BIG DATA AND BEHAVIOURAL TARGETING IN THE CONTEXT OF THE GENERAL DATA PROTECTION REGULATION

IT Law Lab research paper

Tartu
2018
# Table of contents

## Abbreviations

## Introduction

1. Definitions of big data and behavioural targeting
   1.1. Definition of big data
   1.2. Definition of behavioural targeting
   1.3. Sources of big data in behavioural targeting
      1.3.1. First party data
      1.3.2. Second party data
      1.3.3. Third party data

2. The legal framework of data protection, big data and behavioural advertising
   2.1. The regulation of big data in international legislation
   2.2. The regulation of big data in the European Union
      2.2.1. Overview of the history of data protection legal framework in the European Union
      2.2.2. The legal framework of big data in the EU
   2.3. The regulation of big data in Estonian legislation

3. The issues with the GDPR, big data and behavioural targeting and the hypothetical solutions
   3.1. Applicability of the GDPR to big data and behavioural targeting
      3.1.1. How does big data containing behavioural data relate to personal data
      3.1.2. Are big data analytics and behavioural targeting in the scope of the GDPR?
   3.2. Privacy loss and transparency issues under Articles 13 and 14 of the GDPR
      3.2.1. Unprecedented privacy loss
      3.2.2. Transparency issues

## Summary

## Bibliography

  List of literature
  List of legal acts
Abbreviations

AI – artificial intelligence
Art – article
CJEU – the Court of Justice of the European Union
ECHR – European Convention of Human Rights
EU – European Union
GDPR – General Data Protection Regulation
MB - Megabyte
OECD – the Organization for Economic Cooperation and Development
PDPA - Personal Data Protection Act
Introduction

In the midst of the digital age, involvement with technology and data processing is a ubiquitous reality in everyone’s everyday life. It is estimated, that by the year 2020 for every person on the Earth, 1,7 MB of data will be created every second.\(^1\) The catch phrase of the digital age is that instead of oil, the world’s most valuable resource is personal data.\(^2\) Data has become a commodity, which has lead the society to a data-driven economy, in which companies rely heavily on the collection and analysis of data in its business activities.

On May 6, 2015, the European Commission adopted “A Digital Single Market Strategy for Europe” which was planned to be a “key enabler of the Digital Single Market that should allow European citizens and businesses to fully benefit from the digital economy.”\(^3\) However, one of the key building blocks of the strategy,\(^4\) the recently established general data protection regulation (EU) 2016/679 (“GDPR”), is being constantly challenged by the evolvement of the privacy issues rooting from the disruptive technologies, including big data.

European Commission has stated that the big data is the backbone of prosperous economies.\(^5\) Considering the fact that big data sector is growing by 40% per year, which is seven times faster than the IT market,\(^6\) the big data is definitely acute. Big data processing, which mainly entails collection, storage and analysis of vast amounts of information to extract targeted value, for example behavioural predictions, is an indispensable part of nowadays personalised marketing.

Behavioural targeting involves tracking data subject’s online activities for the purposes of delivering tailored advertising to potential customers. The most common method for carrying out behavioural targeting is cookies, which track data subject’s activities by associating those to a particular device. The collected information provides companies with valuable information, which allows to align their advertisements with the preferences of the potential customer and

\(^2\) K. Varshney. Data is the world’s most valuable resource - so who will control it? Digital Directions, 10.05.17. Available at: https://digitaldirections.com/data-worlds-valuable-resource-will-control/ (05.12.18).
eventually purchase. The new added element to behavioural targeting is utilising big data in advertising. Formerly, big data was used for guessing data subject’s preferences for goods or service, nowadays, the profiler guesses information about the data subject. Advertisers use big data analysis in order to save money and reach the most potential customers with the least effort. By learning user behaviour, big data can be used to create targeted and personalized advertisements (targeting advertising).

Technological developments and immense data sharing have allowed companies and advertisers to collect greater amounts of personal data than ever before. While this has the potential to benefit consumers as well as marketers, it is undoubtedly coupled with concerns regarding transparency and informed consent.

The purpose of this research paper is to analyse the challenges of data protection in the context of utilising big data in behavioural targeting and to provide possible solutions to ensure data subject’s privacy. The challenges and risks mainly concern the lawfulness, e.g. obtaining a valid consent for the processing of personal data and data subject’s awareness of the right of withdrawal, and transparency, e.g. informing the data subject in accordance with Arts 13 and 14 of the GDPR, of processing personal information.

The research paper consists of three chapters, which are divided into subchapters. The first chapter consists of three subchapters, which pertain to the nature of big data, the concept of behavioural targeting and the sources of behavioural targeting respectively. The second chapter consists of three subchapters as well. In this chapter, the authors analyse the legal framework of big data in international legislation, European Union legislation and Estonian national legislation. The subchapter about European Union legislation is divided into two smaller subchapters to address the history of and current legislation separately. The third and last chapter is dedicated to the issues arising with the GDPR, big data and behavioural targeting and possible solutions to such problems. In two subchapters the authors analyse issues pertaining to applicability of the GDPR and two problems deriving from the GDPR, such as privacy loss and transparency issues. The first subchapter consists of two subchapters which answer the following questions respectively: is big data personal data and are big data analytics and behavioural targeting in the scope of the GDPR. The second subchapter is directed to solve the problems of privacy loss and transparency issues under Articles 13 and 14 of the GDPR. Each issue is addressed in a separate subchapter. Unprecedented privacy loss in turn consists of two

subchapters about anonymisation and pseudonymisation. Transparency issues are discussed in two subchapters as well. The first entails the provision of accurate information and receiving a valid consent from the data subject, the second addresses data subject’s rights.

This research paper was written by the co-authors as a joint effort. Stella Raudsepp was the main author of the first chapter, whereas Gerli Kuldre was the main author of the second chapter. The third Chapter was written together. Both authors contributed to writing, formatting and proofreading every chapter of the paper.
1. Definitions of big data and behavioural targeting

1.1. Definition of big data

Big data does not have a single universal definition, as the interpretation of it differs depending on the specific discipline and field of application. Big data has been defined as a technology, a type of data processing, a form of analytics, a research paradigm, and a phenomenon. One of the most wide-known definition describes big data through three V-s, as the dimensions of challenges in data management: “Big data is a 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 automatization”. High-volume refers to the size of the dataset, high-variety refers to the structural heterogeneity of a dataset (multiple repositories, domains, or types of data), and high-velocity to the speed and consequently, the pace of generating data. The foregoing definition has emerged as an acknowledged framework to describe big data. After the main “three V-s”, a few more dimensions have been added as characteristics of big data, such as veracity, variability and low-
value density. Veracity refers to the unreliability of data sources, variability refers to variation in flow rate of data deriving from inconsistency in the big data velocity, and low-value density refers to the fact that before analysing the data it is substantially unusable. As the widely accepted approach is to describe big data through the “three V-s”, variability and low-value density are brought out here for illustrative purposes, however the veracity is reflected in the third party data processing (see Chapter 1.3.3.).

The aforementioned dimensions make the processing of big data impossible with traditional processing software, which are not designed to manage vast amounts of data within a reasonable time, thus big data represents a fundamental change in the architecture needed to efficiently handle the processing of current datasets. This means, that in the context of data protection, the main challenges do not solely concern the “three V-s”, but also “the analysis of the data using software to extract new and predictive knowledge for decision-making purposes regarding individuals and groups”. Big data analytics is the process of examining large and varied datasets to uncover hidden patterns, correlations, market trends, customer preferences and other useful information. Therefore, for the purposes of this research paper, “big data” as a term shall encompass big data and big data analytics.

1.2. Definition of behavioural targeting

Behavioural targeting entails monitoring data subject’s online behaviour over time and using the collected information to present to the same data subject advertisements matching his/her inferred interests. To give a simplistic example, a company might assume that an Internet user who often visits online book store X science fiction section is enthusiastic about reading books in this genre. If the same user visits a new website, he/she is prompted with an advertisement for a book club, new launch of a science fiction book or book stores nearby. In this instance, the book store X might have an arrangement with a network advertiser, to provide advertising to its website visitors, thus it placed a cookie on the user’s device. As the new website is also

24 Ibid.
part of the advertising network, the relevant advertisements are shown. A bit more sophisticated advertiser may use similar technique combined by other users’ data. Having access to habits, likes, and propensities of many users, the advertiser may predict unknown information about the data subject in the example by comparing his/hers data to other users known data. The latter example is the process that Amazon and Netflix apply for generating suggestions for books, movies and other goods.\(^{30}\)

Behavioural targeting (collecting data and displaying advertisement) can be carried out utilizing various technologies. The primary approaches are monitoring IP addresses, cookies, javascripts, supercookies, deep packet inspection, device or browser fingerprinting, and spyware.\(^{31}\) This research paper focuses solely on the “cookie approach” as currently, this is the most widely used technique in digital marketing.\(^{32}\)

The cookie approach entails placing a cookie on data subject’s device for the purposes of tracking the data subject’s behaviour across websites visited in the Internet. “A cookie is a named piece of data that a website sends to a web browser, along with a request that the data subject’s web browser retains it.”\(^{33}\) It is possible to indicate the time period for how long the web browser should retain the cookie and for which websites the data in the cookie is intended.\(^{34}\) By assigning each web browser a cookie with a unique number, the owner of the cookie can later differentiate and profile the activities of many users visiting the website.\(^{35}\)

1.3. Sources of big data in behavioural targeting

The main sources of information used for behavioural targeting derives from web tracking, using the same technologies as mentioned in Chapter 1.2. of this research paper. Data collected via these methods and used as an input for big data analytics can be categorised into three: (i) first party data, (ii) second party data, and (iii) third party data.


\(^{34}\) Microsoft Dev Center. HTTP Cookies. 31.05.2018. Available at: https://docs.microsoft.com/en-us/windows/desktop/WinInet/http-cookies (01.12.2018)

1.3.1. First party data

First party data involves behavioural targeting based on data collected at and by a single website, CRM, and subscription data. For example, online shopping websites deploy first party cookies (aka data collected through these is solely accessible to the website owner) to remember the contents of a shopping cart. First party data is valuable, as it allows the company to get specific insights regarding its already existing customers and provides an opportunity to utilize super precise behavioural targeting.

1.3.2. Second party data

Second party data entails behavioural targeting based on another company’s first party data, meaning that all of the second party data originates from one certain source. In order to obtain second party data, it is bought directly from the other entity. For example, an entity has an agreement with AdWords, which receives keyword data from Google. The first entity receives that data regardless of the users not having agreed to the entity’s privacy policy, meaning that if these users happen to come across the entity’s website and blocks any cookies the website uses, the entity will still have received the keyword data based on the agreement with AdWords.

1.3.3. Third party data

Third party data refers the data one entity purchases from an outside source, who is not the original collector of the data. Usually, the sellers have scraped and bought these data from various websites and other data vendors. Scraping as such implies to using cookies, which the advertising networks (“Profiler”) place and read to learn online behaviour of Internet users (“third party cookies”). As the third party data is bought and sold rapidly and on a large volumes, it is the primary input for big data analytics.

---

2. The legal framework of data protection, big data and behavioural advertising

2.1. The regulation of big data in international legislation

Some legal acts regulate the fundamental protection of privacy, for example the Article 12 of the Universal Declaration of Human Rights\textsuperscript{44} states that no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence.\textsuperscript{45} The Organization for Economic Cooperation and Development (\textquotedblleft OECD\textquotedblright) took a step further and set forth Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data\textsuperscript{46}, which according to its Article 2 of Part I apply to personal data, whether in the public or private sector, which, because of the manner in which they are processed, or because of their nature or the context in which they are used, pose a danger to privacy and individual liberties.

The emergence of information communication technology brought forth the need to further regulate the protection of privacy. As a result, in 1981 the Council of Europe signed a document, called “Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data”, also known as the Convention 108 and it is the first binding international instrument on data protection.\textsuperscript{47} The Convention 108 entered into force on January 28, 1985.\textsuperscript{48} Even though the convention was signed by the Council of Europe, it is open for accession to non-state members as well. In addition to Member States, the Convention has been ratified by Uruguay, Senegal and Mauritius.\textsuperscript{49} The aim of the Convention 108 is to extend the safeguards for every person’s rights and fundamental freedoms arising from the European Convention of Human Rights\textsuperscript{50}(\textquotedblleft ECHR\textquotedblright). More specifically, it serves as a protection to respect for privacy and at the same time balancing it with the freedom of information.\textsuperscript{51} Pursuant to the Preamble it is necessary to reconcile the fundamental values of the respect for privacy and the free flow of information between peoples.\textsuperscript{52} Pursuant to Article 1 of the Convention 108, the purpose of Convention 108 is to secure the respect for every individual’s rights and fundamental freedoms, and in particular their right to privacy, with regard to automatic processing of personal data.

\textsuperscript{44} Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, Europ.T.S. No. 5; 213 U.N.T.S. 221.

\textsuperscript{45} UN General Assembly, Universal Declaration of Human Rights, 10 December 1948, 217 A (III). Art 12.


\textsuperscript{47} Ibid.


\textsuperscript{49} Ibid.


\textsuperscript{51} Ibid.

\textsuperscript{52} Council of Europe, Convention for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, 28 January 1981, ETS 108. Available at: https://rm.coe.int/1680078b37 (21.11.18).
relating to them. The Convention sets forth principles for processing of personal data and regulates the storing of such data.\textsuperscript{53} Out of all the aforementioned legal acts, internationally speaking, Convention 108 is the most important instrument in regulating data protection on an international level, as it set forth basic terms and conditions for data processing.\textsuperscript{54}

2.2. The regulation of big data in the European Union

2.2.1. Overview of the history of data protection legal framework in the European Union

The European Convention of Human Rights was the first legal act applicable to every Member State of the European Union (\textbf{``EU''}) to state that personal data is a fundamental human right and it needs explicit protection.\textsuperscript{55} The ECHR entered into force in 1953 and declared in Article 8 that everyone has the right to respect for their private and family life, their home and their correspondence.\textsuperscript{56} As complying with the ECHR was mandatory to every Member State of the EU in 1959, the European Court of Human Rights was founded for the purposes of ensuring their compliance. The Court has clarified throughout its practice, that Article 8 in addition to refraining from violating the right to privacy also obligates the states to actively protect it.\textsuperscript{57}

Data Protection Directive 95/46/EC was the direct result of the increasing awareness of data protection importance. The Directive entered into force on December 12, 1995 and it set strict limits to collecting and using personal data. The aim of the Directive was to protect the rights and freedoms of people with regard to the processing of personal data by laying down ground rules for lawful processing and the principles of data quality.\textsuperscript{58} The Member States of the EU were under the obligation to harmonise the Directive with legislation, that was previously adopted to regulate privacy and data protection.\textsuperscript{59} The entity to ensure the Member States’ compliance to the Directive is the Court of Justice of the European Union (\textbf{``CJEU''}).\textsuperscript{60} Pursuant to Article 28 of the Directive every Member State is obligated to provide a supervisory authority

\begin{flushleft}
\textsuperscript{53} I. M. Zunko. \textit{Op. cit.} \\
\textsuperscript{54} \textit{Ibid.} \\
\textsuperscript{55} I. M. Zunko. \textit{Op. cit.} \\
\textsuperscript{57} I. M. Zunko. \textit{Op. cit.} \\
\textsuperscript{59} I. M. Zunko. \textit{Op. cit.} \\
\textsuperscript{60} \textit{Ibid.}
\end{flushleft}
to monitor the application of the Directive within its territory. The Data Protection Directive was effective until May 24, 2018.

The Data Protection Directive was repealed by the General Data Protection Regulation once it entered into force on May 25, 2018. Compared to the Data Protection Directive, the GDPR is set to guarantee higher level of harmonisation as it is in the form of a regulation and therefore applies directly in all Member States. The GDPR brought along several changes for the data subjects as well as the controllers. For example, instead of the “right of erasure” the data subject now has the right to be forgotten. Pertaining to the controllers, one significant change the GDPR introduced, is the data protection impact assessment which must be conducted to identify high risks to the privacy rights of individuals when processing their personal data. These are only two examples to indicate how the data protection has evolved with the GDPR.

2.2.2. The legal framework of big data in the EU

Considering the legal instruments, which are currently in force in EU regulation, none of them address big data per se. Privacy and data protection matters are regulated in the ECHR, which sets forth the most basic fundamental rights; in Directive (EU) 2016/680, which addresses processing of personal data regarding criminal offences; in Directive 2002/58/EC, which regulates privacy in electronic communications; and in the GDPR.

Directive 2016/680 entered into force on May 5, 2016 and repealed the Council framework decision 2008/977/JHA on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters. Considering the fact that the Directive pertains to data processing in criminal matters, it does not apply to private actors and companies.

---

62 Ibid.
63 Ibid.
66 GDPR Art 35.
69 Ibid. Art 64.
Directive 2002/58/EC, also known as the ePrivacy Directive entered into force on July 31, 2002 and repealed Directive 97/66/EC concerning the processing of personal data and the protection of privacy in the telecommunications sector. Pursuant to Article 1 of the ePrivacy Directive, it aims to ensure an equivalent level of protection of the right to privacy with respect to the processing of personal data in the electronic communication sector. On January 10, 2017, a proposal for a new ePrivacy Regulation was drafted by the European Commission in order to implement the Digital Single Market Strategy. At first, it was meant to enter into force alongside of the GDPR on May 25, 2018 but failed to do so. For now, the current ePrivacy Directive stands, though in the near future changes in the EU privacy law are expected.

From the point of view of this research paper, the GDPR is the most important legal act in the EU data protection field, as it harmonises the rules for data processing and makes it more transparent. In the scope of this work, the most relevant provisions of the GDPR are Art 3(2), 6 and 22, as these address behavioural data processing and automated processing respectively. The question of the regulation of big data in the GDPR is addressed in Chapter 3.

2.3. The regulation of big data in Estonian legislation

Estonia as a Member State of the EU, is obligated to harmonise the EU legislation to the national legislation. The implementing act of the GDPR and Directive (EU) 2016/680 to the Estonian legislation is currently in a draft form and is in the third reading round in the Parliament. The proposed enforcement date of the GDPR implementing act is January 15, 2019. The proposed act solely reflects derogations from the GDPR, thus contains no big data or profiling specific sections.

The current Personal Data Protection Act (“PDPA”) in force was implemented into national legislation based on the Data Protection Directive 95/46/EC. Given, that the directive was repealed by the GDPR leaving the PDPA in a situation where it is based on a void directive and has not adopted new rules set forth by the new regulation, then the PDPA cannot be considered to be up to date any more. All the discords between the current PDPA and the GDPR will be removed once the new version of PDPA enters into force.

---


71 Isikuandmete kaitse seadus. RT I, 06.01.2016, 10. Available at: https://www.riigiteataja.ee/akt/106012016010 (05.12.18).
In conclusion, the current legal framework does not address big data extensively as big data analytics is a relatively new means being used to analyse big datasets, thus when automated processing of personal data became a subject of legal regulations, the datasets being processed were considerably smaller than they are now and it probably was not foreseen as a subject to be extensively regulated. However, nowadays, tracking data subjects and collecting information through their online activities across various websites in order to send them targeted advertisements based on their inferred interests is an active point of concern for consumers and regulators alike. While there is no legislation that speaks specifically about “big data” and “behavioural targeting”, the privacy laws and regulations of general application are applicable.72

Considering the aim of this research paper, the following chapters approach the issues pertaining to big data and behavioural targeting in the light of the GDPR as the main instrument with a wide territorial scope providing the protection of data subject’s privacy in the context of big data processing and behavioural targeting.

---

3. The issues with the GDPR, big data and behavioural targeting and the hypothetical solutions

In the context of big data and behavioural marketing, GDPR has several shortcomings. In this research paper, the authors address two of them: (i) applicability of the GDPR to big data and behavioural targeting, and (ii) privacy loss and transparency issues under Articles 13 and 14 of the GDPR.

3.1. Applicability of the GDPR to big data and behavioural targeting

3.1.1. How does big data containing behavioural data relate to the personal data?

Pursuant to Article 4(1) of the GDPR, personal data refers to any information which directly or indirectly allows to identify a natural person. Thus, a data set contains personal data if (i) an individual is directly identifiable by this information (e.g. name, address); or (ii) an individual, while not identified, can be singled out by this information in a way which makes it possible to find out the identity by further research. Behavioural data entails the behaviour of the data subject on the Internet, for example all the clicks they make and all the “likes” they give, as well as the websites they visit. The core of behavioural targeting lies in big data processing, which involves looking for behavioural patterns that reflect “characteristics of a type of personality”. Therefore, behavioural data must be deemed as personal data, even if an entity cannot tie a name to the data it holds about an individual and behavioural targeting entails personal data processing, as a data subject can be singled out by the characteristic drawn out from the big data.

3.1.2. Are big data analytics and behavioural targeting in the scope of the GDPR?

Article 2(1) of the GDPR, which regulates the material scope of the GDPR, does not expressly refer to behavioural targeting nor big data, however it refers that the GDPR is applicable to “personal data” which is processed by “automated means”. The material scope is stipulated in Article 3 of the GDPR, making a reference to behavioural data. Article 3(2)b) of the GDPR states that the GDPR applies to processing of personal data of data subjects who are in the EU,

---

77 Article 2(1) of the GDPR.
by a controller or processor not established in the EU, where the processing activities are related to the monitoring of data subject’s behaviour as far as their behaviour takes place within the EU. Recital 24 to the GDPR stipulates the following: “In order to determine whether a processing activity can be considered to monitor the behaviour of data subjects, it should be ascertained whether natural persons are tracked on the internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes.” Regardless the fact that the GDPR does not mention expressly the volume of the data processed, it can be deduced that that behavioural data processing with automated methods is in the scope of the GDPR, if it makes the natural person identifiable to the data possessing entity.

In order to determine, whether a natural person is identifiable to the data controller, account should be taken of all the means reasonably likely to be used and objective factors, such as the costs of and the amount required for identification, by the controller to identify the natural person directly or indirectly.78 In the context of behavioural targeting, profiling is a primary tool for extracting value from big data.79 Profiling is an algorithmic inference drawn from data about an individual. Profiling is mostly carried out by an artificial intelligence algorithm, which takes vast datasets (big data) as input and generates output according to predetermined patterns or factors in the algorithm.80 Pursuant to Article 4(4) of the GDPR, profiling refers to any kind of automated personal data processing that analyses or predicts certain aspects of an individual’s economic situation, movements, preferences, and so on.81 As the essence of big data analytics is aggregating the gathered data (including behavioural data) for the purposes of drawing out patterns of it, the process itself continuously produces new personal data regarding a certain data subject. The foregoing refers that big data used for behavioural targeting allows to identify individuals for the data controllers. Therefore, it can be concluded that processing behavioural data of the data subject through big data analytics, for the purpose of behavioural targeting is governed by the GDPR, as the controller has the means to identify the data subject through its behaviour, even without primary characteristic data, such as name.

It must be noted that, automated processing methods, including profiling are governed by Article 22(1) of the GDPR, pursuant to which data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal

78 Recital 26 of the GDPR.
81 Article 4(4) of the GDPR; recital 71 of the GDPR.
effects concerning him/her or significantly affect him/her. However, the referred article is problematic in the context of big data used for the purposes of behavioural targeting. The obstacle the Article 22 of the GDPR and the scope of it proposes is that it solely governs such automated processing, which produces legal or similar effects to the data subject. In case of behavioural targeting, legal effects as a result of profiling is a highly unlikely outcome.\textsuperscript{82} Thus, regardless of the fact that behavioural targeting and big data containing personal data are within the scope of the GDPR, it does not per se regulate them.

Even though, the GDPR does not regulate or foresee special regulation for the volume of the data sets and big data analytics, the general principles of the GDPR apply. Thus, each data processing activity relating to personal data in the context of behavioural targeting and big data must be in compliance with the GDPR.

3.2. Privacy loss and transparency issues under Articles 13 and 14 of the GDPR

The necessity for bulks of data as an input for big data in behavioural targeting raises several privacy problems. This research paper analyses the following two: (i) unprecedented privacy loss; (ii) fulfilling the obligation of the controller to notify data subject under Article 13 of the GDPR.

3.2.1. Unprecedented privacy loss

Big data processing for the purposes of behavioural targeting results in unprecedented loss of privacy, which is indicated in the compilation of a sizable array of personal data about the data subjects’ that exists outside data subjects’ ability to protect, control or monitor it.\textsuperscript{83} By using the Internet, data subject unknowingly loses control over which details about his/hers private life are known, and data subject has substantially no control over who gets access to his/hers data after it passes to a data profiler.\textsuperscript{84} Data profiler in this context usually is the advertising network provider (see Chapter 3.2.2.), however, it may also be any other stakeholder in the Internet with the necessary processing capacity, for example website owners as data controllers, big data analytics providers as data processors, or sub-processors engaged for providing IT-services. The following subchapters analyse the prospect of tackling unprecedented privacy loss with either anonymisation or pseudonymisation.

3.2.1.1. Anonymisation

Anonymisation is an effort to eliminate any characteristics from a dataset that could enable someone to discern the identities of the users described in the dataset.\textsuperscript{85} Pursuant to Recital 26 of the GDPR the principles of data protection should not apply to anonymous information, specifically information that does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. It follows that, the GDPR does not pertain to the processing of anonymised personal data. Therefore, if the Profilers could render their databases of data subjects’ information completely anonymous, they could process such data as they see fit and without having to worry about complying with conditions set forth by the GDPR. Thus, at first, anonymisation seems like an instant solution for the Profilers.

On the other hand, the Profilers have attempted to mitigate the risk of privacy loss of data subjects’ through anonymisation long before the GDPR came into effect. Some advertising entities cut off direct ties from users’ profile to their identity in order to anonymise their data, but this has proven to not be a solution, as computer scientists determined that it was still remarkably easy to identify specific users.\textsuperscript{86} This implies, that anonymisation is not be the best solution even today in the era of the GDPR.

Analysing user behaviour is a method almost every advertising company has embedded in their working strategies and more often than not, it relies on cookies\textsuperscript{87}. Article 29 Data Protection Working Party has found, that using first party cookies, privacy risks are not so likely, insofar as that first party uses such data for their own statistical purposes, they have provided clear information about these cookies and appropriate safeguards are implemented.\textsuperscript{88} By using first party data, the issue of data subject’s consent could easily be overcome, as their privacy policy contains all the necessary information and the data subject is able to give their informed consent. Appropriate safeguards would entail a user friendly mechanism to opt-out from data collection and moreover, comprehensive anonymisation mechanisms that are applied to other collected identifiable information (for example IP addresses).\textsuperscript{89} In case of third party cookies, the user’s privacy risks cannot be avoided using such safeguards as the origin or even accuracy of data cannot be confirmed.

\textsuperscript{86} \textit{Ibid.}
\textsuperscript{88} \textit{Ibid.}
\textsuperscript{89} \textit{Ibid.}
There are two main approaches for anonymising data: randomization and generalization. The first is a family of techniques that alters the veracity of the data to remove the link between the data and the data subject. On its own, randomization will not reduce the singularity of a record as that record will still be derived from a specific data subject. Still, it is a good way to protect the data from interference attacks or risks. This means, that in order to ensure complete anonymity, additional measures must be implied. The second method consists of generalizing or diluting the attributes of data subjects by modifying the respective scale or order of magnitude. While it can be an effective means to prevent singling out a data subject, it still requires specific and sophisticated quantitative approaches to prevent linkability and interference.\(^{90}\)

Though anonymisation may be considered to be an appropriate safeguard, it must be emphasized, that firstly, it would work only with data collected through first party cookies and secondly, it must be determined without a fault, that the data is in fact completely anonymised, meaning a specific data subject cannot be identified or re-identified by exercising reasonable effort.\(^{91}\) From that, however, derives an issue. The purpose of big data analysing is to create a profile of a user and use that profile to present targeted advertisements and predict future movements or needs of that user. Even if the most basic personal information, such as name or social security number is removed from that data, then the user’s behavioural patterns, which are deemed as personal data (see Chapter 3.1.1.) will be known to the data controller and to the processor, given that these two are not the same entity.

Moreover, as mentioned in Chapter 3.1.1., the Profilers may create personal data and even data belonging to special categories (such as health data) during the process of big data analytics. For example, the Massachusetts Institute of Technology researchers analysed over 4 000 students’ Facebook profiles and were able to predict with 78 percent accuracy, whether a profile belonged to a gay male.\(^{92}\) The more recent example derives from Cambridge Analytica case, whereas Cambridge academic Michal Kosinski built a model, which is able to reverse engineer a personality profile from one’s Facebook activity. The model is able to predict someone’s personality profile (including predicting the race, religion and sexual orientation) with the same accuracy as a spouse just based on 300 “likes”. Later on, this model was further developed by another academic and a deal was cut with Cambridge Analytica, which built personality profiles

---


for more than 100 million registered US voters and then used these profiles for targeted advertising.\footnote{M. Wade. Psychographics: the behavioural analysis that helped Cambridge Analytica know voters’ minds. The Conversation. 21 March 2018. Available at: http://theconversation.com/psychographics-the-behavioural-analysis-that-helped-cambridge-analytica-know-voters-minds-93675 (05.12.2018).}

Considering the aforementioned, the fact that in order to reach complete anonymisation, the Profilers will have to implement additional techniques, which means that the anonymisation process will be more costly. This, however, poses a threat that many companies choose to save money instead of protecting personal data of data subjects. In addition, for the same reason fair number of companies use third party data instead of obtaining the information from data subject directly. Using third party data is faster and cheaper and all in all fits in with the companies’ mentality of achieving as much as possible with as little resources as possible. However, it may be impossible to determine the source and lawful basis of obtaining the third party data, likely leading to breaching data subject’s rights under the GDPR.

3.2.1.2. Pseudonymisation

Pseudonymisation is a technique, which entails replacing on unique piece of information (e.g. name) with a pseudonym.\footnote{European Union Agency for Fundamental Rights and Council of Europe. Op. cit., p 94.} Pseudonymisation does not result in anonymous dataset, but reduces the linkability of such dataset with a specific data subject. The most popular methods of pseudonymisation are for example, encryption with a secret key, hash function, salted-hash function, and keyed-hash function with stored key. Pseudonymisation is an effective way to protect the data from malicious users, e.g. if the data gets stolen, it is unusable without an encryption key or some other necessary piece of information needed to read the data.\footnote{Article 29 Data Protection Working Party. Opinion 05/2014. Op. cit., p 20.}

Due to the fact, that pseudonymisation does not provide anonymity to data subjects, it cannot be considered as a pertinent solution to privacy loss. Pursuant to Recital 28 of the GDPR applying pseudonymisation can reduce the data privacy risks and help controllers and processors to meet their data-protection obligations. Thus, if the Profilers were to use first party data, meaning that they are provided with such data by the client under the permission of data subjects and solely use it for behavioural targeting without aggregating it with other client’s data (thus second and third party data), pseudonymisation would be an appropriate means for protecting the data. In any other case for the purposes of big data analytics, pseudonymisation by itself is not sufficient protection of the data subject’s private information.
In conclusion, given the fact, that anonymisation and pseudonymisation are not suitable solutions to the privacy issues described above, as these cannot be considered to be effective ways to prevent privacy loss. It would be wrong to state that the GDPR limits the evolution of big data by setting forth limitations to the processing of personal data, as the problem of unprecedented privacy loss stems from the data subjects’ online behaviour and the technical solutions often applied in the behavioural targeting and big data analysis. Thus, it can be concluded that the GDPR effectively fulfils its aim at protecting the rights of the data subjects. However, the GDPR certainly forces the Profilers to come up with new technical solutions which are as effective for the purposes of big data processing and behavioural targeting, without breaching data subject’s privacy. If the Profilers aim to process the data lawfully, it would result in using less third party data, and more first or second party data. Even though acquiring the data firsthand or using second party data might be more time consuming and expensive, it would mean more high-quality data as an input for big data processing as well as a more trustworthy relationship between the Profiler, the website owner (“Publisher”, see also Chapter 3.2.2 for the specific allocation of roles) and data subjects.

Furthermore, the protection provided to the GDPR and restrictions applicable to the Profilers and Publishers may even induce innovation and evolvement of big data techniques, as one solution for the above mentioned issues would be providing behavioural targeting without using personal data as input for big data processing, because obtaining a valid consent as a lawful basis from the data subject may render difficult and in the future, when data subjects have gained more knowledge about their privacy rights, giving a consent may be an unpopular choice. The foregoing method is yet to be invented.

3.2.2. Transparency issues

Data subject has no substantial control over the methods the Profilers use to mine compiled data in order to construct the behavioural profile of the specific data subject. Data subject is not able to determine the risks related to participating in activities in the Internet (inter alia the risk of unprecedented privacy loss as described in Chapter 3.2.1.). However, the GDPR intends to tackle these problems by establishing as one of the six lead principles of the GDPR, a principle of lawful and transparent processing under Article 5(1)a) of the GDPR, supported by data controller’s obligation to provide data subject with thorough information pursuant to Article 13 and 14 of the GDPR.

Transparency is an obligation arising from the GDPR which pertains to three specific areas: (i) the provision of information to data subjects related to fair processing; (ii) how data controllers communicate with data subjects in relation to their rights under the GDPR (discussed together
in Chapter 3.2.2.1.); and (iii) how data controllers facilitate the exercise by data subjects of their rights.\textsuperscript{96} The definition of transparency is not provided by the GDPR, but Recital 39 of the GDPR explains the essence of the principle of transparency, according to which the principle requires that any information and communication relating to the processing of the personal data be easily accessible and easy to understand and presented in clear and plain language.

Prior to analysing three specific aspects of transparency, it is crucial to determine the lawful basis of processing and who is the data controller towards the data subject in the context of behavioural targeting, as this determines the responsible party for notifying the data subject under Article 13 of the GDPR.

\textbf{Roles}

In allocating the roles of the relevant parties, it is important to take into account all three parties: (i) data subject, (ii) website owner ("\textbf{Publisher}"), (iii) advertisement network provider ("\textbf{Profiler}" as defined in Chapter 1.3.3.). From the data protection perspective, the Publisher should be deemed as an autonomous data controller, as it collects personal data (IP address, location, language of operating system etc) from the data subject for its own purposes.\textsuperscript{97} The Profiler should also be deemed as a separate data controller insofar as it determines the essential means of the processing of data\textsuperscript{98} (for example big data analytics). In addition, depending on the agreement between the Publisher and the Profiler, the parties could be deemed as joint controllers in the context of providing targeted ads on the Publisher’s website. Deriving from the allocation of roles as described above, the information flow between the parties should be as follows:


Graphic showing information flow and consents required for the basic functioning of behavioural targeting provided by the Profiler at the request of the Publisher.

Lawful basis of the processing

The only relevant basis for providing a data subject with advertisements based on behavioural targeting is data subject’s explicit consent.\(^{99}\) For a consent to be valid under the Article 4(11) of the GDPR, it must be freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he/she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him/her.\(^{100}\) This emerges two problems: (i) obtaining the consent for the processing activities of the Profiler as an independent data controller, (ii) providing the data subject with specific and informed information regarding profiling activities (including big data analytics)(will be discussed in Chapter 3.2.2.1.).

When the Publisher wishes to engage a Profiler for the purposes of providing advertisements based on behavioural targeting, it must obtain several consents from the data subject (as shown in the Graphic 1). The second consent for processing activities solely carried out by the Profiler indicates the situation whereas the Publisher substantially obtains the lawful basis for the Profiler to carry out its processing activities as a separate controller. This means that the Publisher has the burden of acquiring the consent in compliance with the Articles 4(11) and 7 of the GDPR and the guidelines published by the European Data Protection Board (formerly known as Article 29 Data Protection Working Party). However, it is likely that the Publisher

---


does not know in detail how the data is being processed by the Profiler, thus compliance with transparency obligations may render impossible and might result in a breach of Articles 6 and 13 of the GDPR by both - the Profiler and Publisher. Regardless of the fact that the GDPR does not oblige separate controllers, who exchange data, to conclude a data processing agreement, it could be an effective tool to mitigate the mentioned risk, by establishing strict rules on the extent of each party’s notification obligation and liability between them.

On one hand, the above mentioned issues may escalate further once the ePrivacy Regulation becomes into force containing the proposed obligation that browsers and devices must present users with a choice of what degree of tracking they will accept and makes the choice enforceable. Assuming this survives the final regulation text then it is likely that most users in Europe will not choose tracking. On the other hand, the proposed rule will be more user-friendly as browser settings will provide for an easy way to accept or refuse tracking cookies.\(^\text{101}\)

3.2.2.1. The provision of accurate information by data controller to data subject under Article 13 of the GDPR and communicating to the data subject under Article 12 of the GDPR

As was mentioned in Chapter 3.2.2., this chapter analyses the provision of information to data subjects related to fair processing, and the communication between the data controllers and data subjects in relation to their rights under the GDPR.

For a consent to be a lawful basis for providing advertisements based on behavioural targeting, it must be accompanied with specific information, \textit{inter alia} regarding the processing activities. Pursuant to Article 13 of the GDPR, at the time personal data are obtained, the data controller shall provide the data subject with all information required in the Article 13(1) and (2) of the GDPR, including the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least, in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject. As discussed in Chapter 3.1.2., big data processing for the purposes of behavioural targeting cannot be deemed as automated individual decision-making under Article 22 (1) or (4) of the GDPR. Thus, if the Article 13(2)f) is interpreted narrowly and linguistically, neither Publisher or Profiler is obligated to provide data subject with meaningful information about big data analytics. However, the European Data Protection Board (formerly known as Article 29 Data Protection Working Party) has stated in its guideline as a reference to Article 5(1)a) of the GDPR, that given the core principle of transparency underpinning the GDPR, controllers must

ensure they explain clearly and simply how the profiling works. Thus another challenge in obtaining consent is providing specific, yet meaningful information about big data analytics carried out in the course of behavioural targeting, especially taking into account the continuously developing AI, which is usually utilised in big data analytics.

Next to the data subject’s consent, one other important prerequisite for transparency is the communication between the data controller and the data subject. Article 12(1) of the GDPR sets forth the rules for communication stating, that any communication provided to the data subject must be in a concise, transparent, intelligible and easily accessible form, using clear and plain language, provided in writing (or by other means, where appropriate). Pursuant to Article 12(5) of the GDPR, any communication must be provided free of charge. Thus, any communication that takes place between the Publisher or Profiler and the data subject regarding the consent of the data subject and the conveyed information, must follow the foregoing rules set forth by Articles 12(1) and 15 of the GDPR.

In a situation, where the Publisher and the Profiler are the same entity, meaning first party data is used, no significant issues can derive from the communication obligation, provided that the data controller shares all the relevant information required by Article 12 of the GDPR. Communicating to the data subject in case of third party data is the most problematic. As was mentioned in the previous chapter, the data subject may not be aware of the identity of the Profiler. It is also questionable, if the Profiler upon receiving the big data dataset will actually contact every data subject separately in order to exercise any communication at all. Given that such a dataset contains information on an immense number of consumers, such an undertaking will certainly prove to be costly, thus it is rather unlikely, that the data subject’s receive proper communication in case of processing third party data.

One way to tackle these challenges is that the Publisher includes into its own privacy notice some information about the Profiler, fulfilling the bare minimum of transparency obligation, by providing basic information about how the Publisher has engaged the Profiler (e.g. mentions a contract between them), whether the cookies deployed for that purpose are first party cookies or Profiler’s cookies (i.e. does the data move through the Publisher or it becomes accessible solely to the Profiler), making a clear indication of Profiler as a separate controller, its contact details and a reference to the Profiler’s privacy notice, which ideally has several precision levels, thus the data subject can choose the level of precision he/she would like to read about the technicalities of big data processing and behavioural targeting.

---

If the Profiler does not receive the data through the Publisher (i.e. third party cookies provided by Profiler’s domain are deployed on Publisher’s site), it must provide the data subject with a privacy notice in compliance with Article 14 of the GDPR, although it may render impossible, as the Profiler might not have any contact point with the data subject.

3.2.2.2. Exercising data subject’s rights

The GDPR sets forth several rights the data subjects possess, *inter alia* right of access (Art 15), right to be forgotten (Art 17), right to object (Art 21) and right to withdraw the consent (Art 7). Article 21(2) of the GDPR stipulates the data subject’s right to object to processing of personal data for direct marketing purposes. Article 22 states, that data subjects have the right not to be a subject to automated decision making. Exercising these rights presumes conveying the intention to exercise a right to the Profiler.

Pursuant to Article 21(2) of the GDPR, data subject has the right to object direct marketing, which includes profiling. Exercising this right is fairly straightforward and easy, if the party, who collected the data from data subjects, is in fact the party who processes the data. As previously described, using aggregated third party data for big data analytics is rather common. In case of third party data, when the data collecting party has sold the dataset to another entity, who is now using the data how they see fit, exercising the right to object direct marketing is more difficult. The main issue in this situation is that the data subject has no information about who is the real data controller and how to contact them to notify them about the objection. If the data subject were to convey their objections to the party who collected the information, it would still be difficult to let the new data controller know, that this specific data subject no longer agrees to processing of their data. Moreover, as the data subject’s personal data was sold as a part of a dataset, it may prove to be difficult to extract specific data from that dataset.

Similarly, to the obligation of communication, exercising data subject’s rights is much easier in case of first party data. The data subject can inform the Profiler and the Publisher at the same time as the two are the same entity. In case of processing of third party data, conveying the intention to exercise a right to the Profiler may prove to be impossible due to lack of knowledge about the identity of the Profiler.

Pursuant to Recital 59 of the GDPR, modalities for facilitating the exercise of the data subject’s rights under the GDPR should be provided to the data subject. These modalities include mechanisms to request and obtain, free of charge, access to and rectification or erasure of personal data and the exercise of the right to object.103 The Profiler should therefore provide

---

103 Recital 59 of the GDPR.
these modalities on their webpage, for example as electronic forms which the consumer can fill in order to exercise their right. The data subject’s ignorance regarding the identity of the Profiler implies, that the Profiler has breached their obligation to notify the consumer about the Profiler’s identity under Article 14(1)a) and most likely has failed to provide the data subject with other important information as well set forth by Article 14. The foregoing leads to a conclusion, that the consumers data is at a significantly greater risk when processed as a third party data as the integrity of the data as well as proper handling cannot be ascertained by the data subject.
Summary

The data revolution, big data analytics and the GDPR in concurrence have brought along several challenges to data subject’s privacy and data protection in the online world, as the emergence of new technologies has allowed entities to collect and process enormous amounts of personal data, including behavioural data which is used for behavioural targeting. In this research paper we analysed the challenges of data protection in the context of utilising big data in behavioural targeting and provided possible solutions to ensure data subject’s privacy. In addition, we gave an overview of big data and behavioural targeting and their legal framework.

It became evident that big data and behavioural targeting as such are not expressly regulated on international nor European level (including in Estonia). The reason may be that big data analytics is a relatively new means being used to analyse big datasets, thus when automated processing of personal data became a subject of legal regulations, the datasets being processed were considerably smaller than they are now, and it probably was not foreseen as a subject to be extensively regulated. Similarly, there has been a shift in utilising big data for the benefit of behavioural targeting. Formerly, big data was used for guessing data subject’s preferences for goods or services, now the profiler guesses information about the data subject. It was concluded that, while there is no legislation that specifically addresses “big data” and “behavioural targeting”, the privacy laws and regulations of general application are applicable. Furthermore, it was found that even though, the GDPR does not regulate or foresee special regulation for the volume of the datasets and big data analytics, the general principles of the GDPR apply. Thus, each data processing activity relating to personal data in the context of behavioural targeting and big data must be in compliance with the GDPR.

Big data processing for the purposes of behavioural targeting results in unprecedented loss of privacy, which is reflected in the collection of enormous amounts of personal data about data subjects that visit the Internet, without the actual ability to protect, control or monitor the information provided to the stakeholders. The authors of the research paper argued that commonly used anonymisation and pseudonymisation are not suitable solutions to cure the unprecedented loss of privacy in the online world. This is due to the fact, that pseudonymisation does not provide anonymity and anonymisation is not always effective, as a specific data subject can still be identified based on the information contained in big data, even after the primary identifiers, such as name or address, have been removed from the dataset, as the core of big data analytics is drawing out data subject’s behavioural patterns, which makes the data subject re-identified to the controller. In order to reach complete anonymity, it would require implementing additional means, which the advertising networks may not do. It was concluded
that the GDPR effectively fulfils its aim at protecting the rights of the data subjects and it induces innovation and evolvement of big data techniques in a long run. For example, one possible solution would be to provide behavioural targeting without using personal data, because obtaining a valid consent as a lawful basis from the data subject may be impossible and, in the future, unpopular choice among data subject.

Considering the fact, that such a solution is yet to be invented, the authors propose as a short term solution a motion to encourage the Profilers to prefer first and second party data over third party data, as this would mean more transparency to the data subjects and an easier way for the advertising networks to comply with the conditions set forth by the GDPR. Though this might prove to be costlier to the advertising networks, it would ensure higher quality data and better protection to the data subjects.

In addition to the foregoing unprecedented privacy loss issue, the authors analysed the challenges posed by the transparency principle to the big data analytics and behavioural targeting. This issue concerns the website host (the publisher), advertising network provider (the profiler) and the data subject, whereas the publisher and the profiler are usually separate data controllers, thus the publisher is required to obtain the data subject’s consent as a lawful basis for the Profiler’s independent processing activities.

The authors proposed, that the publisher includes into its own privacy notice some information about the profiler and thereby fulfils the bare minimum of transparency obligation by providing basic necessary information, e.g. how the publisher has engaged the profiler, whether the cookies deployed for that purpose are first party cookies or profiler’s cookies, a clear indication of profiler as a separate controller, its contact details and a reference to the profiler’s privacy notice. Such a notice would ideally have several precision levels, thus the data subject can choose the level of precision he/she would like to read about the technicalities of big data processing and behavioural targeting. However, it does not resolve the possible complexity of the description of the big data processing as a mechanism. It was concluded that it is unlikely, this problem will be completely solved, but it is essential to implement every possible means to provide sufficient information to data subjects, for instance through simplistic description and examples.
Bibliography

List of literature


Gandomi, A., Haider, M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015/35, pp 138-144.


36. Varshney, K. Data is the world’s most valuable resource - so who will control it? Digital Directions, 10.05.17. Available at: https://digitaldirections.com/data-worlds-valuable-resource-will-control/ (05.12.18).
List of legal acts

6. Isikuandmete kaitse seadus. RT I, 06.01.2016, 10. Available at: https://www.riigiteataja.ee/akt/106012016010 (05.12.18).