29 Data Protection Working Party 2014) WP126 9.
[175] ibid 10; Voigt and von dem Bussche (n 14) 13.
[176] 'Opinion 05/2014 on Anonymisation Techniques' (n 174) 12; Voigt and von dem Bussche (n 14) 13.
[177] 'Opinion 05/2014 on Anonymisation Techniques' (n 174) 12.
[178] ibid 16; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 49.
[179] 'Opinion 05/2014 on Anonymisation Techniques' (n 174) 16.

Both of these methods are not automatically entirely reliable[180] and one still has to be careful that the remaining data cannot be used to still identify a certain person despite the taken measures.[181]

(2) Pseudonymization, on the other hand, separates the data that can be used to identify the person, from the data that is supposed to be processed for analysis etc. While the data is still personal, pseudonymization makes it easier for the person responsible to comply with the GDPR. For pseudonymization to be effective, the data that allows identification needs to be kept separate and secure by using technical or organizational measures, according to recital 29 of the GDPR.[182]

4.1.3.4 Application to Big Data

In Big Data Profiling, the purpose of the data collection is to evaluate the user by observing behavior over time. This is only possible if the service has a way to recognize the user. This can, as written above, be done via cookies, a loyalty card or an app that uploads information about a user and stores it online. In most of these cases, the data is linked to some form of identifier, such as a name, an e-mail address or an online identity such as the IP-address. Target, for example, was able to link the data to a postal address to send advertising to the pregnant girl.

Since the data in Big Data Profiling is directly linked to an identifier, the person in question can be identified, and it therefore falls within the scope of the GDPR.

Based on the broad concept of identity in the GDPR, even if a name or e-mail address is not collected by an online service, the data might still be considered personal: If the service creates a profile of, for example, hobbies or social or cultural identity of a person, linked to a user-account, the data will be considered personal despite not being linked to the real-world identity of the person. Therefore, any information linked to a cookie is likely to be considered personal information. Even shadow profiles created by Facebook or Google of users that are not logged in or have no user account, by the placing of cookies, are to be seen as Personal Data, since they allow the companies to single out individuals and reach decisions about them, such as which ads to show to whom. New under the GDPR is that online identifiers and location data are mentioned as an identifier. While online identifiers already seem to be covered under the DPD, a controller holding the location data of an individual is now also likely enough to identify a person.

[180] ibid 10.
[181] Voigt and von dem Bussche (n 14) 13–15; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 51.
[182] Voigt and von dem Bussche (n 14) 15; Schantz (n 117) 1843; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 40-47; Schild (n 137) Rn. 68-70.

This means that data held by companies and used in Big Data Profiling is in principle always to be seen as Personal Information. As a result, companies are bound to the purpose for which they collected the data (see chapter 5) and they have to delete the data once the purpose is complete, due to the principle of Storage Limitation.[183] This collides with the interest of the companies within Big Data Profiling to store data because it has a latent value. As mentioned above, it could be advantageous for a company to store data, to use it later to identify correlations and build models for future users.

One possibility to allow the continuous storage and further processing of collected data unencumbered by the GDPR is therefore the *anonymization* of data. Anonymized data falls completely outside of the GDPR, because it cannot be used to identify an individual according to the definitions of Article 4(1) GDPR. The question is whether it is possible to remove enough pieces of the data that a re-identification of the individual is impossible (meaning that the individual is not "identifiable" under the GDPR), while still allowing the controller to extract meaningful correlations.

The risk of reidentification of a supposedly anonymous dataset seems to increase in a Big Data Profiling context.[184] All three of the features of Big Data Profiling facilitate the identifiability of data. The *latent value of data* means that companies will be interested in keeping anonymized data as long as possible to be able to utilize it to improve the business at a later stage. During the time the data is kept, new technologies could be developed that enable the reidentification of the data, requiring the controller to keep track of and try to predict technological developments.[185] Another possibility is that the controller might get access to a new auxiliary dataset that allows the identification of the data. These occurrences would have the effect of dramatically lowering the price and time effort required to identify the data, and thus make the previously thought anonymous data identifiable. The longer the data is supposed to be stored, the higher the possibility of identifying the data at a later stage.[186] The GDPR foresees the increased possibilities of new technologies and requires the controller to take them into consideration when ascertaining if means are reasonably likely to be used to identify a natural person, according to recital 26 GDPR.

The second property of *large data volume* also contributes to the ease of identifying data. The more granular the data the controller collects, the more likely it is to be able to single out an individual. As we have seen, as little as four data points linked to a unique identity are often enough to identify the person behind the identity.[187] In today's world, thousands of data points can easily be collected.[188] Even if the data itself might seem devoid of

[183] See chapter 3.5.
[184] Zarsky (n 4) 1000.
[185] Voigt and von dem Bussche (n 14) 13.
[186] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 15.
[187] Sweeney (n 157).
[188] Dylan Curran, 'Are You Ready? This Is All the Data Facebook and Google Have on You' *the Guardian* (30 March 2018)

identifying information, the combination is often enough to single out individuals. The volume of collected data also increases the availability of data sets about individuals. There are many data brokers that actively collect data on individuals and sell that data to third parties. The datasets seem very large – some firms reportedly have thousands of data points per citizen.[189] The availability of such auxiliary datasets means that it is a lot more likely that controllers will be able to identify people in their own datasets in connection with other datasets.

People themselves are also active in contributing to the publicly available information about them. Users post 500 million tweets[190] and 350 million pictures to Facebook per day.[191] People readily share their location, which products they enjoy, which books they've read and other personal details with the world. Since a lot of this data is publicly accessible to anyone, it does in my opinion have to be considered being part of the auxiliary information that is available to identify an individual.[192] An example is a user who buys a book that is rarely sold and tweets about it. If the book store publishes anonymous sale statistics for all books, such as "number of purchases in the last 24 hours", a determined attacker would be able to make the connection and thus identify that the user bought the book at this particular store. The tweet is publicly and easily accessible. According to the test employed by the Working Party and European Court of Justice, the data in question from twitter should therefore fall into the "means reasonably likely to be used" definition. A simple search on twitter for the title of the book would be enough to find the tweet and thus identify the data subject.

The question is then whether it is enough that a single data subject can be identified, to turn the entire dataset into Personal Data. According to the definition of Personal Data in the GDPR, only the data that is actually linked to an individual is Personal Data.[193] However, it seems impossible for a controller to monitor the individual data subjects and all the auxiliary data available worldwide on other sources about them. Since it is impractical to distinguish the data, the controller will have to treat all of it as Personal Data in order to be safe.[194] The Working Party has led a similar reasoning in an example relating to an Internet Service Provider not being able to distinguish between (anonymous) IP-addresses at an internet café and those

<http://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy> accessed 3 April 2018.

[189] Chih-Liang Yeh, 'Pursuing Consumer Empowerment in the Age of Big Data: A Comprehensive Regulatory Framework for Data Brokers' [2017] Telecommunications Policy 4 <http://linkinghub.elsevier.com/retrieve/pii/S0308596117304743> accessed 21 February 2018.

[190] 'Twitter Usage Statistics' (*Internet Live Stats*) <http://www.internetlivestats.com/twitter-statistics/> accessed 3 April 2018.

[191] Cooper Smith, 'Facebook Users Are Uploading 350 Million New Photos Each Day' (*Business Insider*, 18 September 2013) <http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9> accessed 3 April 2018.

[192] compare Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 11.

[193] General Data Protection Regulation (n 13) article 4.1.

[194] compare Wendleby and Wetterberg (n 87) 36.

tied to a home address, therefore having to treat all the addresses as Personal Data.[195]

The third property of *advanced analysis methods* also contributes to the identifiability of datasets. As proven by researchers at Cambridge, it is possible to construct sophisticated models from innocuous activity, such as Facebook likes, that are able to efficiently profile people.[196] These models can then be used to for example identify the cultural and social identity of a person, such as ethnicity and political affiliations. Due to the definition of "identified" under the GDPR, which does not necessarily require a person's name or other direct identifier, the combination of such information and a way of singling a user out is enough to consider a user being identified, and thus the data to be personal. Take the example of *reddit*, a forum that allows the creation of anonymous user accounts. If a logged in user subscribes to a forum about a certain political party, the user is already identified, despite the account being completely anonymous. The political identity of the user together with the cookie that is used to keep the user logged in is enough to identify the user, and for example show targeted ads. The data is therefore Personal Data.

Additionally, the models used to profile users might reveal enough information about them to make it possible to establish a link to a real-life identity. Therefore, even if a dataset has been anonymized by removing all possible identifiers and there is no auxiliary information, these identifiers might be restored by using these models. Consider the release of the anonymous data by Netflix. Even if the data by the Internet Movie Database was not available, it is possible that a model could be built to correlate liked movies and attributes such as age, ethnicity, political affiliations, interests and location. This could potentially be enough to create a connection to a real person. Thus, even data that is anonymized and is not identifiable even by using auxiliary data, could still become identifiable by using the advanced analysis methods prevalent in a Big Data context.

It might be tempting for the controller to use anonymization to hold on to the data and use it for future analysis. However, as we have seen, anonymization is difficult to perform properly. True anonymization has to go further than just removing some few fields, because the other fields that are stored about the user can still be used to distinguish that person. Even if the controller believes that a dataset is anonymous, he must be aware of the release of correlating datasets or new technologies that might still be able to identify the individual, thus turning the data into Personal Data again.[197] Even if a dataset seems to be anonymized, it may become Personal Data if identification succeeds.[198]

---

[195] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 17.
[196] Michal Kosinski, David Stillwell and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 Proceedings of the National Academy of Sciences 5802.
[197] Voigt and von dem Bussche (n 14) 15.
[198] Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) 13.

Since through Big Data methods more and more datapoints can be linked to an individual, especially in conjunction with other such datapoints, a larger and larger portion of the dataset will have to be removed in order to prevent a re-identification of an individual. Such substantial removal of information decreases the value of the dataset for the controller. Especially the large volumes of data make the datasets feasible to mine for correlations. Also, the granularity of the data, which is the very thing that makes the data valuable for analysis, makes it easier to identify individuals in the dataset. Therefore, anonymizing the dataset to make it possible to use under the GDPR is often not desirable for the market players in Big Data Profiling.

4.1.4 to a natural person

The final criterium to determine if data is Personal Data according to Article 4(1) GDPR is that the data must be related to a natural person. This means that data about dead people is usually exempt. However, it might be hard for the controller to discern between living people and dead people, making it easier to apply the protection to all data. Data about deceased people might also indirectly reveal information about living people, such as heritable diseases.[199] Legal persons are also generally excluded. In cases where the name of a legal person is derived from the name of a natural person, this too can be Personal Data.[200]

One potential effect of Big Data about deceased people could be the uncovering of details that also affect children of a deceased individual. For example, the data broker Axciom is able to detect if a person has diabetes based on purchase history.[201] This could be seen as also revealing that the children have a higher risk of the disease. Data about the location of the deceased individual while alive could reveal information about family trips and the family life of the child. Both of this could be considered personal information about the child. Social check-ins or messaging history of an individual is likely to contain information about other individuals who were with that person. The feature of large volume for Big Data means that datasets about individuals are likely to contain a lot of data, some of which might also concern other individuals. This can be extracted using the advanced analysis methods of Big Data Profiling.

In conclusion, I therefore believe that Big Data Profiling will lead to controllers having to be careful about releasing or reusing data of dead individuals. If data is collected about an individual that later dies, it is likely that there could be traces of other individuals in that data. The controller should therefore err on the side of caution and treat the data as still being personal.

---

[199] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 22; Schild (n 137) Rn. 11.
[200] 'Opinion 4/2007 on the Concept of Personal Data' (n 124) 23; Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 5.
[201] Moerel and Prins (n 11) 56.

4.2 Conclusion

In conclusion, it seems like Big Data has the effect of drastically lowering the threshold for what is Personal Data. The Personal Data used in Big Data Profiling is usually stored in an electronic form, which makes it fall under the "any data" requirement of the definition of Personal Data. As for the "relating to" requirement, the large volume of data could make it more likely that one of the pieces of data describes an individual.

The largest effect of Big Data Profiling can likely be found on the requirement of "identifiable". Big Data controllers are likely to want to keep the data longer than the initial purpose to extract all the possible value. This makes it possible that new technologies or auxiliary datasets are released that turn the dataset identifiable, even if the data has been anonymized to a certain extent that may have been sufficient at the time. Further, in a Big Data Profiling context, large volumes of data are collected. This significantly increases the possibility of using correlations in the data to single out an individual. The age of Big Data, internet and social media also means that there are likely a lot more datasets that can be used to identify an individual, whether provided by the data subject itself on social media, contained in public datasets or available for sale from data brokers. Since this makes it easier for anyone to identify individuals in the dataset, it lowers the threshold for data to become Personal Data.

There is also the effect of Big Data on the Natural Person requirement: Since more data is collected, it is likely that a profile of a person will contain not just data concerning that data subject, but also data relating to other individuals. This means that this data will be considered Personal Data even after the death of the first individual.

All of this means that even data that might be considered anonymous by the controller, might in fact have to be regarded as still being Personal Data. Therefore, it will be significantly harder for controllers to anonymize data that they want to keep, for example for later data mining. Even data that does not contain any identifiers today, might be possible to tie to an individual using future auxiliary datasets. Removing enough data to make the dataset efficiently anonymous in a future-proof way is likely to ruin the value for the controller.

5 Purpose Limitation

One of the fundamental principles for the collection and processing of Personal Data is that of Purpose Limitation. It can be found in Article 5.1(b), where it states that Personal Data shall be:

"collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');"[202]

More or less the same provision can be found in Article 6.1(b) of the DPD. In general, the Purpose Limitation Principle aims to prevent the use and reuse of collected data in ways not expected to the data subject, while still allowing controllers to further process the data for other useful purposes as long as they are compatible.[203] Purpose Limitation is seen as a cornerstone of the data protection regime.[204] It serves as a basis for many of the other provisions.[205] The Purpose Limitation Principle originates from Convention 108 of the Council of Europe[206] which entered into force in 1981 and is currently ratified or ascended to in 51 countries, including Sweden.[207] It thus predates even the DPD.

The regulation contains two building blocks: The first being the requirements of collection of Personal Data (Purpose Specification) and the second being the limitations on further use (Compatible Use).[208] The effect of Purpose Specification on Big Data Profiling will be handled in chapter 5.1. Chapter 5.2 will analyze the limitations on the further use of Personal Data.

[202] General Data Protection Regulation (n 13) article 5.1(b).
[203] 'Opinion 03/2013 on Purpose Limitation' (n 115) 4.
[204] ibid; Peter Schantz, 'DS-GVO Artikel 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2017) Rn. 12.
[205] Voigt and von dem Bussche (n 14) 88–89; 'Opinion 03/2013 on Purpose Limitation' (n 115) 4.
[206] 'Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Adopted 28 January 1981, Entered into Force 1 October 1985) ETS No.108' <https://rm.coe.int/1680078b37> accessed 4 May 2018 article 5(b); Forgó, Hänold and Schütze (n 23) 24.
[207] 'Chart of Signatures and Ratifications of Treaty 108' (*Council of Europe - Treaty Office*) <https://www.coe.int/en/web/conventions/full-list> accessed 4 May 2018.
[208] 'Opinion 03/2013 on Purpose Limitation' (n 115) 11–12.

5.1 Purpose Specification at the time of collection

5.1.1 Introduction

In this section, I will examine the question to which extent the purpose of a processing has to be specified at the point of data collection in a Big Data Profiling context. The requirement of collection for a specified purpose seems to clash with the process used in Big Data Profiling, which relies on collecting data first and then finding a purpose for it. [209] However, depending on how broad the specification of a purpose can be, it might give the controller flexibility to decide how the data should be processed, while still falling under the initial purpose for collection. In other words, the question is if a Big Data Profiling company can simply use a very broad purpose for collection of the data, and then later decide how the data should be used inside the wide scope of the initial purpose.

For example, if Facebook could collect data for the purpose of "tailoring the user experience", it could potentially use this data both to decide which ads to show to the user and/or to suggest friends the user is likely to know. You could say that both of these cases fall under the purpose of "tailoring the user experience". Yet they have very different implications for the data subject.

5.1.2 Requirements under Purpose Specification

5.1.2.1 Introduction

According to Article 5.1(b), the Principle of Purpose Specification contains three requirements on the purpose at the point of collection. The purpose must be:

- Specified
- Explicit
- Legitimate

5.1.2.2 Specified

The first requirement for collection of data is that the purpose is specified.[210] Specified relates to the internal process of the controller, not so much to specifying in the meaning of communication to the data subject. It can be seen of a sort of self-regulation, requiring the controller to have a specified

[209] Mayer-Schonberger and Padova (n 21) 6.
[210] General Data Protection Regulation (n 13) article 5.1(b).

purpose before beginning the collection, and thus considering what the purpose of a certain collection should be.[211]

The purpose has to be specified before the collection of the data takes place.[212] Further, it has to be specified enough that it is possible to precisely determine which kinds of processing fall within the purpose and which do not. Vague formulations, such as "marketing purposes" or "future research" are unlikely to hold.[213] The level of detail required depends on the context of the processing, such as the number of affected individuals.[214] The important thing seems to be that a normal reasonable individual is able to understand which kinds of processing will be done on the data.[215] If Personal Data is collected for multiple purposes, each should be specified separately.[216]

5.1.2.3 Explicit

While the requirement of "specified" seems to focus on the fact that the data must be specified in the mind of the controller, the "explicit" requirement adds that this must be made explicit at the time of collection,[217] meaning that it has to be communicated to the affected parties in a way clear enough that everyone will have the same idea of what the purpose encompasses. This increases transparency and predictability.[218] The purpose can be made explicit by a notice to the data subject or the notification to a supervisory authority. While it does not have to occur in writing, this can be helpful.[219]

5.1.2.4 Legitimate

The requirement for the collection to be "legitimate" sounds like it could be a reference to article 6, concerning the legal bases for processing. However, the "legitimate" requirement is generally seen to be broader than this. Not only does in encompass the grounds for legitimate processing, but also all

---

[211] 'Opinion 03/2013 on Purpose Limitation' (n 115) 15; Eike Michael Frenzel, 'DS-GVO Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 27; Schantz (n 204) Rn. 15; Forgó, Hänold and Schütze (n 23) 26.
[212] 'Opinion 03/2013 on Purpose Limitation' (n 115) 15; Schantz (n 204) Rn. 14; Forgó, Hänold and Schütze (n 23) 26; Wendleby and Wetterberg (n 87) 366.
[213] 'Opinion 03/2013 on Purpose Limitation' (n 115) 15–16; Judith Rauhofer, 'Of Men and Mice: Should the EU Data Protection Authorities' Reaction to Google's New Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle' (2015) 1 Eur. Data Prot. L. Rev. 5, 8; Forgó, Hänold and Schütze (n 23) 27.
[214] 'Opinion 03/2013 on Purpose Limitation' (n 115) 16; Voigt and von dem Bussche (n 14) 89; Forgó, Hänold and Schütze (n 23) 27.
[215] Frenzel, 'DS-GVO Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' (n 211) Rn. 27; Voigt and von dem Bussche (n 14) 89.
[216] 'Opinion 03/2013 on Purpose Limitation' (n 115) 16; Voigt and von dem Bussche (n 14) 89; Forgó, Hänold and Schütze (n 23) 27.
[217] 'Opinion 03/2013 on Purpose Limitation' (n 115) 17; Forgó, Hänold and Schütze (n 23) 28.
[218] 'Opinion 03/2013 on Purpose Limitation' (n 115) 17; Schantz (n 204) Rn. 16; Forgó, Hänold and Schütze (n 23) 28.
[219] 'Opinion 03/2013 on Purpose Limitation' (n 115) 18.

other data protection laws and laws in general.[220] It can be necessary to consider even decrees, customs and contractual arrangements.[221]

5.1.3 Application to Big Data

5.1.3.1 Views in the literature

As discussed previously, the most determining quality of Big Data is the latent value of the data. This flips the process of data collection of use from the previous model. While previously, data was collected after identifying a purpose, the Big Data model consists of first collecting all kinds of data and then finding a purpose for this data, also known as letting the data speak for itself.[222] This seems to be incompatible with the Purpose Limitation principle, which requires the controller to have a specific, explicit and legitimate purpose for the collection before it occurs. A number of authors have dealt with determining the effects of this on Big Data:

Zarsky[223] believes that the Purpose Limitation Principle is at odds with Big Data analyses. Since many of the uses data in a Big Data context might only be revealed after a while, Zarsky believes this makes it complicated or even impossible to perform Big Data analyses.[224] He also does not believe in the possibility of creating a wide purpose that would allow for different types of later processing, as this would be in conflict with the rule of specification.[225] Mayor-Schonberger and Padova[226] agree with this and mention how Google and Facebook typically request very broad purposes.[227] They believe the use of such broad purposes will be more difficult with the requirements on consent in the GDPR.[228] Forgó et al[229] also agree that vague purposes will not suffice, and that this hinder open-ended analysis of Big Data.[230]

The GDPR itself mentions the possibility of a purpose only being revealed after collection, only in scientific research purposes in recital 33. It allows data subjects to consent to extra broad processing for scientific research purposes.[231] Schantz believes that the fact that the legislator saw it necessary to create this possibility as an exception from the rule speaks for a strict

[220] ibid 19; Frenzel, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' (n 109) Rn. 28; Schantz (n 204) Rn. 17; Forgó, Hänold and Schütze (n 23) 28; Wendleby and Wetterberg (n 87) 49.
[221] 'Opinion 03/2013 on Purpose Limitation' (n 115) 20; Forgó, Hänold and Schütze (n 23) 28.
[222] Mayer-Schonberger and Padova (n 21) 6.
[223] Zarsky (n 4).
[224] ibid 1006.
[225] ibid.
[226] Mayer-Schonberger and Padova (n 21).
[227] ibid 322.
[228] ibid 326.
[229] Forgó, Hänold and Schütze (n 23).
[230] ibid 31–32.
[231] General Data Protection Regulation (n 13) recital 33.

requirement on the purpose specified for any other, non-scientific data collection.[232]

From Article 5.1(b) and from what has been said above it follows that data collection without a purpose is prohibited.[233] Operations that, for example, track a user and then later determine the purpose of the data are therefore illegal under the GDPR. This means that letting the data speak, to identify novel uses of Personal Data after broad collection, will under normal circumstances be very difficult in practice.

A softer approach could be to specify a purpose that is broad enough to allow the controller the flexibility he wants in analyzing the data after collection, like the example from Facebook mentioned above.

This question is more nuanced, and likely more applicable in practice. One could imagine a situation where a controller specifies an extremely broad purpose and then proceeds to collect a lot of data within that purpose. Such a purpose could in my opinion be, for example, "To create a profile based on the Personal Data to target advertising". This would allow the controller to change the measures used to target the advertising and the Personal Data used while still remaining within this purpose. Both Zarsky and Mayer Schonberger and Padova mention this as a possibility, but both are critical of its reconcilability with the Purpose Specification principle.[234]

A case that illustrates the issue is the one of Google changing their privacy policy in 2012. Google decided to only have one single privacy policy covering all its services such as the Google Search Engine, Youtube, Maps and Chrome.[235] Therefore the policy made the purposes for which the data was collected more general and allowed Google to combine user data between the different services.[236] The Article 29 Working Party asked Google to hold off on integrating the new Privacy Policy until the data protection aspects could be determined. It tasked the French Commission Nationale de l'Informatique and de Liberté (CNIL), to conduct this investigation.[237] In its final letter to Google, the CNIL found that the purposes that Google used to combine the data were too broad, thereby not complying with the "specified" and "explicit" requirements in the DPD.[238] However, CNIL did not further delve into this issue but instead focused on the information requirements.[239] In a similar situation surrounding Facebook

[232] Schantz (n 117) 1844; compare Forgó, Hänold and Schütze (n 23) 34.
[233] General Data Protection Regulation (n 13) article 5.1(a).
[234] Zarsky (n 4) 1006; Mayer-Schonberger and Padova (n 21) 325–326.
[235] Rauhofer (n 213) 6.
[236] ibid 7.
[237] ibid 5.
[238] 'Google Privacy Policy: Main Findings and Recommendations' (CNIL) 2–3 <https://www.cnil.fr/sites/default/files/typo/document/GOOGLE_PRIVACY_POLICY-_RECOMMENDATIONS-FINAL-EN.pdf> accessed 13 April 2018.
[239] Rauhofer (n 213) 10.

that updated its Privacy Policy in 2014, the focus seems to be yet again on transparency and consent rather than Purpose Limitation.[240]

The key feature of the "specified" requirement is that a reasonable person will have to determine precisely which processing operations will fall under the purpose. How specified the purpose must be, additionally depends on the context – for example, a local company will have to be less specific than an international online business. The more advanced analysis methods are used, the more specified the purpose has to be.[241] A multinational corporation selling goods and using analytics to personalize offers will have to specify in a detailed and comprehensive way the methods used for processing the data, and the criteria to profile the user.[242] General purposes have to be broken down into sub-purposes in order to comply. For example, the Working Party suggest breaking down data processing connected to a social benefit claim into sub-purposes such as identification, eligibility check etc.[243]

The Working Party also mentions the example of an algorithm that is able to tell the pregnancy status of a person, based on purchasing patterns, and using this information for the targeting of advertisements. According to the Working Party, this does not fall under the broad purpose of marketing, due to the unexpectedness and the secrecy of the algorithm.[244]

Borgesius and Poort have applied the Purpose Specification principle to price-based discrimination.[245] They argue that the "specified" requirement means that a controller that offers different prices to different customers has to inform them of this fact in a more detailed fashion than "using the data to personalize experiences", even if such information would upset the customers.[246]

Moerel and Prins argue that data collection and analysis in *itself* can be a purpose for data collection. This would make the specified and explicit test meaningless.[247]

5.1.3.2 Analysis

[240] 'Common Statement by the Contact Group of the Data Protection Authorities of The Netherlands, France, Spain, Hamburg and Belgium' (2017) <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/common_statement_16_may_2017.pdf> accessed 13 April 2018.
[241] 'Opinion 03/2013 on Purpose Limitation' (n 115) 51; Forgó, Hänold and Schütze (n 23) 27.
[242] 'Opinion 03/2013 on Purpose Limitation' (n 115) 52.
[243] ibid 53.
[244] ibid 61.
[245] Zuiderveen Borgesius and Poort (n 50).
[246] ibid 359.
[247] Moerel and Prins (n 11) 44.

Before beginning the analysis of how specified a purpose for data collection must be, it is important to define how "collection" of data is to be interpreted. If "collected" is interpreted in a narrow sense, only data that is specifically collected from the data subject would be included. However, Big Data Profiling allows the creation of *new* Personal Data, such as age, interests, etc, from existing collected Personal Data. Is created Personal Data therefore excluded from the requirement for a specific purpose, as it is not "collected"? For example, if an interest of an individual is inferred, can it be used freely for any purpose by the controller? [248]

Looking at the GDPR, it is clear that this is not the case. Collection is always the very first step of processing of data.[249] If data is acquired through inferring from other data, therefore, this operation should be seen as collection. Additionally, the purpose of Personal Data plays a crucial role in the system of Data Protection. It can be seen as an anchor for all processing.[250] It would not work with the rest of the GDPR to have Personal Data that is exempt from the need for a purpose due to being inferred from existing information, and not "collected". Therefore, I believe collection encompasses all methods of acquiring data, including inferring. Inferred data, such as interests and likely future behavior, therefore falls under the requirement of Purpose Limitation, just as data collected directly from the data subject.

Based on the remarks above, it does not seem legally possible for a Big Data Profiling company to specify a broad purpose, encompassing a large amount of possible future analyses of Personal Data. As we have seen, Big Data analysis typically deals with a huge volume of data. This means that they are likely to process a lot of Personal Data of many different people. This is one of the factors requiring an extra specific purpose. Another factor seems to be the type of analysis, and the sophistication of the methods used to process the data. Big Data Profiling companies are likely to use extremely sophisticated and advanced analysis methods, again increasing their need for specificity. These factors together show that Big Data processing will require a very high level of specificity of purpose.

I will use an example to illustrate this. If a social network uses profiling to determine attributes of an individual and target ads based on this profile, it will have to have a specific purpose: Using the purpose for example "we will create a profile of you to target advertisements to you" would likely not be specific enough. This is due to the fact that the data subject would not be able to properly assess which kinds of profiling the social network would engage in. Under this purpose, the controller could introduce new methods

[248] compare Norjihan Abdul Ghani, Suraya Hamid and Nur Izura Udzir, 'Big Data and Data Protection: Issues with Purpose Limitation Principle.' (2016) 8 International Journal of Advances in Soft Computing & Its Applications.
[249] 'Opinion 03/2013 on Purpose Limitation' (n 115) 21; compare Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' (n 147) Rn. 23.
[250] Frenzel, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' (n 109) Rn. 23.

of processing the data that could involve unexpected factors, such as if the individual is single, pregnant or affected by a certain disease.

However, a purpose could also be formulated in a different way, for example: "We will use your location data, which websites you visit and which links you click to build a profile on you. This profile will include your approximate age, your gender and your interests. We will use this profile to show you ads in your newsfeed." This purpose description gives a clear indication of which data is used, what the profile of the user will look like and what the profile will be used for. The data subject would be more able to determine what kind of processing the data will be used for, and decide whether he is willing to consent to such a use of the data. If the data controller wishes to include new kinds of data in the processing, or incorporate new details into the profile, such as relationship status, he will have to ask the data subject for renewed consent, thereby allowing the data subject to understand the changed consequences.

Even with this example, the assessment will have to be made on a case-by-case basis. In some instances, maybe even some of the logic involved will have to be specified. Depending on what is covered by "interests", maybe this should be elaborated on as well. The precise level of detail required is something that will have to be developed by application of the law once it has come into force.

We have seen before that "data is the new oil". This means that companies will want to try to collect as much of it as possible, and then try to find a way to extract value from it later. Under the principle of Purpose Specification, this is not possible. Likewise, collecting data for a very broad purpose and then determining how the data can be used later is also unlikely to be legal, subject to a case by case decision with regard to the different factors mentioned above. However, due to the specifics of Big Data Profiling, a broader purpose is unlikely to be acceptable. It is likely that the Purpose Limitation Principle makes the "Collect now, analyze later" attitude illegal.

Moerel and Prins argue that the analysis of the data in itself can be a Purpose.[251] I disagree with this. A purpose has to be specified enough that a data subject is able to assess precisely how the data will be processed. Having analysis and exploration of the data as the purpose would not allow the data subject to determine this, as it would open the door for any kinds of processing. The fact that the legislator saw the need to allow for a broader possibility to specify a purpose at scientific collection as exception speaks for the fact that such a purpose is generally not allowed as a rule.

It is important to note that the principle of Purpose Specification is not new in the GDPR, but was already included in the DPD in Article 6.1(b). The

[251] Moerel and Prins (n 11) 44.

Purpose Limitations discussed in this analysis therefore already apply and applied even before the GDPR enters into force.

The prevalent method of circumvention used by the large companies seems to be that they specify a purpose that is as broad as possible.[252] Under the DPD there is not much case law concerning the application of the principle of Purpose Specification to these companies. A reason may be that the protection of Personal Data was not as strong under the DPD as it will become under the GDPR. Therefore, it is very possible that the increased powers of the National Supervisory Authorities, including the increased sanctions, the fact that the GDPR is directly applicable in all member states and does not have to be implemented, and the general strengthening of the Personal Data protection by the GDPR will lead to increased activity in enforcing Purpose Limitation.

It will be interesting to see how the Supervisory Authorities and courts will rule on for example the new Privacy notice by Facebook that Facebook supposedly adjusted to the GDPR. Facebook specifies as a purpose for the collection of Personal Data to do "Product research and development", in order to "develop, test and improve" their products.[253] According to my analysis, this is unlikely to be specified enough under DPD and GDPR, since it allows Facebook to conduct almost any research on the Personal Data. Data subject therefore cannot assess how their data will be used.

However, companies often *do* have a purpose to collect some data about their customers, for specific purposes. For example, Facebook allows users to upload images for the purpose of showing these images to Facebook friends. As long as this purpose is specified, explicit and legitimate, the collection is legal.

If Facebook however wants to use these pictures to determine the age of a data subject, in order to target ads, this is another issue: In the next part, I will analyze to which extent it is possible for data controllers to go beyond the initial purpose and unlock the latent value of the data that they already possess.

[252] Mayer-Schonberger and Padova (n 21) 322.
[253] 'Data Policy' (*Facebook*, 19 April 2018) <https://www.facebook.com/about/privacy?ref=new_policy> accessed 22 May 2018.

5.2 Change of Purpose after collection

5.2.1 Introduction

In this section, I will analyze if a controller can change the purpose of processing Personal Data after the initial collection. I will therefore assume that the data has already been collected, under a specific purpose, as described in the previous chapter.

A typical situation in Big Data Profiling could be: A social network has collected Personal Data in the form of webpages that a user has liked. This data was collected with the purpose of showing ads from the liked pages on the user's newsfeed. We assume that this purpose was specified, explicit and legitimate, and that the collection had a valid legal basis. Now, at some point, the social network realizes that it could potentially also use the data of liked pages to find potential romantic matches for the user, based on common interests with other users. The social network therefore wants to mine the collected Personal Data of all users in order to identify correlations and then apply this data to show users people that they might be interested in dating.

In this case, the Personal Data has already been collected in compliance with the Purpose Specification principle. The question is now if and under which circumstances the controller can use the same data for the new, unrelated purpose. This is answered by the second part of the Purpose Limitation principle in Article 5.1(b) GDPR, called Compatible Use. This is the text that specifies it:

"and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');"[254]

According to the Working Party, every act of processing after the initial collection is further use. Thus, the controller has to make a determination for each processing step whether it is a use that is not incompatible with the purpose for which the data was originally collected.[255] The double negative formulation in Article 5.1(b) ("not be considered incompatible") is likely a way to increase the leeway available to controllers.[256]

[254] General Data Protection Regulation (n 13) article 5.1(b).
[255] 'Opinion 03/2013 on Purpose Limitation' (n 115) 21; Forgó, Hänold and Schütze (n 23) 29.
[256] Schantz (n 204) Rn. 18; Frenzel, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' (n 109) Rn. 30; Forgó, Hänold and Schütze (n 23) 29.

If a further processing is encompassed by the initial purpose, the further processing is obviously compatible.[257] However, there are some instances where a controller might want to process data for a different purpose than it was initially collected for. This situation is very relevant in a Big Data Profiling situation due to the latent value all data possesses. One piece of data is likely to have many different uses.

The GDPR has in Article 5.1(b), Article 6(4) and according to recital 50 four possibilities of further processing data beyond the initial collection:

- Consent
- A law that safeguards certain principles
- Assumed compatibility due to privileged purposes, such as statistics
- The processing is compatible with the initial purpose

While the compatible use principle exists in the DPD, there are some new provisions in the GDPR here compared to the DPD. The possibility of further processing under consent and laws that safeguard certain principles are new under the GDPR.[258] Under the DPD, consent was taken into consideration in the compatibility assessment but did not automatically guarantee compatibility.[259] The GDPR thus makes it easier to process data further if consent is obtained for the new processing.

While the DPD requires the implementation of safeguards for archiving, research and statistical processing in order for the exemption to be valid.[260] This has been replaced with a reference to Article 89 in the GDPR, which elaborates on required safeguards and allows member states to introduce deviations from the GDPR in these types of processing.[261] However, this does not seem to be a big difference in practice.[262]

The GDPR has also explicitly incorporated the factors in assessing compatibility into the regulation.[263] However, they overlap with the factors specified by the Working Party on how to interpret the compatibility assessment under the DPD.[264] Therefore, there does not seem to be a change in the method from the DPD here.

---

[257] 'Opinion 03/2013 on Purpose Limitation' (n 115) 22; Forgó, Hänold and Schütze (n 23) 29.
[258] Data Protection Directive (n 19) recital 29-30; General Data Protection Regulation (n 13) recital 50.
[259] 'Opinion 03/2013 on Purpose Limitation' (n 115) 27.
[260] Data Protection Directive (n 19) article 6.1(b).
[261] General Data Protection Regulation (n 13) article 89.
[262] Detlev Gabel and Tim Hickman, 'Chapter 6: Data Protection Principles – Unlocking the EU General Data Protection Regulation' (*White & Case*, 22 July 2016) <//www.whitecase.com/publications/article/chapter-6-data-protection-principles-unlocking-eu-general-data-protection> accessed 14 May 2018.
[263] General Data Protection Regulation (n 13) article 6(4); Forgó, Hänold and Schütze (n 23) 35.
[264] Moerel and Prins (n 11) 52; Forgó, Hänold and Schütze (n 23) 35.

5.2.2 Consent

The first possibility of processing data for a new purpose after collection is in cases where the data subject has consented to processing for a new purpose.[265] If the data subject has consented to processing for a new purpose, it will be aware of how the data can be used, and thus needs no additional protection by the Purpose Limitation Principle.[266] Consent needs to be freely given, specific, informed and unambiguous.[267] These requirements are a lot stricter than under the DPD.[268]

Freely given implies that the controller cannot deny the user service if he refuses to give consent for a processing, if a specific processing is not necessary for the service. While this is still unclear, it could potentially mean that online services would have to offer service even to people who refuse the processing of their data.[269]

Specific means that consent needs to be given for a purpose that is specific enough to comply with the requirements in the Purpose Specification principle.[270] It also means that consent cannot be grouped together. If there is data that is to be collected for different purposes, the data subject has to be able to give consent to each of these separately.[271]

Informed means that the data subject has to be made aware of a number of things, including the identity of the controller, the purpose of the processing and what type of data will be used.[272] It also includes that the language used is clear enough that the average person can understand it and that it is not buried in a long terms and conditions document.[273]

Unambiguous means that consent needs to be given through a clear affirmative act, such as ticking a box or otherwise opting in. Silence or passivity cannot be seen as such an act.[274]

In some cases, when the processing is performed on data belongs to the special categories of Personal Data according to Article 9 GDPR (such as

---

[265] General Data Protection Regulation (n 13) article 6.4; Voigt and von dem Bussche (n 14) 108; Schantz (n 204) Rn. 22; General Data Protection Regulation (n 13) recital 50; Forgó, Hänold and Schütze (n 23) 35.
[266] Schantz (n 204) Rn.22.
[267] General Data Protection Regulation (n 13) article 4(11).
[268] Wendleby and Wetterberg (n 87) 52.
[269] Voigt and von dem Bussche (n 14) 95–96; Schantz (n 117) 1845; Wendleby and Wetterberg (n 87) 53.
[270] 'Guidelines on Consent under Regulation 2016/679' (Article 29 Data Protection Working Party 2018) wp259rev.01 11–12 <http://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030> accessed 7 May 2018, see below under chapter 6.
[271] ibid 12; Wendleby and Wetterberg (n 87) 53.
[272] 'Guidelines on Consent under Regulation 2016/679' (n 270) 13.
[273] ibid 14; General Data Protection Regulation (n 13) article 7(2).
[274] General Data Protection Regulation (n 13) recital 32; Voigt and von dem Bussche (n 14) 94; 'Guidelines on Consent under Regulation 2016/679' (n 270) 16–17.

data revealing political opinion or health life), normal consent is not enough (Article 9.2(a) GDPR). In these cases, the data subject has to give *explicit* consent. Explicit consent can be given, for example, through a written statement or through filling out an electronic form.[275]

The data controller bears the burden of proof that consent was obtained.[276] The data subject shall also have the right to withdraw consent at any time.[277]

5.2.2.1 Analysis

The first question to answer is whether the company can let the user consent to any further use of the user's Personal Data. Consent for a new purpose means that processing is allowed under the Purpose Limitation principle. Therefore, if a Big Data Profiling company receives consent from the user, this should make it possible for the corporation to process the data in the new way.

As mentioned, consent needs to be freely given, which prevents the controller from denying the user further use of the service if he refuses to agree to the further use. It is possible that a large number of users might refuse the processing. In a Big Data Profiling context, this might be problematic. Big Data Profiling relies on having massive amounts of data available to create accurate models. An exclusion of a large portion of the data available to a Big Data Profiling company could therefore lower the value of such an analysis.

Consent also needs to be specific. This includes that the purpose of the use of the data needs to comply with the requirements described above, about Purpose Specification. Again, the data subject needs to understand which kinds of further processing will be performed. If the Big Data Profiling company knows what the data will be used for, for example for assessing interests of an individual, asking for consent to cover the further processing should be possible. However, one of the features of Big Data is that the data often speaks for itself. Before a preliminary analysis of the data, it might not be possible to see in which ways the data could be used. However, in order to conduct this preliminary analysis, specific consent is required. Since the data controller does not know how the processing will look, he would not be able to request specific consent.[278]

In some instances, explicit consent might be required by the data subject. This is the case if the data processed reveals intimate details about the data subject according to Article 9 GDPR. Since Big Data Profiling is able to infer even sensitive attributes from Personal Data, a lot of data could be seen to reveal data belonging to the special categories, which puts high

[275] 'Guidelines on Consent under Regulation 2016/679' (n 270) 18–19.
[276] Wendleby and Wetterberg (n 87) 51.
[277] General Data Protection Regulation (n 13) article 7.
[278] compare Zarsky (n 4) 1006; Forgó, Hänold and Schütze (n 23) 32.

requirements on Big Data Profiling companies for obtaining consent. For an exploration of the effect on Big Data Profiling on Personal Data in the special categories, see chapter 5.2.5.2 (under "Potential Impact").

In conclusion, this means that consent is a strong possibility for a data controller to enable further use. This is the case if the controller knows in which ways the data will be used and is able to specify this. However, it might be hard to ask for consent if the data itself is needed to determine the purpose of the processing, which can often be the case in a Big Data Profiling company.

5.2.3 A law that safeguards certain principles

The second case where the change in purpose is allowed is when it is motivated by Union or Member state law that safeguards principles in Article 23. These include national security, defense and the enforcement of civil law claims.[279]

I see this being difficult to use to motivate profiling of individuals. First of all, a Member state or the Union has to create a law that legitimizes a change of purpose of the processing. Secondly, the objectives seem mostly concerned with the protection of the public. 23.1(i), which is concerned with the protection of the data subject, could be used to motivate psychological profiling, similar to the one performed by Facebook, aimed at detecting people with risk for suicide. However, while these are uses of Big Data Profiling, they cannot be used to allow the profiling for purposes of marketing. They also rely on additional legislation. Therefore, this legal ground will not apply to Big Data Profiling companies.

5.2.4 Privileged purposes

Processing for a new purpose is, according to Article 5.1(b) GDPR also allowed when the processing is compatible with the initial purpose. The GDPR posits that further processing shall be considered compatible if it is performed "for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes".[280] In the Big Data Profiling context, the statistical purposes exception seems to be the most relevant.[281]

Statistical purpose implies that the result of the processing is aggregated and that the result is not used to support measures or decisions regarding any

[279] General Data Protection Regulation (n 13) article 6.4; Schantz (n 204) Rn. 22; Voigt and von dem Bussche (n 14) 109; General Data Protection Regulation (n 13) recital 50.
[280] General Data Protection Regulation (n 13) article 5.1(b).
[281] Forgó, Hänold and Schütze (n 23) 30.

particular data subject.[282] The working party believes these measure or decisions to cover any impact on an individual, whether positive or negative.[283] The statistical purpose can however be used by Big Data firms in a commercial sense, for market research.[284] The important factor is that "functional separation" is maintained. This is the concept that data used for statistical purposes cannot be used in regards to individuals. A key tool in achieving this is pseudonymization or partial automation of the data before statistical processing.[285] Additionally, other safeguarding measures, such as encryption and a limitation of who can access the data, should be taken to ensure the separation.[286]

The GDPR does not itself specify how the safeguards should be implemented, but it leaves it up to further regulation from the Union or member states to introduce them.[287] Unlike the directive, which requires the safeguards to be in place for the statistical exception to be valid,[288] the GDPR does not make the exception conditional. It even allows member states to introduce legislation derogating from some GDPR principles during statistical processing.[289]

5.2.4.1 Views in the Literature

The Working Party considered two separate situations in Big Data processing.

(1) In the first situation, the controller wants to find general trends in the data. Here, the processing can fall under the statistical exception. It is then crucial that the functional separation guarantees that the data is not used to affect decisions relating to individuals, which has to be guaranteed by organizational and technical measures, such as pseudonymization. As long as this is done, further processing should be allowed.[290]

(2) If, on the other hand, the data is used to predict attributes about an individual and then inform measures regarding that decision, such as which

[282] General Data Protection Regulation (n 13) recital 162; Forgó, Hänold and Schütze (n 23) 36.
[283] 'Opinion 03/2013 on Purpose Limitation' (n 115) 30; Forgó, Hänold and Schütze (n 23) 30.
[284] 'Opinion 03/2013 on Purpose Limitation' (n 115) 29.
[285] ibid 30.
[286] ibid 32.
[287] General Data Protection Regulation (n 13) article 89, ibid recital 162; Forgó, Hänold and Schütze (n 23) 37.
[288] Data Protection Directive (n 19) article 6.1(b).
[289] General Data Protection Regulation (n 13) article 89.2.
[290] 'Opinion 03/2013 on Purpose Limitation' (n 115) 46; Paolo Balboni and Theodora Dragan, 'Big Data - Legal Compliance and Quality Management' in Kuan-Ching Li, Hai Jiang and Albert Y Zomaya (eds), *Big Data Management and Processing* (CRC Press 2017) section 1.3.1.2; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 38.

ads to show, the statistical purpose exception does not apply. The controller then has to use one of the other ways of changing the purpose, such as consent (see 5.2.2). The Working Party also believes that data subjects have to be given access to their profiles, including the derived data, and an explanation of how the logic behind the decision works. The Working Party also recommends the data to be made portable, so that users are able to export their data and use it with another service. They also have to be given the option to correct the data.[291]

Zarsky believes that the statistical use exception does not cover the applications of Big Data to individuals. This is due to the fact that the statistical use exception cannot be used to tailor experiences to individuals.[292] Mayor-Schonberger and Padova as well as Forgó et al also acknowledge that the statistical purpose does not cover measures directly affecting individuals.[293]

5.2.4.2 Analysis

The third possible ground for further use of data is the statistical purpose in the GDPR. If a processing is performed for a statistical purpose, it is automatically considered compatible further processing. The important point here is that the result of the data cannot be used to tailor experiences to individuals. It can, however, be used to perform market research and find general correlations.

These general correlations can be very useful and are surely an important use of Big Data. The Working Party mentions using Big Data to perform market analytics etc. An example of this could be a site implementing analytics systems that allow it to track how users interact with the site. This data can then be used to determine issues in the page flow or where users get confused. The important thing here is functional separation. Through security measures such as pseudonymization, it has to be ensured that the data cannot be used to perform measures regarding individuals. Mayor-Schonberger and Padova are optimistic about this way of using Big Data.[294] Classification under the statistic exception is not a silver bullet. The controller still needs to identify a legal basis for the processing and comply with the safeguards mentioned to ensure functional separation and with any potential additional legislation implemented by the member states. It will be interesting to see how the member states handle this and what effect the differing legislation might have on cross-border Big Data processing.

---

[291] 'Opinion 03/2013 on Purpose Limitation' (n 115) 46–47; Balboni and Dragan (n 290) section 1.3.1.2; 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (n 6) 38.
[292] Zarsky (n 4) 1008.
[293] Mayer-Schonberger and Padova (n 21) 327; Forgó, Hänold and Schütze (n 23) 39.
[294] Mayer-Schonberger and Padova (n 21) 331.

I agree with Zarsky, who is pessimistic about the use of the exception. The most desirable use of Big Data is creating profiles that are able to distinguish users and tailor experiences to them. This could be, for example, a site implementing an analytics system and then using it to suggest products a user might be interested in. This is clearly a measure targeted at individuals and would therefore fall outside of the statistical purpose.

An interesting question is whether it is allowed for controllers to perform data mining under the statistical exception without applying this to individual users. While complying with the appropriate safeguard, this could enable a controller to build a model from Personal Data. Once it is clear how the model works, the controller could then request specific consent from users and apply the model to create profiles of them for example to target ads. One could see the model created as a statistical result, correlating some values with others. "People who have visited Site A are more likely to be interested in Science" could be seen as a statistical result, but could also be used to create a profile about a particular person. The Working Party seems to hint to the possibility for individuals to specifically authorize the use of statistical results with regards to them.[295] However, recital 162 seems to close this door, since it states that the statistical purpose itself implies that Personal Data are not used in measures regarding a person. Even if the result will only be used on people after their consent, it therefore seems like the eventual use of data in this way falls outside of the statistical purpose. Nonetheless, I find this possibility the most plausible future way to enable the experimentation and creation of models that could be valuable both for the user and the controller.

In conclusion, while the statistical exception seems to be able to support some Big Data processing, such as the general detection of market trends, it is explicitly designed to not cover data that can be used to affect measures or decisions taken towards individual users, i.e. it is not applicable to the Big Data Profiling that is being examined in this thesis.

5.2.5 Compatible processing

If none of the exceptions above are fulfilled, a general compatibility assessment has to be performed. The Working Group suggested a number of factors that should be taken into account to determine compatibility. These factors have been incorporated into the GDPR in Article 6(4) in a slightly modified form.[296] However, they are only factors, which means that there is

[295] 'Opinion 03/2013 on Purpose Limitation' (n 115) 30.
[296] General Data Protection Regulation (n 13) article 6.4, see also ibid recital 50; Forgó, Hänold and Schütze (n 23) 34–35.

considerable legal uncertainty in how these will be applied in individual cases.[297]

(1) The first factor to be considered is the connection between the initial purpose for collection and the purposes for further processing. In cases where the processing is already encompassed by the initial purpose, this makes the purposes more likely to be compatible. If the purposes are very divergent, they are more likely to be incompatible.[298]

(2) The second factor that has to be considered is the context of the collection and the reasonable expectations of the data subject on what will be done with the data later on. If processing is surprising or objectionable to the data subject, it is more likely to be incompatible. Here, the nature of the relationship between the data subject and the controller also has to be considered. If the processing goes beyond what is customary, or there is a strong imbalance of power between the controller and the data subject, the further processing is less likely to be compatible.[299] It should be noted that the reasonable expectations requirement has been specifically removed from the GDPR as compared to the guidelines set out in the WP29 Opinion. This has been interpreted as being an attempt to make the analysis more objective.[300] However, according to the recitals, the reasonable expectations of the data subject should still be taken into account.[301]

(3) The third factor is the nature of the data and the potential impact that further processing could have on the data subject. The more sensitive the data is, the more restrictive the allowed further processing is likely to be. If the processing can have a large impact on the data subject, such as by public disclosure of the data or combining it with other data in profiling, processing is less likely to be allowed.[302] According to the GDPR, the sensitivity of the data shall be pursuant to Article 9, which contains a set of special categories of data that is deemed extra sensitive due to the potential to harm the individual if revealed.[303] Data belonging to these categories is

"personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the

[297] Eike Michael Frenzel, 'DS-GVO Art. 6 Rechtmäßigkeit Der Verarbeitung' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018) Rn. 46, 48; Marion Albers, 'DS-GVO Artikel 6 Rechtmäßigkeit Der Verarbeitung' in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2017) Rn. 69.
[298] 'Opinion 03/2013 on Purpose Limitation' (n 115) 23–24; General Data Protection Regulation (n 13) article 6.4(a); Voigt and von dem Bussche (n 14) 109.
[299] 'Opinion 03/2013 on Purpose Limitation' (n 115) 24–25; General Data Protection Regulation (n 13) article 6.4(b); Schantz (n 204) Rn. 21; Voigt and von dem Bussche (n 14) 109.
[300] Moerel and Prins (n 11) 52–53.
[301] General Data Protection Regulation (n 13) recital 50.
[302] 'Opinion 03/2013 on Purpose Limitation' (n 115) 25–26; General Data Protection Regulation (n 13) article 6.4(c)-(d); Frenzel, 'DS-GVO Art. 6 Rechtmäßigkeit Der Verarbeitung' (n 297) Rn. 49; Voigt and von dem Bussche (n 14) 109.
[303] Frenzel, 'DS-GVO Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' (n 211) Rn. 6; Voigt and von dem Bussche (n 14) 111.

processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"[304]

(4) The fourth and final factor that should be considered in the opinion of the Working Group is the additional safeguards that the controller has implemented for further processing. If the safeguards are very strong, this can in some ways compensate for the incompatibility of the purpose. These safeguards can be, for example providing additional information to the data subject through notices and implementing privacy and security enhancing measures such as encryption, pseudonymization or aggregating the data.[305]

This list is not supposed to be exhaustive.[306]

5.2.5.1 Views in the Literature

The Working Party has set out a number of examples on how the compatibility assessment could be applied to a Big Data Profiling controller: In the first example of the Working Party, the controller uses customer purchasing data to target ads based on the pregnancy status of a customer. The Working Party believes that there is a strong indication of incompatibility here. This is due to the fact that many customers would find the analysis unexpected, inappropriate and objectionable. The algorithm used is secret and objectionable, and there are no safeguards in place, such as transparency around the algorithm or a clear consent from the user.[307]

The Working Party also describes a case where similar processing could be seen as compatible: In this example, a lawn mower company uses previous purchasing patterns to send promotional material relating to lawn mowers. The company uses an algorithm incorporating data about purchasing patterns to send offers right around the time when a previously purchased product is likely to need replacement. The company is very open with the way it plans to process this data and offers the customer insight into the way the data is collected and processed. Since the processing falls within the reasonable expectations of the customer and the customer is given the option of opting out, this processing is likely compatible. Gardening equipment purchases is also less sensitive data than the pregnancy status of a person.[308]

---

[304] General Data Protection Regulation (n 13) article 9.1.
[305] 'Opinion 03/2013 on Purpose Limitation' (n 115) 26–27; Frenzel, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' (n 109) Rn. 31; General Data Protection Regulation (n 13) article 6.4(e); Frenzel, 'DS-GVO Art. 6 Rechtmäßigkeit Der Verarbeitung' (n 297) Rn. 50; Voigt and von dem Bussche (n 14) 109.
[306] 'Opinion 03/2013 on Purpose Limitation' (n 115) 27.
[307] ibid 61.
[308] ibid 62–63.

In another opinion about behavioral advertising, the Working Group points out that data collected for one purpose cannot be enriched by other information about that individual. In these cases, additional consent needs to be obtained.[309]

Zarsky believes that the general compatibility assessment in article 6(4) of the GDPR is not sufficient to allow the use of Big Data.[310] Since Big Data is collected in a wide variety of contexts, he argues that it is difficult to assess compatibility based on the context factor. The nature of the Personal Data, which is also part of the assessment, is constantly in flux with Big Data. Lastly, pseudonymization, which is the suggested safeguard, removes much of the information contained in the data. All this contributes to making the Purpose Limitation principle a hinder to Big Data use. Zarsky advises applying the limitation narrowly.[311]

5.2.5.2 Analysis

In light of the above, the question is whether the notion of compatible processing can allow the use of Big Data Profiling. Of course, a specific determination will to a large degree always depend on the facts in the individual case. There is considerable legal uncertainty how exactly this analysis will be applied. However, I will try to provide a general overview over how the Big Data Profiling properties will affect the compatibility assessment.

Proximity of purposes

The first factor is as we have seen the proximity of the two purposes of the initial collection, and the further processing of Personal Data. The closer they are together, the more likely that the purposes will be compatible. This factor obviously depends on which purpose was initially specified, and to which level. The broader the purpose is, the easier it is for a new processing to be covered by the initial purpose, or to be found compatible with it. However, as I have described above, the nature of Big Data Profiling means that the purpose will initially have to be specified to a very detailed level. Simply having "marketing" as a purpose will likely not be sufficient, but the exact ways in which the data will be fed into algorithms and how these algorithms will work has to be described by the controller. This means that it is more difficult for further processing to fall under the same purpose.

Additionally, one of the ideas behind Big Data Profiling is that the same data can be reused in many different contexts and ways. Data collected to ship an order might be used to predict other products a user might be

[309] 'Opinion 2/2010 on Online Behavioural Advertising' (n 29) 20.
[310] Zarsky (n 4) 1008.
[311] ibid 1008–1009.

interested in. These are two very different purposes, and if the purpose of advertising was not specified at the start, this indicates incompatibility.

Context

The second factor in the compatibility assessment is the context of the collection and the reasonable expectations of the user. Despite being removed from the GDPR main text, I believe that the reasonable expectation of the user has an important role to play in determining the scope of further use. This is clear in the example of the store using purchasing patterns to predict the pregnancy status of an individual. It lies far outside of the expectations of the customer that such data would be re-used with the new purpose to predict intimate details. When, in the other example, data used by a lawn mower company to send new offers at a time when the products were likely being needed, this was not surprising to the customer. It was also clear from the context that such an analysis was performed.

When applying the same kind of analysis to Big Data Profiling companies such as Google and Facebook, we see that the data is often used in ways that might be considered surprising. For example, when searching for something on Google, there will be ads related to that search on the side of the results. This does not seem very surprising, as the connection between the data (the search query) and the ads is very close. However, in cases where a user visits a website that contained a hidden Google Tracker, and later receives ads for products by that company when visiting another website, this processing seems a lot more surprising. The context for the data collection was the initial visiting of that website, and it was likely not clear to the customer that this data was collected at all and that it would be continued to be used to show ads on another website. The processing goes beyond what is customary for data generated when visiting a site, which increases the chances of the further processing being incompatible.

Facebook performs similar processing related to ads and also uses information about people to suggest connecting to people they are likely to know. The suggestions are often eerily accurate and seem to be based on location data or appearing in other people photos etc. This is also a form of processing that can be very surprising. When checking Facebook at a meeting, the user would not expect the location data to be used to correlate with all other people who are at that meeting and estimate probabilities of them knowing each other.

Another aspect of the context question is whether there is a power imbalance between the user and the controller. In the cases of Google and Facebook, both of these companies have huge market powers and likely stand between the user and human interaction with their peers, via Facebook chat or Gmail. It is very hard to move away from these services or change to another. This also speaks for a stricter approach regarding incompatibility of a further processing of collected data.

Zarsky argues that the context aspect is hard to apply to Big Data companies, since Big Data calls for the processing of a lot of different data from different contexts.[312] However, this does not seem to be an issue to me. The fact that data is collected and combined between different contexts makes it harder for the data subject to assess how the data is used. It therefore makes sense to place restrictions on such processing.

Potential impact

The third factor is the sensitivity of the data and the potential impact that further processing could have on the data subject. This shall be assessed with regards to the special categories of personal information in Article 9. Not only data which directly contains the information in the protected categories is included, but also data which "reveals" such data. The Working Party means that not only data that directly contains one of the special categories is protected, but also data from which data in the special categories can be concluded.[313]

Zarsky believes that Big Data will render the distinction between regular Personal Data and the special categories meaningless, since it is possible to deduce the special categories from the regular data.[314] He mentions the example of health data being revealed by purchase history. This means that the special categories will grow and grow to consumer more and more regular data categories.[315] Additionally, the discrimination that the special categories aim to prevent does no longer occur along the straight lines of the categories, but instead in a data-driven way, making the distinction moot even as a symbol.[316] Frenzel also argues that the distinction is very hard to make due to ways of processing data in context.[317]

Moerel and Prins have a similar view on the special categories of Personal Data: They mention the example of Facebook being able to predict sexual preferences, and the data broken Axciom being able to predict diabetes based on purchasing patterns.[318] As we have seen previously, Big Data is very efficient at profiling individuals. Kosinski and other researchers were in 2012 able to prove how a simple analysis of Facebook likes was able to determine very intimate details about the users. They were able to predict sexual orientation to an accuracy of 88%, skin color with 95% accuracy and

---

[312] ibid 1008.
[313] 'Advice Paper on Special Categories of Data ("Sensitive Data")' (Article 29 Data Protection Working Party 2011) 6 <http://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf> accessed 17 May 2018.
[314] Zarsky (n 4) 1013.
[315] ibid.
[316] ibid 1014.
[317] Frenzel, 'DS-GVO Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' (n 211) Rn. 8.
[318] Moerel and Prins (n 11) 56.

political affiliations to an accuracy of 85%.[319] They also hint that even more analyses, based on other user patterns and revealing other attributes, would be possible given the right training data.[320]

I believe this means that almost any Personal Data in Big Data Profiling can have a connection to the sensitive special categories of Personal Data according to Article 9. Both Facebook and Google process data that could, using the Big Data Profiling tools available today, give information about attributes of people in the special categories of Personal Data. Therefore, any further processing of this data will make processing less likely to be compatible, since the potential impact of such data could be disastrous for the data subject.

Big Data Profiling companies are also likely to want to combine data from different sources to paint a more complete picture of a user. This is mentioned by the Working Party opinion as also increasing the threshold for compatibility.

Compensation through safeguards

The first three factors suggest that it might be very difficult for a Big Data Profiling company to argue that their further processing is compatible with the initial purpose of collection. However, the Working Party suggests additional safeguards and transparency measures implemented can make processing more likely to be compatible with the initial processing, and in some ways compensate for incompatibility. These measures can include organizational and technical procedures, such as partial anonymization, as well as measures designed to enhance transparency with regards to the data subject. It can also include the option of allowing the individual to opt-out of the processing.

How this can work in practice can be seen in the example of the lawn mower company explained by the Working Party: Here, a company uses Big Data Profiling to offer personalized coupons. However, the company has implemented safeguards that seem to make the processing compatible. For example, it gives the customers access to a portal that allows them to view the data about them and how the algorithm works. It also offers customers the possibility of downloading the data about themselves.

The key factor seems to be that the company has made significant efforts to ensure transparency to customers and give them a choice. I believe this applies to all Big Data Profiling ventures If the customer can always see how his data is being collected and used, this guarantees that processing will not be surprising for the individual, and that the data subject has the possibility of objecting if the processing goes too far. In a Big Data Profiling world, I therefore believe companies will have to find new consumer-friendly ways of explaining the complicated algorithms involved

[319] Kosinski, Stillwell and Graepel (n 196) 1.
[320] ibid 4.

in the decision making, and which pieces of the huge data collections have which effects on the outcome. Due to the previously discussed factors, it seems that such measures will have to go very far to compensate for the other factors that lean towards t further processing being incompatible.

As mentioned before, deep learning makes transparency as a safeguard to enable further processing of collected data even more difficult. Due to the black-box nature of the models created, it is near impossible even for the data controller to understand the sophisticated and complicated correlations and connections between the Personal Data. This makes it difficult for the controller to provide the needed transparency to the data subjects to make further processing compatible.

The GDPR explicitly mentions pseudonymization as an additional safeguard.[321] Since the mining of data for correlations does not require the data subject to be identified, it is feasible to implement this safeguard for data mining. This would increase the likelihood that a processing is compatible.

5.2.6 Legal basis – alternative or cumulative requirement?

Now that we have assessed the ways of further processing according to the GDPR, we should look at the effects of a legality of further use under the principle of Purpose Limitation. If a further processing is compatible with the purpose of the initial collection, can the controller perform it without obtaining a legal basis according to Article 6 GDPR[322] that covers the further processing? Likewise, can an incompatible further processing still be performed if the controller obtains a new legal basis, such as being able to motivate the processing under a legitimate interest?

The question can be illustrated with an example: Say that Facebook collects the age of users for the purpose of limiting the content that young users may see on Facebook pages. Facebook later wishes to perform statistics on the age of users in a certain area. Since this is processing for a statistical purpose, it is assumed to be compatible. However, the consent that the user gave was only for the purpose of restricting content, and therefore does not cover this statistical use case. If the requirements are alternative, the compatibility of the processing means that no new legal basis will need to be identified. If the requirements are cumulative, Facebook will have to identify a new legal basis in order to perform the statistical processing. For example, it could argue that the further processing is necessary to fulfill a legitimate interest of the controller.

[321] General Data Protection Regulation (n 13) article 6.4(e).
[322] See chapter 3.6.

For the DPD, the answer is clear: In all opinions, the Working Party has confirmed that the requirements for compatibility of further processing and a legal basis for the processing are cumulative. Only if both are fulfilled can the processing proceed.[323]

During the legislative process around the GDPR, the commission introduced an exception from the Purpose Limitation Principle that would allow further processing even for incompatible purposes if the processing was covered under a legal basis.[324] This proposal was widely criticized by data protection activists and the Article 29 Working Party, who suggested this would hollow out the Purpose Limitation Principle and lower the GDPR protection level to under that of the DPD.[325] The Working Party suggested the removal of the provision and the introduction of a section detailing the test for incompatibility.[326] This has occurred in the final version of the GDPR.[327] It therefore seems like the legislator has stepped away from making the requirement for compatible processing and a legal basis alternative and aimed to maintain the cumulativeness of the requirements.

However, recital 50 in the GDPR still contains a provision saying that a new legal basis is not required if processing is compatible with the purpose.[328] Some authors argue that this is a leftover of the scrapped approach of the commission, and therefore an editorial error.[329] Others take the recital at face value and thus accept that no legal basis is required if the purpose for further use is compatible.[330]

---

[323] 'Opinion 03/2013 on Purpose Limitation' (n 115) 41; ibid 27.
[324] Moerel and Prins (n 11) 51; 'Opinion 03/2013 on Purpose Limitation' (n 115) 41–44; Albers (n 297) Rn. 73; Schantz (n 117) 1844.
[325] 'Press Release on Chapter II of the Draft Regulation for the March JHA Council' (Article 29 Data Protection Working Party 2015) <http://ec.europa.eu/justice/article-29/press-material/press-release/art29_press_material/2015/20150317__wp29_press_release_on_on_chapter_ii_of_the_draft_regulation_for_the_march_jha_council.pdf> accessed 24 April 2018; 'Key Aspects of the Proposed General Data Protection Regulation Explained' (European Digital Rights) <https://edri.org/files/GDPR-key-issues-explained.pdf> accessed 24 April 2018 section 2; Joe McNamee, 'Letter to President Juncker' (21 April 2015) <https://edri.org/files/DP_letter_Juncker_20150421.pdf> accessed 24 April 2018; 'Opinion 03/2013 on Purpose Limitation' (n 115) 41; 'Data Protection under Threat from EU Council Agreement' (*EDRi*, 15 June 2015) <https://edri.org/press-release-privacy-and-data-protection-under-threat-from-eu-council-agreement/> accessed 16 May 2018.
[326] 'Opinion 03/2013 on Purpose Limitation' (n 115) 41–44.
[327] General Data Protection Regulation (n 13) article 6(4).
[328] ibid recital 50.
[329] Schantz (n 117) 1844; Albers (n 297) Rn. 72-73.
[330] Inga-Lill Askersjö and others, 'Ny Dataskyddslag - Kompletterande Bestämmelser till EU:S Dataskyddsförordning' (Dataskyddsutredningen 2017) SOU 2017:39 235; Voigt and von dem Bussche (n 14) 110; Stefan Lövfen and Morgan Johansson, 'Regeringens Proposition 2017/18:105 - Ny Dataskyddslag' (Justitiedepartementet 2018) Prop. 2017/18:105 111 <http://www.regeringen.se/492373/contentassets/561c615d11104ad38c42b59cda9c33bc/ny-dataskyddslag-prop.-201718105> accessed 24 April 2018; ibid 123; Forgó, Hänold and Schütze (n 23) 37.

I agree with the assessment that the requirements are cumulative. This seems clear from the removal of a previous version of article 6(4) that placed the factors as alternatives. It also works better together with the system of the GDPR, since a legal basis is required for all processing.[331] It seems unlikely that the legislator would make such a significant change with only a recital. If this is the intention, it is very unfortunate that the sentence in the recital has been left in. This is likely to cause a lot of confusion around this important issue.

One exception from the requirement for a compatible purpose is consent. If the user consents to a further processing of his data, the new purpose does not need to be compatible, according to recital 50 and Article 6(4) GDPR. Consent also serves as a legal basis for processing. Therefore, in this instance, only consent is needed to enable further processing.

5.3 Conclusion

In this chapter, I have discussed how the principle of Purpose Limitation is likely to affect Big Data Profiling companies, such as Google and Facebook.

First, I focused on the first part of the principle, Purpose Specification. It requires controllers to only collect Personal Data for a specified, explicit and legitimate purpose. The requirement most likely being problematic for Big Data Profiling companies is that the purpose has to be specified in advance. This must allow a reasonable person to determine precisely which kind of processing is likely to be performed on the data. Big Data Profiling firms usually collect large amounts of information about a large number of people and after the collection decide on different purposes for the use of the Personal Data. Due to the size and power of the companies, their global activities and their large amounts of collected data, the requirements for specificity are even higher. It is likely not enough that such a controller specifies that the data will be used for marketing or to personalize experiences. Instead, a detailed break-down of how the algorithm will use the data and which factors are considered will have to be provided to the data subjects for their information at the time of collection.

Purpose Specification runs contrary to the concept of Big Data Profiling, which relies on collecting data and then using that data to create new algorithms and models to better target users. Much of this kind of purposeless collection, or collection for vague purposes, will likely have to stop.

However, in many situations Big Data Profiling companies do collect Personal Data covered under a specific purpose, but later realize that they may be able to use the same data for other purposes. If the controller was entirely bound by the initial purpose, such further processing would be

[331] General Data Protection Regulation (n 13) article 6.1.

impossible. The GDPR mentions three ways that allow the controller to further process the data in a way that is potentially not covered by the initial purpose.

The first such way is data subject consent to the further processing. In my opinion, this is the best way for Big Data Profiling companies to continue their operations under the GDPR. By asking for further consent in a clear and specified way, and also allowing the user to deny the request without any repercussions, the further processing is legal. However, this only works if the data controller already knows how the data will be processed to create a profile. Letting the data speak to discover new correlations is unlikely to be possible with the consent of the user, since the controller then is unable to request a permission that is specific enough.

The second way, interest of the public as specified in Article 23 is unlikely to apply to the common Big Data Profiling companies.

The third way is the notion of compatible processing: Some processing purposes, such as the statistical purpose, are assumed to be compatible. However, while general statistics can be fit under the statistical purpose as one example of compatible processing according to Article 5.1(b), it cannot be used to support measures regarding individuals. While certainly useful in some Big Data contexts, general statistics does not cover Big Data Profiling, which is intended for measures regarding individuals.

In other cases, a general compatibility examination requires taking a number of factors into consideration to determine if further processing is allowed without the data subject's consent. In my opinion, it is difficult for Big Data Profiling companies to fulfill this compatibility requirement. The controller is likely to want to use the data in many different context and for many purposes. Many of these might be considered surprising, such as ads for recently viewed products appearing. Finally, the data carries the potential to reveal intimate and sensitive information about the individual. All of these make compatibility less likely. To some extent, safeguards can be implemented by the controller to increase the chance for compatibility. These can be, for example, increasing transparency and predictability of the processing for the data subject by offering a detailed view of the data collected about the individual, explaining how and why the data will be further processed or pseudonymizing the data.

6 Concluding remarks

Some authors have argued that the GDPR will throw a wrench in the wheels for Big Data Profiling companies. Based on my analysis in this thesis, I tend to agree with this statement. While most of the provisions of the GDPR that I have analyzed (such as the Purpose Limitation Principle) are not new compared to the DPD, they are likely to be enforced much more efficiently and with severe consequences for companies that are not compliant.

First, I analyzed the effect of Big Data Profiling on the definition of Personal Data. I found that Big Data radically expands the notion of Personal Data. Even data that is only linked to an online identifier, such as a cookie placed in the user's browser, is Personal Data. Further, data is a lot easier to trace back to an individual using Big Data methods. This means that even data that might seem anonymous can possibly be traced back to the individual behind the data in combination with other available data or by developing intelligent algorithms that can infer Personal Data from available information. If a Big Data Profiling company wishes to anonymize Personal Data in order to be able to keep it and use algorithms to extract value from it, it therefore has to be very careful that the data cannot be reidentified. This might involve destroying parts of the data, thereby limiting its usefulness.

Then, I looked at the effect of the Purpose Limitation Principle on the collection of Personal Data in Big Data Profiling. I found that Big Data Profiling companies have to be very specific about for which purposes Personal Data is collected. This likely includes which data is used, and which attributes of the data subject are inferred. The data subject needs to be able to precisely tell which operations will be performed on the data and for which purpose.

I also looked at possibilities for the controller to deviate from this initial purpose when further processing previously collected data. There are several possibilities provided in the GDPR. One of these, and the most powerful, is consent of the data subject. If the controller obtains consent for the new purpose from the individual, further processing is allowed. It could however be difficult for Big Data Profiling companies to obtain consent before the specific purpose of the further processing is clear enough, since consent has to be specific.

Finally, a change of purpose for the further processing can be allowed if the further processing is seen as compatible with the initial purpose. The example of statistical purposes will not apply to Big Data Profiling which is typically intended for measures regarding individuals. The last possibility is to make a general compatibility assessment. This can be difficult for Big Data Profiling companies because they need to consider a number of different factors that can vary from case to case. The specifics of Big Data

Profiling also make it difficult and only leave a small area of compatible further processing.

It should be noted that even if processing is compatible with the Purpose Limitation principle, a legal basis has to be identified for the processing to be allowed. For example, if Personal Data is initially collected based on consent, and the controller wishes to further process the data in a compatible way, a legitimate interest of the controller could be identified to allow further processing. Due to what is likely an editorial error, this is difficult to understand from reading the GDPR. If the user consents to a further processing, no compatibility assessment has to be performed, since the processing is automatically compatible.

It is important to note that the GDPR has not yet entered into force, and that this analysis is thus slightly speculative in nature. While most of the provisions were already in force under the DPD, they do not seem to have found a wide-spread application. This could be due to varying implementations in member states and a lack of sanctions. It is likely that courts, legal doctrine, data protection agencies, maybe the legislator and other stakeholders over the next few years will develop a more comprehensive picture of how the Purpose Limitation Principle should be applied to Big Data Profiling.

However, the result of my analysis seems to paint a bleak picture for many Big Data Profiling ventures. Those who rely on a "collect before select" model of first collecting data, and then figuring out what to do with it, will have to change their practice and business model concerning Personal Data, or face the heavy sanctions of the GDPR. The purpose has to be clearly specified at the time of collection. Personal Data that has been collected under a specified purpose can also not be further analyzed, to identify new ways to use the data. This might be difficult even with consent, because the consent requires detailed information about how and why the Personal Data will be collected or further processed. The other possibilities offered by the GDPR seem unlikely to work in a Big Data Profiling operation.

Many of the current operations that focus on Personal Data are likely in breach of the Principle of Purpose Limitation. They will have to adapt or shut down or face the heavy sanctions of the GDPR. It will affect especially those that collect aimlessly and then try to find a purpose for the Personal Data. Operations that have a clear purpose for their profiling and are able to explain this to their users are not affected to the same degree. However, they are still hampered in their ability to reuse that data for other purposes. It will be very interesting to see the effect of this on the industry of targeted advertising. Some have claimed that the regulation will "pop the bubble" of advertising based on tracking. [332] The German chancellor Angela Merkel

[332] David Searls, 'GDPR Will Pop the Adtech Bubble' (*Doc Searls Weblog*, 12 May 2018) <http://blogs.harvard.edu/doc/2018/05/12/gdpr/> accessed 14 May 2018.

warned that a too strict regulation could hamper the development of Artificial Intelligence, turning it into "cows that are not fed".[333]

On the other hand, it will also lead to a world where users are able to clearly see which of their Personal Data will be used for which purposes, and where they are able to consent to or refuse specific uses of their Personal Data. No longer will people have to worry about their Personal Data being used for unexpected purposes and to invade their privacy by building intimate profiles of all aspects of their lives without their knowledge or comprehension. This transparency and assuredness that their privacy is protected could turn into an increased feeling of safety and control for the users, leading to increased trust and good-will for the companies running these operations. While maybe far-reaching, these regulations could therefore have a positive effect not only for the individuals, but also for the Big Data Profiling industry.

[333] Dietmar Neuerer, 'Datenschutzgrundverordnung: Merkel torpediert neue EU-Datenschutzregeln' *Handelsblatt* (11 May 2018) <http://www.handelsblatt.com/politik/deutschland/datenschutzgrundverordnung-merkel-torpediert-neue-eu-datenschutzregeln/21268426.html> accessed 14 May 2018.

7 Bibliography

Literature

Legal

Abrams M, ‘The Origins of Personal Data and Its Implications for Governance’ [2014] SSRN Electronic Journal <http://www.ssrn.com/abstract=2510927> accessed 9 May 2018

Albers M, ‘DS-GVO Artikel 6 Rechtmäßigkeit Der Verarbeitung’ in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2017)

Balboni P and Dragan T, ‘Big Data - Legal Compliance and Quality Management’ in Kuan-Ching Li, Hai Jiang and Albert Y Zomaya (eds), *Big Data Management and Processing* (CRC Press 2017)

‘Big Data, Artificial Intelligence, Machine Learning and Data Protection’ (Information Commissioner’s Office) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 20 February 2018

Borocz I, ‘Clash of Interests-Is Behaviour-Based Price Discrimination in Line with the GDPR’ (2015) 153 Studia Iuridica Auctoritate Universitatis Pecs Publicata 37

Ernst, ‘DS-GVO Art. 1 Gegenstand Und Ziele’ in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, ‘DS-GVO Art. 2 Sachlicher Anwendungsbereich’ in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, ‘DS-GVO Art. 3 Räumlicher Anwendungsbereich’ in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, ‘DS-GVO Art. 4 Begriffsbestimmungen’ in Paal and Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

Forgó N, Hänold S and Schütze B, 'The Principle of Purpose Limitation and Big Data', *New Technology, Big Data and the Law* (Springer, Singapore 2017) <https://link.springer.com/chapter/10.1007/978-981-10-5038-1_2> accessed 15 May 2018

Frenzel EM, 'DS-GVO Art. 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, 'DS-GVO Art. 6 Rechtmäßigkeit Der Verarbeitung' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, 'DS-GVO Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

——, 'DS-GVO Art. 83 Allgemeine Bedingungen Für Die Verhängung von Geldbußen' in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (2nd edn, 2018)

Ghani NA, Hamid S and Udzir NI, 'Big Data and Data Protection: Issues with Purpose Limitation Principle.' (2016) 8 International Journal of Advances in Soft Computing & Its Applications

Kamarinou D, Millard C and Singh J, 'Machine Learning with Personal Data' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2865811 <https://papers.ssrn.com/abstract=2865811> accessed 1 March 2018

Korling F, *Juridisk metodlära* (Mauro Zamboni ed, Studentlitteratur AB 2013)

Mayer-Schonberger V and Padova Y, 'Regime Change: Enabling Big Data through Europe's New Data Protection Regulation' (2015) 17 Colum. Sci. & Tech. L. Rev. 315

Moerel L and Prins C, 'Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2784123 <https://papers.ssrn.com/abstract=2784123> accessed 27 March 2018

Murphy MH, ‘Algorithmic Surveillance: The Collection Conundrum’ (2017) 31 International Review of Law, Computers & Technology 225

Rauhofer J, ‘Of Men and Mice: Should the EU Data Protection Authorities’ Reaction to Google’s New Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle’ (2015) 1 Eur. Data Prot. L. Rev. 5

Rubinstein IS and Hartzog W, ‘Anonymization and Risk’ 91 WASHINGTON LAW REVIEW 59

Schantz P, ‘DS-GVO Artikel 5 Grundsätze Für Die Verarbeitung Personenbezogener Daten’ in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2017)

——, ‘Die Datenschutz-Grundverordnung – Beginn Einer Neuen Zeitrechnung Im Datenschutzrecht’ [2016] Neue Juristische Wochenschrift 1841

Schild H-H, ‘DS-GVO Artikel 4 Begriffsbestimmungen’ in Heinrich Amadeus Wolff and Stefan Brink (eds), *BeckOK Datenschutzrecht* (23rd edn, 2018)

Temme M, ‘Algorithms and Transparency in View of the New General Data Protection Regulation’ (2017) 3 European Data Protection Law Review 473

Tene O and Polonetsky J, ‘Big Data for All: Privacy and User Control in the Age of Analytics’ (2013) 11 Northwestern Journal of Technology and Intellectual Property 38

Wendleby M and Wetterberg D, *Dataskyddsförordningen GDPR: förstå och tillämpa i praktiken* (Första upplagan, Sanoma Utbildning 2018)

Voigt P and von dem Bussche A, *The EU General Data Protection Regulation (GDPR)* (Springer International Publishing 2017) <http://link.springer.com/10.1007/978-3-319-57959-7> accessed 20 February 2018

Yeh C-L, ‘Pursuing Consumer Empowerment in the Age of Big Data: A Comprehensive Regulatory Framework for Data Brokers’ [2017] Telecommunications Policy <http://linkinghub.elsevier.com/retrieve/pii/S0308596117304743> accessed 21 February 2018

Zarsky TZ, ‘Incompatible: The GDPR in the Age of Big Data’ (2016) 47 Seton Hall L. Rev. 995

Zuiderveen Borgesius F and Poort J, ‘Online Price Discrimination and EU Data Privacy Law’ (2017) 40 Journal of Consumer Policy 347

Other

Datta A, Tschantz MC and Datta A, ‘Automated Experiments on Ad Privacy Settings’ (2015) 2015 Proceedings on Privacy Enhancing Technologies 92

Degeling M and Herrmann T, ‘Your Interests According to Google - A Profile-Centered Analysis for Obfuscation of Online Tracking Profiles’ [2016] arXiv:1601.06371 [cs] <http://arxiv.org/abs/1601.06371> accessed 17 May 2018

Hasan O and others, ‘A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case’ (IEEE 2013) <http://ieeexplore.ieee.org/document/6597115/> accessed 9 May 2018

Hurley M and Adebayo J, ‘CREDIT SCORING IN THE ERA OF BIG DATA’ (2016) 18 Big Data 70

Kosinski M, Stillwell D and Graepel T, ‘Private Traits and Attributes Are Predictable from Digital Records of Human Behavior’ (2013) 110 Proceedings of the National Academy of Sciences 5802

Kugler L, ‘The War over the Value of Personal Data’ (2018) 61 Communications of the ACM 17

Mayer-Schonberger V and Cukier K, *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight* (New and expanded edition, John Murray 2017)

Najafabadi MM and others, ‘Deep Learning Applications and Challenges in Big Data Analytics’ (2015) 2 Journal of Big Data 1

Narayanan A and Shmatikov V, ‘Myths and Fallacies of “Personally Identifiable Information”’ (2010) 53 Communications of the ACM 24

Sweeney L, ‘Simple Demographics Often Identify People Uniquely’ [2000] Carnegie Mellon University, Data Privacy Working Paper 3 34

Public Press

Article 29 Data Protection Working Party

‘Advice Paper on Special Categories of Data (“Sensitive Data”)’ (Article 29 Data Protection Working Party 2011) <http://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf> accessed 17 May 2018

‘Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679’ (Article 29 Data Protection Working Party) WP251rev.01 <http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826> accessed 28 February 2018

‘Guidelines on Consent under Regulation 2016/679’ (Article 29 Data Protection Working Party 2018) wp259rev.01 <http://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030> accessed 7 May 2018

‘Opinion 2/2010 on Online Behavioural Advertising’ (Article 29 Data Protection Working Party 2010) WP171

‘Opinion 03/2013 on Purpose Limitation’ (Article 29 Data Protection Working Party) WP 203 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> accessed 28 February 2018

‘Opinion 4/2007 on the Concept of Personal Data’ (Article 29 Data Protection Working Party) WP136 <http://collections.internetmemory.org/haeu/content/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 20 March 2018

‘Opinion 05/2014 on Anonymisation Techniques’ (Article 29 Data Protection Working Party 2014) WP126

‘Press Release on Chapter II of the Draft Regulation for the March JHA Council’ (Article 29 Data Protection Working Party 2015) <http://ec.europa.eu/justice/article-29/press-material/press-release/art29_press_material/2015/20150317__wp29_press_release_on_on_chapter_ii_of_the_draft_regulation_for_the_march_jha_council.pdf> accessed 24 April 2018

Sweden

Askersjö I-L and others, ‘Ny Dataskyddslag - Kompletterande Bestämmelser till EU:S Dataskyddsförordning’ (Dataskyddsutredningen 2017) SOU 2017:39

Lövfen S and Johansson M, ‘Regeringens Proposition 2017/18:105 - Ny Dataskyddslag’ (Justitiedepartementet 2018) Prop. 2017/18:105 <http://www.regeringen.se/492373/contentassets/561c615d11104ad38c42b59cda9c33bc/ny-dataskyddslag-prop.-201718105> accessed 24 April 2018

Legislation

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) [2016] OJ L119/1

Cases

Criminal proceedings against Bodil Lindqvist (Reference for a preliminary ruling from the Göta hovrätt (Sweden)) [2003] European Court of Justice Case C-101/01

Patrick Breyer v Bundesrepublik Deutschland (request for a preliminary ruling from the Bundesgerichtshof — Germany) [2016] European Court of Justice (Second Court) Case C-582/14

Treaties

Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Adopted 28 January 1981, Entered into Force 1 October 1985) ETS No.108 <https://rm.coe.int/1680078b37> accessed 4 May 2018

Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A(III) (UDHR)

Online Sources

Balakrishnan A, 'Facebook Earnings Q4 2017: ARPU' (*CNBC*, 31 January 2018) <https://www.cnbc.com/2018/01/31/facebook-earnings-q4-2017-arpu.html> accessed 3 May 2018

'Browser Tracking' (*me and my shadow*, 16 February 2017) <https://myshadow.org/> accessed 2 May 2018

Cadwalladr C, '"I Made Steve Bannon's Psychological Warfare Tool": Meet the Data War Whistleblower' *The Guardian* (18 March 2018) <http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> accessed 2 May 2018

Cadwalladr C and Graham-Harrison E, 'How Cambridge Analytica Turned Facebook "Likes" into a Lucrative Political Tool' *The Guardian* (17 March 2018) <http://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm> accessed 2 May 2018

Chadha R, 'Ad Trackers Are on More than 75% of Websites' (*eMarketer*, 8 January 2018) <https://www.emarketer.com/content/ad-trackers-are-on-more-than-75-of-web-pages> accessed 22 February 2018

‘Chart of Signatures and Ratifications of Treaty 108’ (*Council of Europe - Treaty Office*) <https://www.coe.int/en/web/conventions/full-list> accessed 4 May 2018

‘Common Statement by the Contact Group of the Data Protection Authorities of The Netherlands, France, Spain, Hamburg and Belgium’ (2017) <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/common_statement_16_may_2017.pdf> accessed 13 April 2018

Curran D, ‘Are You Ready? This Is All the Data Facebook and Google Have on You’ *the Guardian* (30 March 2018) <http://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy> accessed 3 April 2018

‘Data Protection under Threat from EU Council Agreement’ (*EDRi*, 15 June 2015) <https://edri.org/press-release-privacy-and-data-protection-under-threat-from-eu-council-agreement/> accessed 16 May 2018

Neuerer D, ‘Datenschutzgrundverordnung: Merkel torpediert neue EU-Datenschutzregeln’ *Handelsblatt* (11 May 2018) <http://www.handelsblatt.com/politik/deutschland/datenschutzgrundverordnung-merkel-torpediert-neue-eu-datenschutzregeln/21268426.html> accessed 14 May 2018

‘Data Policy’ (*Facebook*, 19 April 2018) <https://www.facebook.com/about/privacy?ref=new_policy> accessed 22 May 2018

Davies J, ‘Know Your Cookies: A Guide to Internet Ad Trackers’ (*Digiday*, 1 November 2017) <https://digiday.com/media/know-cookies-guide-internet-ad-trackers/> accessed 22 February 2018

D’Onfro J, ‘Alphabet Earnings Q1 2018’ (*CNBC*, 23 April 2018) <https://www.cnbc.com/2018/04/23/alphabet-earnings-q1-2018.html> accessed 3 May 2018

‘EU GDPR Information Portal’ (*EU GDPR Portal*) <http://eugdpr.org/eugdpr.org.html> accessed 2 May 2018

Gabel D and Hickman T, ‘Chapter 6: Data Protection Principles – Unlocking the EU General Data Protection Regulation’ (*White & Case*, 22 July 2016) <//www.whitecase.com/publications/article/chapter-6-data-

protection-principles-unlocking-eu-general-data-protection> accessed 14 May 2018

——, ‘Chapter 5: Key Definitions – Unlocking the EU General Data Protection Regulation’ (*White & Case*, 13 September 2017) <//www.whitecase.com/publications/article/chapter-5-key-definitions-unlocking-eu-general-data-protection-regulation> accessed 14 May 2018

Searls D, ‘GDPR Will Pop the Adtech Bubble’ (*Doc Searls Weblog*, 12 May 2018) <http://blogs.harvard.edu/doc/2018/05/12/gdpr/> accessed 14 May 2018

‘Google Chrome Privacy Notice’ (*Google Chrome*, 6 March 2013) <https://www.google.com/chrome/privacy/> accessed 2 May 2018

‘Google Privacy Policy: Main Findings and Recommendations’ (CNIL) <https://www.cnil.fr/sites/default/files/typo/document/GOOGLE_PRIVACY_POLICY-_RECOMMENDATIONS-FINAL-EN.pdf> accessed 13 April 2018

Hill K, ‘How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did’ [2012] *Forbes* <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> accessed 22 February 2018

——, ‘How Facebook Figures Out Everyone You’ve Ever Met’ (*Gizmodo*, 11 July 2017) <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691> accessed 22 February 2018

‘How Google Uses Your Data for Ads’ (*Google Ads*) <http://privacy.google.com/intl/en-GB_ALL/how-ads-work.html> accessed 14 March 2018

Issenberg S, ‘Cruz-Connected Data Miner Aims to Get Inside U.S. Voters’ Heads’ *Bloomberg.com* (12 November 2015) <https://www.bloomberg.com/news/features/2015-11-12/is-the-republican-party-s-killer-data-app-for-real-> accessed 15 May 2018

‘Key Aspects of the Proposed General Data Protection Regulation Explained’ (European Digital Rights) <https://edri.org/files/GDPR-key-issues-explained.pdf> accessed 24 April 2018

Krogerus M and Grassegger H, ‘The Data That Turned the World Upside Down’ [2017] *Vice Motherboard* <https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win> accessed 14 March 2018

Kwon D, ‘Can Facebook’s Machine-Learning Algorithms Accurately Predict Suicide?’ [2017] *Scientific American* <https://www.scientificamerican.com/article/can-facebooks-machine-learning-algorithms-accurately-predict-suicide/> accessed 15 March 2018

Lapowsky I, ‘Facebook Exposed 87 Million Users to Cambridge Analytica’ [2018] *WIRED* <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/> accessed 2 May 2018

Marr B, ‘What Is The Difference Between Artificial Intelligence And Machine Learning?’ [2016] *Forbes* <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/> accessed 22 February 2018

——, ‘How To Use Analytics To Identify Trends In Your Market’ [2016] *Forbes* <https://www.forbes.com/sites/bernardmarr/2016/08/16/how-to-use-analytics-to-identify-trends-in-your-market/> accessed 2 May 2018

McNamee J, ‘Letter to President Juncker’ (21 April 2015) <https://edri.org/files/DP_letter_Juncker_20150421.pdf> accessed 24 April 2018

Michael P and Bread W, ‘8 Ways Retailers Are Tracking Your Every Move’ *Time.com* (23 September 2016) <http://time.com/money/4506297/how-retailers-track-you/> accessed 2 May 2018

Plummer L, ‘This Is How Netflix’s Top-Secret Recommendation System Works’ [2017] *WIRED UK* <http://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like> accessed 13 May 2018

Rotella P, ‘Is Data The New Oil?’ [2012] *Forbes* <https://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/> accessed 21 February 2018

Smith C, ‘Facebook Users Are Uploading 350 Million New Photos Each Day’ (*Business Insider*, 18 September 2013) <http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9> accessed 3 April 2018

Solon O, ‘Google’s Ad Tracking Is as Creepy as Facebook’s. Here’s How to Disable It’ *the Guardian* (San Fransisco, 21 October 2016) <http://www.theguardian.com/technology/2016/oct/21/how-to-disable-google-ad-tracking-gmail-youtube-browser-history> accessed 22 February 2018

Suich A, ‘Getting to Know You’ [2014] *The Economist* <https://www.economist.com/news/special-report/21615871-everything-people-do-online-avidly-followed-advertisers-and-third-party> accessed 22 February 2018

Sutton J, ‘Demystifying the Role of Artificial Intelligence in Marketing and Advertising’ (*eMarketer*, 22 April 2016) <https://www.emarketer.com/Article/Demystifying-Role-of-Artificial-Intelligence-Marketing-Advertising/1013864> accessed 11 May 2018

‘Twitter Usage Statistics’ (*Internet Live Stats*) <http://www.internetlivestats.com/twitter-statistics/> accessed 3 April 2018

Weisberg J, ‘Bubble Trouble’ [2011] *Slate* <http://www.slate.com/articles/news_and_politics/the_big_idea/2011/06/bubble_trouble.html> accessed 4 May 2018

‘What Is Big Data?’ (*Gartner IT Glossary*) <https://www.gartner.com/it-glossary/big-data> accessed 21 February 2018