Univerzita Karlova v Praze

Právnická fakulta

Dana Marečková

Profilování a právní úprava ochrany soukromí

Diplomová práce

Vedoucí diplomové práce: doc. JUDr. Zdeněk Kühn, LL.M., Ph.D.

Katedra teorie práva a právních učení

Datum vypracování práce (uzavření rukopisu): 5. května 2016

Charles University in Prague

Faculty of Law

Dana Marečková

Profiling and Legal Regulation of Privacy Protection

Master's Thesis

Master's thesis supervisor: doc. JUDr. Zdeněk Kühn, LL.M., Ph.D.

Department of Legal Theory and Legal Doctrines

Date of finishing of the thesis: 5 May 2016

Prohlašuji, že předloženou diplomovou práci jsem vypracovala samostatně a že všechny použité zdroje byly řádně uvedeny. Dále prohlašuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

Dana Marečková

Obsah

1. Introduction

The 21st century we are now living in is called the "information age." Information is becoming more and more important assets in nowadays economics. Sometimes it is called the oil of the 21st century. Businesses collect information about their users' background, interests, search queries, buying habits and anything else they might find useful. They do it by means of collecting and analysing a vast amount of data. The technology and lower costs of data storage made it possible to collect immense amounts of information about the customers. There are businesses that have no other business plan than expanding their customer base and collecting data. Sometimes they do not even plan to be profitable and despite that they have huge market value.[1] This shows how much value data have.

Being overloaded by data these days, there is a need for the help of technology (hardware) and techniques (software) to enable businesses to gain some useful knowledge out of it. The method helping with that is called data mining. It is a process of looking for correlations among huge amount of data. When the gained knowledge is used to build profiles, we speak about profiling. Each profile includes suggestions for how to treat the subject that fits in respective profile, e.g. suggestions for customized advertising. The knowledge of the customers' interests and desires is therefore a big competitive advantage. Businesses want to fine-tune their services to their clients' needs so that they want to use them. Customers' attention is a scarce commodity, which causes a big competition among businesses for getting the biggest part of it. Modern technology gives them the possibility to communicate personally with their customers for low costs so in case of succeeding in gaining the attention by targeting the customer with the right content, larger revenues can be expected on the side of the businesses.

On the side of the customer there are certain profits as well. He sees more content that is relevant to him, which saves him time he would have to spend with searching. Or he might get a special discount on the product he likes to buy. In this case, he probably does not mind the fact that all of this is a result of processing vast amounts of data about him. However, if he is charged more for products he is in a dire need of, he might stop feeling so comfortable about that and might start doubting whether he wants to share all the data about him or be it used that way. And there are other less apparent problems that can raise doubts, e.g.

[1] Grassegger, H.: *Das Kapital bin ich*, Zürich, Kein & Aber, 2014, p. 34, 37

regarding right for self-development or enjoying individual autonomy. Whereas the businesses want the customer to be transparent and provide as much data as possible, what they do with the data afterwards, which data is used for what purposes, where they are stored, how they are secured or who they further provide it to is often concealed.

If the user does not want to be subject to those practices, he has only one possibility: not using the services collecting data at all, thus loosing possibility to take part in almost any online life. But mostly, users prefer to use their personal data as a currency and trade future privacy risks for a short-term convenient service.

This thesis summarizes what data is used for creating profiles, how it is collected, what are the concerns regarding privacy, what is the actual legislation in the field of profiling (especially in data protection) in the European Union, what are its shortcomings and what is being proposed to ensure better privacy protection. Finding the right solution means finding the right balance between the interests of individuals and the society as a whole on privacy protection and the interests of private entities on conducting a profitable business and everyone's interest on innovations and good functioning of the market avoiding or correcting market failures. Because the legislation has to take into account the impacts on economics, it is necessary to describe also todays' business models. The main focus will be on the question how to put through the interest on taking part in the online world and using the online services without having the only possibility to pay for it with losing control over our data.

2. What is Profiling?

Profiling is the process of creating profiles and the following application of them. There are many different profiling methods that are used for various purposes. What all those methods have in common is that they are a set of technologies using algorithms[2] or other techniques to transform data into knowledge (creating profiles) and individuating a subject or identifying a subject as a member of a certain group so that the right profile is applied to them.[3] Profiling methods help us to cope with the information overload, to gain useful knowledge out of it and to address a large number of customers individually.

This work will focus on automated profiling, which is done by machines using the methods of data mining. Profiling includes five steps: data collection, data preparation, data mining, interpretation and determine actions.[4] Further, the crucial steps (collection of data and data mining) are described in more detail. But first of all will be explained what profiles are.

Profile

The simplest category of profiles is an individual profile. It consists of the factual part, including information about who the customer is, and the behavioural part, including rules about the customers' behaviour. The factual part is based on the collected factual data (age, gender etc.) that may be obtained directly from the consumer or derived from transactional data (buys cat food regularly → has a cat). The behavioural part is based only on transactional data consisting of records of the customer's previous purchases or transactions.[5]

---

[2] An algorithm is a set of orders that starts by taking an input (e.g. length and width of a rectangle) and ends after a finite number of steps by producing an output (e.g. the area of a rectangle; the steps would be: check if the input are two positive numbers and if not, let the user know that the input is wrong and end the computation, otherwise multiply the length by the width, return this value to the user and end the computation).

[3] Hildebrandt, M: 'Defining Profiling: A New Type of Knowledge?', in: Hildebrandt, M., Gutwirth, S. (eds.): *Profiling the European Citizen*, Springer, 2008, p. 17-18

[4] These steps were mentioned in the following book in relation to Knowledge Discovery in Databases (KDD) – a process of gaining useful knowledge out of a database, and they are applicable to profiling as well. Custers, B.: 'Data Dilemmas in the Information Society: Introduction and Overview', in: Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 8

[5] Adomavicius, G. and Tuzhilin, A.: 'Using Data Mining Methods to Build Customer Profiles', *Computer*, 2001, 34 (2), pp. 74-82, at p. 74

Group profiles contain behavioural rules for certain groups of individuals sharing at least one common attribute, e.g. left-handed people or students living in dormitories. Groups are identified especially by these two data mining methods: classification (used to map data into several predefined groups) and clustering (forming groups with similar properties)[6]. Within a group, either all of its members share the same attributes, e.g. a group of car driving license holders who are older than 15 (these groups are called distributive), or the profile does not apply to all of its members, e.g. people living in a certain area having average earnings of X; some earn X, some X+1 or X-1000 (these groups are called non-distributive). Most of the group profiles, except for the attribute they are defined by, are non-distributive. This distinction between group profiles is important for understanding the risks of profiling described in Chapter 3.

Collection of Data

Data is collected either directly from the user/customer or automatically by using technical means.

Directly it is gained from the users of Internet services by asking them to provide information about themselves – name, address, e-mail address, date of birth, phone number, credit card number etc. either by a voluntary registration or by a 'mandatory' one, in order to be allowed to use the service. This means that if they want to use the service, they have no other choice than providing their data.

But a lot more of other data may be collected while browsing the web without the user even being aware, e.g. what pages the user has visited, how much time he has spent there, from which site he has entered to the current site, what he does like etc. The collection methods are different, some of them are more accurate than others. The tools helping to gain more information about subjects are: requests including user's IP address, cookies, web bugs, JavaScript (majority of websites use this technology used e.g. for animations or interactive content and browsers support it; if user disables JavaScript, he cannot usually see all the website's content) or Flash (used sometimes on websites for displaying animations and videos). Now it is going to be explained how.

[6] Calders, T. and Custers, B.: 'What Is Data Mining and How Does It Work?', in: Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 31

IP address

IP addresses are unique identifiers of devices connected to the Internet enabling the communication between those devices. If a user wants to look at a website, his computer must send a request to the server hosting the website with a piece of information telling where to send the required content back to. Even though the IP address identifies a single device, monitoring from which IP address the request has come is not a very reliable method for identifying specific users in a long-term perspective, because firstly, the IP addresses can change on the side of the user often (when using mobile devices and changing locations or in case of dynamic IP addresses when the user is assigned a new IP address from a pool of addresses that their internet service provider has available each time they log on), and secondly, IP addresses specify rather the device than the concrete user. The device may be used by more family members or it may be a proxy server or a router serving many clients and sending the requested content further back to one of those clients.

Cookies

Cookies are small data files that are saved in user's browser which are sent along with the loading request to the server hosting the website every time the user wants to load it again. So they give information about whether the user had already visited the website before, what his language settings or his activities were, e.g. what he had purchased or looked for before or if he had logged in. They can be used to facilitate the proper function of information society services, e.g. remembering the items added to the shopping cart on an online shopping website, or they may be used for the purposes of monitoring the users' behaviour. The latter purpose have the third-party cookies, especially. If a web page includes links, images, HTML Iframe (presenting another HTML document within the same window) or similar components that are stored on some other servers than the respected web page and if the user hasn't blocked acceptance of these third-party cookies, than these cookies will be saved in his browser as well. The providers of these cookies are companies like AOL Advertising, DoubleClick (subsidiary of Google) and so on, that penetrate a huge number of sites. If a user surfs on different websites that include content from these companies, the users' browser always sends them the same cookie. That's why they have information from more websites he has visited. This presumes that the cookies belong to the category of persistent cookies which are not deleted after leaving the website – in comparison to session cookies, that are deleted when the user leaves the website.

Because it is possible to block the acceptance of the third-party cookies (in the settings of the browser), new methods of tracking the user's behaviour have been developed.

Web Bugs

A web bug is an object on a webpage, usually an invisible one pixel gif image, retrieved from a third-party server. In the moment of sending request for the webpage content, another request is sent also to that third-party server for this one pixel image. In that moment it is enabled to pass information to this third-party website about which device asked for it and when.

The most famous web bug nowadays is the Facebook like button (designed as a hand giving "thumbs up"). Any content provider can include this button on his website if he wants to make use of its promotion functions – because any Facebook user clicking on it will share the link to it with its friends. Every time when loading a page that includes this button, the website sends a request to Facebook's servers, handing him over information about what page it comes from, the visitor's IP address and the Facebook ID if the user is logged in. And because it is sending this request along with a cookie, Facebook is able to create a database of all the websites visited from a certain device or even by a concrete person (without needing to even have a Facebook account).

Clickstream

Every time when a user is browsing the web, he leaves a trail of data – a so called log file giving information about their IP address, date, time, the webpage that provided the link to the website and so on. Aggregating this data, it is possible to create a clickstream giving information on what websites and what particular webpages on that website a user has visited, in which order and how much time the user has spent there, when he clicked the 'back' button, what was the search query that got the user to the website etc. Clickstream gets stored on the servers hosting the website (monitoring the behaviour on its webpages), in the browser, on the routers of the internet service provider (ISP) or on the servers of advertising networks.

Fingerprinting

Fingerprinting is a technique based usually on JavaScript and Flash. The information collected are the type and version of browser, operating system, installed plugins, screen size, installed fonts, language, time zone and even hardware configurations. Tracking firms

use sophisticated methods to find out these unique characteristics.[7] Because of many different possibilities of settings, there is a high chance that the settings will be unique for each device.[8]

Data Mining

Data mining is a process of analysing a vast amount of data by algorithms looking for patterns and relations between the data. The relations can be trivial (revealing just facts that everyone knows), spurious (there is no actual causality), irrelevant (revealing correlations of no utility) or useful (revealing rules about customers' behaviour). As an observer you do not have to know beforehand, what kind of relation you are looking for. You are not just verifying if your theory is right, you are actually discovering all possible relations you need not have thought of before. After that follows a step called validation, when a human validator looks at the results and decides which ones are useful. It means that by doing data mining you are actually generating hypotheses and only after an interpreter sees them and finds them interesting (novel, useful and nontrivial to compute) and certain enough (data are accurate and complete, the sample is big enough and the results are significant), they become knowledge.[9] The correlations do not reveal any reasons for themselves. This would be the task of a further (human) interpretation, eventually. However, knowing the causes or reasons is not really of importance. Important is making good decisions about what to do next. For example in the case of ad networks: what ad to send to appear on the user's site.

The profiling technologies and techniques enable applying the profile as well. Making decisions about the business strategy towards consumers according to the knowledge gained by data mining is a very important step for businesses. Profiles are used either for cross-selling (trying to make a customer buy additional product or service on the base of the knowledge that a similar customer has made this decision) by means of making

[7] More information on those methods can be found here: Nikiforakis, N. and Acar G.: 'Browser Fingerprinting and the Online-Tracking Arms Race', available at: http://spectrum.ieee.org/computing/software/browser-fingerprinting-and-the-onlinetracking-arms-race (31 December 2015)

[8] A study by Peter Eckersley has proven this fact. The study has shown that from a sample of around half a million distinct browsers, 84 % had unique configurations. Among browsers that had Flash or Java enabled, 94% were unique. https://panopticlick.eff.org/static/browser-uniqueness.pdf (26 January 2016)

[9] Custers, B., 'Data Dilemmas in the Information Society: Introduction and Overview', in Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 10-12

recommendations (e.g. “other customers who bought this brand have also shopped for” or “frequently bought together”), one-on-one marketing (individualizing the virtual store or the offers for a concrete person) or banner targeting (tailoring the advertisements the user sees on the websites he browses on). What problems these practices may cause regarding privacy will be described in the following chapter.

3. Risks of Profiling in Relation to Privacy

As a first step, profiling requires collecting a lot of data. The huge amount of data (on browsing behaviour or online transactions) that is being collected is not a problem itself. The problems may start with analysing and finding meaning in it[10] that is intimate or sensitive for the respective subjects, something what they would like to keep private. Richard Posner does not agree that the process of data mining could invade privacy in any way because it is done by machines and privacy can be invaded only when human scrutiny occurs.[11] He makes this observation regarding analysing information by intelligence agencies to find a couple of people who are a potential threat for national security. However, by the private entities the situation changes a little because decisions are being made upon the gained knowledge, affecting everyone – both the individuals whose data has already been analysed as well as the potential future clients. Decisions that are being made might be considered unethical or unwanted within society (like excluding some individuals or groups from certain services).

Information provided by customers are provided with a certain intention or under certain circumstances. Many people do not object to the use of their personal data within that sphere or context they intended to share it in. However, if the data is used for completely other purposes, people do mind it and feel like their privacy has been breached. This is because the contextual norms of distribution and appropriateness (as Nissenbaum calls them) have been breached. Norms of distribution govern the flow of information, like who with and under what circumstances the information can be shared (e.g. it is expectable that my doctor shares some information about my health with his colleagues to discuss the right treatment for me but it is not expectable that he shares it with my boss). Norms of appropriateness govern what information is appropriate to reveal in a certain context (e.g. when with friends in a bar or at a job interview).[12] These spheres may be breached e.g. by third-party cookies, clickstream data collected across websites, mergers of companies, selling of data etc. Nissenbaum comes with a theory of contextual integrity which is

[10] Much of the raw data does not have to be personal at all and although lead to very "personal" revelations, e.g. regarding relationships, diseases or beliefs.
[11] Posner, R.: 'Our Domestic Intelligence Crisis', *The Washington Post*, December 21, 2005, available at http://www.washingtonpost.com/wp-dyn/content/article/2005/12/20/AR2005122001053.html (16 February 2015)
[12] Nissenbaum, H.: 'Privacy as Contextual Integrity', *Washington Law Review*, 2004, 79 (1), pp. 101-139, at p. 120

maintained when both of the norms, of distribution and of appropriateness, are upheld. These privacy norms vary from place to place, culture to culture or period to period.[13] Contextual integrity is the definitive value protected by the right to privacy.[14]

This is a new aspect in comparison to other earlier concepts of privacy that we will look at now. In general, it is hard to explicate what privacy is. The debates have started after publishing the famous article by Warren and Brandeis 'The Right to Privacy' and there is still no unity on that point until these days. To Warren and Brandeis it is primarily the "right to be let alone," protecting individuals from invasions of other parties.[15] It is interesting that their article was a reaction to the then development of tabloid media as nowadays we are facing the same question – what to cover under the term 'right to privacy' in the current information age to be able to protect it. Westin describes privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others."[16] His approach is marked by the new possibilities for tapping of telecommunication systems.[17] Agre defines privacy as "the freedom from unreasonable constraints on the construction on one's own identity."[18] Similarly, Cohen says that privacy is "more generally about preventing the seamless imposition of patterns predetermined by others,"[19] ensuring people the possibility to build their own identity without the pressure of what others think. In that sense privacy is understood as an 'intermediate' rather than a 'final' value.[20] In my view, all these definitions are a reaction to a then present state of technological development. As Kühn points out, nowadays we are facing privacy interferences that are almost invisible, but frequent and ubiquitous, which is different from the earlier interferences like search warrants,

[13] The problem is how to define the spheres as they may vary in time or according to culture and as they are influenced by technical innovations as well. Nissenbaum, H.: 'Privacy as Contextual Integrity', *Washington Law Review*, 2004, 79 (1), pp. 101-139, at p. 120, 138

[14] Dumortier, F.: 'Facebook and Risks of "De-contextualization" of Information', in: Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010, p. 129

[15] Warren, S. and Brandeis, L.: 'The Right to Privacy', *Harvard Law Review*, 1890, 4, available at: http://www.gutenberg.org/files/37368/37368-h/37368-h.htm (20 February 2016)

[16] Westin, A. F.: *Privacy and Freedom*, New York, Atheneum, 1967, cited in: Solove, D. J. et al.: *Information Privacy Law*, New York, Aspen Publishers, 2006, p. 41

[17] Kleve, P. and De Mulder, R.: 'Privacy Protection and the Right to Information: In Search of a New Symbiosis in the Information Age', in: Mercado Kierkegaard, S. (ed.): *Cyberlaw, Security & Privacy*, International Association of IT Lawyers, 2007, p. 206

[18] Agre, P. E. and Rotenberg, M. (eds.), *Technology and Privacy: The New Landscape*, Cambridge, MIT Press, 1997, p. 7

[19] Cohen, J. E.: *Configuring the Networked Self*, New Haven, Yale University Press, 2012, p. 150

[20] Rouvroy, A. and Poullet, Y.: 'The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy to Democracy', in: Gutwirth, S. et al. (eds.): *Reinventing Data Protection?*, Springer, 2009, p. 53

surveillance or communication tapping that are very clear and significant. The single interferences are not remarkable on its own but when combined together they can form a big privacy threat.[21] Earlier, it used to be the state or mass media who had represented the threat to privacy. Now this power shifted to the online content providers and big technological enterprises.

As we see, a lot has been written about the term privacy. Even though there is no precise definition of privacy, it cannot prevent us from having a discussion on how harmful the consequences of massive data collection and profiling can be. And neither from looking for the solutions to that.

However, privacy is not the only value worth protecting. It opposes to other public goods such as innovation and efficiency.[22] Businesses are not primarily interested in sniffing out information about one particular person. They are interested in coming with new ways how to gain attention of customers and their willingness to buy their goods or services. Innovations can lead to more efficiency in many different ways – starting producing goods consumers really want, not interrupting them with advertisements on what they are not interested in (thus saving them time) etc. When enjoying the benefits people often behave like they do not care about their privacy at all. They are either trading their privacy for some convenient services or they are sharing lots of details about their lives online just for the pleasure of sharing. Some people argue then that there is no need for protecting privacy when the people voluntarily share immense amounts of data about themselves with others. This results in the opinions that privacy is dead. But the voluntariness of sharing the data is questionable as will be described later and therefore the argument of the 'dead privacy' cannot be accepted.

In my view, although many people treat their privacy as a commodity and trade it for services, they want to do it consciously and want their data to be used in the context they provided it in. Even people using every social network possible do not like to be surprised by what data is collected and processed without them knowing it.

In addition to the possible disrespect of contexts, there is a problem with the bargaining power strength. The collectors are making use of the large market share they have or of the

[21] Kühn, Z.: 'Ochrana soukromí v internetové době', in: Šimíček, V. (ed.): *Právo na soukromí*, Brno, Masarykova univerzita, 2011, p. 111
[22] Schermer, B. W.: 'The Limits of Privacy in Automated Profiling and Data Mining', *Computer Law and Security Review*, 2011, 27 (1), pp. 45-52, at p. 49

technological architecture that enables them to collect data without obstacles. These are the conditions for success. The more customers and the more useful data you have, the more valuable your collection is. Exactly these things make finding solutions to the privacy protection problem so difficult. Either the collectors have to be deprived of that power or deprived of the possibilities of using it for whatever purposes they want.

After covering general problematic concepts related to profiling, the following part lists all kinds of specific risks emerging in the context of data mining and profiling, describing them in more detail.

Inaccuracy

If the raw data (i.e. data before processing; might not be meaningful by itself) is inaccurate or unreliable, it may be problematic if decisions towards groups or individuals are being made on its grounds. Other risks like discrimination can have more severe consequences if based on inaccurate data.[23] It is the task of data preparation step (preceding the data mining step) to assure a good quality of the raw data. In spite of that, mistakes can still occur. It can happen, that there is not enough data and individuals may be classified as members of a group even if they do not fit in it or may not be classified as members of a group they do fit in. These are called false positives and false negatives. The consequences may be either to the favour or to the detriment of the customer; the severance of the mistakes may vary. Although there are tools to fight against this issue like giving an individual a right to know what data is being held about him and a right to correct inaccuracies in the raw data or (again) data mining used for correcting mistakes, none of these methods is perfect, nor does any of them guarantee getting rid of all inaccuracies. On the contrary, they involve the same risk. Another problem is, if the conclusion (the result of the data mining), although based on true data, is wrong. However, so can be a human decision. Any protection against this is probably not possible.

Discrimination

Dividing individuals into groups and making different decisions towards them on this ground is the essential quality of profiling. Customers get different offers or promotions according to their date of birth, what they had purchased in the past or who they are. There

[23] Schermer, B. W.: 'The Limits of Privacy in Automated Profiling and Data Mining', *Computer Law and Security Review*, 2011, 27 (1), pp. 45-52, at p. 48

are opinions (like the one of Zarsky),[24] that these practices raise neither any legal issues, nor have any bad impact on the society. They just confirm how business works. Other opinion, as the one of Lessig, states that these profiling practices reconstitute the system of status and destroy the environment of equality that had been developed.[25] He sees that as harmful for the society. In my opinion, there is nothing wrong with trying to build a good relationship with customers who you want attract or keep. I do not see anything bad in this way the market works or any market failure that should be corrected except for price discrimination based on misusing information asymmetry to the detriment of consumer, which will be described further below.

However, sometimes can be the differentiation illegal – if someone is treated less favourably and it is based on religion or belief, gender, sexual orientation, race, colour, ethnic or social origin, genetic features, language, political or any other opinion, membership of a national minority, disability, property, birth or age (Article 21 of the Charter of Fundamental Rights of the European Union). The problem is how to recognize such illegal discrimination. It is not necessary that the database includes the field of nationality or race, if it is possible to deduce that information from other fields, e.g. a location, and thus discriminate indirectly. Proving a discriminatory behaviour presupposes knowing the decision making procedure. However, the concrete steps that are taken before making a decision are not a public thing. They may even form a trade secret that deserves protection as well.

According to Zarsky, data mining might offer a solution. As anti-discrimination laws should protect people against bigotry and subconscious resentment of other people, then in the case of data mining, which is done purely by machines, no such thing can happen. The only problem is that if decisions are made by the method of classification, it means, that the machine already got some hint on what the sorting criteria are. So the other method, clustering, should be used.[26] But what if data mining methods find correlations that, as a result, differentiate on the basis of 'forbidden criteria'? Is not the goal of anti-discriminatory laws to protect society against decisions based on these criteria? If the answer is yes, than I

[24] Zarsky, T. Z.: '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion', *Yale Journal of Law & Technology*, 2002-2003, 5, pp. 1-56, at p. 25

[25] According to Lessig, the system of hierarchy disappeared after the society became more mobile and flexible and it was no longer possible to track enough detailed information on people to make subtle distinctions between them according to their rank. Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, p. 221-222

[26] Zarsky, T. Z.: '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion', *Yale Journal of Law & Technology*, 2002-2003, 5, pp. 1-56, at p. 27-29

cannot see how this problem is solved in Zarsky's description. I think that if we want to protect society against these decisions, than it does not help if we say that machine analysis does not constitute a problem. I agree though, that the right place for regulation is within the field of anti-discriminatory laws and does not have to be repeated or specified differently in the data protection laws.

Price Discrimination

Consumer profiling combined with dynamic pricing may lead to price discrimination. Given enough information about consumers' demand and preferences (especially out of their transactions history or information on their background), it may be possible to find out what is the highest price a consumer would be willing to pay for a good. This is exactly the amount the seller wants from every customer to maximize his revenues. If the price was higher, customer would not be interested in the purchase anymore. If it was lower, the profit would be not only on the seller's side anymore, but on the customer's as well, as he would have been ready to pay even more.

Customers may even be manipulated by asking to pay a higher price for goods they urgently need or if he is in a hurry. Or the seller may remember that the customer was interested in a service yesterday, when the exchange rate was less advantageous for the seller than today, and therefore wanting the same (higher) price even a day later, when the exchange rate got better for the customer. Thanks to one-on-one marketing it is hard to reveal such manipulation. Even in case of revealing it, if the seller's position on the market is strong enough (monopolistic), he does not have to be aware of losing customers because of his behaviour.

Presumably, if the users knew that their data would be used in such a way, they would not provide it. Or at least not voluntarily. Or just for a proper compensation. The problem is that consumers are often not aware of what can happen with their data (they are uninformed, e.g. because of high transaction costs) or if they are, they have often only two options – be allowed to use the service and agree with providing the data or not providing the data and not being allowed to use the service, which raises other problems described later.

De-individualisation and Stereotyping

By using group profiles, people are being judged by characteristics that are true only with some probability for a specific individual (especially if the conclusion is made on the basis

of more than one factor). They are being treated as members of a group and not as individuals. Not all of the group characteristics have to be valid for them (as described earlier in Chapter 2). Individuals may thus feel injustice or being stigmatized. Moreover, if we count that people see themselves as developing morally, then judging them according to what has been collected about them so far cannot be totally reliable as the reality dynamically changes.

A similar problem is stereotyping. Individuals are categorized into a few predefined categories that do not reflect all nuances of every personality. The profiles then work as stereotypes. This preempts individuals to present themselves. However, in my opinion, a stereotype must be known to a broad public to have these effects and in my opinion it is not the case of online profiling activities by businesses. It could have been the case, if the profiles got known to a broader public.

Individual Autonomy

Making a border between what is personal information people want to keep for themselves and what they want to share with public or businesses is crucial for a free development of their thoughts and identities. According to Velleman, persons "have a fundamental interest in being recognized as a self-presenting creature."[27] This means deciding in which circle of people and what information they find appropriate to disclose about themselves. The reason is that people are influenced by what other people think of them and what behaviour is socially desirable.

The Federal Constitutional Court in Germany has acknowledged the importance of informational self-determination in his decision from December 15, 1983 (BVerfGE 65, 1 – Volkszählung) where it stated, that people must be able to ascertain who knows what about them and when. If they are unable to ascertain it, they may avoid behaviour that might cause harm to them and thus restrict their self-development.[28] The court decision is orientated on

[27] Velleman, J. D.: 'The Genesis of Shame', *Philosophy and Public Affairs*, 2001, 30, pp. 27-52, cited in: van den Hoven, J.: 'Information Technology, Privacy and the Protection of Personal Data', in: van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009, p. 316

[28] In original version the courts' statement is: "Mit dem Recht auf informationelle Selbstbestimmung wären eine Gesellschaftsordnung und eine diese ermöglichende Rechtsordnung nicht vereinbar, in der Bürger nicht mehr wissen können, wer was wann und bei welcher Gelegenheit über sie weiß. Wer unsicher ist, ob abweichende Verhaltensweisen jederzeit notiert und als Information dauerhaft gespeichert, verwendet oder weitergegeben werden, wird versuchen, nicht durch solche Verhaltensweisen aufzufallen. Wer damit rechnet, daß etwa die Teilnahme an einer Versammlung oder einer Bürgerinitiative behördlich registriert wird und daß ihm dadurch Risiken entstehen können, wird abweichende Verhaltensweisen jederzeit notiert und als Information dauerhaft

data that the government had been collecting (the Court was deciding about the constitution conformity of data collection by the state during census). Does the same apply to data collected by private entities? Can we compare the power of a state and the power of a private entity? The state is definitely interested in collecting data for the purposes of tax and criminal investigation. After finding out the individuals the state can use its powers against them to enforce laws. Private entities are interested in increasing revenues and are able to target individuals according to the decisions they made themselves. In that sense, it is the same. Only the 'legislative' and the 'executive' power is not really divided in case of private entities in comparison to the state. The state must obey certain procedural rules to do enforcement lawfully. For the private entities there are some rules as well, however, their effectiveness or what happens in the case of not obedience is questionable (see Chapter 4). Private entities act as 'legislators' as well. They determine what data will be collected, for what purposes and how it will be used. As the other party (an individual) does not have any bargaining power it is something like a law for them with which they comply without even willing to. All in all, private entities with these powers are more powerful than governments, but it is governments that are more restricted than businesses. This imbalance has historical grounds and should be redressed. It may seem as an easy task as it is the state that would restrict someone else than itself. But because of the technology architecture it is not that easy to solve this problem as the legal regulation is not enough powerful regulatory tool.

The profiling methods make manipulation of individuals by targeting advertising or information that users are looking for according to their interests (like in the case of search engines or media sites) easier. Manipulating and influencing the public is the essence of advertising. Its persuasive abilities can be even strengthened by targeting. The problems of this personalized advertising and informing may be: one-sided information, limiting the options (and thus normalizing) or pushing one to do things he would normally not do. Lessig

---

gespeichert, verwendet oder weitergegeben werden, wird versuchen, nicht durch solche Verhaltensweisen aufzufallen. Wer damit rechnet, daß etwa die Teilnahme an einer Versammlung oder einer Bürgerinitiative behördlich registriert wird und daß ihm dadurch Risiken entstehen können, wird möglicherweise auf eine Ausübung seiner entsprechenden Grundrechte (Art. 8, 9 GG) verzichten. Dies würde nicht nur die individuellen Entfaltungschancen des Einzelnen beeinträchtigen, sondern auch das Gemeinwohl, weil Selbstbestimmung eine elementare Funktionsbedingung eines auf Handlungsfähigkeit und Mitwirkungsfähigkeit seiner Bürger begründeten freiheitlichen demokratischen Gemeinwesens ist." Bundesverfassungsgericht, *Volkszählung*, BVerfGE 65, 1 (43), Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83

asks, "When the system seems to know what you want better and earlier than you do, how can you know where these desires really come from?"[29]

Another example of manipulation could be the case when a person suddenly changes their interests to the detriment of the seller (e. g. stops buying cigarettes and orders a nicotine patch) and he is pushed to change his decision back (getting special offers for cigarettes).

These practices may have impact not only on the individual but on the whole society as well. Several scholars warn that too much personalizing may threaten the democracy and freedom itself. [30] As the private entities have the power to affect what content a user sees, people might be confronted only with issues and opinions they support or, in the worst scenario, they might be provided only with a content that the provider likes. Although people naturally tend to surround themselves with the information they like or agree with (e.g. buying left-wing or right-wing newspapers), the ease with which they can do it thank to personalization technologies is new. They do not even have to make an effort, the online content providers do that automatically as it gets visitors coming back to their websites. In my view, this is the difference from newspapers that try to catch attention of a broader public and therefore the probability of being confronted with information challenging the readers' views is bigger. Not being confronted with other opinions or topics might lead to forming groups of people that no longer understand each other. This argument may sound a little far-fetched but it is necessary to have it in mind for the case that we would witness more specific dangers of that kind. Up until now, we witness rather the problems connected with making money, rather than trying to gain power and control public opinions.

Richard Posner points out that people invoking right to privacy often want to hide discreditable information and manipulate people around them. The other party should not be prevented from revealing these deceptions. It is fully legitimate that both parties want information about the other one to protect itself against disadvantageous transactions.[31] I think that what businesses are doing could be seen from this perspective as well. It is natural for both sides to search for all the possible relevant information and to reveal about themselves just the amount of information that is the most beneficiary for them, especially if it is about business transactions. Whereas the transaction costs of businesses of getting to

[29] Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, p. 220
[30] E.g. Sunstein, C.: *Republic.com 2.0*, Princeton, Princeton University Press, 2009; Pariser, E.: *The Filter Bubble: What the Internet Is Hiding from You*, New York, Penguin Press, 2011
[31] Posner, R. A.: 'The Right of Privacy', *Georgia Law Review*, 1978, 12 (3), pp. 393-422, cited in: Solove, D. J. et al.: *Information Privacy Law*, New York, Aspen Publishers, 2006, p. 41

know masses of their customers into the detail have got lower, the possibilities of consumers to reveal how they are being manipulated are still very constrained. If their only possibility not to be subject of data mining is not using hardly substitutable services, there is a question if there exists a market failure as customers do not really have a choice. Or they do only for the price of limiting themselves in full application of their right to self-development. These two arguments, i.e. right to self-development and information asymmetry problem (described below) seem to me to be the most significant.

Information Asymmetries

Information asymmetry is a well-known market failure leading to power imbalance between the parties that requires to be corrected.

As businesses collect huge amount of data about their customers and analyse them afterwards, they gain a lot of knowledge about them. This knowledge they may (mis)use for price discrimination, cutting certain groups from specific offers, precise targeting of advertisements or other media content and so on (as described above). All that with the purpose of increasing their revenues. On the other hand, the consumer hardly knows what knowledge the businesses have. They have no clue why such a decision is being made towards them, on what grounds. It is too hard for a consumer (if not impossible) to realize what data he is providing and how they contribute to his overall profile.[32] Consumers do not have enough information to know what the fair price of information they are providing is and even if they had, they would hardly be in a position to negotiate about it. There is no transparent and fair market for the commodity of identity-relevant information until now.

Abuse and Misuse of Profiles

One of the things a person may consider before providing their data is the chance of misuse for fraudulent purposes or for damaging their reputation. Data protection law includes rules on confidentiality and security of processing data. Without proper data security data protection law would lose its meaning. However, data protection law applies only to processing of personal data, i.e. information relating to identified or identifiable natural person. Anonymized data are not subject to these rules. This may be problematic if the subject to whom this data relates to can be easily identified anyway. Issues like selling or other handovers of data and combining them with another databases could be seen as

[32] Zarsky, T. Z.: '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion', *Yale Journal of Law & Technology*, 2002-2003, 5, pp. 1-56, at p. 34

belonging under this sub-headline as well because the eventual misuse could have more severe consequences.

Although the risk of misuse is real and the consequences of such conduct could be very unpleasant, data security will not be discussed in the following parts. It was mentioned briefly just to see the whole picture.

Understanding the problems data mining presents is vital for the analysis of what solutions can be accepted. But first of all, we will look at the current legal framework.

4. Current Legal Framework and Its Drawbacks

The previous chapters attempted to show that there is a strong incentive for businesses to collect as much data as possible to be able to mine them and use the gained knowledge as a tool for a further profit increase. Unless there is a strong and effective demand on the side of consumers for ceasing these practices, the only possibility to protect their interests is by legal regulation enacted by the states or better (to avoid obstacles to the single market), by the European Union. Now, we will look at the current legal framework. This work is going to focus on the EU legal framework. The following text will mainly describe data protection law (applied already in the stage of data collection) complemented by anti-discrimination law (that applies in the stage of profile application). Consumer protection law and e-commerce law are not going to be covered as they do not contain important provisions on online profiling in the European Union, although they theoretically could.

Data Protection Law

The two main legal bodies in force in the EU are the Data Protection Directive 95/46/EC (DPD) regulating processing of personal data and free movement of such data and the ePrivacy Directive 2002/58/EC regulating data protection and privacy in the digital age. For this thesis, the second Directive is relevant because of its Article 5 (3) regulating "cookies." Although the reason for this provision is that the information (e.g. cookies) is saved in user's terminal equipment that belongs to his private sphere and not because it is (or it is not) personal data, the description of the provision will be included in this chapter. These two legal frameworks were inspired by the OECD Privacy Principles (which are part of the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data) which were developed in the 1970s and adopted in 1980 and the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) from 1981. Because not all of the EU member states had ratified and implemented the Council of Europe Convention 108 and there were differences in the regulation between the member states, the DPD was proposed in 1990.[33]

[33] Nouwt, S: 'Towards a Common European Approach to Data Protection: A Critical Analysis of Data Protection Perspectives of the Council of Europe and the European Union', in: Gutwirth, S. et al. (eds.): *Reinventing Data Protection?*, Springer, 2009, p. 275

DPD sets rules under which it is possible to process personal data of individuals. The main principles (articulated in the OECD Privacy Principles) are that personal data (i) must be processed fairly and lawfully, (ii) must be collected for specified, explicit and legitimate purposes and processed only in a way compatible with those purposes, (iii) must be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed, (iv) must be accurate, complete and where necessary kept up to date and (v) can be collected and processed only if the data subject[34] has unambiguously given his consent or in other situations when it is 'necessary.' The data controller (vi) must inform the data subject of the purposes of the processing for which the data are intended (not later than at the time of collection) and of the recipients or categories of recipients of the data and (vii) he is responsible for the security of processing and the security of the personal data. The data subject (viii) has a right to obtain a confirmation from the controller whether or not data relating to him are being processed, has right to have the data rectified, erased, completed or amended if appropriate ("the individual participation principle") and is entitled to obtain the knowledge of the logic involved in any automatic processing of data concerning him, at least in the case of the automated decisions.

In January 2012, due to differences in the implementation of the DPD on the national level and technological progress leading to uncertainty how to deal with the risks associated, notably, with online activity, a new proposal on the General Data Protection Regulation (GDPR) was presented.[35] In December 2015 the outcome of the final trialogue of the European Commission, the Council of the EU and the European Parliament was published. The final form was accepted by the European Parliament in April 2016. The GDPR will replace the DPD. The ePrivacy Directive will remain in effect. The goal of the legal regulation is to balance the privacy of data subjects and the interest on free flow of information.[36] Since 2007, due to the Lisbon treaty, the right to protection of personal data has become one of the fundamental rights – for the first time in legal history. It is set in Article 8 of the Charter of Fundamental Rights of the EU after the right to respect for private and family life, making

[34] Data subject is an identified or identifiable natural person. (Article 2 (a) DPD as well as Article 4 (1) GDPR)

[35] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM/2012/011 final - 2012/0011 (COD)

[36] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM/2012/011 final - 2012/0011 (COD)

it something more than just an aspect of this general right to privacy.[37] After publication in the Official Journal of the European Union GDPR will enter into force and after two year transition period the Regulation will apply.

In the following part it is going to be described how the DPD and ePrivacy Directive regulate different stages of online profiling, what are its drawbacks and if the GDPR helps solving the current problems and how.

Collection Stage

DPD applies when personal data are being processed. Not all of the data used for creating profiles must be personal. Thus, the question that has to be answered is whether personal data are being processed or not. This question seems easy to answer (e.g. in comparison to the question if there is a discrimination which is crucial for determination whether anti-discrimination law applies). However, finding the answer is not always simple.

The problem is the definition of personal data. DPD states in its Article 2(a) that "'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity." The definition of personal data in the GDPR is the same like the one in DPD but includes more examples of identifiers, namely: "in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to his physical, physiological, genetic, mental, economic, cultural or social identity of that person."

According to the words of the GDPR definition, it could seem that all information connected to a technical identifier are personal data. However, there are exceptions if data are anonymized. Recital 23 states that "the principles of data protection should not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable." If the data is anonymized to the extent that it is no longer possible to identify an individual, it is not personal data. Anonymization must be distinguished from pseudonymisation. This is defined in Article 4 (5) GDPR as "processing

[37] Blume, P.: 'It Is Time for Tomorrow: EU Data Protection Reform and the Internet', *Journal of Internet Law*, 2015, 18 (8), pp. 3-13, at p. 5

of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person." Pseudonymised data stays personal data. Recital 23 states that "data which has undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered as information of an identifiable natural person." It is likely that in many cases online identifiers such as IP addresses, cookies and other tracking technologies are going to be subject to the GDPR. Recital 24 states in the current version that "individuals may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses, cookie identifiers or other identifiers such as Radio Frequency Identification tags. This may leave traces which, in particular when combined with unique identifiers or other information received by the servers, may be used to create profiles of the individuals and identify them." In the version from 2012 (Commission's proposal) this text was followed by sentence: "It follows that identification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances." This sentence was deleted in the current version and the former version did not include the words "in particular when", meaning that online identifiers only when combined with unique identifiers and other information received by the servers constitute personal data.[38] The current wording thus lost its unambiguity but, in my view, still means that those online identifiers are not personal data as such. Hartung thinks there is a risk that the Recitals will not be taken into consideration by the appliers and that an interpretation will be used that they are personal data, which would have significant consequences for the businesses.[39]

Another question is by whom the natural persons should be identifiable, if only by the controller or by any other person as well. The Recitals of both DPD and GDPR talk about the latter version but e.g. German courts have been deciding in favour of the former version. The risk of the latter approach is that almost everything could be considered personal data as it is almost always possible that someone exists who would be able to identify a person whom the data relates to. Custers points out the same risk because of technology providing

[38] Council of the European Union, General Data Protection Regulation, Version 21/04/15, Council's consolidated version of March 2015, available at: http://www.bvdw.org/fileadmin/downloads/mepo/Synopse_EUDSGVO_march-2015.pdf (15 March 2016)

[39] Hartung, J.: 'Neue Regulierungsaspekte in der EU-Datenschutzreform', in: Weber, R. H. and Thouvenin, F. (eds.): *Neuer Regulierungsschub im Datenschutzrecht?*, Zürich, Schulthess, 2012, p. 41

more and more means of linking data and increased dissemination of such data.[40] According to the Recital 26 of the DPD determining whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify the said person. The Recital 23 of the GDPR states the same and mentions in addition that the available technology, time, money and manpower should be taken into consideration as well. This is a clear attempt to limit the definition so that it is not too broad. If the definition was too broad, there would be a danger that data protection law would be too cumbersome and unable to really protect individual's rights and freedoms.

According to the possibility to identify an individual we can divide data to these categories: referential (data refers to a specific person, not just any person) and attributable (describe a situation or a fact without a reference to a specific person). Attributable data could go unprotected under the DPD definition.[41] However, it may happen that attributable data will become referential. Schermer describes three situations when it can happen:

1) adding referential data to attributable data,
2) a profile becomes so unique that it fits only one individual,[42]
3) linking a profile to an individual by means of unique identifiers (other than name, address or date of birth, e.g. an end device like a mobile phone or computer used by the individual).[43]

Van den Hoven opines that even data that seem to have no meaning should be protected because if combined with other data or after applying new tools on them they may start making sense.[44] This is why van den Hoven and Manders suggest to define the object of protection so that it includes not only referential data but attributable data as well, in terms of the notion of "identity relevant information."[45]

---

[40] Custers, B.: *The Power of Knowledge. Ethical, Legal and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Nijmegen, Wolf Legal Publishers, 2004, at p. 148

[41] van den Hoven, J.: 'Information Technology, Privacy and the Protection of Personal Data', in: van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009, p. 309

[42] A famous example is the 'AOL searcher 4417749'. AOL published an anonymized dataset of search queries but the researchers managed to find out the real name of a person behind those data.

[43] Schermer, B.: 'Risks of Profiling and the Limits of Data Protection Law', in Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 143-144

[44] van den Hoven, J.: 'Information Technology, Privacy and the Protection of Personal Data', in: van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009, p. 301

[45] van den Hoven, J.: 'Information Technology, Privacy and the Protection of Personal Data', in: van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009, p. 310

Data protection law of the EU fulfils this request with its concept of personal data (data relating to identified or identifiable natural person). According to the Article 29 Working Party Opinion 4/2007 on personal data, in the first situation described by Schermer an individual is indirectly identifiable, in the second situation an individual is indirectly identified and in the third he is directly identifiable because he can be distinguished from other individuals. As the Opinion states, "a name may itself not be necessary in all cases to identify an individual."

If a person is identifiable depends on a specific context. As already mentioned, for the data collectors it is often not important to know the exact identity (e.g. a name) of an individual. It is enough to recognize him as a member of a group and target him with a certain content. But the possibility to identify a person in combination with a cookie or similar tool makes data protection law to apply. If it is sufficient to single out a person (and it does not have to be a name), the associated profile is considered to be personal data.[46]

According to the Article 29 Working Party's Opinion 2/2010 on behavioural advertising "the behavioural advertising methods described in this Opinion often entail the processing of personal data as defined by Article 2 (a) of Directive 95/46/EC and interpreted by Article 29 Working Party. This is due to various reasons: i) behavioural advertising normally involves the collection of IP addresses and the processing of unique identifiers (through the cookie). The use of such devices with a unique identifier allows the tracking of users of a specific computer even when dynamic IP addresses are used. In other words, such devices enable data subjects to be 'singled out', even if their real names are not known. ii) Furthermore, the information collected in the context of behavioural advertising relates to (i.e. is about) a person's characteristics or behaviour and it is used to influence that particular person. This view is further confirmed if one takes into account the possibility for profiles to be linked at any moment with directly identifiable information provided by the data subject, such as registration related information. Other scenarios that can lead to identifiability are mergers, data losses and the increasing availability on the Internet of personal data in combination with IP addresses." Based on that we can come to a conclusion that most if not all profiling exercises fall within the scope of the DPD.[47]

[46] Schermer, B.: 'Risks of Profiling and the Limits of Data Protection Law', in: Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 144
[47] Schermer, B.: 'Risks of Profiling and the Limits of Data Protection Law', in: Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 144

Data subjects whose personal data are being processed must be (among others) provided information about the purposes of the processing for which the data are intended and the recipients or categories of recipients of the data and the existence of the right of access and the right to rectify the data concerning them (Article 10 (b) and (c) DPD and Articles 13 (1)(c), (e) and (f) GDPR). These rights are of the biggest importance in the sphere of profiling. The problem is that the purposes are sometimes specified too broadly or it is not clear what data are used for them. Only if personal data are further processed for historical, statistical or scientific purposes, it is not considered incompatible with the given purposes so the data subject does not have to be informed of these, but only provided that the data are not used to support measures or decisions regarding any particular individual (Recital 29). So this legal base cannot be used for the group profiling purposes.[48]

Article 6 (1)(c) DPD (and similarly Article 5 (1)(c) GDPR) states that personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed. On one hand, there is a requirement on data minimisation so that only the data that is really necessary is processed. On the other hand, requiring that the data is adequate means that there has to be enough information so that the profile is accurate.[49] These two requirements are thus a little bit contradictory. By limiting the range of data that may be processed the risk raises that the data mining will come to false results and will no more be as useful as expected.[50]

We can see that a series of tricky questions has to be answered to decide whether DPD (or GDPR) applies. The wording of the notion "personal data" has not changed much so the difficulty with answering this question if something is personal data or not is still the same. If the question is answered affirmatively, the collector must not forget about his information duties and must keep a certain quality of the data.

Construction of Profiles and Anonymization

Profiles are generally created by using data that has been anonymized. Anonymization is one of means of processing according to Article 2 (b) DPD (according to Article 4 (2) GDPR

[48] Schreurs, W. et al.: '*Cogitas, Ergo Sum.* The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector', in: Hildebrandt, M., Gutwirth, S. (eds.), *Profiling the European Citizen*, Springer, 2008, p. 248
[49] Schermer, B.: 'Risks of Profiling and the Limits of Data Protection Law', in: Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013, p. 147
[50] Schermer, B. W.: 'The Limits of Privacy in Automated Profiling and Data Mining', *Computer Law and Security Review*, 2011, 27 (1), pp. 45-52, at p. 49

as well).[51] If there is no connection of the data to an identifiable person, data protection law does not apply. The step of anonymization looks appealing for privacy advocates but in fact, it may mean that individuals are less protected. If profiles are based on anonymized data, neither the subject whose data were used for constructing the profiles, nor the subject to whom the profile is applied have the rights of individual participation because the DPD does not apply.

Because anonymization falls within the term "processing", the rights of the data subject in relation to anonymization are: right to be informed of it (Articles 10 and 11 DPD and Articles 13 and 14 GDPR)[52], provide consent to it (Article 7 DPD and Article 6 (1) GDPR) and object to the processing (Article 14 DPD and Article 21 GDPR).

According to Article 7 DPD and Article 6 (1) GDPR, the processing of personal data is lawful either if the subject has given consent to it or if it is "necessary" according to one of the other five conditions listed in the Articles. The problems of giving a consent that is freely given and informed will be described later. Another problem is that even if consent is not given and data are anonymized, it is impossible for the subject to find out that his data were used. Processing of data by means of anonymization without a consent could probably be legitimized also according to Article 7 (f) DPD (and similarly according to Article 6 (1)(f) GDPR) which states that personal data may be processed if "processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)."

According to Article 14 (a) DPD, at least in the situation that data is processed on the grounds of Article 7 (f) DPD, the data subject has a right "to object at any time on compelling legitimate grounds relating to his particular situation to the processing of data relating to him, save where otherwise provided by national legislation. Where there is a justified objection, the processing instigated by the controller may no longer involve those data." The

[51] The broad definition covers practically everything. Article 2 (b) DPD states that processing "shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure or transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction."

[52] Users should be informed that their data can be anonymized and used for construction of profiles. Schreurs, W. et al.: '*Cogitas, Ergo Sum.* The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector', in: Hildebrandt, M., Gutwirth, S. (eds.), *Profiling the European Citizen*, Springer, 2008, p. 247

GDPR includes such right in Article 21 (1) but puts the data subject in a better position. The data subject "shall have the right to object, on grounds relating to his or her particular situation, at any time" and "the controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject." In that case it is the controller that has to demonstrate compelling legitimate grounds, not the data subject.

In any case, even if consent is given, Article 14 (b) DPD grants the data subject right "to object, on request and free of charge, to the processing of personal data relating to him which the controller anticipates being processed for the purposes of direct marketing, or to be informed before personal data are disclosed for the first time to third parties or used on their behalf for the purposes of direct marketing, and to be expressly offered the right to object free of charge to such disclosures or uses." The GDPR enhances this protection because apart from Article 21 (2) that gives the data subjects "right to object at any time to the processing of personal data concerning him or her for such (direct) marketing, which includes profiling to the extent that it is related to such direct marketing," it adds in Article 21 (3) that "where the data subject objects to the processing for direct marketing purposes, the personal data shall no longer be processed for such purposes" under any circumstances.

Both of the rights in Article 21 (1) and in Article 21 (2) "shall be explicitly brought to the attention of the data subject and shall be presented clearly and separately from any other information." However, the fact that it is the data subject who has to take the initiative is probably the reason why the right to object to processing of personal data is seldom exercised.

Regarding construction of profiles one more topic will be discussed: processing of "sensitive data." Article 8 (1) DPD and Article 9 (1) GDPR prohibit the processing of sensitive attributes (ethnicity, religion etc.) if explicit consent is not given or other conditions are not met (e.g. data manifestly made public by the data subject etc.). Moreover, GDPR prohibits in Article 22 (4) profiling based only on sensitive data (unless the data subject has given explicit consent to it). It was already mentioned that current profiling technologies make it possible to deduce those information from other data, e.g. a postal code, a level of education etc., but this case is not explicitly regulated.

All in all, even if data is anonymized, the data subject must be informed that their data can be anonymized and used for construction of profiles, he must provide consent to it unless the processing is necessary for the purposes of the legitimate interests of the enumerated

subjects and depending on the basis of processing (consent or necessity) he has the respective right to object to the construction of profiles. In case of sensitive data the rules are slightly stricter.

Application of Profiles

After the profile is constructed the individuals to whom it is applied should have a right to know what the mechanism of the profile, that has been applied to them, is. Article 12 (a), third dash of the DPD gives every data subject the right to obtain knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15 (1) DPD. The problem is that, out of a definition of 'data subject,' if no personal data are processed, no information has to be provided. Further, the data controller may simply add a feature to make it seem like that the decision is not fully automated and avoid the obligation to provide knowledge of the logic.

Another article related to profiling is Article 15 DPD. It gives protection to "every person," giving them "right not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct etc." There are some ambiguities in this formulation. Firstly, what decisions are and secondly, which of them have legal effects or significantly affect an individual. Recital 58 of the GDPR gives examples of "automatic refusal of an on-line credit application or e-recruiting practices without any human intervention." Bygrave asks if advertising banners adjusting their content according to a visitor involve a decision being made as well.[53] Most probably, if it was considered as a decision, it would not have significant effects on the website visitor.

Article 22 (1) GDPR gives the same right to every data subject (not every person) and broadens the grounds when the data subject is not granted a right not to be subject to an automated decision. Not only when it is necessary for entering into, or performance of a contract and in cases authorized by law which also lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests, but also if he gave his explicit consent.

[53] Bygrave, L. A.: 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling', *Computer Law & Security Report*, 2001, 17 (1), pp. 17-24, at p. 19

In this case, it is again the natural person or the data subject who has to take the initiative to protect himself. A problem could be that in some situations the individual does not need to know that a profile is being applied to them. There is only a general information obligation in the case of existence of automated decision making that produces legal effects concerning him or significantly affects him (Article 13 (2)(f) GDPR).

Articles 7 (f) DPD and 6 (1)(f) GDPR state that personal data may be processed without the consent of the data subject if it is necessary for the legitimate interests pursued by the controller. Zuiderveen Borgesius analyses the possibility of basing processing of personal data for behavioural targeting on this provision and comes to a negative conclusion. Only in exceptional circumstances, like a bookstore processing the data to provide recommendations on its website, it might be allowed. But in other cases, even if the ad network had a legitimate interest (because of the freedom to conduct a business) and the processing would be necessary (they would prove that the large-scale tracking of people's behaviour is substantial for targeting individuals), the data subject's fundamental rights would in the end prevail.[54] On the other hand, Kotschy, when analysing the consequences of the new Regulation, points to formulation that European Parliament proposed to include in Article 6 (1)(f) and what is now included in Recital 47 GDPR. It states that the processing must meet the reasonable expectations of the data subjects that processing for this purpose might take place in the context of the relationship between the data subject and the controller. According to Kotschy, everybody knows nowadays that the free internet services are paid by providing data used for marketing purposes. This further use would then be lawful (as everybody can reasonably expect it) and the protection of customers would not be enhanced.[55]

To sum up, application of profiles is connected with certain rights of individuals (natural persons or data subjects): right to obtain knowledge of the logic on which the automated decision is based, right not to be subject to an automated decision which produces legal effects concerning him or significantly affects him. Applying these rights presupposes knowledge of the fact that an automated decision is being made towards the individual. There are different opinions on the question whether consent of the data subject is needed for using personal data for marketing purposes.

---

[54] Zuiderveen Borgesius, F. J.: 'Personal Data Processing for Behavioural Targeting: Which Legal Basis?', *International Data Privacy Law*, 2015, 5 (3), pp. 163-174, at pp. 167-169

[55] Kotschy, W.: 'The Proposal for a New General Data Protection Regulation – Problems Solved?', *International Data Privacy Law*, 4 (4), 2014, pp. 274-281, at p. 280

Consent to the Processing of Personal Data

Consent means according to Article 2 (h) DPD "any freely given specific and informed indication of his (data subject's) wishes by which the data subject signifies his agreement to personal data relating to him being processed." According to Article 7 DPD the data subject has to unambiguously give his consent to the processing of his personal data. It means that silence is not enough. An opt-out system is usually criticised as not sufficient for obtaining a consent either.[56] The data controller must be able to provide evidence that the data subject consented.

According to Article 7 (1) GDPR, consent should be given by a clear affirmative action. According to the Recital 25 a clear affirmative action means ticking a box when visiting an Internet website, choosing technical settings for information societal services or any other statement or conduct that clearly indicates the data subject's acceptance of the processing of their personal data. It means the opt-in system. Silence, pre-ticked boxes or inactivity should not constitute consent. Giving consent should not be hidden in the general terms any more. Article 7 (4) GDPR states that "when assessing whether consent is freely given, utmost account shall be taken of the fact, whether, among others, the performance of a contract … is made conditional on the consent to the processing of data that is not necessary for the performance of this contract." This wording sounds promising and it will be interesting to see what effect it will have.

Svantesson describes the thought of everyone giving free and informed consent as a "fairytale concept".[57] First of all, no one reads long descriptions of things he does not understand. Secondly, use of many free services that one may need for work or anything else important, is based on giving consent. Thirdly, it can happen so many times a day that it would take too much time to read everything carefully. And lastly, if a person reads what he is giving consent to, it does not mean he is able to really understand all consequences tied with it. Giving informed consent is almost impossible.

In the case of profiling exercises, when an identifier is used to read or write information to the terminal equipment, e.g. a cookie, Article 5 (3) of the ePrivacy Directive applies. It states that "storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber

[56] Zuiderveen Borgesius, F. J.: 'Personal Data Processing for Behavioural Targeting: Which Legal Basis?', *International Data Privacy Law*, 2015, 5 (3), pp. 163-174, at p. 170
[57] Svantesson, D.: 'The (Uncertain) Future of Online Data Privacy', *Masaryk University Journal of Law and Technology*, 2015, 9 (1), pp. 129-286, at p. 148-149

or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing." This requires a prior consent, i.e. an opt-in system. Consent may be given by settings of a browser.[58] Despite this consent, another one is needed if a company uses a tracking cookie to process personal data because Article 5 (3) of the ePrivacy Directive does not automatically provide legal basis for processing of personal data.[59]

We can see that the concept of informed and freely given consent is stricter and less ambiguously described in GDPR in comparison to DPD. A consent is needed for processing personal data of a data subject and for storing cookies on the user's terminal equipment unless other conditions are met that allow to do these activities without a consent. The concept of a consent is criticised as providing no real protection.

General Data Protection Regulation – Important News

Newly, GDPR includes in its Article 4 (4) the definition of profiling as following: "'profiling' means any form of automated processing of personal data consisting of using those 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." The processing has to be automated and according to the Recital 24 it is particularly an activity that leads to taking decisions or predicting the data subject's behaviour. The data subject has a right to be informed of the consequences of profiling decisions (Articles 13 and 14 GDPR). Recital 70 indicates that profiling can be related to direct marketing but does not have to be at whole extent. In the case that it is, the data subject has the right to object to profiling according to Article 19 (2) GDPR. Always when there is an automated decision based on profiling which produces legal effects or significantly affects him, the data subject must be informed about the logic involved as well as the significance and the envisaged consequences of such processing for them (Article 13 (2)(f) GDPR).

GDPR takes into account the possible discriminatory effects of decisions based on profiling, i.e. if based on race or ethnic origin, political opinions, religion or beliefs, trade union membership, genetic or health status, or sexual orientation, and states in Recital 58 that a

[58] According to Recital 66 of the Cookies Directive 2009/136/EC that amended the ePrivacy Directive 2002/58/EC.
[59] Zuiderveen Borgesius, F. J.: 'Personal Data Processing for Behavioural Targeting: Which Legal Basis?', *International Data Privacy Law*, 2015, 5 (3), pp. 163-174, at p. 173

controller should use adequate mathematical or statistical procedures and implement technical and organisational measures to prevent those discriminatory effects. However, in the binding text of the Regulation there is no such obligation.

In comparison to DPD the GDPR enhances in its Article 3 (2) the territorial scope of the EU data protection law as it will apply every time when personal data of data subjects residing in the Union is processed, where the processing activities are related to (a) the offering of goods or services to data subjects in the Union or (b) the monitoring of their behaviour as far as their behaviour takes place within the European Union, even if the controller is established outside of the EU. This clearly encompasses the profiling activities.

Other new provisions contained in GDPR that are enhancing the protection of individuals are "right to be forgotten" (Article 17), right to data portability (Article 20) and the obligation of a controller to process data in accordance with the principles of privacy by design and privacy by default. The last obligation will be described in more detail in the following chapter.

The form of regulation (and not directive as it has been so far) means stronger harmonization because regulation is directly part of the national law. However, practical applications are then on national data protection supervisory authorities and courts (unless CJEU gives their opinion on a subject). Legal terms and standards that are too general leave space for possible different applications of the same provision. The future will show if the intention of stronger harmonization succeeds.

All in all, we may see that the current data protection legislation is based on ex ante protection, not ex post protection. It is based on the presumption that it is possible to hide the information from observation easily, which is no more the case. Also, it is no more the input but the output (determining the decisions) that matters. As data collecting is here to stay because in many cases it is inevitable or wanted, Zarsky points out that the regulation should focus more on how data are used, i.e. on the later stage than the collection stage.[60]

[60] Zarsky, T.: 'Responding to the Inevitable Outcomes of Profiling', in: Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010, p. 57

Another common criticism of data protection law is that it includes only procedural rules. The regulated subjects only look at if they comply with those rules but do not really care if there are any real privacy threats as a consequence of their conduct.

Anti-discrimination Law

Article 14 of the European Convention on Human Rights (ECHR) provides protection against discrimination when enjoying other rights protected by the Convention. Protocol No. 12 to the Convention broadens the protection against discrimination to all rights ensured, not only to those mentioned in the Convention.

Other international treaties containing prohibition of discrimination are: the International Convention on the Elimination of All Forms of Racial Discrimination and the Charter of Fundamental Rights of the European Union.

These treaties provide protection against public authorities. They do not apply directly if discriminatory behaviour occurs between private parties. However, it is considered that the protection might be used in case of clear omission to protect individuals against discrimination by setting no obligations to private parties not to discriminate. It is the task of the national legislator to enact suitable laws. Theoretically, this may be applied to profiling practices as well.[61]

The national anti-discrimination laws are binding for private parties. In the Czech Republic, discrimination is forbidden regarding access to goods and services that are offered to a broad public. In the European Union such behaviour is forbidden by the Directives if based on sex, race, or ethnicity. Cases of such discrimination happened in the United States.[62]

The conditions of discriminatory behaviour are: (i) different treatment between two persons or groups (ii) in analogous or relevantly similar situations (iii) without an objective and reasonable justification, i.e. pursuing a legitimate aim.[63] It has to be proved that a person or a group was treated differently because of different characteristics. If the burden of proof is on the claimant, he is in a hard position as the decision mechanisms are not public. Only

[61] Schreurs, W. et al.: '*Cogitas, Ergo Sum.* The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector', in: Hildebrandt, M., Gutwirth, S. (eds.), *Profiling the European Citizen*, Springer, 2008, p. 259

[62] This behaviour was called redlining, i.e. denying services (such as banking or insurance) because of living in certain areas where usually certain ethnic group lives.

[63] Schreurs, W. et al.: '*Cogitas, Ergo Sum.* The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector', in: Hildebrandt, M., Gutwirth, S. (eds.), *Profiling the European Citizen*, Springer, 2008, p. 260

in some cases, e.g. in the Czech Republic if discriminated on the grounds of ethnicity, race or sex in access to goods and services, the burden of proof shifts to the defendant.[64]

However, there is no general agreement on prohibition of discrimination between private parties when it comes to providing services as the freedom of choosing the contractual party is a very important principle of private law.

[64] Article 133a of the Code of Civil Procedure.

5. Solutions

We have seen that there are many different interests that have to be balanced and that especially the interest of individuals on not having to choose between being excluded from using services or giving up control over huge amount of data about themselves is not sufficiently put through. Some people are downplaying this issue, saying that individuals should simply stop worrying about privacy. In 1999 the former CEO of Sun Microsystems Scott McNealy said: "You have zero privacy anyway. Get over it."[65] Advocates of this view say that there is too much information flowing around to regulate it and all rules would be unenforceable anyway. You cannot stop the technology. However, as described in Chapter 3, the data mining techniques can lead to consequences that require intervention. It is true, though, that emerging technologies complicate finding the right solution.

Regulation of this issue does not have to be done only by laws. On the contrary, it is almost necessary to use other means as well. Lawrence Lessig describes in his book four types of regulation: law, norms (e.g. ethical ones), market and architecture.[66] States can impose obligations and sanctions by laws, businesses can voluntarily regulate themselves by setting norms for their conduct, the demand for certain behaviour can be so strong that the invisible hand of the market will make it happen or whoever is in the power of designing the architecture of systems can enable only certain behaviour.

I have collected suggestions for improvements from "cosmetic measures" to more radical ones. I think the best solution would be the one that does not require drastic changes of legal or market environment but, at the same time, would eliminate the disadvantages of the current situation. I will try to introduce the solutions in this order while respecting the category of problems they fall into. As Zarsky describes, in order to find a suitable solution we have to make three inquires: (i) whether the proposal really solves the problem, (ii) what its side effects are and (iii) whether the outcome is fair and achieves equity among the participants in the information market.[67] These questions will be asked by each proposition together with the question if there are not too big obstacles to its realization.

---

[65] Sprenger, P.: 'Sun on Privacy: Get Over It', *Wired*, 26 January 1999, available at: http://archive.wired.com/politics/law/news/1999/01/17538 (17 March 2016)
[66] Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, p. 138
[67] Zarsky, T.: 'Desperately Seeking Solutions: Using Implementation-based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society', *Maine Law Review*, 2004, 56 (1), pp. 14-59, at p. 16

Fairytale Consent

The first problem to be discussed is the fairytale, i.e. uninformed and/or not freely given consent. Kotschy criticises this concept and gives suggestions on how to ensure that the consent is really informed. He acknowledges that most people do not have enough time and knowledge to understand the long privacy policies and suggests that several expert institutions would check the legitimacy of processing. He sees a potential in Articles 36 and 42 GDPR that go in this direction.[68]

Articles 35 and 36 say that the data controller shall carry out a data protection impact assessment in cases where data are processed for taking decisions based on profiling those data. If the result of the data protection impact assessment shows that there are high risks to the rights and freedoms of individuals, the supervisory authority should be consulted prior to the processing of personal data. It is the supervisory authority's task to use its powers to correct the measures that should be taken to protect individuals' rights and freedoms.

Article 42 presumes establishment of data protection certificates (seals and marks) demonstrating compliance with GDPR. It would be relatively easy to understand who complies with certain standards and if consumers got used to pay attention to it, businesses would be motivated to apply those standards to get certified.

The certification mechanisms have not been done yet so it is premature to assess if they are really able to protect individuals. The efficiency of these certification mechanisms would depend a lot on its final form. The GDPR involves so many open formulations that it would depend on their interpretation how broad the protection of individuals' rights would be in the end.[69] But generally, the advantage of this approach would be the user-friendliness and unburdening of individuals when assessing the level of their data protection. On the other hand, there would be not much space for adjusting the protection to individual wishes as the level of protection would be uniform. All the data collectors would have to be treated the same to ensure the equity among them and to ensure that individuals can rely on the results of these procedures.

---

[68] Kotschy, W.: 'The Proposal for a New General Data Protection Regulation – Problems Solved?', *International Data Privacy Law*, 2014, 4 (4), pp. 274-281, at p. 280

[69] It reminds of the competition law, which is so complex that there are just a few general rules in the statutes and then there are institutions that look at single cases and decide if a certain conduct is anti-competitive or not and prohibit such conduct or allow it under certain conditions.

Svantesson points to the fairytale concept of an informed and freely given consent as well and suggests a "nanny state" approach to data protection that would ban certain types of contract terms in certain types of contracts.[70] Individuals would not have the possibility to agree to certain terms and there would be no need for the fairytale consent. For answering the questions if this proposal would solve the problem and if the outcome would be fair would be decisive how the state would control and enforce the rules. The particular wording of the provisions would be decisive for assessment if it is not too strict and if it does not enshrine the business activities too much to the detriment of the consumer – these might be the negative side effects.

Overcharging As a Consequence of Information Asymmetry

Price discrimination, when used for overcharging customers by misusing information asymmetry, is unfair. Zarsky suggests solving the price discrimination problem by promoting secondary markets where customers would communicate with each other, creating a flow of information about the products and thus enabling revealing price discrimination practices. To implement this idea, an architecture for communication is needed that is independent on the seller, e.g. a "chat room" affiliated to the website.[71] Ensuring that these "chat rooms" are really independent would be probably hard to realize. But if that would be managed, it would be an elegant solution using market forces. However, I am not sure if it would solve the whole problem – customers that are in a hurry would probably not check the "chat rooms" as it would still take time to the customer to find the relevant information, not speaking of sharing the information with others.

Not relying on this market solution, a legal solution could be a ban on the price discrimination to the detriment of consumer (overcharging in case of dire need etc.) that could find place in the consumer protection law. The question would be than how to find out that this behaviour occurred to be able to enforce it. We can see that the market solution with its "chat rooms" could help with the first problem but again, making them work is the weakest point.

---

[70] Svantesson, D.: 'The (Uncertain) Future of Online Data Privacy', *Masaryk University Journal of Law and Technology*, 2015, 9 (1), pp. 129-286, at p. 149

[71] Zarsky, T.: 'Desperately Seeking Solutions: Using Implementation-based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society', *Maine Law Review*, 2004, 56 (1), pp. 14-59, at p. 54

Intransparency

The absence of knowledge and information on what data are collected and what happens with them brings some problems that require intervention. One of the things that remains private are the algorithms used for data mining, i.e. for the revelation of correlations and for the application of profiles. After the GDPR is in effect, the logic of the algorithms leading to taking decisions that are significant towards the individual will have to be revealed already at time when data are obtained. However, it does not mean that the user will be notified that the content he sees has been personalized. The obligation to do that could help individuals to be more aware of the fact what their data are used for and in what situations decisions are being taken towards them. This measure would need just a change of legislation.

Weitzner et al. have the opinion that we should design information infrastructure differently in order to enable society to control that the data collected about individuals are correct and that the automated decision processes towards them are logically grounded on permissible uses of personal information. This "transparency by design" architecture would help controlling whether the data is used only in legally-approved ways. The legal requirements would have to make such mechanisms available and effective, i.e. set rules on how broad the individuals' access rights would be, what would be the mechanism for correction of data, what would happen if incorrect data was used although the data subject had pointed out to that fact etc.[72] It would probably be difficult for most people to comprehend how exactly the data have been used if they wanted to check it. However, this is the same thing as with legal documents. It would probably be enough that at least some people would be able to understand that and be able to point to the eventual failures.

Few Control Over Data by the Data Subjects

The current technology architecture and market forces give businesses much power when it comes to data processing. In this section will be described how the users could gain more control over their data.

[72] Weitzner, D. J. et al.: 'Transparency and End-to-End Accountability: Requirements for Web Privacy Policy Languages', available at: https://www.w3.org/2006/07/privacy-ws/papers/34-weitzner-transparency-accountability/ (18 March 2016)

Platform for Privacy Preferences (P3P)

According to Lessig, nowadays we have architectures that deny individuals control over their data. He sees the way to respond to that in the Platform for Privacy Preferences (P3P) that would enable machine-to-machine negotiations.[73]

P3P is a platform that stores the user's privacy preferences in a machine readable form and checks whether a website that expresses its policy in the same form complies with the user's ones. Thus, reading of long legal documents to check whether the website the user is visiting is consistent with his requirements or not is no more needed. E.g. the user would want to be notified if the website asked for his e-mail address that can be given to third parties, but not if the website asked for it to pursue a contract with the user. The legislation would have to make it an obligation to include those machine readable privacy policy descriptions on the websites and make them enforceable.[74]

As a result, the user would have an immediate overview of the website's privacy policy, thus having more control over who knows what about him. This technological solution enables to take into account the different individual's wishes. The outcome is fair for the participants because the privacy policy applied is negotiated between them beforehand. However, the question is if users would not be forced to have their settings in compliance with the website's settings to be able to use the services offered. If so, everyone would have their settings according to the requirements of the website. This problem is not solved by P3P.

Privacy Enhancing Technologies (PETs)

Other tools helping to strengthen the protection of user's privacy are PETs. PETs are information and communication technologies enhancing individuals' control over collecting and processing their personal data or preventing too extensive processing of data. Examples of these techniques are anonymizers (hiding the real online identity), encryption tools, privacy proxies etc. Generally, they have narrow field of application. They can help to protect certain aspect that is needed to protect privacy as a whole, meaning that they function for the purpose of privacy protection rather if more of them are combined together and if the person knows what he is doing. E.g. data confidentiality (through cryptography and secure systems) and anonymous communication (that the real identity of the person that is leaving traces is hidden) are not enough to protect privacy. The anonymizers are

[73] Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, p. 232
[74] Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, p. 228

limited by how big the anonymity set (the group out of which the user can be distinguished) is and the power of observer (how easy is for him to analyse the information he has to identify the subject). Moreover, in the online environment using categorization, where the consumer's digital identity is his real identity, anonymizers do not protect against unwanted behaviour that is not based on the unique identity. Hence, in this case, PETs offer customers only a false perception of autonomy.[75] In spite of that, as already said, they can be a useful tool to protect privacy when combined together, which will be shown in the following section.

Privacy by Design

The main goal of Privacy by Design is that information technology systems are designed as protecting privacy from the beginning. The difference from PETs is that it is a whole system architecture that is created. To implement the architecture, appropriate PETs are used. The tools must cooperate on the level of the system so that it is able to protect privacy. The architecture should comply with all the data protection principles, as formulated by OECD. The users should have the possibility to set the systems to their wanted level of protection as different users have different wishes and expectations.[76]

In GDPR we can see support for these systems. Article 25 states that the controller shall implement appropriate technical and organisational measures which are designed to implement data protection principles and protect the rights of data subjects. Of course having regards to the state of the art and the costs of implementation. The controller may get a certificate that he complies with these Regulation requirements. So officially, there is a requirement for Privacy by Design but practically we cannot exactly say what it is, which makes the enforcement difficult.

Data Ownership and the Real Data Market

The fact that the customers do not participate, or just for a small countervalue, on the huge profits that customers' data brings businesses leads Grassegger to the idea of "retransforming" the market pretty radically. He suggests that individuals save most of their

[75] Zwick, D. and Dholakia, N.: 'Whose Identity Is It Anyway? Consumer Representation in the Age of Database Marketing', *Journal of MacroMarketing*, 2003, 24 (1), pp. 31-43, cited in: Gürses, S. and Berendt, B.: 'PETs in the Surveillance Society: A Critical Review of the Potentials and Limitations of the Privacy as Confidentiality Paradigm', in: Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010, p. 309

[76] Le Metayer, D.: 'Privacy By Design: A Matter of Choice', in: Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010, p. 325

data on their own servers or clouds to which would have access only those willing to pay for it. Every individual could have their own General Terms that would have to be accepted by the accessing party. He suggests creating a market for data where data would be provided only if reasonable equivalent value is offered back.[77] This would reduce the power imbalance between the parties on the information market. According to Cohen, trading information on this market would have a good side effect as it would develop the capacity for autonomous choice by individuals and thus bring benefit to the whole democratic society.[78]

This step would need a legislation constituting data about an individual their own property. Cohen says that property is simply how we talk about important things. If something is not owned, it is presumed to be accessible to all.[79] Whoever would use the data in an incompatible way could be sued on the basis of property rights. This measure would give the individuals the whole control over their data.

Personal Data Stores

The last technological solution that is going to be introduced combines features of previously mentioned solutions. It was invented on the Massachusetts Institute of Technology[80], taking into account the current political and legal context and offering its users to have their data being collected in their own "Personal Data Stores (PDSs)." Special software would accept requests (the code to be run against the data) from applications on the end-device of the user and then would send back an answer to the question (the output of the code) that the application wants, e.g. it would send back in which geographic zone the user is, but not the raw data (i.e. the exact location he spends time at).[81] There would be higher chance that the answers are anonymized. The user would save his personal data store on his own server or in a cloud of his choose. Data collection and processing would happen there and the user would have full control of it. One of the side effects would be

[77] Grassegger, H.: *Das Kapital bin ich*, Zürich, Kein & Aber, 2014, at pp. 69-73
[78] Cohen, J.: 'Examined Lives: Informational Privacy and the Subject as Object', *Stanford Law Review*, 2000, 52 (5), pp. 1373-1438, at p. 1426
[79] Cohen, J.: 'Examined Lives: Informational Privacy and the Subject as Object', *Stanford Law Review*, 2000, 52 (5), pp. 1373-1438, at p. 1379
[80] The paper describing this project was published in 2014. de Montjoye, Y.-A. et al.: 'openPDS: Protecting the Privacy of Metadata through SafeAnswers', *PLoS ONE*, 2014, 9 (7), available at: http://dx.doi.org/10.1371/journal.pone.0098790 (23 March 2016)
[81] A study has shown that mobility traces are very unique for individuals, thus even if anonymized it is possible to link the data to an individual. de Montjoye, Y.-A. et al.: 'Unique in the Crowd: The Privacy Bounds of Human Mobility', *Scientific Reports*, 3, 2013, available at: http://www.nature.com/articles/srep01376 (26 February 2016)

lowering the market entry barriers for new businesses because they would have the data always available in PDSs and would not have to collect them again on their own.[82] This is why businesses might be interested in this solution as well. To realize this idea would either mean using the market forces and persuading some large businesses to encourage it or to make a brand new legal regulation that would enact it.

Summing up, we can see that there are two different approaches – either the "nanny state" one, where the state aims to protect individuals whether they want or not or the one giving more powers to individuals to be able to have control over their data. Many of the propositions solve only part of the problems relating with profiling. To me, the new project of the PDSs have the biggest potential to solve all of the problems. However, bigger discussion of broader public about the practical sides of its realization would be needed.

[82] de Montjoye, Y.-A.: 'On the Trusted Use of Large-Scale Personal Data', *IEEE Data Engineering Bulletin*, 2012, 35(4), pp. 5-8

6. Conclusion

In this thesis I tried to explain why it is important to protect privacy in the era of new technologies that make possible deducing a lot of details about individuals that are identifiable with a high probability despite not using the "traditional personal data" (name, address etc.). The reasons were: possible discrimination, de-individualisation, manipulation and other misuses of information asymmetry or the severe consequences that misuse of data might have.

So far, when it came to privacy protection, the EU data protection law used to concentrate on the collection stage, especially. The logic was that to protect the data, they should be kept private as much as possible and can be collected only if necessary or if the data subject consented. However, nowadays when it is possible to collect vast amounts of data and get consent to its processing easily, it is more important to focus on how the data is used. The weaknesses of legal regulation have led to approach that technology must be involved as well. GDPR came with data protection by default, data protection by design and other requirements that show attempt to better respond to the current problems with privacy and data protection but unfortunately affirms that fighting with the problems by legal regulation is not easy. It is hard to describe exactly what the controller's obligations are and enforce it. The open and sometimes vague terms are not enough to make things change. They are necessary though in complementing other forms of regulation (norms, market and architecture).

The solutions proposed in this thesis are mostly mixtures of these different forms of regulations. They either incline to giving more power and responsibility to individuals that can decide what happens with their data or they give the task to ensure a good standard of protection to the state. That means that either the users will carry the burden of controlling their data or they will not have choice, e.g. to provide their data because of strict rules on data minimisation or to demand stricter level of privacy than the state would require.

Solutions like data ownership and storing data on some kind of personal data stores enabling control over what is being done with the data look the most realistic to me. I think the users should have the option to use something like personal data stores and not be excluded from using information society services. The consequence would probably be that not so many services would be for free but nowadays, with the possibility to have so many

customers, the prices would probably not be very high. However, the reactions to the idea of realizing the concept of PDSs are hard to predict. It has not been a subject of much debate. It would be interesting to hear the opinions on that from broader public.

This thesis did not cover all aspects of online profiling due to capacity reasons. Another large relating issue are e.g. territoriality of data protection law and transborder data flows. An eventual following work could include those topics as well as more detailed insight into the solution using PDSs and what would be needed to do if the PDSs solution should really start operating in our physical and virtual worlds.

Teze v českém jazyce

(dle požadavku čl. 43 odst. 3 Pravidel pro organizaci studia na PF UK)

Profilování a právní úprava ochrany soukromí

V médiích i vědeckých publikacích se často mluví o tom, že žijeme v tzv. informační době. Informace se v dnešní ekonomice stávají stále důležitější komoditou a někdy se označují jako ropa 21. století. Pro spoustu společností je shromažďování a analyzování dat základem úspěchu jejich podnikání. Sbírají informace o svých uživatelích, o jejich zázemí, zájmech, nákupních zvycích a o všem, co by jim pomohlo zjistit o nich co nejvíce. Často mají již samotná data obrovskou hodnotu, od níž se pak odvíjí i hodnota společnosti, případně jsou data ještě zpracovávána a zisk je tvořen z využívání výsledků jejich analýzy.

Ke zpracování velkého množství dat slouží metody data miningu. Data mining je proces hledání korelací mezi daty. Pokud jsou jeho výsledky použity pro tvorbu profilů, mluvíme o procesu profilování. Každý profil obsahuje např. návrhy, jaký obsah reklamy zacílit na jednotlivce, který do tohoto profilu patří. Je tedy jasné, že znalost zákazníkových zájmů a přání představuje velkou konkurenční výhodu, jelikož se získáním zákazníkovy pozornosti je spojena větší pravděpodobnost tvorby zisku.

Zákazník může pozorovat výhody tohoto procesu např. v tom, že nepotřebuje tolik času na vyhledání obsahu, který ho zajímá, nebo v získání slevy na produkt, který často kupuje. Čeho si již všimnout nemusí, jsou nevýhody s tímto procesem spojené, jako je užívání dat k úplně jiným než zákazníkem předpokládaným účelům, jejich sdílení se subjekty, které zákazník původně neočekával apod. Může se dokonce i stát, že informace o tom, že momentálně nějaký produkt nutně potřebuje, bude zneužita k nabídnutí vyšší ceny za tento produkt než obyčejně. Odpovědi na otázky, jaká data jsou o zákaznících sbírána, jak jsou dále zpracovávána, kde jsou ukládána a jak jsou chráněna, jsou pro subjekty údajů často nezjistitelné.

Údaje jsou sbírány pomocí různých technických prostředků, jako jsou IP adresy, cookies, web bugs či fingerprinting. Díky těmto prostředkům je možné identifikovat uživatele, aniž by si toho on sám musel být vědom. Často je přenos dat od uživatele směrem k poskytovatelům online služeb nutný k umožnění technického přenosu vyžádaných dat

zpět k uživateli (např. „sdělení" IP adresy uživatele, aby na ni mohl být zaslán obsah webové stránky, kterou si uživatel přeje zobrazit). Identifikace je dále možná i pomocí vědomého poskytnutí údajů při registraci apod.

Zpracovávání údajů za účelem vytvoření profilů a jejich aplikace může zasahovat do soukromí uživatelů. Ti poskytují své údaje za určitých okolností či s určitým záměrem. Pokud jsou údaje použity v úplně jiném kontextu, než ve kterém byly poskytnuty, jedná se podle teorie kontextuální integrity popsané Helen Nissenbaum[83] o zásah do soukromí.

Debaty o tom, co je to soukromí, začaly již v roce 1890 článkem Warrena a Brandeise Právo na soukromí. Ti tehdy reagovali na pokrok v oblasti fotografie a vznik bulvárního tisku a definovali právo na soukromí jako „the right to be let alone".[84] Westin popsal soukromí jako „nárok jednotlivců, skupin či institucí, aby určovali, jak a do jaké míry budou informace o nich šířeny mezi ostatními".[85] Reagoval tím na nové snadné možnosti odposlouchávání telefonní komunikace. V posledních zhruba dvaceti letech s rozšířením nových technických prostředků definuje Agre soukromí jako „svobodu od neodůvodněných omezení při budování vlastní identity"[86] a podobně Cohen tvrdí, že soukromí je „spíše o zabránění nepřerušeného utvrzování vzorů předurčených jinými".[87] Zdá se, že všechny pokusy o definování soukromí vždy reagují na nějaký aktuální vývoj (technologický, společenský), který přináší nové výzvy pro zajištění ochrany soukromí.

Rizika profilování lze spatřovat v mnoha ohledech. Jedním z nich je, že správnost aplikace profilů závisí na správnosti vstupních dat. Nejsou-li tato data správná, může to vést např. i ke zhoršení účinků diskriminace, což je jedno z dalších rizik profilování. Rozdělování jednotlivců do skupin a odlišné zacházení s nimi podle toho, do které skupiny patří, je základní esencí profilování. Nicméně některé rozlišování může být nežádoucí či dokonce nezákonné, a sice v případě, pokud je založené na pohlaví, rase, etnicitě, sociálním původu apod. Takováto kritéria mohou však být v profilech obsažena i jen nepřímo, proto je velmi těžké zjistit, zda k profilování na základě těchto kritérií dochází. Diskriminace může být i cenová, a to v případě, že prodávající upraví cenu zboží nebo služby na míru přesně tak, aby

[83] Nissenbaum, H.: 'Privacy as Contextual Integrity', *Washington Law Review*, 2004, 79 (1), str. 101-139, na str. 120
[84] Warren, S. a Brandeis, L.: 'The Right to Privacy', *Harvard Law Review*, 1890, 4, k dispozici zde: http://www.gutenberg.org/files/37368/37368-h/37368-h.htm (20. února 2016)
[85] Westin, A. F.: *Privacy and Freedom*, New York, Atheneum, 1967, cited in: Solove, D. J. a kol.: *Information Privacy Law*, New York, Aspen Publishers, 2006, str. 41
[86] Agre, P. E. a Rotenberg, M. (eds.), *Technology and Privacy: The New Landscape*, Cambridge, MIT Press, 1997, str. 7
[87] Cohen, J. E.: *Configuring the Networked Self*, New Haven, Yale University Press, 2012, str. 150

se rovnala tomu, kolik je zákazník ve své momentální situaci ochoten maximálně zaplatit. Toto chování hraničí s manipulací, kdy zákazník může být i motivován kupovat si zboží, které by jinak za obvyklou, pro všechny stejnou, cenu nechtěl. Dalším problémem profilování je jistá stereotypizace a určování chování vůči uživatelům na základě určitých zobecnění, která jsou pro jednotlivce platná jen s určitou pravděpodobností, a na základě údajů, které se mohou měnit s tím, jak se mohou měnit zájmy a preference lidské bytosti.

Panuje shoda na tom, že soukromí je velmi důležitým předpokladem svobodného rozvoje identity jednotlivce. Podle Vellemana mají osoby „zásadní zájem na tom, aby byli uznáváni jako sebeprezentující bytosti“,[88] čili bytosti, které samy rozhodují, jakému okruhu lidí o sobě sdělí jaké informace. I německý ústavní soud uznal tuto podmínku jako zásadní pro právo na vlastní rozvoj.[89] Ústavně je tedy toto právo zaručeno vůči státu, jak je to však vůči privátním subjektům? Vůči nim takovéto právo zakotveno není, přestože již nyní mají větší moc než státy, a tudíž představují větší hrozbu. S tímto souvisí i již zmiňované možnosti manipulace. Teoreticky bychom mohli být svědky i toho, že lidem bude nabízen jen obsah, který je výhodný pro jeho poskytovatele. Zatím jsme zvyklí na to, že obsah je přizpůsobován podle preferencí uživatele tak, aby se rád na stránky poskytovatele služeb vracel. Konfrontování lidí jen s obsahem, se kterým souhlasí nebo který se jim líbí, přináší také svá úskalí. V případě informací o dění ve světě by mohl mít za následek, že lidé nebudou mít navzájem ponětí o názorech jiných skupin, než do kterých patří, což je obecně nebezpečné pro demokracii.

Na druhou stranu, Posner upozorňuje na to, že je legitimní, aby smluvní strany zjišťovaly o sobě navzájem co nejvíc s cílem vyhnout se tomu, že budou druhou stranou podvedeny či zneužity. S tím lze jistě souhlasit, nicméně zákazníci či uživatelé jsou při zjišťování takovýchto informací v mnohem slabší pozici. Jejich jedinou obranou proti takovémuto chování ze strany poskytovatele zboží či služeb je často vůbec jejich nabídku nevyužívat, což opět ústí v omezování využívání svého právo na vlastní rozvoj. Problémy v oblasti práva na vlastní rozvoj a informační asymetrie jsou ze všech problémů, které profilování přináší, dle mého názoru ty nejvýznamnější.

---

[88] Velleman, J. D.: ‘The Genesis of Shame’, *Philosophy and Public Affairs*, 2001, 30, str. 27-52, cited in: van den Hoven, J.: ‘Information Technology, Privacy and the Protection of Personal Data’, in: van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009, str. 316

[89] Bundesverfassungsgericht, *Volkszählung*, BVerfGE 65, 1 (43), Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83

Informační asymetrie patří mezi způsoby selhání trhu. Při profilování se projevuje tak, že uživatelé často nemají ponětí, co o nich poskytovatelé služeb vědí, na základě jakých údajů a proč je s nimi zacházeno tak, jak je. Nevědí ani, jakou hodnotu jejich data pro podnikatele mají, tudíž neexistuje žádný „trh s informacemi o vlastní osobě". Jakožto selhání trhu by však tento problém měl být napraven. Návrhy řešení však budou popsány až po stručném popisu současné právní úpravy.

Nyní účinnou úpravu představuje směrnice 95/46/ES o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů (DPD) a směrnice 2002/58/ES o zpracování osobních údajů a ochraně soukromí v odvětví elektronických komunikací (ePrivacy směrnice), která je významná kvůli svému ustanovení čl. 5 odst. 3 týkající se cookies.

Hlavní principy úpravy jsou následující: (i) osobní údaje musí být zpracovávány za použití legálních a čestných prostředků, (ii) musí být shromažďovány pro specificky vymezené a legitimní účely a zpracovávány v souladu s těmito účely, (iii) shromážděné osobní údaje musí být relevantní a přiměřeně odpovídat účelům, pro který byly shromážděny nebo pro který jsou zpracovávány, (iv) osobní údaje musí být přesné, úplné a průběžně aktualizované, (v) mohou být shromažďovány pouze se souhlasem subjektu údajů nebo v jiných případech, kdy je to „nezbytné", (vi) správce údajů musí informovat subjekt údajů o účelech zpracování nejpozději v době shromažďování údajů a také o příjemcích nebo kategoriích příjemců údajů, (vii) subjekt údajů má právo obdržet potvrzení o tom, že jeho údaje jsou zpracovávány a má právo na opravu, výmaz nebo blokování údajů a je oprávněn získat oznámení o postupu automatického zpracování údajů, které se ho týkají, alespoň v případě automaticky přijímaných rozhodnutí.

Vzhledem k rozdílné implementaci DPD v různých státech a technologickému rozvoji zejména v oblasti online činností byl v lednu 2012 představen návrh nového obecného nařízení o ochraně osobních údajů (GDPR). Finální verze tohoto nařízení byla přijata Evropským parlamentem v dubnu 2016. V účinnost nařízení vstoupí v roce 2018.

Důležitým pojmem pro aplikaci těchto předpisů je pojem „osobní údaj". Podle článku 2 písm. a) DPD i článku 4 bodu 1) GDPR je osobním údajem „každá informace o identifikované nebo identifikovatelné osobě (subjekt údajů); identifikovatelnou osobou se rozumí osoba, kterou lze přímo či nepřímo identifikovat, zejména s odkazem na identifikační číslo nebo na jeden či více zvláštních prvků její fyzické, fyziologické, psychické, ekonomické, kulturní nebo

sociální identity“. GDPR přidává další identifikátory: lokační údaje, síťový identifikátor či prvky genetické identity.

Předpisy na ochranu osobních údajů se však neaplikují na anonymní údaje, tj. údaje, které jsou anonymizovány do té míry, že subjekt údajů již není identifikovatelný. Jednoznačně určit, kdy již nelze subjekt údajů identifikovat, však nemusí být jednoduché. Mohou nastat následující situace: a) údaje, podle kterých osobu identifikovat nelze, se propojí s údaji, které již tuto možnost dávají, b) údaje lze spojit s konkrétní osobou vzhledem k jejich jedinečnosti, c) údaje lze spojit s konkrétní osobou za pomoci jedinečných identifikátorů (jiných než jméno a adresa, např. koncové zařízení).

Identifikovatelnost osoby závisí na tom, kým má být identifikovatelná, zda jen správcem údajů nebo jakoukoliv třetí osobou. Dále záleží na tom, jak náročné (na čas a peníze) by identifikování osoby bylo. Důležitý je také kontext – pro poskytovatele služeb není významné, jak se osoba, na kterou cílí, skutečně jmenuje apod. – hlavní je rozpoznat ji a zacílit na ni správný obsah. Proto lze dojít k závěru, že na většinu, pokud ne na všechny činnosti související s profilováním, se vztahují předpisy upravující ochranu osobních údajů.

Anonymizování údajů je jedním ze způsobů zpracování údajů. Proto je k němu zapotřebí souhlasu subjektu údajů (nebo jiné okolnosti, která k takovémuto zpracování správce opravňuje). Další zpracovávání takto anonymizovaných údajů již předpisům na ochranu osobních údajů nepodléhá.

Všechny osoby mají právo „nestat se subjektem rozhodnutí, které vůči nim zakládá právní účinky nebo které se jich významně dotýká, přijatého výlučně na základě automatizovaného zpracování údajů“. Tato formulace obsahuje prostor pro výklad toho, co je to rozhodnutí, a která z nich zakládají právní účinky nebo se osob významně dotýkají. Je např. rozhodnutím i zobrazování personalizované reklamy na webových stránkách, které si uživatel zrovna prohlíží?[90] (Odpoveď zní pravděpodobně ne, a to vzhledem k požadavku významného dotčení subjektu.)

Ke zpracovávání údajů je buď zapotřebí souhlas, nebo musí být podle článku 7 písm. f) DPD, případně podle článku 6 odst. 1 písm. f) GDPR „nezbytné pro uskutečnění oprávněných zájmů správce nebo třetí osoby či osob, kterým jsou údaje sdělovány, za podmínky, že nepřevyšují zájem nebo základní práva a svobody subjektu údajů“. Tato podmínka bude

[90] Bygrave, L. A.: ‘Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling’, *Computer Law & Security Report*, 2001, 17 (1), str. 17-24, na str. 19

moci být právním důvodem zpracování osobních údajů pro reklamní účely spíše zřídkakdy. Proto se nyní zaměříme na souhlas subjektu údajů. Udělení souhlasu znamená podle článku 2 písm. h) DPD a podobně podle článku 4 bodu 11) GDPR „svobodný, výslovný a vědomý projev vůle, kterým subjekt údajů dává své svolení k tomu, aby osobní údaje, které se jej týkají, byly předmětem zpracování". Jeho největší nevýhodou je to, že se v reálném světě prakticky nevyskytuje.[91] Za prvé, nikdo nečte dlouhé popisy něčeho, čemu nerozumí. Za druhé, i kdyby je četl, musel by tak učinit i několikrát denně, což by zabíralo příliš mnoho času. A za třetí, i kdyby je přesto četl, neznamená to ještě, že z nich bude schopen pochopit veškeré následky. Dát skutečně informovaný souhlas je téměř nemožné.

Mezi nejdůležitější novinky obecného nařízení o ochraně osobních údajů patří rozšíření místní působnosti oproti DPD. Nařízení se bude aplikovat i v případě, že zpracovávání údajů probíhá mimo Unii, pokud jsou zpracovávány osobní údaje subjektů nacházejících se v Unii a pokud zpracovávání souvisím s monitorováním chování, k němuž dochází na území Unie. Další novinkou je definice profilování. Tím se rozumí dle článku 4 bodu 4 GDPR „jakákoli forma automatizovaného zpracování osobních údajů spočívající v jejich použití k hodnocení některých osobních aspektů vztahujících se k fyzické osobě, zejména k rozboru nebo odhadu aspektů týkajících se jejího pracovního výkonu, ekonomické situace, zdravotního stavu, osobních preferencí, zájmů, spolehlivosti, chování, místa, kde se nachází, nebo pohybu". Jedná se především o činnost směřující k přijímání rozhodnutí nebo k analýze či odhadu preferencí, postojů a chování osoby (recitál 24).

Ochrana osobních údajů je tedy založena na ochraně ex ante. Je založena na předpokladu, že osobní údaje lze snadno před ostatními skrýt, což již ale neplatí. Také již tolik nezáleží na vstupu, nýbrž na výsledcích analýzy dat a určení následného chování vůči subjektu, čili na výstupu zpracování osobních údajů. Proto by se spíš právní úprava měla soustředit na stadium užívání dat, nikoliv na stadium jejich shromažďování.[92]

Vedle soukromí však existují i jiné hodnoty, se kterými je třeba právo na soukromí poměřovat, neboť jsou důležité zase z jiných hledisek. Např. zájem na inovacích a zefektivňování a právo na podnikání. Nicméně je vidět, že především zájem na tom, aby se jednotlivec nemusel rozhodovat pouze mezi poskytnutím obrovského množství údajů o

[91] Svantesson, D.: 'The (Uncertain) Future of Online Data Privacy', *Masaryk University Journal of Law and Technology*, 2015, 9 (1), str. 129-286, na str. 148-149
[92] Zarsky, T.: 'Responding to the Inevitable Outcomes of Profiling', in: Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010, str. 57

sobě a mezi celkovým odmítnutím služby, kterou by jinak chtěl nebo kterou je dokonce „nucen“ využívat, není dostatečně obhajován.

K jeho prosazování není však právní regulace jediným prostředkem. Lessig popisuje ve své knize čtyři způsoby regulace: právo, normy (např. etické či dobrovolné samoregulační), trh (řízený pravidly poptávky a nabídky a neviditelnou rukou trhu) a architektura (ve smyslu jaké chování umožňují technologie).[93]

Pokud jde o právní regulaci, mohla by pomoci v případě řešení problému s informovaným souhlasem. Některé instituce by mohly hodnotit zásady ochrany soukromí určené uživatelům a udělovat certifikáty zaručující soulad těchto zásad s předpisy na ochranu osobních údajů. Trh by poté zajistil, že by uživatelé dávali přednost poskytovatelům služeb s těmito certifikáty, a ti by byli více motivováni si je opatřit. Svantesson pak navrhuje, aby některá ujednání ohledně zásad ochrany soukromí byla v zájmu ochrany subjektů údajů (jakožto slabší strany) zakázána.[94]

Jedním z dalších opatření je posílení transparentnosti za účelem odstranění informační asymetrie. Jedná se jak o zavedení smysluplného informování o postupu při automatickém rozhodování, které je zakotveno v článku 14 odst. 2 písm. g) GDPR, tak o možnost zavedení notifikací, že zobrazovaný obsah byl personalizován. Takováto povinnost zatím nikde zakotvena není. Architektonické řešení by spočívalo v zavedení systémů „transparency by design“, které by společnosti umožňovaly kontrolovat, zda údaje nashromážděné o subjektech jsou správné a zda postupy automatického rozhodování jsou logicky odůvodněné a založené na dovolených způsobech zpracování údajů.

Lepší možnost kontroly jednotlivců nad vlastními údaji by mohla přinést tzv. Platform for Privacy Preferences (P3P). Ta by umožňovala uživatelům nastavit požadavky ohledně ochrany soukromí, které musí webová stránka, kterou chce navštívit nebo služba, kterou chce využít, splňovat. Pokud by je nesplňovala, byl by uživatel notifikován a mohl by se rozhodnout o dalším postupu. Odpadlo by tak čtení dlouhých dokumentů, čímž by se posílila i informovanost případného souhlasu se zásadami ochrany soukromí. Nevyřešila by se tím ale slabá vyjednávací síla jednotlivce tak, aby nemusel vždy pouze jen přistupovat na požadavky poskytovatele služeb.

[93] Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006, str. 138
[94] Svantesson, D.: ‘The (Uncertain) Future of Online Data Privacy’, *Masaryk University Journal of Law and Technology*, 2015, 9 (1), str. 129-286, na str. 149

Architektura vystavěna na základě „Privacy by Design“ představuje systém užívající různých jednotlivých „Privacy Enhancing Technologies (PETs) tak, aby odpovídal principům ochrany osobních údajů a který umožňuje, aby si uživatelé nastavili svou úroveň ochrany. Podoba takovéhoto systému závisí na technickém vývoji a nákladech realizace již vynalezených technických řešení. Požadavek na takovouto architekturu je v GDPR již zakotven, nicméně s ohledem na předchozí větu bude jeho dopad na úroveň ochrany osobních údajů spíše nevýznamný.

Jako nejnadějnější způsob řešení se mi jeví rozšíření tzv. Personal Data Stores (PDSs). Tento nástroj by umožnil uživatelům ukládat veškerá data o své osobě na svém vlastním serveru nebo cloudu, čili ve vlastním úložišti dat. Aplikace třetích stran chtějící využít tato data by jen poslaly na toto úložiště svůj požadavek s algoritmem, který by na tomto úložišti provedl výpočet a poslal zpět jen výsledek výpočtu. Subjekt údajů by měl tedy pod kontrolou jak svá data, tak jejich zpracování. Pravděpodobnost zachování anonymity při takovémto procesu by byla mnohem vyšší než za současné situace, kdy data jsou shromažďována na serverech poskytovatelů služeb a výpočet probíhá tamtéž. Data už jednou nashromážděná by navíc takto byla dostupná všem, což by snížilo bariéry vstupu na trh nových poskytovatelů služeb. Zavedení tohoto řešení by obnášelo buď přemluvení nějakého velkého poskytovatele, aby zavedení tohoto řešení podpořil, nebo přijetí úplně nové právní úpravy. Ještě předtím by ale jistě měla proběhnout širší debata o následcích zavedení tohoto řešení.

References

Books

Agre, P. E. and Rotenberg, M. (eds.), *Technology and Privacy: The New Landscape*, Cambridge, MIT Press, 1997

Cohen, J. E.: *Configuring the Networked Self*, New Haven, Yale University Press, 2012

Custers, B.: *The Power of Knowledge. Ethical, Legal and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Nijmegen, Wolf Legal Publishers, 2004

Custers, B. et al. (eds.): *Discrimination and Privacy in the Information Society*, Berlin, Springer, 2013

Gutwirth, S. et al. (eds.): *Reinventing Data Protection?*, Springer, 2009

Gutwirth, S. et al. (eds.): *Data Protection in a Profiled World*, Dordrecht, Springer, 2010

Grassegger, H.: *Das Kapital bin ich*, Zürich, Kein & Aber, 2014

Hildebrandt, M., Gutwirth, S. (eds.): *Profiling the European Citizen*, Springer, 2008

Lessig, L.: *Code version 2.0*, New York, Basic Books, 2006

Mercado Kierkegaard, S. (ed.): *Cyberlaw, Security & Privacy*, International Association of IT Lawyers, 2007

Pariser, E.: *The Filter Bubble: What the Internet Is Hiding from You*, New York, Penguin Press, 2011

Solove, D. J. et al.: *Information Privacy Law*, New York, Aspen Publishers, 2006

Sunstein, C.: *Republic.com 2.0*, Princeton, Princeton University Press, 2009

Šimíček, V. (ed.): *Právo na soukromí*, Brno, Masarykova univerzita, 2011

van den Hoven, J. and Weckert, J. (eds.): *Information Technology and Moral Philosophy*, New York, Cambridge University Press, 2009

Weber, R. H. and Thouvenin, F. (eds.): *Neuer Regulierungsschub im Datenschutzrecht?*, Zürich, Schulthess, 2012

Periodical Articles

Adomavicius, G. and Tuzhilin, A.: 'Using Data Mining Methods to Build Customer Profiles', *Computer*, 2001, 34 (2), pp. 74-82

Blume, P.: 'It Is Time for Tomorrow: EU Data Protection Reform and the Internet', *Journal of Internet Law*, 2015, 18 (8), pp. 3-13

Bygrave, L. A.: 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling', *Computer Law & Security Report*, 2001, 17 (1), pp. 17-24

Cohen, J.: 'Examined Lives: Informational Privacy and the Subject as Object', *Stanford Law Review*, 2000, 52 (5), pp. 1373-1438

de Montjoye, Y.-A.: 'On the Trusted Use of Large-Scale Personal Data', *IEEE Data Engineering Bulletin*, 2012, 35(4), pp. 5-8

de Montjoye, Y.-A. et al.: 'Unique in the Crowd: The Privacy Bounds of Human Mobility', *Scientific Reports*, 3, 2013, available at: http://www.nature.com/articles/srep01376 (26 February 2016)

de Montjoye, Y.-A. et al.: 'openPDS: Protecting the Privacy of Metadata through SafeAnswers', PLoS ONE, 2014, 9 (7), available at: http://dx.doi.org/10.1371/journal.pone.0098790 (23 March 2016)

Kotschy, W.: 'The Proposal for a New General Data Protection Regulation – Problems Solved?', *International Data Privacy Law*, 4 (4), 2014, pp. 274-281

Nissenbaum, H.: 'Privacy as Contextual Integrity', *Washington Law Review*, 2004, 79 (1), pp. 101-139

Schermer, B. W.: 'The Limits of Privacy in Automated Profiling and Data Mining', *Computer Law and Security Review*, 2011, 27 (1), pp. 45-52

Svantesson, D.: 'The (Uncertain) Future of Online Data Privacy', *Masaryk University Journal of Law and Technology*, 2015, 9 (1), pp. 129-286

Zarsky, T. Z.: '"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion', *Yale Journal of Law & Technology*, 2002-2003, 5, pp. 1-56

Zarsky, T.: 'Desperately Seeking Solutions: Using Implementation-based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society', *Maine Law Review*, 2004, 56 (1), pp. 14-59

Zuiderveen Borgesius, F. J.: 'Personal Data Processing for Behavioural Targeting: Which Legal Basis?', *International Data Privacy Law*, 2015, 5 (3), pp. 163-174

Documents Obtained from the Internet

Eckersley, P.: 'How Unique is Your Web Browser?', available at: https://panopticlick.eff.org/static/browser-uniqueness.pdf (26 January 2016)

Nikiforakis, N. and Acar G.: 'Browser Fingerprinting and the Online-Tracking Arms Race', available at: http://spectrum.ieee.org/computing/software/browser-fingerprinting-and-the-onlinetracking-arms-race (31 December 2015)

Posner, R.: 'Our Domestic Intelligence Crisis', *The Washington Post*, December 21, 2005, available at http://www.washingtonpost.com/wp-dyn/content/article/2005/12/20/AR2005122001053.html (16 February 2015)

Sprenger, P.: 'Sun on Privacy: Get Over It', *Wired*, 26 January 1999, available at: http://archive.wired.com/politics/law/news/1999/01/17538 (17 March 2016)

Warren, S. and Brandeis, L.: 'The Right to Privacy', *Harvard Law Review*, 1890, 4, available at: http://www.gutenberg.org/files/37368/37368-h/37368-h.htm (20 February 2016)

Weitzner, D. J. et al.: 'Transparency and End-to-End Accountability: Requirements for Web Privacy Policy Languages', available at: https://www.w3.org/2006/07/privacy-ws/papers/34-weitzner-transparency-accountability/ (18 March 2016)

Official Documents

Bundesverfassungsgericht, *Volkszählung*, BVerfGE 65, 1 (43), Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83

Council of the European Union, General Data Protection Regulation, Version 21/04/15, Council's consolidated version of March 2015, available at:

http://www.bvdw.org/fileadmin/downloads/mepo/Synopse_EUDSGVO_march-2015.pdf (15 March 2016)

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 L 281/31

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L 201/37

Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws [2009] OJ L 337/11

Position of the Council at first reading with a view to the adoption of a Regulation of the European Parliament and of the Council 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), ST 5419 2016 INIT - 2012/011 (OLP)

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM/2012/011 final - 2012/0011 (COD)

Summary

The purpose of this thesis is to elucidate what online profiling is, what happens with users' or customers' personal data during this process, how these activities interfere with the individuals' right to privacy, what the legal regulation in this field is, whether the privacy interests of individuals are sufficiently protected and if not, how the situation might be improved.

The thesis starts with description of todays' business practices that are based on collecting data about customers, analyzing it and creating profiles suggesting the most profitable behaviour of businesses towards customers. It is followed by explanation of the technological tools enabling data collection and the method of data mining that is the key enabler of creating profiles.

The text continues with description of risks of profiling in relation to privacy, i.e. the issues of discrimination, de-individualisation, restriction of individual autonomy, information asymmetries and possible misuse of profiles. The notion of right to privacy is explained and other interests that have to be balanced with privacy are mentioned as well.

After that follows a critical description of the current legal framework in the European Union. It consists of Data Protection Directive, ePrivacy Directive and since 2018 of the General Data Protection Regulation replacing the Data Protection Directive. The drawbacks of the legal regulation resulting in insufficient privacy protection are pointed out and then their possible solutions are introduced.

The proposed solutions are mostly based on the following forms of regulation: law, market and architecture. The advantages and disadvantages of each solution are described. As the most promising solution is suggested using so called "Personal Data Stores" that enable users having better control over their data and its processing.

Shrnutí

Cílem této diplomové práce je ozřejmit co je online profilování, jak je při něm nakládáno s osobními údaji zákazníků či uživatelů, jak tyto činnosti kolidují s právem na ochranu soukromí jednotlivců, jaká je právní úprava této oblasti, zda zájem na ochraně soukromí je dostatečně chráněn a pokud ne, jak by se tato situace dala napravit.

Práce začíná popisem dnešních obchodních praktik založených na shromažďování údajů o zákaznících či uživatelích, jejich analyzování a tvorbě profilů obsahujících návrhy na chování poskytovatelů služeb vůči svým zákazníkům tak, aby poskytovatelům přinášelo co největší zisk. Následuje technický popis nástrojů umožňujících shromažďování údajů a metody data miningu, která je klíčová pro vytváření profilů.

Text pokračuje popisem toho, jaká rizika pro ochranu soukromí profilování představuje, tj. riziko diskriminace, deindividualizace, omezování práva na osobní rozvoj, informační asymetrie a možné zneužívání profilů. Při této příležitosti je vysvětlen i pojem práva na soukromí a jsou zmíněny ostatní zájmy, které by měly být se zájmem na ochraně soukromí vyvažovány.

Následuje kritický popis současné právní regulace v Evropské Unii. Ta je tvořena směrnicí o ochraně osobních údajů, ePrivacy směrnicí a od roku 2018 obecným nařízením o ochraně osobních údajů, které nahradí dosavadní směrnici o ochraně osobních údajů. Je poukázáno na nevýhody těchto právních úprav vedoucích k nedostatečné ochraně soukromí a poté jsou navržena řešení zmíněných nevýhod.

Tato řešení užívají většinou následující formy regulace: právo, trh a (systémovou) architekturu. U každého z řešení jsou popsány jeho výhody a nevýhody. Jako nejslibnější řešení je navrhováno rozšíření užívání tzv. „Personal Data Stores“, které by umožnily uživatelům mít lepší kontrolu nad údaji o své osobě a nad jejich využíváním.

Dana Marečková

Profiling and Legal Regulation of Privacy Protection

Keywords: data protection law - privacy protection – profiling

Dana Marečková

Profilování a právní úprava ochrany soukromí

Klíčová slova: ochrana osobních údajů - ochrana soukromí – profilování

2016