INFORMATIONAL PRIVACY AND THE INTERNET

Nina Louise Arberg

Information does not exist in isolation. There are consequences. Context.

Abstract

We live in the Age of Information, where personal information has become a currency, where companies use aggressive tactics in gaining access to the individual's personal information for their own profit. This thesis focus on the prevalence of collecting and using personal information in an online setting in relation to the individual's *informational privacy*. The thesis will explore different approaches to privacy as a concept and provide a definition of *informational privacy* according to Floridi and ontological interpretation of *informational privacy.* Where the individual's *informational privacy* is a function of the ontological friction in the infosphere and the degree of ontological friction is determined in relation to factors within the infosphere. Furthermore, the thesis will explore the ways information and communication technologies has re-ontologized the infosphere and changed the individual approach to *informational privacy* in relation to sharing personal information in an online setting and the collection and use of personal information. Using Floridi's theory of *informational privacy* to analyse the impact of internet cookies and data mining on the individual's *informational privacy*. Furthermore, Facebook and Google are two of the biggest collectors of personal information, as such they will be used as cases and analysed in relation to how their collection and use of the individual's personal information, impacts *informational privacy* in the infosphere. Additionally the thesis will explore benefits for EU citizens brought on by the implementation of the GDPR on May 25 of 2008, in relation to collection, use and protection of personal information in an online setting.

Contents

Introduction

We live in the Digital Age, where using the Internet has become an integrated part of our daily lives, from communication and information seeking to shopping online in virtual stores. We create profiles on Social Network Sites (SNS) to share our lives and personal thoughts with one another, staying connected with friends, sharing pictures of our kids, friends we spent time with and what we had for dinner last night (Wellman & Haythornthwaite, 2002; Boyd, 2007). Our mobile phones have evolved into small pocket computers, able to connect to the Internet via mobile data or wi-fi networks (Schneier, 2015). This means that we can now be online almost anywhere and at any time and with 3.57 billion Internet users in 2017[1], the result is that personal information about the individual has now become a commodity to be sold and traded (Angwin, 2014). Al Gore stated in 2014, *"We now have a stalker economy where businesses are finding out everything about you."*[2] A very accurate statement, when looking at the amount of personal information being collected on Internet users. For example, Facebook collects around 600 terabytes of data daily from the 2.16 billion active users[3] (Sumner & Rispoli, 2016).

Personal information as a commodity, is also reflected in the use and development of tracking techniques and information gathering tools being used in an online setting (Sumner & Rispoli, 2016; Schneier, 2015). What started as simple internet cookies, used to keep track of passwords and running websites, has evolved into super-cookies and browser fingerprinting tracking users across websites, while collecting as much personal information about users as possible. Alongside collecting any personal information, publish willingly by users on social media sites, such as Facebook and Twitter (Castelluccia & Narayanan, 2012).

The main reason is profit, by using data mining on the personal information collected, it can be used to provide companies with strategies regarding online advertisement and the patterns in users behaviours can be used in the development of websites. The profiles on internet users and their interests, is also used for optimising search engines and development of products (Brock, 2016; Dixon & Gellman , 2011; Sumner & Rispoli, 2016). Some companies have specialised in the collection and aggregation of information from various sources, either

[1] https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/
[2] https://www.usatoday.com/story/tech/2014/06/10/al-gore-tech-southland-conference/10299753/
[3] https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

using the techniques mentioned above or buying large set of data from other companies, by merging all the information in to massive databases and selling these to be used in data mining (Etzioni, 1999). Because information has become a currency and is being sold to whomever is willing to buy both governments and private companies. The result is a growing concern about what exactly is being collected about the individual, how this information is being used and shared with (Angwin, 2014). This concern can be justified by looking at the recent case with Cambridge Analytica and Facebook, where Cambridge Analytica using questionable means, obtained personal information about around 50 million Facebook users. Many users were seemingly unaware that Cambridge Analytica was collecting their personal information, this was due to their friends using the app gave Cambridge Analytica access to their friends-list, with this access Cambridge Analytica was able to collect public personal information about the profiles of users not using the app. The information was gathered by a third-party company under the pretence of being used for academic purposes but was late sold to another company to be used for data mining. Some of the practices used here, have since been banned by Facebook[4], however this doesn't mean that it can't happen again. This shows that when personal information is collected by third-party companies, the individual can very easily lose control over what happens to the information. Resulting in an apprehension about the violation of an individual's *informational privacy*, even if the information collected is anonymous, it can be used to correctly presume other information that the individual wish to keep private (Etzioni, 2015).

The aim of this thesis is to examine the concept of *informational privacy* using Floridi's (2005) theory as presented in the research paper, *The ontological interpretation of informational privacy*. The impact of this collection and use of personal information in an online setting has on the individual's level of *informational privacy*, the focus will be on the commercial side of collection and use. With companies becoming more aggressive in collecting personal information about an individual to use for their own gain, it has become more important to understand how this is done. From this perspective *informational privacy* has become an important issue to be discussed, due to the various forms of personal information being collected, some very sensitive in their nature. Furthermore, I will examine how the re-ontologization of the infosphere has impacted the individual's *informational privacy* and can

[4] https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

both enhance and erode the *informational privacy* by focusing on the tools and methods used by companies in the collection and use of personal information. Additionally with the implementation of the GDPR in the EU, that seeks to providing EU citizens better protection and a greater control over of their personal information in an online setting. I therefore believe that researching the methods of collecting and using personal information and their impact on the individual's *informational privacy* is needed to gain a better perspective of what can be done to best ensure the protection of one's *informational privacy* and personal information.

Research questions

1. How is informational privacy defined?
   1.1. How has ICTs effected the individual's informational privacy?
2. How does the gathering of personal information effects the individual's informational privacy in an online setting?
3. How does Google's collection and use of personal information effect the informational privacy?
4. How does Facebook's collection and use of personal information effect the informational privacy?
5. How does the EU General Data Protection Regulation benefit the citizens in the EU in relation to informational privacy?

Structure of the thesis

The first section will start with defining and highlighting different approaches to privacy as a concept. The focus will be on Floridi's (2005) *ontological interpretation of informational privacy*, identifying some key elements in his theory. The aim here is to explain in which way Floridi's theory can be used in analysing the level of *informational privacy* had by the individual in an online setting. Furthermore, the section will explore how information and communication tools (ICTs) has re-ontologized the infosphere in becoming more digital and how this re-ontologization has changed the environment of the infosphere in relation to accessibility of an individual's personal information.

The second section will concentrate on the collection and use of personal information in an online setting. The focus will be on analysing how personal information is being collected and

used in relation to internet cookies, browser fingerprinting and data mining to build profiles on individuals and used in online targeted advertisements, to enhancing user experience and predict new trends and their effect on *informational privacy*. This section will also outline some of the tools and practices that can be employed to protect an individual's *informational privacy* and personal information in an online setting, as well as highlight concerns that has arisen in

The third section will analyse Facebook's collection and use of personal information in relation to the impact on the individual's *informational privacy* by using Floridi's theory of *informational privacy* (2005) to identify factors effecting the ontological friction in the Facebook region of the infosphere. As well as highlight cases where Facebook's approach to the individual's *informational privacy* can be viewed as concerning.

The fourth section will analyse Google's collection and use of personal information in relation to the impact on the individual's *informational privacy* by using Floridi's theory of *informational privacy* (2005) to identify factors effecting the ontological friction in the Google region of the infosphere. As well as highlight cases where Google's approach to the individual's *informational privacy* can be viewed as concerning.

The last section will analysis how the new General Data Protection Regulation of the EU benefits the citizens in the EU in relation to protecting their *informational privacy* and personal information in an online setting. Furthermore, the section will explore the compatibility of the GDPR and Floridi's theory of *informational privacy* (2005).

Method

When defining *informational privacy* as a concept, the concept of privacy must first be defined, in this thesis I will explore different approaches to privacy and discuss the ways that they differ from each other and the reason. Additional I will also explore aspects that these different approaches have in common when defining the concept of privacy. The works I will explore is the works of Warren & Brandeis (1890), Thomson (1975), Parent (1983) and Tavani (2009). This is done to gain a better perspective on the various difficulties that arises when defining privacy as a concept and the variety in the approaches to privacy.

In defining the concept of *informational privacy*, I will use the theory of Floridi (2005) as presented in his research paper, *The Ontological Interpretation of Informational Privacy*. Floridi's theory is chosen due to his approach to *informational privacy*, were the focus is on personal information and how the collection of personal information impacts the individual's level of *informational privacy* in an online setting. Furthermore, Floridi (2005) also discusses how information and communication technologies (ICTs) re-ontologized the infosphere and changed the availability of personal information in an online setting. This theory will be used to analyse how various methods of collecting and using personal information in an online setting impacts the level of the individual's *informational privacy* and which factors can help to either erode or enhance the level of *informational privacy*. As well as how various forms of online privacy enhancing tools (PET) can impact the accessibility of personal information and the level of *informational privacy* for the individual. Floridi's (2005) perspective on *informational privacy* is unique due to his belief that information is a constituted part of our being, personal information is not related to our person but a part of our person and identity. His research concerns are primarily Information and Computer Ethics and the Philosophy of Information, as well as the socio-ethical value and implications of digital technologies and their applications in relation to *informational privacy*[5].

In the section on the General Data Protection Regulation (GDPR) of the EU, I will provide an overview and identity some key points in the GDPR and the impact of the regulation for the citizens in the EU in relation to the collection and use of personal information in an online setting. Furthermore, I will explore the GDPR and Floridi's theory of *informational privacy* (2005), the focus will be on the compatibility of the GDPR and Floridi's theory and the challenges that arises by applying Floridi's theory to the GDPR. This exploration will be based on *The EU general data protection regulation (GDPR), a practical guide* by Voigt & Bussche (2017) and Regulation (EU) 2016/679 of The European Parliament and The Council, also known as the GDPR.

In the analysis of Facebook and Google and their collection and use of personal information, the theory of Floridi (2005) will be used to analyse the level of *informational privacy* provided by these two companies to their users. The analysis will focus on identifying the amount and

[5] https://www.oii.ox.ac.uk/people/luciano-floridi/

accessibility of personal information available in the regions of the infosphere, respectively in the Facebook region and the Google region. Furthermore, the theory will be used to determine how various factors such as methods and practices used in the collection and use of personal information impacts the information flow in the infosphere and the ontological friction in the regions of the infosphere. The analysis of Facebook will be based on Facebook's *Data Policy* which was last revised on April 19 of 2018 and as such also includes changes related to the GDPR recently implemented in the EU. Furthermore, the analysis will as draw upon Facebook's current *Terms of Service* and the privacy settings enables by Facebook on their user's personal profile. Examples of privacy settings will be provided from my personal Facebook account and as such some information might be censored to protect my own privacy. The analysis of Google will be based on their latest *Privacy Policy*, that came into effect on May 25 of 2018 and as such includes changes related to the General Data Protection Regulation (GDPR) recently implemented in the EU. I will also draw upon Google's *Terms of Service* which was last modified on October 25, 2017 and the privacy settings enabled by Google's if the users have a Google Account. Examples of privacy settings will be provided from my personal Google account and as such some information might be censored to protect my own privacy. Additional I will also draw upon research paper, newspaper articles and other works that highlights general concerns and previous issues relation to Google and Facebook and their handling of collecting and using personal information.

Defining Informational Privacy

In this section of the thesis I will present the concept of *informational privacy*. In order to define *informational privacy*, the general concept of *privacy* must first be discussed. In the first part of this section I will highlight different approaches to defining *privacy*. *Privacy* is a concept that is still being debated to this day. The approaches to defining *privacy* varies greatly, from viewing *privacy* as a single concept to viewing *privacy* as a concept, divided into several kinds of *privacy*. In this part I will draw upon the works of Warren & Brandeis (1890), Thomson (1975), Parent (1983) and Tavani (2009). These four texts will give an overview of how different the approaches to *privacy* can be, as well as which approach I will take in this thesis. In the second part of this section the theory of *The Ontological Interpretation of Informational Privacy* by Floridi (2005 & 2006) will be presented and how it can be using in analysing *informational privacy* in an online setting.

The concept of Privacy

The article *The Right to Privacy* (1890) by Warren & Brandeis, is often being highlighted as the article that started the discussion of how to define *privacy*, and their article is still being referenced today. This article also shows that *privacy* is not a new concept, it has been discussed for more than a century. Warren and Brandeis (1890) highlight the development of photography and the rise of the newspaper, as being a threat to peoples private and domestic lives. This development has made it easier to invade people's *privacy* and share personal information without a person's consent. In the article Warren and Brandeis (1890) seek to examine if one's right to *privacy* can be covered and protected by the common-law existing at the time. By examining cases regarding a person's ability to prevent publication, of sentiments and emotions expressed through writing or artworks, which was won, Warren and Brandeis (1890) found that the decisions were based on the protection of property. Based on their research Warren and Brandeis concluded the distinguishing attribute of property is the concept that property can be owned. And that writings and artwork expressing sentiments and emotions could be classified as being owned by the creator or another person. Which means that writings and artworks expressing sentiments and emotions would be protected under the common-law of protection of property. Upon closer examination they found that the reasoning was based on the principle of the right to be let alone. Warren and Brandeis' (1890) conclusion was that under the protection of property law, there existed a principle,

the right to be let alone, which could be invoked to protect the *privacy* of a person. Warren and Brandeis were mostly focused on that a person right to *privacy* would be violated, if personal information about him was to be published, and therefore his right to be let alone was being violated. They also noted that this include writing and artworks such as musical compositions. In other word people have a right to *privacy* and this right is protected under the principle of being let alone. Warren and Brandeis (1890) sees information about oneself in which ever form, as being owned or possessed by oneself, and therefore the protection of such information would be protected under property law.

The next approach to *privacy* is by Judith J. Thomson, in her article *The Right to Privacy* (1975), she outlines why she believes that the concept of *privacy* does not need to be defined. Thomson is focused on *privacy* as a right, a right to control what happens to our property or a right to control what happens to our person. Thomson (1975) uses the example of an opera singer, that decides she no longer wish to be listened to and does everything in her power to achieve this. In other words, she exercises her right to control who can hear her sing. If another person uses a listing device to listen to the singer, then her right to *privacy* is being violated, as well as her right to not be listened to and her right over the person. Thomson explains that both the right to *privacy* and the right to not be listened to, exists within the right over the person. Thomson (1975) also uses the example of a man owing a pornographic picture, but the man does not want other people to see the picture, therefore he keeps it locked away in a wall safe. In other words, he exercises his right to control who can look at his property, as well as what happens to his property. Thomson (1975) then argues that if another person were to train an x-ray device on the owner's house and being able to see the photograph, then this is a violation of the man's privacy. Furthermore, the man's right to control who sees the picture as well as his property rights is being violated. Thomson (1975) used this reasoning to argue that a person's right to privacy is protected under rights such as the right of property and the right over the person. Therefore, Thomson (1975) believed that discussing the concept of privacy as redundant, because an individual's privacy rights can be explained and protected by already existing rights, as mentioned above. While Thomson's (1975) and Warren & Brandies' (1890) approaches to privacy might seem similar because both approaches believed that privacy can be protected under existing laws, there is one issue that separate the two. That is their approach to privacy as a concept, Thomson (1975) did not

viewed privacy as a concept, whereas Warren & Brandeis (1890) viewed privacy as the concept of being left alone.

A very different approach to privacy is taken by W. A. Parent in his article *Privacy, Morality and the Law* (1983). Parent (1983) distinguish between the condition of *privacy* and the right to *privacy*. The condition of privacy is to not have undocumented personal knowledge about oneself, possessed by others. The more undocumented personal knowledge about you is possessed by others, the less *privacy* you have. Parent define personal knowledge as facts that you do not want to be widely known and may be concerned if this knowledge is passed beyond a limited circle, such as close friends and family. Parent further define undocumented personal information, as also including fact about oneself, that one finds sensitive and choose not to share with others. Even though most people don't care if the same information is shared about themselves.

Parent (1983) also note that documented personal information is not included in the condition of *privacy*. Parent define documented personal information as recorded public information about one, excluding such records as medical records, on the basis that these are not recorded for public consumption. When Parent (1983) talks about our right to *privacy*, it is from the perspective that we as people place a great value in *privacy*. We do not wish others to have undocumented personal information about us, which can be used to harm us. Parent also note that in a society where the disclosure of information such as sexual orientation or political viewpoints can cause ostracizing from society, our desire for *privacy* is understandable. Parent (1983) further underline that there is also certain information about us, that we believe people, such as strangers or acquaintances are not entitled to know about us. And if undocumented personal knowledge is made public, our condition of *privacy* is breached, and this is seen as a violation of our right to *privacy*.

Parent (1983) also makes a distinction between rightful and wrongful invasion of our *privacy.* A rightful invasion might be a criminal's phone calls being listened upon, to prevent a crime. While a wrongful invasion would be a neighbour bugging your home and listening to your private conversations with your spouse. Parent highlight that the right to *privacy* is not to condemn invasions of *privacy*, but to condemn wrongful invasions of *privacy*. Parent (1983) define *privacy* as only relating to information about oneself, for Parent there is no other kind of *privacy* than *informational privacy*.

The last example of an approach to *privacy* is the approach taken by Tavani (2009) in his article *Informational Privacy: Concepts, Theories and Controversies*. Whereas Warren & Brandeis (1890), Thomson (1975) and Parent (1983) view *privacy* as a single concept, Tavani (2009) divide *privacy* in to four distinct kinds of *privacy*; *Physical, decisional, psychological* and *informational privacy*.

- *Physical privacy* is the notion of *privacy* as physical non-intrusion. It is the freedom from unwarranted intrusion through physical access to a person or through a person's physical possessions. This is very similar to Warren & Brandeis (1890) notion of *privacy* as the right to be let alone.
- *Decisional privacy* is the notion of *privacy* as non-interference involving one's personal choices. It is the freedom from interferences affecting choices such as education, health care, marriage and religion. The focus here is on intrusions which can affect the ability to make decisions, intrusions such as external interference or coercion.
- *Psychological privacy* is the notion of *privacy* as non-intrusion/non-interference involving thoughts and personal identity. It is the freedom from intrusions and interferences, that can affect one's thoughts or personal identity in a negative way. Such as stalking or engaging in brainwashing, as seen in some religious cults.
- *Informational privacy* is the notion of having control over/limiting access to personal information. Personal information is defined as data that can contain information about activities, lifestyle, finances, medical data and academic achievement. Concerns about *privacy* here are often about these forms of information, being stored and communicated between electronic databases. As well as personal information communicated for example using e-mail, text messages and wireless communication devices. Here *privacy* is affected by the amount of personal information collected, how and to whom such information is transferred and amount of time such information is stored.

In other words, when defining *privacy*, Tavani (2009) believe that one must look at what kind of *privacy* is being discussed, that *privacy* cannot be condensed into a simple and all-encompassing concept. Tavani (2009) also highlights that he believes that *privacy* is not static, but an evolving concept being influenced by politics and technology developments.

I have now presented four different approached to *privacy*, the right to be let alone by Warren and Brandeis (1890), *privacy* as a right that can be protected and defined through laws of property and bodily security by Thomson (1975). *Privacy* as a concept only relating to personal information about a person by Parent (1983), as well as *privacy* as being four different kinds of *privacy* by Tavani (2009). These four approaches to *privacy* show that defining *privacy* is not simple and can be a complicated task. However, when looking at these four approaches, there are some characteristics that all four approaches share.

With Warren and Brandeis (1890) *privacy* is being violated when personal information about a person is publicized without consent. By withholding consent, a person chooses which information to be publicized and which to be remain private. By doing this he is excising a control over his *privacy*. While Thomson (1975) does not believe in *privacy* as a concept, she still speaks about limiting access or controlling who has access to one's property or right over the person. As in Thomson example with the opera singer, there singer chooses that she does not wish to be heard when singing or only chooses to allow a limited number of people to listen to her. Parent (1983) also speaks about limiting access to one's personal information, to protect one's privacy. By choosing only to tell a small circle of close friends and family, for example that you are gay, you are choosing to limit who has access to undocumented personal knowledge about you. Tavani (2009) when speaking about *informational privacy*, defines it as have control over who has access and how much access a person has to one's personal information. No matter which approach to *privacy* is taken, there is a common trait that shines through, the right to choose what happens with one's *privacy* and exercise some amount of control over it.

When looking at the approaches of Warren & Brandeis (1890), Parent (1983) and Tavani (2009), *informational privacy* must be discussed when speaking about protecting one's *privacy*. I agree with Tavani (2009) that *privacy* cannot be compromised into a single concept, *privacy* is a to broad concept to be simplified in such a manner. Furthermore, I agree with the notion that when speaking about protecting one's person information, whether in an online-setting or not, we are speaking about *informational privacy*. This is the approach I will take to the concept of *privacy* in this thesis.

Floridi and Informational Privacy

In the previous part the concept of *informational privacy* was touched upon and defined as relating to the protecting of personal information both online and offline. In this part the concept of *informational privacy* will be broaden using Floridi's theory of *informational privacy* and be used to analyse the individual's *informational privacy* in this thesis.

Luciano Floridi approaches *informational privacy* from an Information Ethics (IE) viewpoint. Information Ethics focus on the relationship between creation, organization, distribution and use of information, furthermore the ethical standards and moral codes influencing the individual's behaviour in our society[6]. Floridi (1999) distinguish between four different kinds of privacy like Tavani (2009), *physical privacy*, *mental privacy (psychological privacy), decisional privacy* and *informational privacy*, however Floridi's focus is exclusively on *informational privacy*. While Floridi's (2005) definition of *informational privacy* is very similar to Tavani (2009), that *informational privacy* is affected by the amount of personal information collected, how and to whom such information is transferred and amount of time such information is stored. Floridi's approach to *informational privacy* is very different.

Floridi (2005) views information about an individual, not as related to the individual but as a part of the individual's Being, simply put *"You are your information"* (Floridi, 2005, p. 195). The individual's identity is constituted by personal information about the individual, which is why Floridi (2005) sees the individual's identity and the individual's informational sphere as two sides of the same coin, therefore a violation or breach of the individual's *informational privacy* is seen as a form of aggression towards the individual's personal identity. This perspective is why Floridi (2005) views all information connected to the individual as personal information in his theory and the reason Floridi views *informational privacy* as being extremely valuable and needs to be protected. From an online perspective I believe that Floridi's approach to personal information is the right approach. This is due to while the information collected in an online setting might not depict a full representation of an individual, but it can reveal some sensitive and personal aspects and core characteristics of the individual's identity and personality. Furthermore, seemingly inconsequential information can be used together with other information in identifying an individual, this will

[6] https://www.abc-clio.com/ODLIS/odlis_i.aspx

be elaborated on in the section *Informational Privacy and the Internet*. Therefore, in stands to reason that every piece of information gathered on an individual in an online setting can be classified as personal information.

In the article *The Ontological Interpretation of Informational Privacy* (2005) Floridi describes the individual as an information agent that inhabits the infosphere. The infosphere is a phrase Floridi (2005) uses to describe the world we inhabit, the infosphere is based on the idea of a biosphere. Biosphere referrers to a limited region of our planet able to support life. Floridi's infosphere referrers to the whole information environment comprised of all informational entities including information agents, the informational entities properties, relations with each other, processes and interactions. The infosphere includes both online, offline and analogue spaces of information (Floridi, 2013). Information entities are any kind of artefact that creates a trail of information, such as software programs, data bases, humans and processes, however humans are referred to as information agents (Floridi, 1999; 2005). This means that all humans are a part of the infosphere, but inhabit different regions of the infosphere, some infospheres overlap, while others do not. In an online setting the Internet is referred to as a region of the digital infosphere, the digital infosphere is a region within the whole infosphere (Floridi, 1999). In the infosphere *informational privacy* is affected the information gap between information agents. The information gap is the level of accessibility to personal information, the higher level of accessibility, a lower level of *informational privacy* is had by the individual. Whereas the low level of accessibility, results in a high level of *informational privacy* is had by the individual. The information gap between information agents are influence by the ontological friction. Ontological friction are forces that influence the flow of information and the amount of work required by information agents to obtain personal information about other information agents within the infosphere or a region hereof (Floridi, 2005). Ontological friction can be noise, lack of time, lack of memory space (ex. on a computer), the amount and complexity of the data to be processed, the lack of software or access to the information. In short, any obstacle, condition or policy that can hinder the flow of information and as such widen the information gap between the information agents (Floridi, 2006). Here a high degree of ontological friction is considered a positive thing because the higher degree of friction, the higher level of *informational privacy* can be experienced. Furthermore, a large information gap between two information agents is also considered as

a positive thing, because it means that the amount of personal information one information agent knows about the other information is very small. In short, *informational privacy* is determined by the informational gap between information agents, which is affected by the ontological friction within the infosphere and the ontological factors affecting both the infosphere and the ontological friction. Floridi (2005) formulates this as qualitative sort of equation, which can be used to analyse the relation between factors that affect the ontological friction in the infosphere and an individual's *informational privacy*;

"Given a certain amount of personal information available in (a region of) the infosphere I, the lower the ontological friction in I, the higher the accessibility of personal information about the agents embedded in I, the smaller the informational gap among them, and the lower the level of informational privacy implementable about each of them." (page 187).

Floridi (2006) describes it as "*Put simply, informational privacy is a function of the ontological friction in the infosphere. It follows that any factor affecting the latter will also affect the former.*" (page 110). These factors can vary and change depending on the environment or information agent within the infosphere, the changes can be both permanent or temporary. This means that the ontological friction is not static but dynamic and is influenced by many factors, creating an infosphere that evolves and as such influence the individual's level of *informational privacy*. Floridi (2005 & 2006) discusses some of the factors that can affect the ontological friction within the infosphere, such as technological innovation and social developments. Floridi (2005) focus on Information and Communication Technologies (ICT) as the most influential factors to affect the ontological friction and how ICTs can both enforce and erode *informational privacy*, by affecting the ontological friction. Digital ICTs can be used to protect personal information with encryption, anonymization, password encoding, firewalls, protocols and services such as Privacy Enhancing Technologies (PET). Digital ICTs can help informational agents to control and regulate the use of their data, by facilitating identification of the users involved. And as such help regulate access to their data by these users. Floridi (2005) further states that since digital ICTs can both increase and decrease the individual's *informational privacy*, the developments of PETs and similar technologies must be made to allow the informational agents the ability to design, shape and maintain their identities as informational agents. Furthermore, the development of digital ICTs has in

Floridi's (2005) words re-ontologized the infosphere and created a change in the individual's approach to *informational privacy* and accessibility of personal information.

Re-ontologizing of the Infosphere

When Floridi (2005) speaks of the re-ontologization of the infosphere, he means that radical modification in the very nature (a re-ontologization) of the infosphere, this re-ontologization is brought on by the development of digital ICTs. In the re-ontologized infosphere the conditions of *informational privacy* have changed drastically, due to the high accessibility of personal information provided to the Internet and digital ICTs. Floridi (2007) point to two developments in ICTs that have played a large role in the re-ontologizing of the infosphere; "*the transition from analogue to digital data* and *the growing of digital spaces due to the Internet*." (page 7).

Information and communication technologies covers a wide range of devices and systems used for organising, retrieval, storing and processing of information. Alongside devices and systems used for communication, such as mobile phones, computer and VoIP services. This also include software programs and computer networks, such as email services and the Internet. ICTs cover both analogue and digital technologies. (Butterfield, 2016; Floridi, 2007; Haddon, 2004; Shrum et al., 2007).

The development of digital ICTs has changed the infosphere from a setting where personal information was once hard to obtain to a setting where personal information can be collected with ease and is often given freely in exchange for the use of certain services (Moor, 1997). Furthermore, the ease of access to information has be heighten, information once difficult to obtain, can now be found in a couple of seconds, with the help of an online search engine. For example, finding a telephone number for a person living across the country, before the Internet one would have to call a specific number. When called an operator would be given the known information on the person, one was seeking the telephone number of and the operator would help find the correct number. Now there are numerous online services, where name and location can be entered, and a list of possible matches are returned (Moor, 1997). The infosphere is becoming more and more digital and the infosphere now is vast and infinite (Floridi, 2007), as an example, Floridi (2007) uses the Large Hadron Collider (which was being built at the time Floridi used the example), which was expected to produce around 10

petabytes of data annually. According to the CERN website, the data centre now processes around 1 petabyte of data daily[7]. That the equivalent of around 500 billion standard printed text pages [8]. This example shows that the transition from analogue data to digital data, the increase in storage space on hard disks and the reduction in the cost of digital ICTs has resulted in an explosion of data being collected (Floridi, 2007; Mansell et al., 2007). While the Large Hadron Collider might by an extreme example, it does shows that there is very large amount of data being produced and processed every day and as such the infosphere continues to grow with no end in sight.

Digital ICTs has affected every industry and every service in our society, as well as functions within these industries and services, such as design, production, marketing and distribution (Freeman, 2007). The Internet changed the retail industry by making e-commerce available, consumers can now do most of their shopping online and have access to a much larger range of product, due to the possibility of ordering product from other countries via web shops. Other areas the Internet has facilitated are online banking, gambling and publishing. Digital ICTs has become embedded in the society and have changed the production and consumption of information and media. New forms of media have emerged because of the Internet, such as blogs (text-based blogs), vlogs (video-based blogs), podcasts (episodic audio or video-blogs). Internet radio and tv such as streaming services like Netflix and Hulu, where movies and tv-series are streamed through the website. Real time news sites, online newspapers where news are published as they happen, either as text-based report or as videos. And web portals where information from various sources are presented in a uniform way, such a Google (Introna, 2007). Furthermore, digital ICTs have also changed the way individuals interact with each other, from speaking on the telephone, writing paper-based letters and talking face-to-face to a multitude of different communication platforms. Email services makes it possible to communicate faster than sending a letter through the postal service, where a letter can take several days of reaching the recipient, whereas an email can only take a few of minutes from being sent until it reaches the recipient. Online forums make it possible to connect with people from around the world with the same interests, instant messenger services allow people to communication over the Internet, webcams allow people to

[7] https://home.cern/about/computing
[8] http://whatsabyte.com/

communicate via video chat using applications such as Skype. As well as VoIP (Voice-over Internet Protocol) communication, this allow people to make voice calls over the Internet. Cloud storage, where files can be shared with other regardless of their location, such a Google Drive and Dropbox. Furthermore, digital ICTs have also changed the video gaming industry, from single player or multiplayer to MMO (Massive multiplayer online) games with millions of players, an example is the MMO game World of Warcraft (WoW), which had around 5.5 million subscribers in 2015[9] (Introna, 2007). The development of digital ICTs has created new ways of connecting the offline world with the online world by connecting inert objects to the internet, making it possible to control these objects with an app on one's smartphone (Bunz & Meikle, 2018). An example is the Philips Hue lightbulbs, a wireless lighting system connected to the home's Wi-Fi, making it possible to control the colour and brightness of the lightbulbs from an app or by voice commands using a Google Home or Amazon Alexa[10]. Or fitness wristbands that record every detail about heart rate, movement, sleep cycle and activity level and making this information available on one's smartphone, just to name a few. This network of objects connected to the internet is call the Internet of Things (IoT), the objects are embedded with sensors, software and hardware enabling them to connect and exchange the information collected with servers, applications and computers. This development has drastically changed the accessibility of personal information both for the individual and other parties interested in such information (Bunz & Meikle, 2018). This means that the IoT has made it possible to be under constant surveillance using one's electronic devices and this surveillance can reveal some very surprising and personal information. An example is the case of David Trinidad, his wife's Fitbit was showing some unusual readings, thinking the sensors in the Fitbit was broken he consulted a forum on Reddit dedicated to Fitbit users. One user suggested that the sensors was working fine and that another explanation could be that his wife was pregnant, which a doctor's appointment confirmed (Bunz & Meikle, 2018).

The practices of sharing personal information in an online setting has become the norm with the development of various social media sites, Instagram, Facebook and Twitter are all used

[9] https://www.statista.com/statistics/276601/number-of-world-of-warcraft-subscribers-by-quarter/
[10] https://www2.meethue.com/en-us/products.lamps#filters=LAMPS_SU&sliders=&support=&price=&priceBoxes=&page=&layout=12.subcategory.p-grid-icon

to share information, pictures, locations and activities with others. As sharing everything online has become the norm, people are sharing things and information without concern over whom might see it (Lee, 2013). As such there has been several cases where people have lost jobs, spots on athletic programs or as in the case of youtuber DaddyOFive, a sentence of five years' probation for misdemeanour child neglect. When videos of the couple behind the DaddyOFive channel pranking two of their children started circulating Facebook and Reddit and many users viewed their videos as them being abusive against their children. Leading to the couple being investigated by social services[11]. These different examples show that ICTs are a broad term, spanning over many different types of technologies, furthermore new ICTs are still being develop and new technologies and services are made possible due to new ICTs (Haddon, 2004). Furthermore, they also show how much the world has changed since the beginning of the Internet, the way people communicate, interact and the publication and sharing of personal information using blogs and social media.

Digital ICTs are re-ontologizing the infosphere because they are engineering new environments, that the individual can enter though gateways, gateways such as the Internet. And blurring the line between being offline and online, what Floridi (2007) calls *here* (offline and analogue) and *there* (online and digital), that the digital is merging with the analogue, such as the Internet of Things (IoT). Smartwatches are an example of how a device has be changed fundamentally by being connected to a network and equipped with sensors. Watches use to just tell the time, now smartwatches are linked to smartphones, making the wearer aware of incoming calls, text or emails, it records health information such as steps taken that day, time spent inactive and heartrate to name a few (Bunz & Meikle, 2018). When wearing a smartwatch, the digital is no longer *there*, it has become on part of *here*, the digital is no longer confined to be access through a computer (Floridi, 2007). As Floridi (2007) notes this re-ontologizing of the infosphere and the increasing digitalising of objects and their influence on our social environment, understanding life predigital is becoming increasingly harder to understand. Floridi (2007) argues that due to the convergence of digital resources and digital tools, the infosphere is becoming increasingly digital and in an online setting this convergence is gradually making the infosphere free of ontological friction. This means that the accessibility of personal information keeps increasing and as a result impacts the

[11] https://people.com/crime/daddyofive-youtubers-sentenced-child-neglect-prank-videos/

individual's level of *informational privacy*. However digital ICTs can both erode and enforce the individual's *informational privacy*, where cookies, browser fingerprinting and tracking can erode an individual's *informational privacy*, Privacy Enhancing Technologies (PET) which can help protected the individual from having personal information collected in an online setting (Floridi, 2005;2007; Tavani & Moor, 2001). In the early days of the Internet, users believed that being online meant being anonymous, but with the development of technologies such as tracking software made possible by digital ICTs. Being online is no longer seen as being equal to being anonymous, but as one of the least private things. Personal information is being collected every time the individual is online, sometimes without being known to the individual (Floridi, 2005). The various tools and practices in collecting and using personal information in an online setting and tools and practices to making collecting personal information will be explored in the subsequent section.

Information Privacy and The Internet

This section will explore the collection and use of personal information in an online setting and the impact the methods used has on the individual's *informational privacy*. The analysis will be based on Floridi's (2005) theory on *informational privacy* in the infosphere. The focus will be on the use of cookies and browser fingerprinting as methods for collecting personal information and data and social media mining as methods for use of the personal information. In this section the General Data Protection Regulation (GDPR) will not be considered, because it is limited to the EU, the impact of the GDPR and *informational privacy* will be covered in a separate section of the thesis. The approach in this section will be from a general viewpoint on the collection and use of personal information, additional there will be an exploration on data brokers and their role in the collection and use of personal information.

In the previous section it was explored how digital ICTs has enabled the collection of personal information in an online setting, how the individual's approach to sharing personal information has changes to a somewhat careless and oversharing approach. As a result, the tracking and profiling of individuals on the Internet has become a big business, especially in relation to online advertising. The more personal information to be had on customers, the better advertisements can be developed and used to raise companies' profit. Therefore, more and more companies are interested in knowing as much as possible about their customers, from eating habits, reading habits, websites frequently visited to the number of children, relationship status and age range of their customers to name a few. The result is the development of more persistent and subtler ways of collecting information in an online setting and analysing the personal information in ways that can help predict future trends, buying patterns and discovery new characteristics in a group of individuals (Schneier, 2015).

Data Brokers

Data brokers plays a large role in the collection and use of personal information, especially in the US, whereas in the European data brokers plays a much smaller role. This is due to the individual countries laws and the availability of information and that data brokers tends to be national. However, some major US data brokers have begun to expand into Europe, such as Axciom and Experian (Rieke et al., 2016).

The companies that track internet users and use data mining can be divided into two groups: The first group are companies that tracks users as a by-product of their primary business, these companies gather personal information about purchase history, browsing habits and web pages visited. The information is gathered to be used for recommender services, discounts and to increase sales with target advertisement, as examples. The second group are companies that track users as their main line of business, these companies are often referred to as data brokers (Etzioni, 2012; Martin, 2016). Data brokers information comes from three different sources (Anthes, 2015);

- Public records. Birth records, property records, voter registrations and court records
- Publicly available information. Information gathered from telephone directories, business directories, classified advertisements and social media.
- Non-public information. Information collected from website registrations, contests, surveys and questionnaires and the use of cookies and browser fingerprinting such as IP address, search terms, operating system, purchase histories and web pages visited.

Data brokers collects information from a variety of sources and aggregate the information to create detailed profiles on individuals, in doing so the data brokers can sell the aggregated information to companies, that wishes to target a specific market or group of consumers. This type of aggregated information can be of great value to both the company buying the information and the data brokers by creating a better level of analysis based on as large set of data as possible. Data brokers sell personal information about both individual as well as groups of individuals, an individual's profile is a collection of different characteristics and information about the individual, whereas a group profile is a collection of different characteristics and information about a specific group of individuals (Custer et al., 2013; Martin, 2016). The detailed profiles created by data brokers on individuals are not anonymous/unidentifiable information, it is the individual's name, address, age, income, health conditions, type of car they drive, number of kids including their age and gender, to name just a few. The form of information the data broker might have on the individual are virtually endless, whereas the group profiles are more general in the form of information, for example individuals that live a zip code A buys more organic product that individuals that live in zip code B (Anthes, 2015; Custer et al., 2013).

A concern regarding data brokers, is the massive amount of personal information the data brokers collect on individuals and from various sources, because of this the ontological friction is very low, and the data brokers has a high accessibility to personal information in the infosphere, causing the information gap to be very narrow and a high loss of *informational privacy*. This issue of the individual's *informational privacy* in relation to data brokers has been discussed by several authors; Sumner (2016), Broder (2000), Etzioni (1999), Schneier (2015) and Zarsky (2002), to name a few. While these different authors look at data brokers and informational privacy from different points of view, the general opinion is that the individual has little to no *informational privacy* in this context and that the individual's personal information is being sold to whomever is willing to pay. Leaving the individual with little to no control over their personal information and the context in which it is being used and companies are willing to pay, which can be seen with the 1 billion-dollar annual revenues of Acxiom, one of the largest data brokers worldwide (Sunmer, 2016).

Internet Cookies and Browser Fingerprinting

HTTP cookies, Internet cookie, browser cookie or just cookie they are all the same thing. The HTTP (hypertext transfer protocol) cookie is a cookie that is embedded in the HTML (Hypertext Mark-up Language) file of the web page. When a web browser makes a request to load a web page, a HTML file is received from the domains server, the user is connected to the server via HTTP. Because the user only connects to one server, these requests are called first-party requests. However, the HTLM page can contain additional content known as elements such as JavaScript code, video, audio or pictures, these elements can come from a limitless number of additional parties. These elements are added from an external server and are called third-party elements and as such the downloading of these elements are called third-party requests. This means that third-party elements must be included in the source code of the webpage by the first-party. These third-party elements are capable storing cookies on the user's computer, used for tracking the user across websites. As well as facilitating an advanced form of tracking known as browser fingerprinting (Liebert, 2015a; Peng & Jennifer Cisna, 2000; Puglisi et al., 2017). The flash cookie is yet another type of internet cookie, flash cookies can be embedded in the Adobe Flash Player and has no expiration date and it is stored in multiple locations at the same time. Blocking the flash

cookie can result in not being able to watch content that needs Flash player to run (Bujlow et al., 2017; Sipior et al., 2011).

The use of placing a cookie on a computer, was originally to not having websites constantly asking users for their usernames and passwords. To monitor users most frequent visited websites to learn about their interests. And finally, to use the information collected for personalized offers and advertising (Bayan, 2001; Nikiforakis & Acar, 2014; Miyazaki, 2008). Internet cookies can be divided in to two different types; Transient and persistent. Transient cookies, are cookies that are temporary, meaning that when the browser is closed, the cookie is deleted of the memory of the computer. Persistent cookies are cookies that are "permanent" stored on the computer's hard drive. There are "permanent" in the meaning that, they are not deleted when the browser is closed. This type of cookie is programmed to last from day to months to years (Hinduja, 2004; Miyazaki, 2008). When the cookie is first placed on the computer, it is updated every time the user visits the website or web page again. There it is important to note, that cookies are not only used to collect information on users, but that many websites require cookies to operate effectively. For example, online shopping sites, use cookies to differentiate between their shoppers and their individual virtual shopping carts (Miyazaki, 2008). Third-party cookies are not connected to the website, the user is visiting, these third-party cookies follow the user across websites (Miyazaki, 2008). However, as the use of cookies to gather information and data, not relating to the performance of the websites became widespread, the development of countermeasures to block cookies also started to rise (Bujlow et al., 2017). The result has been more persistent cookies, known as supercookies, where the cookies is stored in multiple locations, such that if deleted one place it is still located another and continues to work. The supercookie is also much more persistent than regular internet cookies because it can rebuild after deletion and some can even reproduce in other browsers used on the computer, making them much harder to permeant delete (Bujlow et al., 2017; Castelluccia & Narayanan, 2012; Ring, 2015).

The forms of personal information collected by the different types of cookies can be divided into two groups;

1. *Personally identifiable information* (PII). This is information is typical asked for when signing up for an account on a website or a service. It can be information such as

name, phone number, email address, gender, age and location (Miyazaki, 2008; Sipior et al., 2001).

2. *Indirectly identifiable information*. The information collected here can as for example be device information, IP address, browser type, operating system and the web pages visited to name a few, which through coupling with PPI makes it possible to identify the individual (Hinduja, 2004; Pierson & Heyman, 2011).

Web browser fingerprinting is another way of tracking Internet user's and identifying them, making it possible to separate users from each other. The personal information collected using browser fingerprinting is installed fonts, which browsers being used, plugins installed in the browser, screen resolution, operating system, graphic card and network information (Upathilake et al., 2015). This type of tracking can be done without the use of cookies and can very difficult to avoid (Acar et al., 2014; Bujlow et al., 2017; Castelluccia & Narayanan, 2012). Web browser fingerprinting does not store any data on the computer as with cookies, which means that they cannot be removed and can be very difficult to avoid. Bujlow et al. (2017) also highlights that even when using incognito browsing or anonymous networks router such as Tor, web browser fingerprinting can't be avoided. The Tor network reroutes users network connection thorough a series of different networks, rather than make a direct connection to the desired network. And as such obscure user's network connection, making it harder to track users[12]. Web browser fingerprinting can be done without the user's knowledge and can be used to track users across websites (Acar et al., 2014; Boda et al., 2011; Nikiforakis et al., 2014).

Web browser fingerprinting can be used for other means than tracking users and collecting uniquely identifying information about them. It can also be used to ensure the user trying to access a website are who they claim and not an attacker trying to compromise the website. Antifraud fingerprinting can be used to protect services against the sharing of credentials, where one paid subscription is shared by several users (Nikiforakis et al., 2014). It is used by online newspapers, who offers a predetermined number of free articles to read over the course of ex. a month. When the limit is reached the user must buy a subscription, this is called a paywall (Nikiforakis et al., 2014). With the growing number of Internet users blocking

[12] https://www.torproject.org/about/overview.html.en

or deleting cookies and the laws in the EU requiring websites to disclose their use of cookies. Browser fingerprinting could easily become the new norm in behavioural advertising, where ads show product for sale relating to previous visited websites (Nikiforakis et al., 2014).

Data and Social Media Mining

All the personal information collected using the techniques described above is often used in data mining. Data mining has become a major business, the ability to analyse massive amounts of data and information, gathered from the Internet and stored in databases, has changed the way many industries does business. Industries such as marketing, advertising, insurance and finance, use data mining to create profiles on their consumers, to more effectively target consumers with new or improved products (Al-Saggaf & Islam, 2015). This means collecting information and data on consumers, has become an important part of various industries and can have an impact on their net income. Data mining is broad terms are finding patterns and correlations between large quantities of data and analysing the data in different ways, to discover new and useful information (Tavani, 1999). These large quantities of data are stored in databases also referred to data warehouses, these data warehouses are highly structured with detailed data entries, providing data mining with a strong foundation for knowledge discovery. The data stored in data warehouse are often data from several different databases, merged together into one huge database (Fulda, 2000 & Zarsky, 2000). The knowledge discovery that comes from data mining, differs from ordinary information retrieval, in that the data extracted is not explicit in the database (Fulda, 2000). The data mining process is performed by different types of algorithms and with no to very little human interference (Zarsky, 2002). Three different types of data mining can be performed, each with focus on different areas of the Internet. These are web structure mining, web content mining and web usage mining (Al-Saggaf & Islam, 2015);

- *Web structure mining*. This type of web mining focus on extracting links and structures of websites, and analyse the data extracted to discover unknown relationships between websites.
- *Web content mining*. The focus here is on extracting content such as text, pictures, audio, video and hyperlinks. Analyse the extracted data to determine the relevance of the context in relation to search queries.

- *Web usage mining*. The focus here is tracking users and analysing the data collected from the tracking cookies, to analyse navigational behaviours of the users. What type of websites do the user visits, how do they navigate from one website to another website and what are the users searching for.

In web structed mining and web content mining the focus is solely on information and data relating to the websites, their structure and content and relationships with other websites. As opposed to web usage mining, where personal information and data of Internet users is collected, which is the type of data mining focused on in this thesis. The information and data gathered by using web usage mining can then be used for knowledge discovery, this is done by employing various methods and practices, each of them specifically suited to a specific task in relation to knowledge discovery (Zarsky, 2002). When analysing the personal information collected in web usage mining, there are three analysis methods which are commonly used. These are *cluster analysis*, *market basket analysis* and *sequential pattern discovery* (Zarsky, 2002);

- *Cluster analysis*. This type of analysis is used to discover unknown correlations between clusters of data. By mapping items into categories, during the discovery process, based on similarities the items share. Meaning the categories are not pre-defined but determine by the patterns found in the data. The data used can be anything from personal information relating to interests, age, gender, location and the likes. Furthermore, it can be data relating to activities on the Internet, the types of websites visited, the amount of time, spend on each web page and types of social media used.
- *Market basket analysis*. Also known as association discovery, which is often in relation to recommender systems. The analysis searches for items bought together or never brought together and uses this information to describe rules relating to purchases of these items.
- *Sequential pattern discovery*. This type of analysis is used to discover patterns and behaviours of long-time customers. Here the analysis focusses on developing rules about data collected about the same objective or behaviour rather than on a single transaction, this type of analysis is often used by banks and credit card companies to detect fraud.

In short data mining can used to uncover unknown information about groups of Internet users, for optimising recommender systems and for providing a level of security for users, when shopping online. Data mining has a broad use and many companies uses data mining, for improving their products, with information and data not collected from Internet users. Vestas is such a company, they use data mining on information and data collected about the weather in improving their windmills and placement of these (Hasselbalch & Tranberg, 2016). Furthermore, data mining is used in the health sector to evaluate the effectiveness of treatments, in predicting the occurrence of diseases in the individual and discover new ways to helping preventing diseases (Shafique et al., 2015). The reason of highlighting these other areas where data mining is used, to illustrate that data mining have many different applications. That not all are based on creating detailed profiles on consumers and using these details for profit.

Another method used in data mining is called social media mining, where personal information is gathered from social media profiles. Social media mining is defined as the process of representing, analysing and extracting actionable patterns from social media profiles (Zafarani et al., 2014). Data mined from social media are very different from the type of information mining in traditional data mining, data mining from social media are generally user-generated content, which are *vast*, *noisy*, *distributed*, *unstructured* and *dynamic* (Barbier & Liu, 2011; Lui, 2014). These five attributes of social media data complictes the data mining and requires a constant development of new techniques and algorithms to produce useable knowlegde discovery. Despite this social media mining has become an popular way of colleting personal information and with 2.46 billion users of social media users worldwide in 2017, the amount of personal information that can be collected are massive (eMarketer, 2017). This is due to the most prominent feature of social media, which are the user profile, a unique webpage displaying information on the user. The information displayed are typical generated from a list of questions the user is asked when first joining the social media site. The question asked are often related to personally identifiable information (PII), such as name, age, gender and location and combined with more general information relating to interests in movies, music and book. Sometimes users can also elect to disclose information on their religion or political views and add a profile picture. Examples of social media sites are Twitter, Facebook, Instagram and YouTube (Boyd & Ellison, 2008; Pierson &Hayman, 2011).

The collection of personal information for social media mining are done by uses serveral methods, such as negligence of the individual in managing privacy setting, through application programming interface (API), using an Facebook account to log into a game application and the use of cookies Al-Saggaf & Islam (2015). Social media mining has several uses, one is a better understanding of opinions about a given subject, another is identifying specific groups in a population and singling out influential people. An example is social media mining showed a correlation between the use of social media by presidential candidates and the winner of the US presidential election of 2008. Which shows that this type of data mining can potential be used in predicting outcomes in other elections (Barbier & Liu, 2011).

Cookies, Data Mining and Informational Privacy

Cookies, browser finger printing and data and social media mining are closely related, all the personal information collected by cookies and browser fingerprinting are used in knowledge discovery. And all play as role in the level of *informational privacy* in an online setting, the region of the infosphere inhabited by the individual. I will apply Floridi's (2005) qualitative equation to analyse the relation between these methods of collecting and using personal information and level of *information privacy* in the infosphere, the equation is;

> *"Given a certain amount of personal information available in (a region of) the infosphere I, the lower the ontological friction in I, the higher the accessibility of personal information about the agents embedded in I, the smaller the informational gap among them, and the lower the level of informational privacy implementable about each of them."* (Floridi, 2005, p. 185).

I have previous shown that the ontological frication is determine by multiple factors, that can either increase or decrease the ontological friction. The personal information used in data and social media mining are mainly gathered using cookies, each cookie used successfully to retrieve personal information decrease the ontological friction in the infosphere, which in turn heightens the accessibility of personal information about the individual within the infosphere. The result is the information gap between the individual and the information agent (for example Google or a data broker) is narrowed, meaning that the information agent has access to more personal information about the individual than before. Which causes the individual's level of *informational privacy* to decrease, leaving the individual's personal information more exposed to be shared with other information agents. As an examples

health information has always been considered as private and protected by law, but with the age of the Internet the search for information regarding health conditions has moved from the doctor's office to WebMD.com, where such information is not protected by law. This means that when searching for information on health conditions there is a high probability that the individual's search query and visits to webpages relating to the condition in question is being collected (Etzioni, 2012; Liebert, 2015b). As an example, Liebert (2015b) use the CDC's (Centers for Disease Control and Prevention) web page on HIV. On the web page a Google Analytic cookie is present, some of the information the cookie collects are device information, location information, type of browser and operating system and the URL of the web page visited. This means that Google can identify users with an intertest in HIV. But Google are not the only ones collecting information relating to health conditions, on Dictionary.com a search word like "depression" results in the installation of 223 tracking cookies, these tracking cookies are used in personalised targeted advertising for antidepressants on other websites (Etzioni, 2012).

Applying Floridi's (2005) theory to the above example, the cookies and the practice of browser fingerprinting used on the website are all factors that effects the ontological friction in the infosphere, in this case the cookies lower the ontological friction and increases the accessibility to personal information. The information gap within the infosphere is narrowed, resulting in a decrease in the individual's *informational privacy*. Each cookie encounter on the Internet, is another factor that effects the ontological friction and aids to increase the accessibility to the individual's personal information. And as such each cookie encountered on the Internet lowers the *informational privacy* of the individual. Resulting in the information gap becoming narrower and narrower, as an example an information agent such as Google tracks individual's on almost every webpage visited and in theory this means that the information gap between the individual and Google is virtual non-existing. Google knows everything about you, especially if you have a Google account and regular uses their services and products. The use of various types of cookies and browser fingerprinting makes it harder and harder to protect personal information in an online setting. The individual, if no countermeasures are taken, has a very low level of *informational privacy* in the region of the infosphere, that is the Internet. This is because when used to track users across websites, a wide variety of personal information can be collected and a very detailed profiled of the user

can be created. Since some cookies also gives the user a unique ID, the third-party collecting the information may not know the individual's full name and address, they may however know the user's sexual orientation, general location, age-range, health information and job title just to name a few. Coupled with information gathered from browser fingerprinting it is possible to identify the specific computer used by the individual and as such identifying which individual is sitting in front of the computer screen, if multiple computers are connected to the Internet from the same network (Acar et al., 2014; Bujlow et al., 2017; Castelluccia & Narayanan, 2012). Tracking individuals across the Internet and collecting information on them are often compared to the practice of surveillance (Bayan, 2001; Hinduja, 2004). Making the information gap very narrow in some cases, depending on the number of cookies and whether browser fingerprinting is used, and the forms of personal information gathered. This is also an example of how dynamic the level of *informational privacy* can be, in some instances the individual might enjoy a high level of *informational privacy* with an information agent, because the information agent has a low number of cookies and does not use browser fingerprinting. Meaning that that specific information agent has a limited access to personal information, whereas an information agent such as Google, as mentioned earlier has a high accessibility to personal information. The level of *informational privacy* is always changing in context to the methods used in collecting and accessing personal information.

Data and social media mining are another factor which effects the ontological friction and as such further impacts the individual's *informational privacy*. Where cookies and other methods of collecting personal information is focused on collecting as much personal information as possible. Data mining is focused on using the personal information to analysis the behaviour, preferences and characteristic of individuals and using the knowledge discovered for profit (Zarsky, 2002). The perhaps most famous example is the case Target and teen pregnancy[13], where a teenager is sent coupons for baby related products and her father accuses Target for encouraging his daughter to have sex and become pregnant. When the manager called to apologise, the father ended up apologising to the manager, as it turns out his daughter was indeed pregnant. The reason that this could happen was due to a unique ID Target assigns to each customer and a prediction model based on results from data mining. The data mining had identified around 25 different products typically used during a pregnancy

[13] https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

and when each product was typically brought during each semester of the pregnancy. This allowed Target to develop a pregnancy predication score, which when applied to the customers unique ID, could predict if a customer was pregnant and estimate how far long in the pregnancy the customer was (Sumner, 2016). This example shows how data mining can contribute to lowering the level of *informational privacy* of an individual, here the collected personal information about the teenage girl, when aggerated showed a high possibility of a pregnancy. Using Floridi (2005) data mining is a factor that lowers the ontological friction in the infosphere, making the teenage girl's personal information more accessible for Target, narrowing the information gap and as such the teenage girl's level of *informational privacy* decreased. The result was the use of data mining on her personal information revealed intimate information relating to her personal health and exposing this information to others (her father).

Besides personal identification and the revelation of health information as seen with the Target example, data mining can affect the individual's level of *information privacy* in other concerning ways such as discrimination. The use of aggregated personal information has the possibility of being used in discriminating groups of individuals that share certain characteristic. For example using personal information about an individual's health, age, location and job title collected by cookies in cluster analysis to identify groups of people both included and excluded in the data set. And using the results of the analysis to raise insurance premiums for that specific group of people (Christl & Spiekermann, 2016; Liebert, 2015b; Zarsky, 2002). The problem here is that not only does the data mining might reveal wrongful conclusions but again the data mining lowers the individual's level of *informational privacy*, giving information agents access to personal information. Personal information the individual might not have intended to share with other information agents and compromising the privacy of the individual. Floridi (2005) would classified this as a violation of the individual's *informational privacy*, due to the personal information revealed by the data mining is not personal information that the individual had given consent to be revealed, shared or collected. This also shows that data mining can lower an individual's level of *informational privacy* without the individual's knowledge. Social media mining is another factor impacting the ontological friction in the infosphere, by mining personal information from social media account, not only new trends can be discovered. An individual's religious and political

interests can be discovered and dividing individuals into groups based on content shared and used to influence their standpoint on such matters. Again mining of personal information can uncover unknown information about an individual or group of individuals, narrowing the information gap between the individual and the information agent. Resulting in a lower level of *informational privacy*, as well as a high accessibility to personal information. The Cambridge Analytica case is an example of how social media mining was used to try and manipulate American voters in the presidential election of 2016. In this case the information agent Cambridge Analytica had a high accessibility to personal information, granted to them by people who used their app, narrowing the information gap and lowering the individual *informational privacy* (Ghosh & Scott, 2018).

Cookies, browser fingerprinting, data and social media mining are all different factors present in the infosphere affecting the ontological friction in various degrees, by making personal information more accessible for information agents, resulting in lowering the individual's level of *informational privacy*. Here it is important to note that for example not all cookies affect the ontological friction and *informational privacy* in the same way, some cookies might collect more intimate personal information than other. As such they each count as an individual factor affecting the ontological friction and the information gap between information agents in the infosphere. The same can be seen with data and social media mining, depending on the method used for analysing the data, different conclusions can be drawn, some providing more intimate personal information than others as seen with the Target example. Whereas, other analysis methods might just show a connection between, for example people who brought The Lord of The Rings trilogy on Amazon also brought the Harry Potter books. The Target and Amazon examples shows how an individual can experience different levels of *informational privacy* within different regions of the infosphere. The Target example provided intimate personal information, that can affect the individual's *informational privacy* to a much higher degree than the Amazon example, where the personal information in form of reading habits affects the *informational privacy* to a lesser degree. As noted earlier the level of *informational privacy* is very dynamic and is affected by many factors relating to the personal information collected, how they are collected, the methods used in analysing the personal information collected and the outcome of the analysis. The level of

informational privacy is contingent on the context of the personal information collected and the use of the personal information.

Protection of Informational Privacy

In the previous part it was shown how the collection and use of personal information are factors that impacts the ontological friction in the infosphere and the accessibility of personal information and as such affects the individual's level of *informational privacy* in an online setting. This part will explore some of the countermeasures that can be used in increasing the level of *informational privacy* in an online setting. The focus will be on three factors, that is used correctly can help in creating a higher level of *informational privacy* for the individual, namely the use of Privacy Enhancing Technologies (PETs), privacy setting and privacy laws.

The use of privacy enhancing technologies are the first factor that can affect the ontological friction in the digital infosphere. PETs are online tools used for controlling the amount of personal information disclosed in an online setting. PETs are often browser extensions, which blocks requests from third-party website such as various types of cookies, blocking ad, pop-ups and social media buttons. As well as VPN services (Virtual Private Networks), which extends the private network and increase privacy, providing an extra layer of security for the individual (Bujlow et al., 2017). This means that PETs can heighten the ontological friction, causing a disruption in the flow of information and lowers the accessibility of personal information. The result is a wider information gap between the individual and other information agents and as such increases the *informational privacy* of the individual.

Privacy setting on social media can provide a further layer of protection, by specifying the visibility of certain information on the individual's profile, such as personal information, posts, friends list or images. Making it harder to mine information from the individual's social media profile to use for profiling. By ensuring the privacy setting are updated on social media can again heighten the ontological friction and lowering the accessibility of information. And further widening the informational gap causing an increase in the individual's *informational privacy*.

The last factor to affect the ontological friction are privacy law. It is important to note that there is a difference in how the EU and the US approach the protection of the individual's personal data. The US have no general theory for privacy and data protection, the approach

has been to deal with issues regarding privacy and data protection when problems arise. The results are different sectors have different regulations relating specifically to the sectors. However, The Federal Trade Commission (FTC) will judge if the consumer has been deceived or treated unfairly by the practices of a company, under the Federal Trade Commission Act, but the FTC focus more on consumer protection than privacy (Jørgensen & Desai, Right to Privacy Meets Online Platforms: Exploring Privacy Complaints against Facebook and Google, 2017). In the EU as of 25 May 2018 the General Data Protection Regulation (GDPR) came into effect, the key points of the regulation are to strengthens existing rights, provide new rights and give the individual an increased control over their personal data. Such as easier access to personal data, the right to promptly be informed of data breaches contain personal data and clarifying the right to be forgotten. The purpose of the GDPR is a single EU-wide law for data protection, before the GDPR each member of the EU had to implement data protection law of their own, based on a set of guidelines. Which led to inconsistencies and legal uncertainty (Questions and Answers - Data protection reform, 2015). The GDPR and its impact on the *informational privacy* of Eu citizens will be explored and analysed in greater detail in the section *The General Data Protection Regulation*. The GDPR will for the citizens in the EU, create additional ontological friction in the digital infosphere, lowering the accessibility of personal information and creating an even greater information gap, providing the individual with a higher level of *informational privacy*.

An example of how these factors effects the level of *informational privacy* can be illustrated by imagine that personal information flows as a river in the infosphere. At the mouth of the river is the individual and their personal information, at the end of the river is an information agent interested in collecting and using the personal information. Each time a PET is used to hinder the collection of personal information, a boulder is placed in the river of information, the more effective PETs that are used, the more boulders appear in the information river, slowing the flow of personal information from the individual to the information agent. Each time a privacy setting is used to block the collection of personal information, more boulders appear, further slowing the river of information. Each privacy law implemented regarding protection of personal information, more boulders appear, slowing the river of information even further. Resulting in the flow of personal information from the individual to the information agent going from a rushing river to a small stream with every little water. Each

stone represents a factor that influence the ontological friction and the accessibility of personal information in the infosphere, changing the level of *informational privacy* of the individual. More stones can be added or removed with for example the development of better PETs or tracking technologies, causing a decrease or an increase of the information flow in the river, allowing for less or more access to personal information. And effecting the level of *informational privacy*, showing again that *informational privacy* is a dynamic concept, that constantly changes in relation to factors present in the environment of the infosphere. However, having total *informational privacy* in the digital infosphere is considered as not being possible, there will always be bits and pieces of personal information about the individual floating around the infosphere. But it is possible to heightening one's *informational privacy* by taking countermeasures against tracking, collection and use of personal information (Schneier, 2015).

Facebook and Informational Privacy

This section of the thesis will explore and analysis Facebook and their approach to collecting and using personal information in an online setting. First, I will present a short overview of the history of Facebook and explore how the platform has transformed since it was first made available worldwide and to the present day. In the second part Floridi's theory of *informational privacy* (2005) will be used to analyse the level of *informational privacy* by the individual in the Facebook region of the infosphere. This is done by identifying and analysing factors in relation to the collection and use of personal information and how these factors affect the ontological friction, the accessibility of personal information and the information gap between the individual and Facebook based on the principles of Floridi's theory (2005). The analysis will mainly be based on the most recent Data Policy by Facebook that became effective on April 19 of 2018, meaning that GDPR has been considered when designing the policy. Additional, examples of privacy setting will be used to illustrate how these can affect the individual's *informational privacy*, these examples will be taken from my personal Facebook account accessed on my desktop, therefore some personal information might be censored. The section will end with highlighting examples of cases where Facebook's approach to the individual *informational privacy* can be viewed as concerning and deceptive.

The Story of Facebook

Facebook was launched in 2004 and was limited to college students in the US, it gave the students the ability to connect with each other in an online setting. By creating personal profiles with their name, hometown, relationship status and profile picture. They could "friend" other users and their friends would be displayed in the use's friend list for other users to see. Furthermore, the users could send messages to each other via the platform and write messages on each other's *walls*. The message would be displayed directly on the user's profile for all to read and comment on. Two years later Facebook became available to everybody over the age of 13 and number of active users has grown from 12 million users in 2006 to 1.45 billion users in 2018[14]. Making it one of the most popular social networks sites in the world. Since the launch of Facebook there has been many changes and additions to the website,

[14] https://newsroom.fb.com/company-info/

some of the features include instant messaging, the “like” button, a location service, which showed users friends located nearby (Jenkins, 2013).

A big part of Facebook from almost the start has been advertising, in the beginning advertisements was students selling their old textbooks and searching for roommates. Today advertisement means targeted ads from business, seeking to sell their products to a specific group of people. A type of advertisements called social ads, are ads where business can tell Facebook, when the ads is shown on Facebook and specify characteristics of the users to target. These characteristics include age, location, relationship status, interests and gender. With the wealth of information on their users, Facebook has become a valuable platform for targeted advertisement (Veer, 2011). Another feature on Facebook are third-party application (apps). Third-party application is developed using Facebook’s Developers Platform and offered to Facebook users through the Facebook app store. This is apps such as Candy Crush Saga and Farmville, which has around 60 million daily users on Facebook (Sumner & Rispoli, 2016). Facebook has evolved from a simple social network site used for connecting students with each other to over a billion daily users and a wide variety of features. The features range from instant messaging, photo-sharing, microblogging, gaming to a marketplace, where users can buy and sell items amongst themselves (Brügger, 2015).

Informational Privacy in the Facebook Region

The infosphere is the whole information environment comprised of all informational entities including information agents, the informational entities properties, relations with each other, processes and interactions. The infosphere includes both online, offline and analogue spaces of information (Floridi, 2013), within the infosphere different regions are located, these regions are inhabited by different information agents, regions within the infosphere can overlap, creating a sub-region inhabited by two or more information agents, as defined in the section *Floridi and Informational Privacy* of this thesis (Floridi, 1999). The expression Facebook region in this thesis refer to a sub-region of the infosphere inhabited by the following information agents; the individual and Facebook.

Determine the level of *informational privacy* for an individual in the Facebook region, any factors and their effect on the degree of ontological friction, as well as the availability and

accessibility of personal information must be identified. As stated by Floridi on his theory of *informational privacy* (2005);

"Given a certain amount of personal information available in (a region of) the infosphere I, the lower the ontological friction in I, the higher the accessibility of personal information about the agents embedded in I, the smaller the informational gap among them, and the lower the level of informational privacy implementable about each of them." (page 187).

The factors affecting the ontological friction can include the use of ICTs as explained in *The Re-ontologization of the Infosphere* section, factors can also include consent to having personal information collected, personal information received from other sources such as data brokers, a change in the environment of the infosphere and the information agent's behaviour with in the infosphere (Floridi, 2005). I have previous highlighted the impact using social media has on the individual degree of *informational privacy*, the publication and the accessibility of personal information and the information agent's behaviour online. This means that creating a Facebook profile is the first factor that can be identified, the creation of a Facebook profile decreases the degree of ontological friction by making personal information more available and accessible for Facebook. This is due to the personal information provided by the individual upon creation of their personal Facebook profile. As such upon creating a Facebook profile the information gap between the individual and Facebook has become smaller and a lowering in the level of *informational privacy* experienced by the individual within the Facebook region of the infosphere.

Collection of information

When collecting personal information Facebook does not distinguish between personally identifiable information and indirectly identifiable information, in Facebook's data policy the phrase information is used without any distinction. However, the data policy highlights that information on religious views, political views, sexual orientation and health is under special protection by the GDPR, but the type of protection these forms of personal information is granted is not explained (Facebook, 2018).

Facebook divided information collected about the individual in to three categories (Facebook, 2018):

- *Things that you and others do and provide*. Information and content provided by the individual, networks and connections, usage of Facebook products and services, transaction information and content, communications and information provided by others about the individual.
- *Device information*. Information about attributes, operations (ex. mouse movement) unique identifiers, Bluetooth signals, Wi-Fi access points, device settings (ex. the use of GPS) and information on mobile operator, connection quality, language, mobile phone number and devices nearby.
- *Information from partners.* Information about games played, purchases made, online and offline activities, including the forms of information mention in category 1 and 2. As well as receiving personal information not only on Facebook users but also on non-Facebook users.

The information collected in these three categories are not only detailed information on the individual such as personally identifiable information (ex. name, age and gender). But also information relating to the type of device used to access Facebook from, hardware and software, IP address, browser type, plugins and battery level, just to name a few. The type of device used, webpages viewed, ads shown, and items purchased, in addition to personal information about the individual from other information agents. The collection of personal information also includes information collected via other information agents related to the use of Facebook products and services, including when other information agents shared or comment on the individual's photos and messages sent between the user and the individual. In short, all information relating to any activity between two information agents on the Facebook platform (Facebook, 2018).

The availability of these forms of personal information is another factor effecting the ontological friction within the Facebook region that can be identified by using Floridi's (2005) theory. The availability of personal information causes a decrease in the degree of ontological friction within the region due to the high accessibility of personal information about the individual. This means that the information flow within the region from the one information agents (the individual) to another information agents (Facebook) flows without any or very few obstructions (Floridi, 2005).

Furthermore, each of the categories can also be identified as a factor that effects the ontological friction within the Facebook region of the infosphere. In the category of personal information collected by Facebook, *information and content provided by the individual*, the willingness of sharing personal information via a Facebook profile and consenting to having personal information relating to activities and connections on Facebook collected. As such this factor decreases the ontological friction in the Facebook by increasing the accessibility of the individual's personal information, in allowing Facebook access to a large amount of information about the individual. Because of the increase in accessibility the information gap between the individual and Facebook becomes narrower, resulting in a lowering of the level of *informational privacy* of the individual within the Facebook region. In the category *device information*, having access to the various devices such as mobile phones, tablets or computers used by the individual decreases the ontological friction, this is due the high availability of personal information about activities, the use of the devices, location, IP address, nearby Wi-Fi access points, browser type and operation system. As such the information gap in the Facebook region grows even narrower, decreasing the level of *informational privacy* had by the individual. In the category *information from partners*, where Facebook receives personal information about individuals from third-parties, Facebook's availability of the individual's personal information, causes a further decrease in the degree of ontological friction and an increase of accessibility. Further, reducing the information gap between Facebook and the individual, causing the individual's level of *informational privacy* to be additional lowered with the Facebook region of the infosphere.

This means that the degree of ontological friction within the Facebook region of the infosphere is very low, allowing the information flow from the individual to Facebook flows without any or very little obstruction. This can be illustrated by imaging a river ending in a lake, at the mouth of the river the individual stands and personal information is flows down the river to the lake. Where the river and lake meet, Facebook is standing and collecting every piece of personal information flowing from the individual, since there are no obstructions in the river, the personal information is not hindered in reaching the lake in any way and the flow of personal information to be very fast. Resulting in Facebook having unrestricted access to the individual's personal information, due to the speed of which personal information reaches the lake and the lack of obstructions, therefore the information gap between the

individual and Facebook is very small causing the individual to experience a very low level of *informational privacy* within the Facebook region of the infosphere.

Use of Information

The personal information collected about individuals by Facebook are used for several purposes to provide, maintain, protect and improved services and in developing new services. As well as for personalising content and ads, for analytical purposes, in communication with the individual and to promote safety, integrity and security on Facebook's products and services. Furthermore, Facebook uses personal information about the individual in research and innovation relating to topics of general social welfare, technological advancement, public interest and health, such as to aid relief efforts during crises (Facebook, 2018). The data policy does not provide much information on the specific methods using for analysing the personal information collected other than Facebook connect personal information collected from different Facebook products and services. This aggregation of personal information is stated to be done to provide a more personalised and consistent experience for the individual across Facebook's products and services including targeted ads, see appendix 1, screenshot no. 1 (Facebook, 2018).

Based on this information on how the personal information is analysed in relation the purpose for collection, as listed above, it is not possible to identify any factors further affecting the ontological friction in the Facebook region. Therefore, according to Floridi's theory (2005) there is no change in the ontological friction and the accessibility of personal information. The result is the information gap between the individual and Facebook stays the same, which is also the case with the level of *informational privacy* experience by the individual in the Facebook region of the infosphere. However, if Facebook was to use methods such as data mining to discover new information about the individual, such a method would be considered as a factor influencing the ontological. This is since information about the individual not previously known is revealed, as described in the *Data and Social Media Mining* section earlier in this thesis.

Privacy Setting

When examining the privacy setting on a profile the only information possible to opt-out off being collected is location, see appendix 1, screenshot no. 2. This is also a setting that is turned

on as default when the account is created and must be manually turned off to stop the collection of location data on the individual. This privacy setting can be identified as a factor that increases the ontological friction within Facebook region, this is due to the setting blocking the collection of location data by Facebook. This factor's effect on the degree of the ontological friction is this case is not very large, because only one specific form of personal information is blocked, in this case the individual's location. So while the location setting can be identified as a factor affecting the ontological friction with the region, the change in accessibility of personal information is very small, causing only a slightly larger information gap between the individual and Facebook. As well as a very small, almost insignificantly, increase in the level of *informational privacy* experienced by the individual.

However, Facebook does provide the individual the option to control personal information in relation to online ads, see appendix 1, screenshot no.3. Here it is important to note that ability to edit ad preferences does not hinder the collection but not the use of personal information in targeted advertisements. In relation to Floridi's theory (2005) the setting used in customising the individual advertisement experience would not be considered a factor effecting the ontological friction within the Facebook region and as such the level of *informational privacy*. This is due to the setting does not restrict or hinder the accesses of the individual's personal information only the use of these information. The access to the personal information has already been provided, as seen in the previous part on the *Collection of Information.*

While Facebook does have a several privacy settings to restrict access to the individual's personal information on their Facebook profile, these setting are only related to other information agents, see appendix 1, screenshot no. 4. These privacy setting is a factor that impacts the level of *informational privacy* of the individual in the infosphere but only between the individual and other information agents excluding Facebook. And as such is not restricted to the Facebook region of the infosphere inhabited by the individual and Facebook, but to a much larger region of the infosphere inhabited by the individual and other information agents including Facebook.

This means that in order to raise the level of *information privacy* in relation to collection of personal information, factors must be introduced outside of Facebook privacy setting. Here the individual can use various PETs, as described in the section *Informational Privacy and The*

Internet, that are developed for blocking social plugins and as such hinder the plugins from collecting the desired information on the individual. An example is Adblock Plus, which is an add-on for browsers, when enabled it disable tracking, block ads and disable social media button (social plugins)[15] and as such raising the level of *informational privacy* for the individual can be raised. Another factor that can introduced and affects the ontological friction, is to exclude information on the individual's Facebook profile, personal information that cannot be found anywhere else on the Internet. This will minimise the amount of personal information that can be collected by Facebook (Young & Quan-Haase, 2013). And as such increase the ontological friction by blocking the accessibility to such personal information, creating a larger information gap between the individual and Facebook. Resulting in a higher level of *informational privacy* of the individual.

Privacy Concerns and Facebook

The information gathering practices used by social media such as Facebook has been criticised by Edward Snowden. In a tweet Snowden compared social media companies how makes money from collection personal information to that of surveillance companies by noting "*Their rebranding as "social media" is the most successful deception since the Department of War became the Department of Defense[sic].*"[16] With Facebook's practice of tracking individuals outside of Facebook platform and collecting as much personal information available, Caitlin Dewey (2016) illustrates this in her article *98 personal data points that Facebook uses to target ads to you*, where she list the 98 data points collected by Facebook and admits that the list might not be complete. The list includes personal information such as education level, field of study, income, expectant parents, relationship status, industry, job title, interests, the type of vehicle owned and charity donations, to name just a few. The connection drawn by Snowden between surveillance and social media becomes easier to understand. The list of personal data points also shows how accessible a large variety of personal information about the individual is and how the degree of ontological friction is every low due to unobstructed flow of this large amount of personal information, resulting in a very low level of *informational privacy* (Floridi, 2005). Furthermore, many individuals do not

[15] https://adblockplus.org/en/features
[16] https://twitter.com/Snowden/status/975147858096742405

read privacy policies or term of service agreements, when creating an account on a social media platform. The main reason for this has been contributed to click-wrap agreements, where the user only needs to click "*I agree"* or "*I accept"* to gain access to the social media sites like Facebook. These users also often rate convenience above privacy (Bechmann, 2014). This means that users often disregard the issue of *informational privacy* when using social media, being connected to friends and using these platforms for communication is more important. Therefore, many users are not actually aware of the amount and forms of personal information being collected. This is an example on how the behaviour of individuals in relation to social media and privacy agreements becomes a factor in decreasing the ontological friction within the infosphere, resulting in a decreasing in *informational privacy* (Floridi, 2005).

Another concerning aspect of social media in relation to *informational privacy* is the sense of false *informational privacy*, where the individual believes that they enjoy a high level of *informational privacy* when this is not true, this is especially relevant in the context of third-party applications. Many users are of the belief that their information is only visible to people on their friends-list, when in truth many third-party apps collect information on the individual through access to their friend's *friend-list* on Facebook. For example when using a Facebook profile as a log-in on a game application on a smartphone, the third-party application might ask to collect the name, profile picture, country, friend-list and email address of a user. This means that the third-party application suddenly has access to the names and profile pictures of individuals appearing on the user's friend-list. As such the user suddenly becomes responsible for and effecting their friend's *informational privacy*, removing the responsibility from Facebook to the user (Symeonidis, et al., 2018). The use of such application acts as a factor decreasing the ontological friction, making it easier to make connections between different users because of the accessibility to names and profile pictures and lowering the level of *informational privacy* according to Floridi's theory (2005). Furthermore, the lowering in *informational privacy* in can happen without the individual's, whose name and profile picture is collected, knowledge.

The Cambridge Analytica case is a perfect example of how users believed that their personal information was protected from being collected by third-party applications. In 2014 the company Cambridge Analytica gained access to personal information about an estimated 50 million Facebook users, the access was gained through a third-party app. The application was

a personality survey developed by Aleksandr Kogan, the app gathered information on the users of the app, as well as people connected to the user through Facebook (Granville, 2018). The practice has since in theory been banned by Facebook and third-party apps can gain access to the user's friends-list, the only real difference is the responsibility has moved from Facebook to the individual users, due to as mention earlier, it is now the user who to grant the access to their friend-list and not Facebook. The estimated 50 million Facebook users' information provided to Cambridge Analytica, was not all users of the app, in fact only about 270.000 users downloaded to app and took the survey. And these 270.000 users were also the only ones whom had consented to having their information gathered, the other 49.7 million Facebook users did not consent or was even aware that their information was being collected (Granville, 2018). A further breach of trust by Facebook in protecting their user's personal information, came to light at a parliamentary committee hearing in the U.K on the 26th of April 2018, attended by Mike Schroepfer, the CTO of Facebook. During the committee hearing Schroepfer admitted that Facebook did not read all the term and conditions regarding the use of the app (Browne, 2018).

The case of Cambridge Analytica is not the only case, where the users of Facebook has been given a sense of false privacy relating to their personal information. The creation and use of groups and their privacy setting on Facebook resulted in 2012 of two individuals being outed as gays. In this case the Queer Chorus of the University of Texas, added Bobbi Duncan and Taylor McCormick to their discussion group, which bypassed both individuals' privacy settings and resulted in a post on their timeline. The post announced that they had join the group to all their Facebook friends which included their parents, apart from McCormick's mother, none of the parents knew that their child was gay. Both Duncan and McCormick had ensured that their privacy setting restricted access to such information by their parents and believed that this personal information was safe. The reason this happened was the Queer Chorus Facebook group was an *open* group, meaning that the group and content from the group is publicly available to all users, which overwrote the privacy setting of Duncan's and McCormick's private profiles. This resulted in a fallout between Duncan and her father, who could not accept that his daughter was gay and while McCormick father was angry at the time, they have since resolved the issue and remains in contact (McCormick, 2012). This is another case where the behaviour of other information agents acts as a factor and affects an

individual's *informational privacy* by decreasing the ontological friction in the infosphere (Floridi, 2005).

The issue of Facebook providing users with a false sense of *informational privacy*, is not the only concerning issue regarding protecting one's *informational privacy* in an online setting. Facebook has been the subject of several complaints both in the US and Europe. In 2012 the FTC (Federal Trade Commission) in US investigated Facebook after receiving a complaint from EPIC (Electronic Privacy Information Center) regarding Facebook's privacy practices as being deceptive and unfair. The FTC found Facebook that misrepresented the users' ability to control their privacy settings, as well as sharing more information with third-party application than stated and failing to delete data on users after they deleted their Facebook account. Additionally, the FTC also found that previously private information was made public after Facebook introduced a set of new privacy features, which could result in substantial injury to their users. The result was a Consent Order because Facebook was found to have changed its promises of privacy without gaining consent from its users, requiring Facebook to obtain consent and inform users on any changes regarding privacy setting changes on Facebook. In 2015 the Tran-Atlantic Consumer Dialogue filed a complaint with the FTC in the US that Facebook was in breach of the 2012 consent order by tracking their users and collecting their browsing activities for advertising purposes and labelling these practices as deceptive. The complaint had yet to investigated by the FTC according to Jørgensen & Dasai (2017). Furthermore in 2016, a complaint was brought before the Brussels Appeals Courts concerning Facebook's practices of tracking non-Facebook users in Belgium without their consent. The case was dismissed due to Facebook's European headquarters is in Ireland and as such was outside of the jurisdiction of the Brussels Appeals Courts (Jørgensen & Desai, Right to Privacy Meets Online Platforms: Exploring Privacy Complaints against Facebook and Google, 2017).

In summation, the analysis of Facebook in relation to *informational privacy*, has shown that it can be difficult to obtain a high level of *informational privacy*, without being conscious about Facebook information gathering practices and understanding how they work. Furthermore, the analysis has shown that while Facebook is considered a social media site, they can just as easily be considered a surveillance company, whose main goal is to collect as much personal information about individuals. Information that is being used for profit in forms of targeted advertising on their platform. And that blindly trusting that Facebook protects the individual's

personal information is not the approach to take, as seen with the Cambridge Analytica case. Furthermore, when personal information is collected through third-party application, any control Facebook or the individual might have had over the information is lost (Symeonidis, et al., 2018). With cases such as the Cambridge Analytica and others mentioned in this thesis, helps to highlight these issues and shows that placing blind trust in the companies that possess detailed personal information about individuals, is not a viable solution. Protecting one's *informational privacy* in an online setting and especial when using social media can be a daunting task, not only in restricting access to one's personal information from the social media site but also from being collected from third-party companies, as seen with the use of social mining practices in the section *Informational Privacy and The Internet.*

Google and Informational Privacy

This section of the thesis will explore and analysis Google and their approach to collecting and using personal information in an online setting. First, I will present a short overview of the history of Google and explore how the platform has transformed since it was first made available worldwide and to the present day. In the second part Floridi's theory of *informational privacy* (2005) will be used to analyse the level of *informational privacy* by the individual in the Google region of the infosphere. This is done by identifying and analysing factors in relation to the collection and use of personal information and how these factors affect the ontological friction, the accessibility of personal information and the information gap between the individual and Facebook based on the principles of Floridi's theory of *informational privacy* (2005). The analysis will mainly be based on the most recent Privacy Policy by Google that became effective on May 25 of 2018. This means that GDPR has been considered when designing the policy. Additional, examples of privacy setting will be used to illustrate how these can affect the individual's *informational privacy*, these examples will be taken from my personal Google account accessed on my desktop, therefore some personal information might be censored. The section will end with highlighting examples of cases where Google's approach to the individual *informational privacy* can be viewed as concerning and deceptive.

The story of Google

Google has become a household name, searching for information is called *googling* and in December 2017 the Google search engine reached a market share of 87.1 %[17], making it the world's leading search engine.

Google Inc. was founded by Larry Page and Sergey Brin in 1998 but the research behind Google began in 1995, when Page and Brin both where working in their Ph.D. in computer science at the Stanford University in California, the first prototype of Google was called BackRub. As their research and developing continued, they realised the importance of their search engine and decided to try and license the Google technology to Internet companies, in which they failed. In 1998 Page and Brin had left Stanford, and with almost $1 million in funding, started Google Inc which they operated out of a garage in Menlo Park in California.

[17] https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

In 1999 Page and Brin obtained $25 million in a second round of funding from different Silicon Valley venture capital firms and moved to Mountain View, California in a new facility which they called the *Googleplex* (Alesso & Smith, 2008). In 2000 Google expanded and created AdWords, a self-service advertisement software, which allows companies to develop ads to be displayed on Google services. And Yahoo! selected Google as their new provider of supplementary search results, as well as America Online (AOL), Netscape and Microsoft Network (MSN). Yahoo! and Microsoft have since their own search technologies. In 2004 Google went public and offered stocks to their investors, this resulted in Page and Brin join the billionaire's club, ranking number 12 and 13 respectively on the Forbes billionaires list as of 2018[18]. Google also introduced as some of new services, Gmail, an email service, Google Earth, a map with satellite images of the earth and Google Talk, a Voice over Internet Protocol (VoIP). And started to compete with Microsoft as the leading computer service provider (Alesso & Smith, 2008).

Googles development and acquiring of new services has continued to grow now including, Google Map, the Google Calendar, the web browser Google Chrome, the mobile operating system Android, Google+, their own social media site, YouTube, Google Drive that includes various text, image and spreadsheet programs, Chromecast and Google Fit, just to name a few[19]. Furthermore, Google has also introduced Google Assistant, a digital assistant that can be connected to other devices and control them with your voice, through a smartphone or the Google Home (a smart speaker and voice assistant). When the Google Assistant is links with other smart devices it can controlling the brightness of lights, tell you about the weather, play music or movies and control the temperature in one's home[20]. Today Google offers around 63 different services not including businesses and developers. As it can be seen Google provides very wide variety of services, all of them can be connected to the individual's Google Account, putting Google in a position to gather a host of personal information about each user.

[18] https://www.forbes.com/billionaires/#7c59c244251c
[19] https://www.google.com/about/products/
[20] https://assistant.google.com/explore?hl=en_us

Informational Privacy in the Google Region

The infosphere is the whole information environment comprised of all informational entities including information agents, the informational entities properties, relations with each other, processes and interactions. The infosphere includes both online, offline and analogue spaces of information (Floridi, 2013). Within the infosphere different regions are located, these regions are inhabited by different information agents, regions within the infosphere can overlap, creating a sub-region inhabited by two or more information agents, as defined in the section *Floridi and Informational Privacy* of this thesis (Floridi, 1999). The expression Google region in this thesis refers to a sub-region of the infosphere inhabited by the following information agents; the individual and Google.

Determine the level of *informational privacy* for an individual in the Google region, any factors and their effect on the degree of ontological friction, as well as the availability and accessibility of personal information must be identified. The degree of ontological friction can be affected using ICTs as explained in *The Re-ontologization of the Infosphere* section, factors can also include consent to having personal information collected, personal information received from other sources such as data brokers, a change in the environment of the infosphere and the information agent's behaviour with in the infosphere (Floridi, 2005). I have previous highlighted the impact collection of personal information can have on the *informational privacy* of an individual in relation to the use of tracking technologies such as cookies, as described in the section *Informational Privacy on The Internet*. In order to access many of Google services, such as Google Drive and the Google Assistant, the creation of a Google account is necessary, this is the first factor that effects the ontological friction in the Google region that can be identified. By creating a Google account and sharing personal information during the creation process, the ontological friction is decreased and an increase in accessibility of personal information is provided. And according to Floridi's theory (2005) this creates a smaller information gap between the individual and the other information agent, in this case Google. Causing the individual's level of *informational privacy* to decrease, leaving the individual personal information less protected within the Google region of the infosphere.

Information Collected

The personal information collected by Google are divided into two categories (Google, Privacy Policy, 2018);

1. *The information you give us*. Information related to name, email, age, gender and credit card number and information Google associate with the user's Google account.
2. *Information we get from your use of our services.* Information related to the use and activities of Googles services, websites visited, device information such as operating system and unique device identifier. Log information, like search queries, phone log details such as duration of calls. IP address, browser information and location information. And interactions with Google services or services offers by partners, includes advertising services such as AdWords, AdSense and Google Analytics and social media buttons.

The information collected in these two categories are not only detailed information on the individual such as personally identifiable information (ex. name, age and gender). But also indirectly identifiable information relating to the type of device used to access Google from, hardware and software, IP address, browser type, plugins and battery level, the type of device used, webpages viewed, ads shown, and items purchased, just to name a few. Google also notes that if the users wish to take full advances of the features offered by them, they might ask the use to create a public Google+ Profile. Applying Floridi's theory of *informational privacy* (2005) it is possible to identify several factors that impacts the ontological friction in the Google region of the infosphere. First is the use of Google's services, second is the use of tracking and collection methods, such as cookies, social media buttons and Google Analytics. Each of these factors results in a lowering of the ontological friction in the Google region and heightens the accessibility of the individual's personal information. As noted previously it is difficult to quantify the exact impact these factors have on the ontological friction, since each factor can affect the ontological friction to a greater or lesser extent. However, it can be concluded that each factor contributes to a lowering in the ontological friction. Use the analogy of the river once again, with the individual at the mouth of the river and Google at the end of the river and the water in the river represents the individual's personal information. The river is filled with boulders and for each factor introduced to the Google region of the infosphere, a boulder is removed, resulting in more and more personal

information flowing towards Google at the end of the river. Giving Google an increasing access to personal information and lowering the individual's level of *informational privacy* with each boulder removed. In short this means that with each Google service used or encountered, an increasing amount of personal information can be collected by Google. The use of third-party cookies on Google service is also a factor the decrease the ontological friction in the infosphere and is not restricted to the Google region. This results in a decrease of the individual's overall level of *informational privacy* in the infosphere.

Use of Information

The personal information collected about individuals by Google are used for several purposes provide, maintain, protect and improved services, as well as develop new services. Furthermore, Google use the personal information collected for personalising content and ads, for analytical purposes, in communication with the individual and to improve the safety and reliability of Googles services. Google combines personal information about the individual from several sources for these purposes (Google, Privacy Policy, 2018). For example personal information collected from YouTube, the Google Assistant and Gmail are combined for the above purposes. Additional personal information from the individual's Google profile and Google account may be used across all of Google's services, this information may include name, profile picture, reviews and comments written and posted on Google+. This information may also be used by Google in a commercial context such as ads. Google also analyse information gathered from the individual's content including emails, this included emails from non-Google users. Google states that this is done to customise the content of Google services to the individual, to avoid spam and for malware detection (Google, Privacy Policy, 2018). Google further explain the methods used in analysing the personal information for the purposes stated are technologies such as automated analysing systems and algorithms for pattern recognising (Google, Privacy Policy, 2018). And gives an example "*For example, if you search for "mountain bikes," you may see an ad for sports equipment when you're browsing a site that shows ads served by Google*." (Google Privacy Policy, 2018 page 6). In other words, they use data mining to analyse the personal information collected and states that sensitive personal information, such as health, race, religion and sexual orientation, are not used in personalised ads (Google, Privacy Policy, 2018). However, they do not state whether this sensitive information is used for other analysis purposes, which can be seen a

problematic. In the part on data and social media mining the analysis showed that such methods had an impact on ontological friction and the accessibility of personal information in the infosphere. The use of data mining by Google on information collected on activities of the individuals while online, can reveal a great deal about the individuals browsing habits, most frequent search topics and favourite website to visit, making it possible to identify groups of individuals with similar interests. This means according to Floridi's theory (2005) that Google's methods for analysing the information collected is another factor that influence the ontological friction in the Google region of the infosphere. Another boulder removed from the river of personal information flowing in the Google region from the individual to Google. Resulting a narrower information gap and a further lowering of the *informational privacy* of the individual within the infosphere.

Privacy Setting and Informational Privacy

In the previous two paragraphs I have analysed how the collection and use of personal information by Google resulted in a low level of *informational privacy* for the individual, in this paragraph I will explore the effects of privacy setting has on level of *informational privacy*. There are several privacy settings that can be identified as factors affecting the ontological frication and in such lowering the accessibility of personal information within the Google region of the infosphere. The examples used in this paragraph are from when accessing the privacy settings on a computer, if examples are accessed from other devices, this will be explicit stated. The two groups of privacy setting focused on here are the activity controls and the customisation of privacy setting related to personalising of online advertisements.

Activity Controls

The first set of factors that can affect the ontological friction in the Google region are activity controls, these controls are as the name implies related to the individual's activities online. The first activity control setting is the *web and app activity* setting, see appendix 2 screenshot no. 1. When the setting is paused Google does not collect information on searches, location, language, IP address, ads clicked, purchase history from advertiser's website, recent apps used, Chrome browsing history and activities on websites and application using Google services (Google, See and control your search activity, 2018). These activities can also be saved when offline, if this setting is paused, activities will not be saved when offline. Applying Floridi's theory (2005) to the web and app activity setting, this setting is a factor that effects

the ontological friction in the Google region of the infosphere. The effect of the setting when paused is that the ontological friction rises, the means that it comes harder for Google to collect personal information in an online setting on the individual, as the accessibility of the personal information is lowered. The setting acts as a barrier between the individual and Google, increasing the information gap and providing the individual with a higher level of *informational privacy*. However, pausing web and app activity does not completely blocks Google from temporarily collecting information related to recent searches to improve the active search session (Google, See and control your search activity, 2018). This means that the individual's level of *informational privacy* can be temporarily lowered, since the temporarily collection of personal information, would according to Floridi's theory (2005) would be a factor that lowers the ontological friction in the infosphere and increasing the information gap between the two information agents, here the individual and Google.

The second factor effecting the ontological friction in the Google region of the infosphere are the *location history setting*, see appendix 2, screenshot no. 2. This setting controls the collection of physical location data, such as places visited, meaning the tracking of the individual in relation to their physical location at a given point in time. When this setting is paused, Google does not collected information on the location of the individual. Setting the location history on pause, is another factor that effects the ontological friction in the Google region of the infosphere, by blocking the access to the individual's location a higher degree of friction is obtained and a lowering of the accessibility of the individual's personal information is provided. This means that when the location history setting is set on pause, the individual increases the information gap and experience a higher level of *informational privacy* than when the setting is turned on.

The third factor that can influence that ontological friction in the Google region of the infosphere are the *device information* privacy setting, see appendix 2 screenshot no. 3. This privacy setting is related to personal information like contacts, calendars, music and device information, such as battery level of computer or smartphone, Wi-Fi connection quality and whether Bluetooth is activated (Google, Manage Device Information setting, 2018). As it can be seen pausing the device information setting is another factor increasing the ontological friction in the Google region, pausing the device information collection, lowers the accessibility of the individual's personal information. Resulting in a wider information gap and

a further raising of the level of *informational privacy* for the individual in the Google region, according to Floridi's theory (2005).

The fourth of the activity settings impacting the ontological friction in the Google region of the infosphere are the *voice & audio activity* setting. This setting is related to the use of voice commands for example when using Google Assistant, see appendix 2, screenshot no. 4. The audio recorded are save to the individual's Google account and can be accessed and deleted any time (Google, Manage Google Voice & Audio Activity, 2018). Pausing the voice and audio activity is yet another factor that impacts the ontological friction by causing an increase in the friction in the Google region by blocking Google from recording the individual's voice and an expanding of the information gap between the individual and Google. Resulting in the individual experiencing an increase in *informational privacy* within the Google region of the infosphere.

The last two activity settings are related to YouTube, see appendix 2, screenshot no. 5, these two setting are related to the individual's search and watched history on the YouTube platform. If these setting are paused, information about search queries and videos watched are not collected by Google. This means that these two setting are two additional factors that increases the ontological friction in the Google region of the infosphere, restricting the accessibility of personal information about the individual's activities on YouTube, which leads to an increase in the information gap and as such a raise in the level of *informational privacy* for the individual when applying Floridi's theory (2005) to the effect of these two setting on *informational privacy* in an online setting.

A final factor is Google Analytics opt-out browser add-on[21], this add-on prevents the individual's data from being collected by Google Analytics, by preventing the Google Analytics JavaScript running on websites and sharing information with Google. This add-on will further increase the ontological friction and the information gap in the Google region between the individual and Google. Allowing the individual to experience a raise in their level of *informational privacy* within the infosphere.

[21] https://tools.google.com/dlpage/gaoptout

When applying Floridi's theory (2005) of *information privacy*, to each of the different activity controls, when paused, is a factor that increases the ontological friction in the Google region, which results in a lowering of accessibility of personal information about the individual. This leads to a larger information gap between the individual and the information agent Google and such the individual will experience an increase in the level of *informational privacy*. Using the analogy with the river, where the individual is at the mouth of the river and Google is at the end of the river and where the water of river is personal information about the individual. If the setting mentioned above is not paused, the personal information about the individual flows without any or very little resistance. When each one of the setting is paused, a boulder is placed in the river and slowing down the flow of personal information from the individual to Google, until very little personal information is flowing from the mouth of the river to the end of the river. However, pausing these setting is not without consequences, each setting paused can cause a degradation in the individual's user experience, when using various Google services, see screenshot no. 2. This means there is a trade-off between *informational privacy* and user experience and the individual must determine which personal information is more important to protect from being collected versus the level of user experience they wish to enjoy when use Google's services online.

While it does seem that Google does everything for the individual to control the accessibility to personal information about them, there have been cases where Google has collected personal information without the knowledge of the individuals who had their information collected. In the next paragraph I will explore cases where Google has acted less honest and more deceptive in their collection and use of personal information.

Privacy Concerns and Google

This paragraph will highlight examples of the privacy concerns raised in relation to Google and their practices regarding collection and use of personal information.

In 2007 Google launched Street View, Google used modified cars to obtain panoramic street-level photograph of residential roads and the structures surrounding the roads. These photographs were used to enhance the Google applications Google Maps and Google Earth. This cause a concern over the individual privacy due to some of the more controversial photograph that was captured by the cars. Such as the man entering an adult video store in

the UK and the photo of a naked woman in Taiwan, while such photos are does not depicted anything that people present at the location wouldn't see. The publication of the photos however made the information available to people not present at the location and as such expanding the number of information agents having access to the information. Resulting in a decrease of the individuals' *informational privacy* according to the theory of *informational privacy* by Floridi (2005). The emergence of the Google Street View and the Google Street View Cars is also an example on how the environment of the infosphere changed the ontological friction, resulting in a decrease in the *informational privacy* within the infosphere (Floridi, 2005).

However, the Google Street View Cars was also capturing, parsing and storing Wi-Fi data, without information anyone outside of Google, the Wi-Fi data was to be in developing a global map of Wi-Fi access points. When data is sent over the Internet, a packet is transmitted which include header information, that contains the source and destination of the packet and the actual data being sent is known as the payload. When the packet reaches its destination, the header information is stripped, and the only data received by the destination system is the payload[22]. In 2010 is was revealed that the Google Street View cars had collected Wi-Fi header data alongside the collection of photographs. Google argued that they had done nothing wrong because the header information was publicly available and did not contain any personally identifying information. Google later admitted that it inadvertently also had collected the payload. This resulted in several regulatory investigations around the world. In UK the Information Commissioner's Office concluded that the collection of payload data was in breached with the Data Protection Act 1998. The Canadian Privacy Commissioner also ruled against Google and determined that that Google was in breach of Personal Information Protection and Electronic Documents Act. In Hong Kong it was concluded that the payload data could not directly identify an individual and the Hong Kong Privacy Commissioner concluded that Google was not in breach of Hong Kong privacy law. The Dutch Data Protection Authority concluded that both the payload and the header information was personal data under the Data Protection Directive in the Netherlands.

22 https://www.dir.ca.gov/dwc/EAMS/EAMSPresentTermSolutionDocumentRepository/Payload%20Definition.pdf

In 2010 the first law suit was brought against Google regarding their email practices of scanning emails for the purposes of creating user profiles or providing targeted advertising was by Keith Dunbar, Dunbar claimed that Google was violating what is known as the Wiretap Act. Since than several more lawsuits have been filed against Google and their practice of scanning the contents of Gmail messages to serve ads to its customers. Google have argued that Gmail users had consented to having their emails scan because they had agreed to the terms of service and privacy policy. Regarding users of other email services communicating with Gmail users, Google argued that all email users understood and accepted the fact the emails are automatically processed. In 2016 the case of Matera vs Google, a lawsuit was brought on the behalf of non-Gmail users how had not agreed to Google's terms of service. Because the scans take place before the messages reaches the Gmail inbox that scans were considered sensitive, even though that scan takes place milliseconds before the messages reaches the inbox. Matera argued that this constituted a breach of both the federal Electronic Communications Privacy Act and the California Information Privacy Act. The result was that Google settled the lawsuit by delaying any advertising-specific scans until the emails reached the inbox. When examining this case using Floridi's theory (2005), the scanning of emails can be identified as a factor that decreases the ontological friction in the infosphere by increasing the accessibility of personal information and lowering the *informational privacy* not only of Gmail users but also individuals using other email services.

In 2014 Google Spain became the centre in a debate on privacy and search engines in the EU, which resulted in the R*ight to be Forgotten*. This came about due to a Spanish citizen named Mario Costeja González, who had requested that Google removed a link to a newspaper article about him. The newspaper article described the foreclosure on his home due to outstanding debt in the late 1990s. However, González had since paid the debt and the proceedings against him had be resolved and González felt that the reference was irrelevant and sought Google remove the link in 2010 and contacted the AEPD (Spain's data protection authority) for help (Belbin, 2018; Brock, 2016). In 2007 the AEPD had issued a paper which argued the search engines should be held liable for the data processing of information about individuals. And that while the publishers of the original material could claim freedom of information rights, search engines could not, it was later clarified that while publishers could refuse erasure requests. Search engines did not have the right to object to de-indexing

(Belbin, 2018; Brock, 2016). In January 2011 five test cases was appealed to the high court in Madrid by Google and González's case was referred to the EU judges for a precedent-setting decision. In May 2014 the Court of Justice of the European Union ruled in favour of González citing that Google fell within EU law and the Data Protection Directive and counted as a data controller. And determined that link that are "*inadequate, irrelevant or no longer relevant, or excessive in relation to the purposes of the processing at issue carried out by the operator of the search engine must be removed from the results by the search engine, pursuant to a request by the individual and an appropriate review*" (Belbin, 2018, p. 21). The ruling limited the *Right to be Forgotten* to domains within the EU, meaning that the link would still appear on non-EU domains, even if the search can from within the EU. This was rectified in 2016 where the *Right to be Forgotten* was to include searches from within the EU (Belbin, 2018; Brock, 2016). Since the ruling in 2014, 2.4 million URLs has been requested for delisting, 89% is from private individual[23]. This case is an example of how a change in legislation can acts as a factor by increasing the ontological friction in the infosphere. Here the possibility to have wrongful and outdated personal information removed from search results, is makes it more difficult to access such forms of personal information, giving the individual an increase in their level of *information privacy* (Floridi, 2005).

The language of privacy policies is a general concern, due to the ambiguous language often used in the description of practices regarding the collection, use and sharing of personal information. The ambiguity of the language undermines the purpose and value of the policies, without clear explicit statement the polices can be quite meaningless for the individual. Two motivations for use of ambiguous language, first the policy may include divergent or separate situations for when an action occurs and do not occurs. In which case the action does occur pending on the situation and that there are foreseeable but unrealised future actions. Here clarifying statements on situations where the action does or does not occur, which would result in a less vague policy (Reidenberg et al., 2016). The second is for the policy to be flexible to accommodate future actions without changes. Here the motivation is to accommodate flexibility can be at best a form of inaccuracy and at worst misleading and misrepresentative. The level of ambiguity in a privacy policy can signal whether a privacy policy is a meaningful notice of an actual policy and practices and a notice that might give rise to a contractual

[23] https://blog.google/topics/google-europe/updating-our-right-be-forgotten-transparency-report/

commitment (Reidenberg et al., 2016). Furthermore, it has become almost customary to create long and complicated consent forms, which can be hard to understand due to the ambiguity and technical terms. Results in a reluctance towards reading the privacy policies and just click on the *"I accept"* and the individual right to complain over the terms of service and the privacy policy is waived and any judiciary will determine that the consent was given voluntary (Reidenberg et al., 2016).

In summation the level of *informational privacy* in the Google region of the infosphere is influenced by many factors, each factor effecting the ontological friction to a higher or less degree. The changes to environment of the infosphere can also have an impact on *informational privacy*, some changes can be permanent, and some can be temporary and either increase or decrease the *informational privacy* in different degrees (Floridi, 2005). This shows that the individual's level of *informational privacy* in the Google region is dynamic and is every changing according to factors introduced or removed for the environment of the region. Protecting one's *informational privacy* can seem as a daunting task, with many factors to be considered, such as the trade-off between privacy and user experience. The cases mentioned in this section shows that placing blind trust in companies like Google to provide adequate protection and not be deceptive when collecting personal information is not the best approach to take if one is concerned about one's *informational privacy*.

The General Data Protection Regulation

The focus in this section is the Regulation (EU) 2016/679, also known as the General Data Protection Regulation (GDPR), the GDPR came into to effect on the 25th of May 2018 in the European Union and as such repealed Directive 95/46/EC, also known as the Cookie Order. The GDPR is a combined effort to create a consistent and homogenous application of rules for protecting the fundamental rights and freedom of the individual citizen in the EU, with an emphasis on the protection of personal data and the processing of this data, in an online setting. This is something that the Directive 95/46/EC failed to do, due to the application and implementation of the Directive being left up to each member state, resulting in a fragmentation of the process of implementation, leading to uncertainty and an increased risk to the protection of personal data (Regulation EU 2016/679).

The creation of the GDPR is also based on the rapid technological developments and globalisation, which has created new challenges and needs relating to informational privacy and the protection of personal data (Regulation EU 2016/679). These challenges are especial connected to the developments of digital Information and Communication Technologies (ITCs), as Floridi (2005) also highlighted as playing a part in the emergence of the digital infosphere. The developments of tools such as the various forms of cookies, social plugin and browser fingerprinting, has made the collection and processing of personal data easier and more unnoticeable than before the Internet (Angwin, 2014; Floridi, 2005). The result is information agents are more vulnerable of having their personal data collected and processed without having any control over this, as also described previously in this thesis. Furthermore, the GDPR is based on providing legal certainty and transparency for economic operators[24] in relation to the processing of personal data on covers both automated and manual processing of the data (Regulation EU 2016/679). The focus in this section of the thesis will be on how the GDPR impacts the collection and use of personal information in an online setting and the impact the GDPR has on the individual *informational privacy* as defined by Floridi (2005).

[24] Economic operators are defined by the Union as any persons and/or entities which offer the execution of works, the supply of products or the provision of services on the market, such as firms, branches, subsidiaries, partnerships, cooperative societies, limited companies, universities, public or private. http://iate.europa.eu/FindTermsByLilId.do?lilId=933932&langId=en

The GDPR, Personal Information and Floridi.

The objective of the GDPR is to provide rules relating to the protection of the individual and the processing of personal data, as well as providing rules for free movement of personal data within the Union. Personal data is defined as "any information relating to an identified or identifiable natural person" (Regulation EU 2016/679, Article 4, section 1, 2016). By identifiable natural person the GDPR means any individual that can be identified, directly or indirectly by referencing identifiers such as name, ID number including online identifiers such as unique ID number like an IP address, location data, as well as one or more factors specific to physical, mental, economic, cultural or social identity. This means that personal information is any form of information related to the individual and can be used either alone or in combination with other information to identify the individual, would warrant protection under the GDPR. Furthermore, the objective of the GDPR, as mention earlier, is to ensure a consistent and homogenous application and implementation of rules across the Union and the protection of the individual and the processing of personal data (Regulation EU 2016/679).

This definition of personal information varies greatly from Floridi's (2005) view on personal information, where the GDPR views personal information as related to the individual, Floridi (2005) views personal information as a part of the individual, a part of the individual's identity. Floridi (2005) argues that viewing the individual's identity as being constituted by that individual's personal information, can help to understand the individual's right to *informational privacy*. Because anything done to the individual's personal information is done to the individual, not the individual's belongings. From Floridi's (2005) perspective personal information cannot be owned or possessed, because the personal information is a part of the individual's Being. An individual's search history and activities on online can show religious beliefs, medical problems and much more, as seen in the section *Informational Privacy on The Internet*, it is basically a printout of what is going on in the individual's brain and the individual is constantly leaving behind a trail of personal information when online. There more an individual's *informational privacy* is protected, the more the individual's identity is safeguarded. In Floridi's (2005) theory the individual's identity and the individual's informational sphere are two sides of the same coin, which is why the protection of *informational privacy* is extremely valuable. While I do agree with Floridi on the point that in

an online setting, personal information about an individual can be constituted as a part of the individual's identity and while the personal information collected might not depict a full representation of the individual, it can reveal some sensitive and personal aspects and core characteristics of that individual's identity and personality. Furthermore, seemingly inconsequential information can be used together with other information in identifying an individual. However, I also believe that this approach to personal information is lacking. This is due to the lack of distinction between which personal information would warrant greater protection than other personal information, furthermore Floridi (2005) also does not specify when a violation of informational privacy occurs in relation to personal information. This means that Floridi's (2005) informational privacy theory as described in this thesis is incompatible with the GDPR.

These challenges are also supported by Tavani (2008) in the paper *Floridi's ontological theory of informational privacy: Some implications and challenges*. Tavani states that Floridi's theory is challenged by the lack of distinction between descriptive and normative aspects of *informational privacy*. Tavani (2008) explains that Floridi (2005) does not differentiates between the loss of privacy from a violation of privacy. Tavani (2008) proposes that the solution could be to incorporate the principles of *naturally private situations* and *normatively private situations*. Where *naturally private situations* are situations where privacy is protected by natural means, such as hiding in a cave, in situations such as this privacy can only be lost and never breached. *Normatively private situations* are situations where privacy is protected by ethical, legal or conventional means. This can be situations such as an appointment with your general physician or a lawyer. In cases as these if there is an unauthorized entry into the situation, *privacy* has not been lost, but has been violated or invaded (Moor, 1997; Tavani, 2008). Furthermore, Tavani (2008) suggests that Floridi contextualizing the infosphere in a way the differentiate segments of it into situations where an information agent may or may not have reasonable expectation of privacy. In the article *Informational Ethics: A Reappraisal* (2009) by Floridi, he discusses this concern of Tavani's (2008). Floridi agree with Tavani (2008) that his theory is lacking regarding distinguishing between the mere loss of *privacy* and the violation of *privacy*. Furthermore, Floridi agrees with Tavani's approach to correct this, by introducing naturally private situations and normatively private situations. And as such to contextualize the infosphere in a way that would differentiate part of the infosphere into

situations to accommodate this. Thus, being able to distinguish between situations where *privacy* is lost and situations where *privacy* is violated. And that work is still to be done and this approach seems to be most promising.

Furthermore, I would propose that Floridi's theory could also benefit from introducing privacy situations as described by Tavani & Moor (2001), it is possible to decide what personal information to keep private and which information is to be shared freely. When working with zones of *privacy* in an online setting, it is important to note that while the individual cannot control the flow of information, personal protection is. This can be done by determine what personal information and activities needs to be protected on the Internet. When defining private situations, it is necessary to define whom has access to the information or activity and under which circumstances (Tavani & Moor, 2001). Furthermore, different parties may have different levels of access and restrictions when gathering and using information or activities. When creating these zones of *privacy,* it is important to inform, what conditions the zones operate under, as well as the level of security that they employ. So that people may decide which zones they wish to use regarding a specific situation (Tavani & Moor, 2001). These zones many offer levels of access ranging from total *privacy* to unabashed publicity. By adopting principles similar to Tavani & Moor's (2001) privacy zones, Floridi's (2005) view of all information being classified as personal information, would remain intact but it would be possible to divide personal information into groups based on which level of protection the information would merit. And as such Floridi's theory would be much more compatible with the GDPR, in relation to defining which personal information such be protected.

Collection and Processing of Personal Information

The collection of personal information has with the implementation of the GDPR, provided a definition on what information gathered should be considered personal information, as seen above. Furthermore, the GDPR states that special categories of personal information are considered particularly sensitive in relation to the rights and freedoms of the individual and thus merits specific protection. Personal information disclosing the individual's ethnic origin, political opinions, religious beliefs, health information or sexual orientation and the disclosure of these personal information might lead to for example discrimination against the individual.

As such the processing[25] of these categories is generally prohibited under the GDPR. However, there are exceptions where processing of these categories is allowed, in an online setting such an exception can be made if the individual gives explicitly consent to allowing the processing of these categories (Voigt & Bussche, 2017). This distinction between personal information and the level of protection they merit, is similar to the privacy zones as described by Tavani & Moor (2001) and as such if adopted by Floridi into his theory regarding *informational privacy* in the infosphere would result in heightening, if consent is not obtained, the individual's degree of *informational privacy*. The information gap between the individual and companies wishing to process such personal information would than broaden and the flow of information would flow less freely. This in one factor in the GDPR that impact the information flow in the infosphere and the individual's degree of *informational privacy.*

The GDPR also applies in situations where a service is provided, regardless of whether a payment is required, for example Facebook provides their platform free of charge. This includes the monitoring of behaviour and tracking of individuals using these services, with the purpose of using the information collected to profile the individual and makes decisions based on the processing of these information. As well as the analysing and prediction of the individual's preferences, behaviours and attitudes. Meaning that any information collected through any form of web tracking such as cookies, social plug-ins, unique IDs and browser fingerprinting is classified as personal information and merits protection under the GDPR (European Parliament & Council of the European Union, 2016; Voigt & Bussche, 2017). Under the GDPR the individual must give consent to the processing of their personal information and is given the right to withdraw consent at any time. The withdrawal of consent means that the individual's personal information may not be further processed, any processing done before the withdrawal is not affected (Voigt & Bussche, 2017). In practices this means that in an online setting the individual must be present with the possibility of giving consent to having their personal information processed, informed on any processing operations performed on their personal information and the purpose hereof. Furthermore, the individual must be informed on the intention to transfer their personal information to a third-party, the retention period for storage of personal data and any future intent to process the personal

[25] Processing means any operation or set of operations which is performed on personal data or on sets of personal data, such as data mining for example.

information in ways other than states upon the collection of the personal information (Voigt & Bussche, 2017). The GDPR focus mainly on the processing of personal information and transparency in relation to the purpose of the processing, whereas the collection of personal information is secondary. However, the GDPR does specify that an individual's personal information shall only be collected for specified, explicit and legitimate purposes (European Parliament & Council of the European Union, 2016; Voigt & Bussche, 2017). Meaning that any collection of personal information that does not fulfil these conditions would be in violation with the GDPR.

In an online setting the GDPR divides the individuals in to two groups, one group where the individual's personal information to some degree is protected by the GDPR and another group where no such protection is provided. Individual under the protection of the GDPR is provide more transparency regarding the purpose behind the collecting and processing of their personal information, the individual outside the protection of the GDPR. This means that individuals covered by the GDPR, should be able to make more conscious choices regarding protecting their personal information and thus be able to obtain a higher level of *informational privacy*. For examples individual within the EU are informed whenever their personal information is being collected for processing purposes and is asked for their consent, giving them a more active role in controlling the use of personal information. This means that individuals in the EU can have a greater influence on the flow of their personal information in an online setting.

Conclusion

Our world is a world of information, the collection, use and sharing of personal information has become an economic factor in everyday business, the evolution of our world to a digital infosphere has changed our approach to informational privacy drastic as opposed to the world before the Internet became available to the everyday person. This thesis has shown that the changes in our infosphere brought on by the re-ontologization of the infosphere, has greatly impacted the individual's *informational privacy*. The developments of digital ICTs have made the collection and use of personal information in an online setting, much easier than before the Internet. The availability and accessibility of personal information online has brought on a new form of currency, the information currency, every piece of personal information about an individual, now has a piece tag. Resulting in a boom of companies collecting and trading in personal information about individuals. In an online setting Floridi's statement that "*you are your information*" (2005, page 195), is very much true, every piece of personal information collected about an individual, reveals some part of the individual's identity. Our search history, websites visited, and videos watch providing snapshots of what is going on inside our brain. Therefore, in an online setting the individual's *information privacy* has become an extremely valuable.

Informational privacy is the concept of privacy concerning the amount of personal information about an individual available in the infosphere, the infosphere is the whole information environment comprised of all informational entities including information agents (us), the informational entities properties, relations with each other, processes and interactions and includes both online, offline and analogue spaces of information (Floridi, 2013). In the infosphere the individual's *informational privacy* is determined by the ontological friction, the accessibility and the size of the information gap between information agents (Floridi, 2005). Here a high degree of ontological friction is considered a positive thing because the higher degree of friction, the higher level of *informational privacy* can be experienced. In relation to the level of *informational privacy,* a large information gap between two information agents is also considered as a positive thing, because it means that the amount of personal information one information agent knows about the other information is very small. The degree of ontological friction is affected by many factors in the infosphere, such as information and communications technologies (ICTs), behaviour of

the information agents and the use of privacy enhancing tools (PETs) when online (Floridi, 2005). This means that the individual's level of *informational privacy* is dynamic and changes in relation to factors introduced or removed from the infosphere, these changes can be permanent or temporary. Floridi (2006) describes this as "*Put simply, informational privacy is a function of the ontological friction in the infosphere. It follows that any factor affecting the latter will also affect the former.*" (page 110). Digital ICTs are re-ontologizing the infosphere because they are engineering new environments, that the individual can enter though gateways, gateways such as the Internet. And blurring the line between being offline and online, what Floridi (2007) calls *here* (offline and analogue) and *there* (online and digital), that the digital is merging with the analogue, such as the Internet of Things (IoT). In the re-ontologized infosphere, digital ICTs can both erode and enforce the individual's *informational privacy*, where cookies, browser fingerprinting and tracking can erode an individual's *informational privacy*, Privacy Enhancing Technologies (PET) which can help protected the individual from having personal information collected in an online setting (Floridi, 2005;2007; Tavani & Moor, 2001).

Cookies, browser fingerprinting, data and social media mining are all different factors present in the infosphere affecting the ontological friction in various degrees, by making personal information more accessible for information agents, resulting in lowering the individual's level of *informational privacy*. Here it is important to note that for example not all cookies affect the ontological friction and *informational privacy* in the same way, some cookies might collect more intimate personal information than other. As such they each count as individual factors affecting the ontological friction and the information gap between information agents in the infosphere. The same can be seen with data and social media mining, depending on the method used for analysing the data, different conclusions can be drawn, some providing more intimate personal information than others as seen with the Target example, where Targets development of a pregnancy predication score, revealed a teenage pregnancy to the girl's father (Sumner, 2016).

When analysing Facebook and Google in relation to the collection and use of personal information and the impact on *informational privacy*, this thesis has shown that the degree of ontological friction in the two regions of the infosphere is every low, if no countermeasures are being utilised, such as the blocking of cookies and privacy setting. This

low degree of ontological friction is caused for example by the willingness of sharing personal information, consenting to having personal information collected, the availability of personal information by creating a Facebook profile and a Google account, disregarding privacy setting and not being aware of the specific forms of personal information being collected. The low degree of ontological friction results in a high accessibility to personal information and a very narrow information gap between the individual and Facebook and Google. The result is that the individual will experience a low level of *informational privacy* within the Facebook region and the Google region within the infosphere. When an individual seeking to better protect their *informational privacy*, a trade-off between *informational privacy* and user experience is encountered, meaning the individual must determine which personal information is more important to protect from being collected versus the level of user experience they wish to enjoy when use Facebook's and Google's services online.

The objective of the GDPR is to provide rules relating to the protection of the individual and the processing of personal data, as well as providing rules for free movement of personal data within the Union. The benefits of the GDPR for the EU citizens is an increase in transparency regarding the collection and use of personal information, this includes the specific purposes for the collection and use of personal information, any personal information that falls out the purpose stated is prohibited from being collected (Voigt & Bussche, 2017). Furthermore, the GDPR states that special categories of personal information are considered particularly sensitive in relation to the rights and freedoms of the individual and thus merits specific protection. Personal information disclosing the individual's ethnic origin, political opinions, religious beliefs, health information or sexual orientation and the disclosure of these personal information might lead to for example discrimination against the individual. As such the processing[26] of these categories is generally prohibited under the GDPR. However, there are exceptions where processing of these categories is allowed, in an online setting such an exception can be made if the individual gives explicitly consent to allowing the processing of these categories (Voigt & Bussche, 2017). Under the GDPR the individual must give consent to the processing of their personal information and is given the right to withdraw consent at

[26] Processing means any operation or set of operations which is performed on personal data or on sets of personal data, such as data mining for example.

any time. The withdrawal of consent means that the individual's personal information may not be further processed, any processing done before the withdrawal is not affected (Voigt & Bussche, 2017). This also means that in an online setting the GDPR divides the individuals in to two groups, one group where the individual's personal information to some degree is protected by the GDPR and another group where no such protection is provided.

References

Acar, G. E. (2014). The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, (pp. 674-689).

Alesso, H. P., & Smith, C. F. (2008). Connecting Information. In H. P. Alesso, & C. F. Smith, *Connections: Patterns of Discovery* (pp. 12 - 29). Hoboken: John Wiley & Sons, Inc.

Al-Saggaf, Y., & Islam, Z. (2015). Data Mining and Privacy of Social Network Sites's Users: Implications of the Data Mining Problem. *Science and Engineering Ethics, 21*(4), 941-966.

Altshuler, Y. (2013). *Security and Privacy in Social Networks.* New York: Springer.

Angwin, J. (2014). *Dragnet nation, a quest for privacy, security, and freedom in a world of relentless surveillance.* New York: Times Books, Henry Holt and Company.

Anthes, G. (2015). Data Brokers Are Watching You. *Communications of The ACM, 58*(1), 28-30.

Baden, R., Bender, A., Spring, N., Bhattacharjee, B., & Starin, D. (2009). Persona: An Online Social Network with User-Defined Privacy. *Computer Communication Review, 39*(4), pp. 135-146.

Bagley, A. W. (2011). Don't be evil: The Fourth Amendment in the age of Google, national security, and digital papers and effects. *Albany Law Journal of Science & Technology, 21*(1), pp. 153–191.

Banks, L. F. (2009). All Friends Are Not Created Equal: An Interaction Intensity Based Approach to Privacy in Online Social Networks. *Proceedings - 12th IEEE International Conference on Computational Science and Engineering* (pp. 970-974). Vancouver: IEEE Computer Society.

Bannister, F., & Moloney, M. (2009). A Privacy Control Theory for Online Environments. *42nd Hawaii International Conference on System Sciences*, (pp. 1-10).

Barbier, G., & Liu, H. (2011). Data Mining in Social Media. In C. C. Aggarwal, *Social Network Data Analytics,* (pp. 327 - 352). New York: Springer.

Bayan, R. (2001). Privacy Means Knowing Your Cookies. *Link-Up, 18*(1), 22-23.

Bechmann, A. (2014). Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook. *Journal of Media Business Studies, 11*(1), 21-38.

Belbin, R. (2018). When Google Becomes the Norm: The Case for Privacy and the Right to Be Forgotten. *Dalhousie Journal of Legal Studies, 26*, 17-35.

Berman, J., & Mulligan, D. (1999). The Internet and The Law: Privacy in the Digital Age: Work in Progress. *Nova Law Review, 23*, pp. 549-927.

Boda, K., Földes, Á., GulyáS, G., & Imre, S. (2012). User tracking on the web via cross-browser fingerprinting. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7161*, 31-46.

Boyd, D., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230.

Brock, G. (2016). Google Spain. In G. Brock, *The Right to Be Forgotten: Privacy and the Media in the Digital Age* (pp. 38- 44). London: I. B. Tauris & Company, Limited.

Broder, A. J. (2000). Data Mining, the Internet and Privacy. *Web Usage Analysis and User Profiling at The International WEBKDD'99 Workshop* (pp. 56-73). San Diego: Springer.

Browne, R. (2018, April 26). Facebook admits it did not read terms of the app that harvested data of 87 million. *CNBC*. Retrieved May 4, 2018, from CNBC: https://www.cnbc.com/2018/04/26/facebook-cto-admits-firm-didnt-read-terms-of-aleksandr-kogans-app.html

Brügger, N. (2015). A Brief History of Facebook as a Media Text: The development of an empty structure. *First Monday, 20*(4), 1-19.

Brunton, F., & Nissenbaum, H. F. (2015). *Obfuscation, a user's guide for privacy and protest.* Cambridge: The MIT Press.

Bujlow, T., Carela-Espanol, V., Beom-Ryeol Lee, P., & Barlet-Ros, P. (2017). A Survey on Web Tracking: Mechanisms, Implication and Defenses. *Proceedings of the IEEE, 105*(8), 1476-1510.

Bunz, M., & Meikle, G. (2018). *The Internet of Things.* Cambridge: Polity Press.

Burdon, M., & McKillop, A. (2013). The Google Street View Wi-Fi Scandal and its Repercussions for Privacy Regulation. *Monash University Law Review, 39*(3), 702-737.

Butterfield, A. (2016). *A Dictionary of Computer Science.* Oxford: Oxford University Press.

Castelluccia, C., & Narayanan, A. (2012). *Privacy considerations of online behavioural tracking.* Athens: The European Network and Information Security Agency (ENISA). Retrieved March 9, 2018, from https://www.enisa.europa.eu/publications/privacy-considerations-of-online-behavioural-tracking

Christl, W., & Spiekermann, S. (2016). *Networks of Control: A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy.* Wien: Facultas.

Crowther, B. T. (2012). (Un)Reasonable Expectation of Digital Privacy. *Brigham Young University Law Review*, 343-369.

Cutillo, L. A., & Molva, R. (2009). Safebook: A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust. *IEEE Communications Magazine, 12*, pp. 94-101.

Dhillon, G., & Moores, T. (2001). Internet Privacy: Interpreting key issues. *Information Resources Management Journal, 14*(4), pp. 33-37.

Dixon, P., & Gellman, R. (2011). *Online Privacy: A Reference Handbook.* Santa Barbara: ABC-CLIO.

eMarketer. (2017, July). *Number of social network users worldwide from 2010 to 2021 (in billions).* Retrieved from https://www.statista.com: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Erhvervsstyrelsen. (2013, April). *Guidelines on Executive Order on Information and Consent Required in Case of Storing and Accessing Information in End-User Terminal Equipment ("Cookie Order").* Retrieved January 5, 2018, from https://erhvervsstyrelsen.dk:

https://erhvervsstyrelsen.dk/sites/default/files/media/engelsk-vejledning-cookiebekendtgorelse.pdf

Etzioni, A. (1999). *The limits of privacy.* New York: Basic Books.

Etzioni, A. (2015). *Privacy in a Cyber Age: Policy and Practice.* New York, NY: Palgrave Macmillan.

Facebook. (2018). *Data Policy.* Retrieved May 22, 2018, from Facebook: https://www.facebook.com/privacy/explanation

Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology, 1*(1), 37-56.

Floridi, L. (2005). The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology, 7*(4), 185-200.

Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology, 8*(3), pp. 109–119.

Floridi, L. (2008). Information Ethics: A Reappraisal. *Ethics and Information Technology, 10*(2), 189-204.

Floridi, L. (2014). *The ethics of information.* Oxford: Oxford University Press.

Freeman, C. (2007). The ICT Paradigm. In R. Mansell, C. Avgerou, D. Quah, & R. Silverstone, *The Oxford Handbook of Information and Communication Technologies* (pp. 34-55). Oxford: Oxford University Press.

Fulda, J. S. (2000). Data Mining and Privacy. *Albany Law Journal of Science & Technology, 11*, 105-359.

Gan, D., & Jenkins, L. R. (2015). Social Networking Privacy: Who's Stalking You? *Future Internet, 7*, 67-93.

Ghosh, D., & Scott, B. (2018, March 19). Facebook's New Controversy Shows How Easily Online Political Ads Can Manipulate You. *Time*. Retrieved 23 April 2018, from http://time.com/5197255/facebook-cambridge-analytica-donald-trump-a/

Goldsborough, R. (2010). Are You Protecting Your Privacy Online? *Teacher Librarian, 5*, pp. 72-88.

Google. (2017, December 18). *Privacy Policy*. Retrieved January 18, 2018, from Google: https://policies.google.com/privacy?hl=en&gl=dk

Google. (2018). *Manage Device Information setting*. Retrieved March 14, 2018, from Google: https://support.google.com/accounts/answer/6135999?p=account_device_info&hl=en&authuser=0&visit_id=1-636684692419730707-2842735672&rd=1

Google. (2018). *Manage Google Voice & Audio Activity.* Retrieved February 12, 2018, from Google: https://support.google.com/websearch/answer/6030020?authuser=1&p=account_voice_audio&authuser=1&visit_id=1-636684831682259512-3218928471&rd=1

Google. (2018, 25 May). *Privacy Policy.* Retrieved May 26, 2018, from Google: https://policies.google.com/privacy#footnote-personal-info

Google. (2018). *See and control your search activity*. Retrieved March 12, 2018, from Google: https://support.google.com/websearch/answer/54068?p=web_app_activity&visit_id=1-636684609929478864-2660099&rd=1

Granville, K. (2018, March 19). Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. *The New York Times*. Retrieved April 21, 2018, from The New York Times: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

Guha, S., Tang, K., & Francis, P. (2008). NOYB: privacy in online social networks. *Proceedings of the First Workshop on Online Social Networks* (pp. 49-54). Seattle: ACM.

Hasselbalch, G., & Tranberg, P. (2016). *Dataetik - Den nye konkurrencefordel.* Valby: PubliShare.

Hinduja, S. (2004). Theory and Policy in Online Privacy. *Knowledge, Technology, & Policy, 17*(1), 38-58.

Introna, L. D. (2007). Making Sense of ICT, New Media and Ethics. In R. Mansell, C. Avgerou, D. Quah, & R. Silverstone, *The Oxford Handbook of Information and Communication Technologies* (pp. 314 - 339). Oxford: Oxford University Press.

Jenkins, B. (2013). Keeping up with Zuck: a brief history of Facebook features. *Techniques, 88*(8), 60-61.

Jerome, J. (2015). Big data: catalyst for a privacy conversation. *Indiana Law Review, 48*(1), pp. 213–242.

Jørgensen, R. F., & Desai, T. (2017). Right to Privacy Meets Online Platforms: Exploring Privacy Complaints against Facebook and Google. *Nordic Journal of Human Rights, 35*(2), 106-126.

Jørgensen, R. F., & Desai, T. (2017). Right to Privacy Meets Online Platforms: Exploring Privacy Complaints against Facebook and Google. *Nordic Journal of Human Rights, 35*(2), 106-126.

Kowalski, M. (2013). Oversight in the era of 'Snowden' and big data: Challenges and opportunities. *Security and Human Rights, 24*, 225-226.

Larose, R., Rifon, N., & Enbody, R. (2008). Promoting Personal Responsibility for Internet Safety. *Communications of the ACM, 51*(3), pp. 71-76.

Le, H., Fallace, F., & Barlet-Ros, P. (2017). Towards accurate detection of obfuscated web tracking. *IEEE International Workshop on Measurement and Networking, Sept. 2017*, (pp. 1-6).

Lee, N. (2013). *Facebook Nation: Total Information Awareness.* London: Springer.

Liebert, T. (2015a). Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. *International Journal of Communication, 9*(1), 3544-3561.

Liebert, T. (2015b). Privacy Implications of Health Information Seeking on the Web. *Communications of the ACM, 58*(3), 68-77.

Lin, J. (2005). Technology and Web User Data Privacy: A Survey of Risks and Countermeasures. *IEEE Security & Privacy, 3*(1), 52-58.

Lipton, J. D. (2010). Mapping online privacy. *Northwestern University Law Review, 104*(2), pp. 477–515.

Luzak, J. A. (2014). Privacy Notice for Dummies? Towards European Guidelines on How to Give "Clear and Comprehensive Information" on the Cookies' Use in Order to Protect the Internet Users' Right to Online Privacy. *J Consum Policy, 37*, 547-559.

Lyon, D. (2010). Surveillance, Power and Everyday Life. In P. Kalantzis-Cope, & K. Gherab-Martín, *Emerging Digital Spaces in Contemporary Society* (pp. 107-120). London: Palgrave Macmillan.

Mansell, R., Avgerou, C., Quah, D., & Silverstone, R. (2007). *The Oxford Handbook of Information and Communication Technolgoies.* Oxford: Oxford University Press.

Margulis, S. T. (2003). On the Status and Contribution of Westin's and Altman's Theories of Privacy. *Journal of Social Issues, 59*(2), pp. 411-429.

Martin, K. (2016a). Data aggregators, consumer data, and responsibility online: Who is tracking consumers online and should they stop? *The Information Society, 32*(1), 51-63.

Martin, K. (2016b). Understanding Privacy Online: Development of a Social Contract Approach to Privacy. *Journal of Business Ethics, 137*(3), 511-569.

Masand, B., & Spiliopoulou, M. (1999). Data Mining for the Web. *Web Usage Analysis and User Profiling* (pp. 1-6). San Diego: Springer.

McCormick, J. P. (2012, October 16). Facebook accidentally outed users to their parents through group permissions loophole. *Pink News*. Retrieved May 3, 2018, from https://www.pinknews.co.uk/2012/10/16/facebook-accidentally-outed-users-to-their-parents-through-group-permissions-loophole/

Miyazaki, A. D. (2008). Online Privacy and the Disclosure of Cookie Use: Effects on Consumer Trust and Anticipated Patronage. *American Marketing Association, 27*(1), 19-33.

Moor, J. H. (1997). Towards a theory of privacy in the information age. *ACM SIGCAS Computers and Society, 27*(3), pp. 27-32.

Moor, J. H. (2007). Towards Informational Privacy Rights. *San Diego Law Review, 44*, 809-845.

Nikiforakis, N., & Acar, G. (2014). Browse at Your Own Risk. *IEEE Spectrum, 51*(8), 30-35.

Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., & Vigna, G. (2013). Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting. *Proceedings - IEEE Symposium on Security and Privacy*, (pp. 541-555).

Nissenbaum, H. (2010). *Privacy in context, technology, policy, and the integrity of social life.* Standford, California: Standford University Press.

Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus, 140*(4), pp. 32-48.

Parent, W. A. (1983). Privacy, Morality, and the Law. *Philosophy & Public Affairs, 12*(4), 269-288.

Peng, W., & Jennifer Cisna. (2000). HTTP cookies – a promising technology. *Online Information Review, 24*(2), 150-153.

Pierson, J., & Heyman, R. (2011). Social Media and Cookies: Challenges for Online Privacy. *Info, 13*(6), 30-42.

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the Horizon, 9*(5), 1-6.

Puglisi, S., Rebollo-Monedero, D., & Forné, J. (2017). On Web User Tracking: How Third-Party Http Requests Track Users' Browsing Patterns for Personalised Advertising. *International Journal of Parallel Emergent and Distributed Systems, 32*(5), 502-521.

Questions and Answers - Data protection reform. (2015, December 21). Retrieved April 2, 2018, from European Commission: http://europa.eu/rapid/press-release_MEMO-15-6385_en.htm

Raab, C., & Mason, D. (2004). Privacy, Surveillance, Trust and Regulation. *Information, Communication & Society, 7*(1), pp. 89-91.

Rachels, J. (1975). Why Privacy Is Important. *Philosophy & Public Affairs, 4*(4), pp. 323-333.

Rakower, L. H. (2011). Blurred Line: Zooming in on Google Street View and the Global Right to Privacy. *Brooklyn Journal of International Law, 37*(1), 317- 347.

Ramlakhan, N. E. (2011). Ethical Implications of Third-party Cookies. *The International Journal of the Humanities, 9*(1), 59-68.

Reidenberg, J. R., Bhatia, J., Breaux, T. D., & Norton, T. B. (2016). Ambiguity in Privacy Policies and the Impact of Regulation. *Journal of Legal Studies, 45*(2), 163-190.

Rieke, A., Yu, H., Robison, D., & Hoboken, J. v. (2016). *Data Brokers in a Open Society.* London: Open Society Foundation.

Ring, T. (2015). Keeping tabs on Tracking Technology. *Network Security, 6*, 5-8.

Robertson, A. (2017, May 10). *Google just acquired on of the most successful VR game studios*. Retrieved March 8, 2018, from The Verge: https://www.theverge.com/2017/5/10/15614274/google-daydream-vr-owlchemy-labs-acquisition-job-simulator

Rotenberg, M., Scott, J., & Horwitz, J. (2015). *Privacy in the Modern Age, The Search for Solutions.* New York: The New Press.

Schinasi, J. (2014). Practicing privacy online: Examining data protection regulations through Google's global expansion. *Columbia Journal of Transnational Law, 52*(2), pp. 569-616.

Schneier, B. (2015). *Data and Goliath, the hidden battles to collect your data and control your world.* New York: W.W. Norton & Company.

Seshagiri, A. (2013, October 1). Claims That Google Violates Gmail User Privacy. *The New York Times*.

Shafique, U., Majeed, F., Qaiser, H., & Ul Mustafa, I. (2015). Data Mining in Healthcare for Heart Diseases. *International Journal of Innovation and Applied Studies, 10*(4), 1313-1322.

Sipior, J. C., Ward, B. T., & Mendoza, R. A. (2011). Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons. *Journal of Internet Commerce, 10*(1), 1-16.

Solove, D. (2011). *Nothing to hide, the false tradeoff between privacy and security.* New Haven: Yale University Press.

Solove, D. J. (2004). *The Digital Person: Technology and Privacy in The Information Age.* New York: New York University Press.

Sumner, S., & Rispoli, M. (2016). *You: For sale, protecting your personal data and privacy online.* Amsterdam: Syngress.

Symeonidis, I., Biczók, G., Shirazi, F., Pérez-Solàc, C., Schroers, J., & Preneel, B. (2018). Collateral damage of Facebook third-party applications: a comprehensive study. *Computers & Security, 77*, 179-208.

Tavani, H. T. (1999). Informational Privacy, Data Mining and the Internet. *Ethics and Information Technology, 1*, 137-145.

Tavani, H. T. (2007). Philosophical Theories of Privacy: Implications for An Adequate Online Privacy Policy. *Metaphilosophy, 38*(1), pp. 1-22.

Tavani, H. T. (2008). Floridi's ontological theory of informational privacy: Some implications. *Ethics and Information Technology, 10*(2), pp. 155–166.

Tavani, H. T. (2008). Informational Privacy: Concepts, Theories and Controversies. In *The Handbook of Information and Computer Ethics* (pp. 131-164). Hoboken, New Jersey: John Wiley & Sons, Inc.

Tavani, H. T., & Moor, J. H. (2001). Privacy Protection, Control of Information, and Privacy-Enhancing Technologies. *ACM SIGCAS Computers and Society, 31*(1), 6-11.

Thomson, J. J. (1975). The right to privacy. *Philosophy & Public Affairs, 4*(4), pp. 295-314.

Tucker, C. (2013). Three findings regarding privacy online. *Proceedings of the sixth ACM international conference on web search and data mining*, (pp. 243-244). Rome.

Upathilake, R., Lin, Y., & Matrawy, A. (2015). A classification of web browser fingerprinting techniques. *7th International Conference on New Technologies, Mobility and Security*, (pp. 1-5).

Veer, E. V. (2011). *Facebook: The Missing Manual.* Sebastopol: O'Reilly Media, Inc.

Voigt, P., & Bussche, A. v. (2017). *The EU General Data Protection Regulation: A pratical guide.* Cham, Switzerland: Springer.

Walker, K. (2000). Everybody Knows Your Name: A Pragmatic Look at the Costs of Privacy and the Benefits of Information. *Stanford Technology Law Review*, 2-50.

Wallace, K. (1999). Anonymity. *Ethics and Information Technology, 1*(1), pp. 21-35.

Warren, S., & Brandeis, L. (1890). The Right to Privacy. *Harvard Law Review, 4*(5), 193-220.

Wellman, B., & Haythornthwaite, C. (2002). The Internet in Everyday Life: An Introduction. In B. Wellman, & C. Haythornthwaite, *The Internet in Everyday Life* (pp. 3-45). Malden, MA: Blackwell Publishers Ltd.

Westin, A. F. (2003). Social and Political Dimensions of Privacy. *Journal of Social Issues, 59*(2), pp. 431-453.

Witte, D. (2014). Privacy deleted: Is it too late to protect our privacy online. *Journal of Internet Law, 17*(1), pp. 1-28.

Young, A. L., & Quan-Haase, A. (2013). Privacy Protection Strategies on Facebook. *Information, Communication & Society, 16*(4), 479-500.

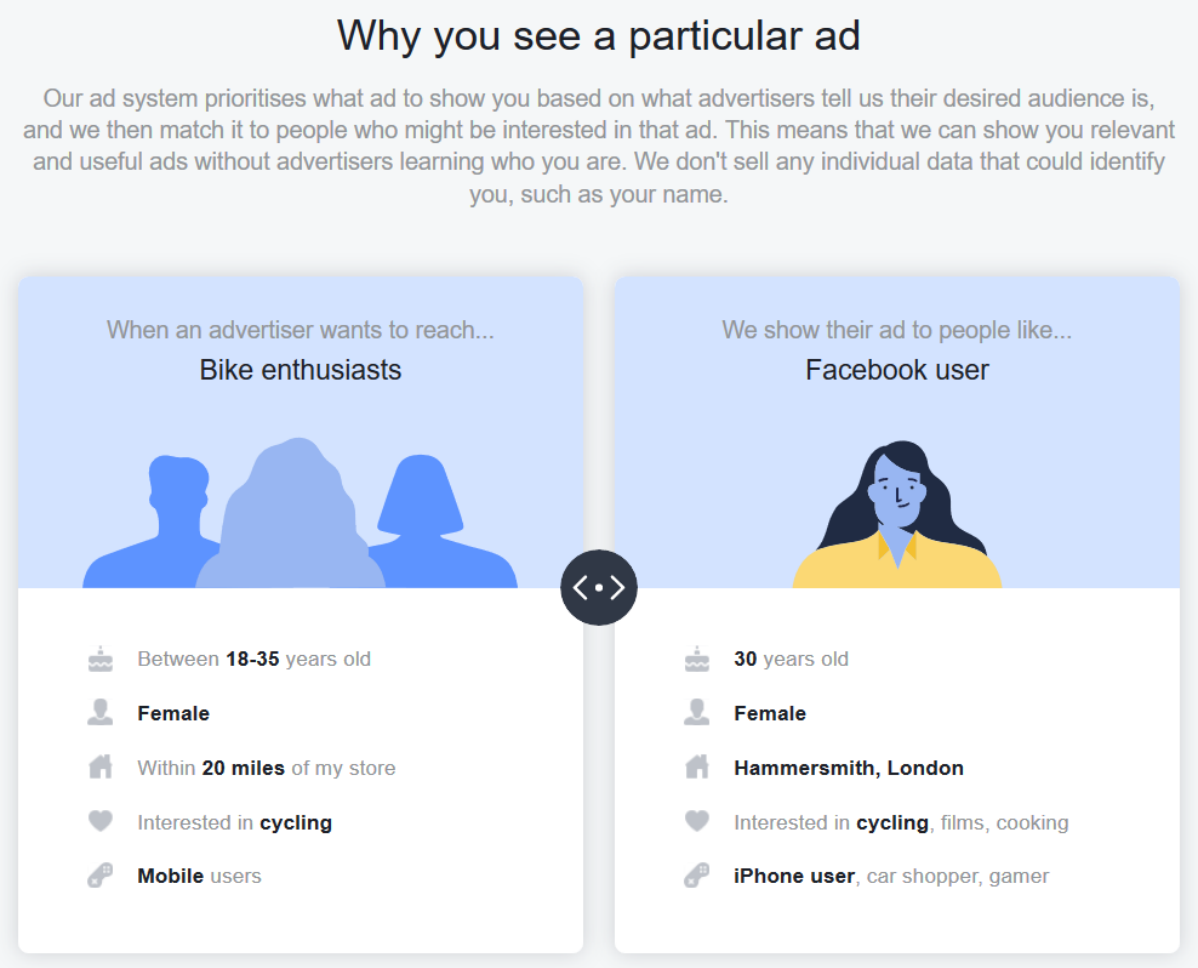
Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social Media Mining: An Introduction.* New York: Cambridge University Press.

Zarsky, T. Z. (2002). 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 and Technology, 5*, 1-141.

Appendix 1 Screenshots from personal Facebook profile

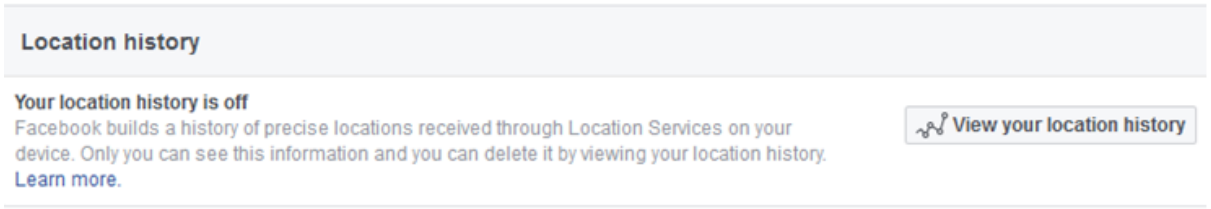
Screenshot no. 1 Why you see a particular ad

Screenshot taken from https://www.facebook.com/about/ads (28/05-2018)



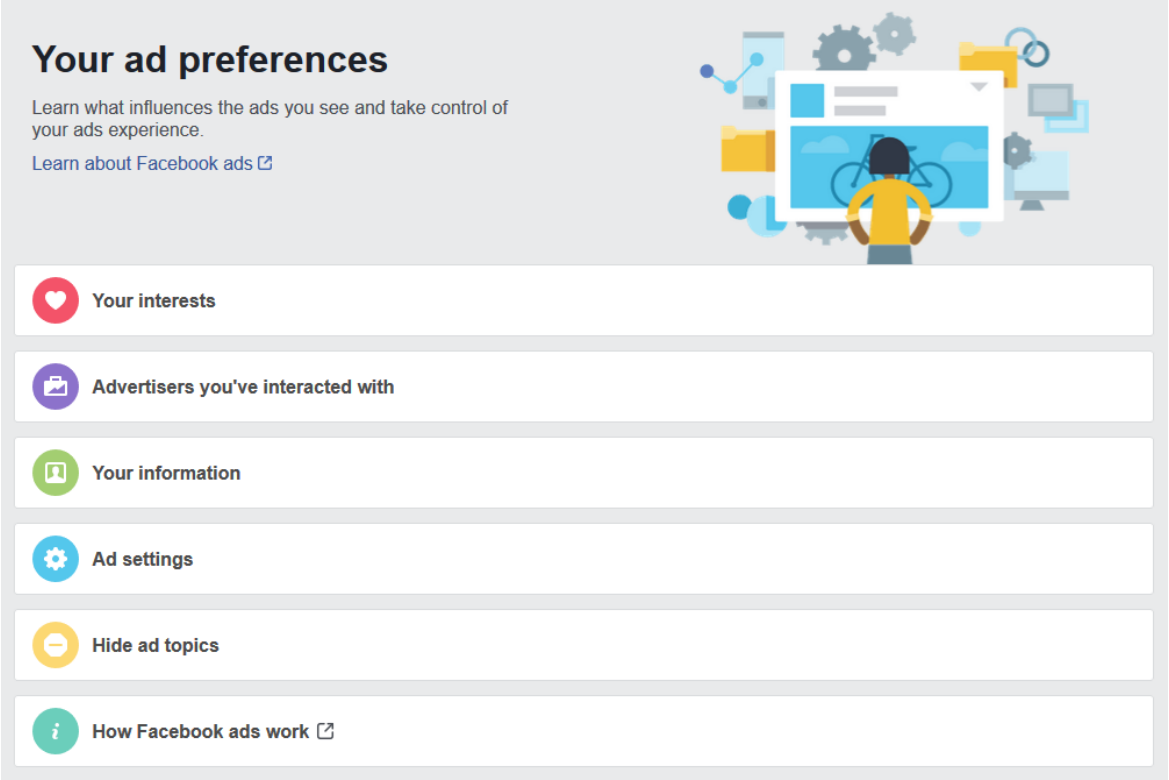
Screenshot no. 2 Location History

Screenshot from private Facebook profile (02/06-2018).



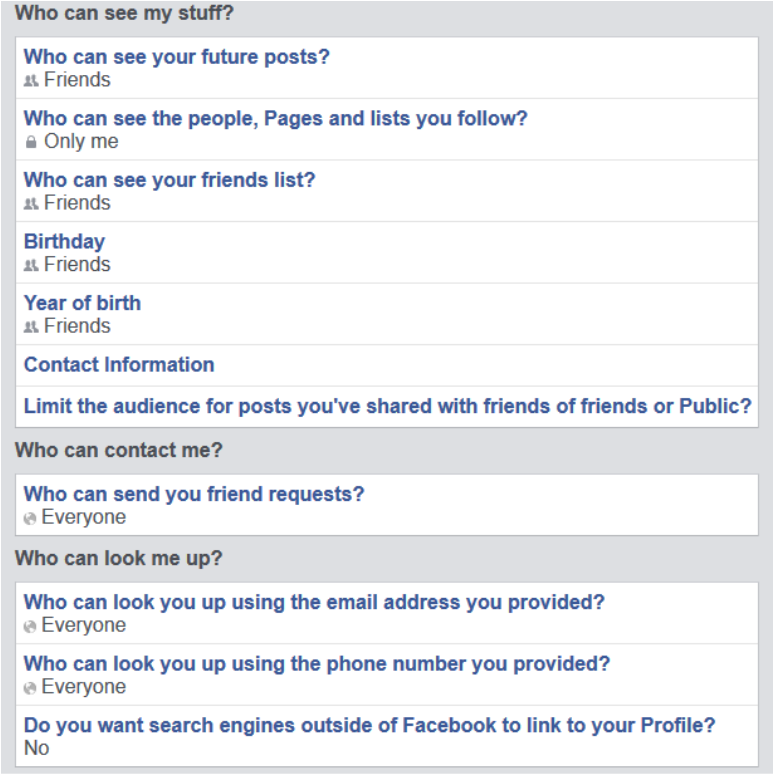
Screenshot no. 3 Your ad preferences

Screenshot from private Facebook profile (02/06-2018).



Screenshot no. 4 Privacy settings

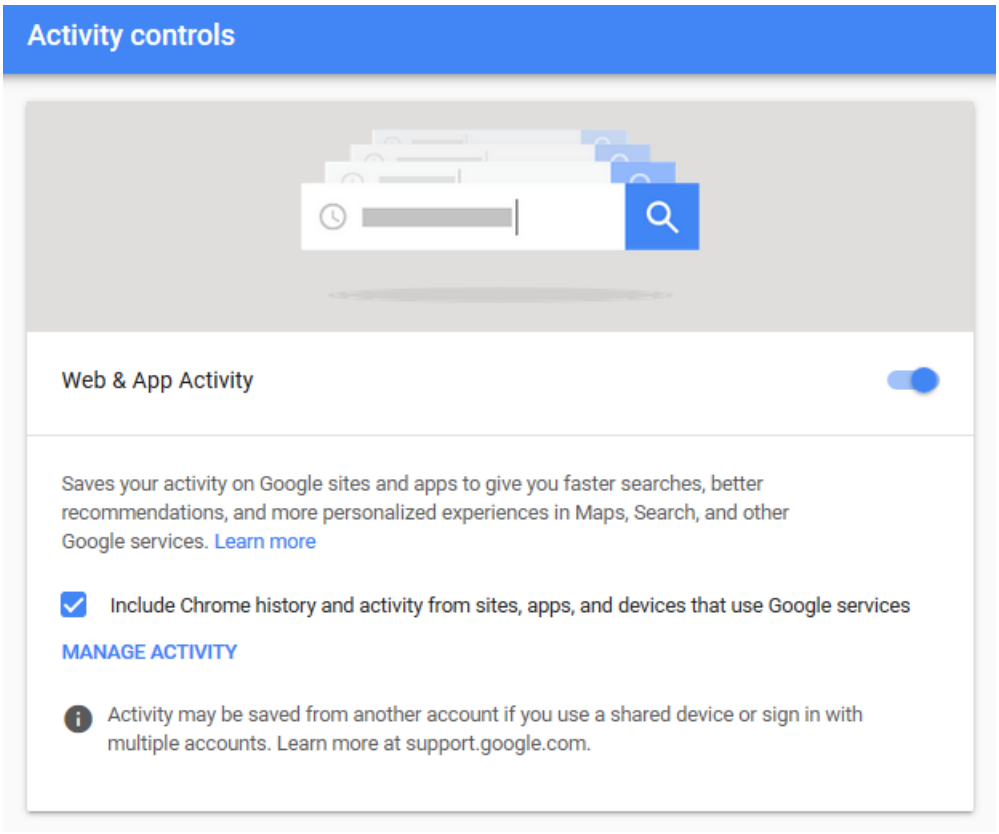
Screenshot from private Facebook profile (28/05-2018).



Appendix 2 Screenshots from personal Google account

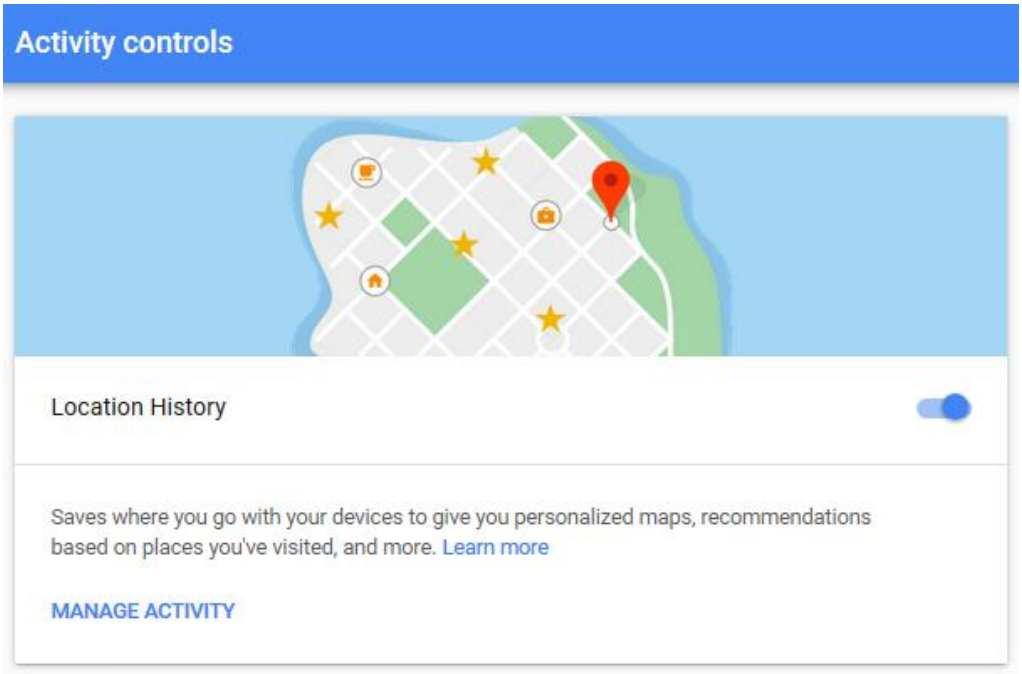
Screenshot no. 1 Web and App Activity

Screenshot taken from personal Google account (24/03-2018)



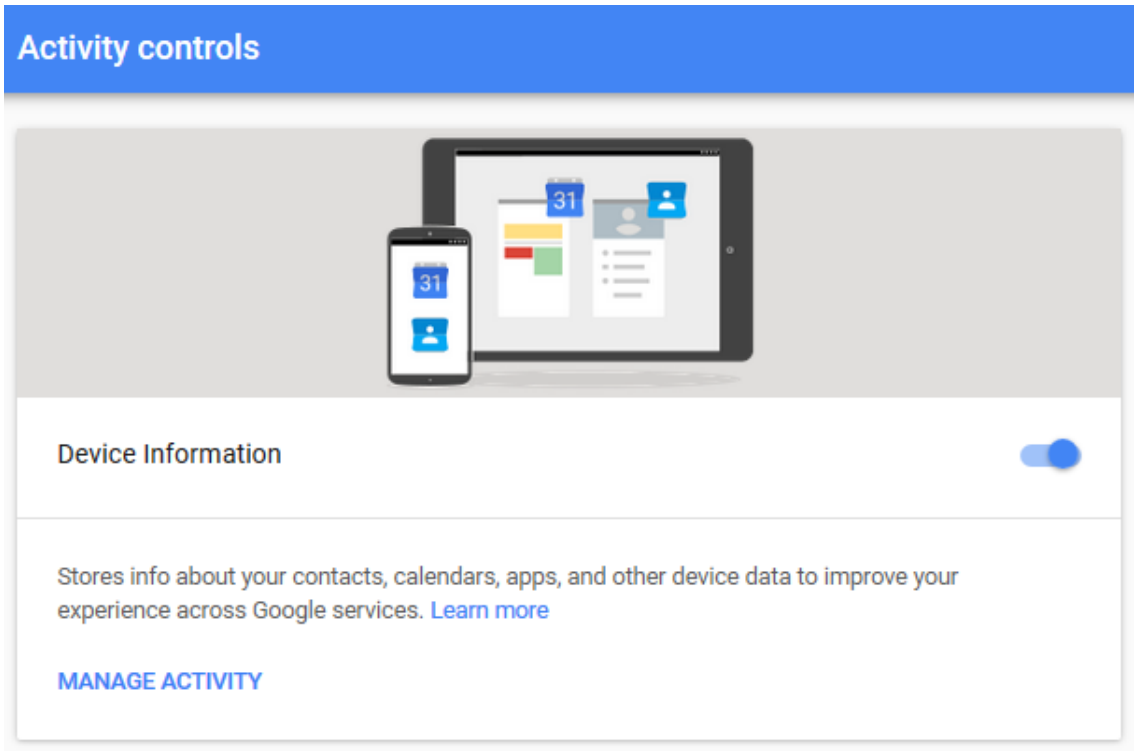
Screenshot no. 2 Location History

Screenshot taken from personal Google account (24/03-2018)



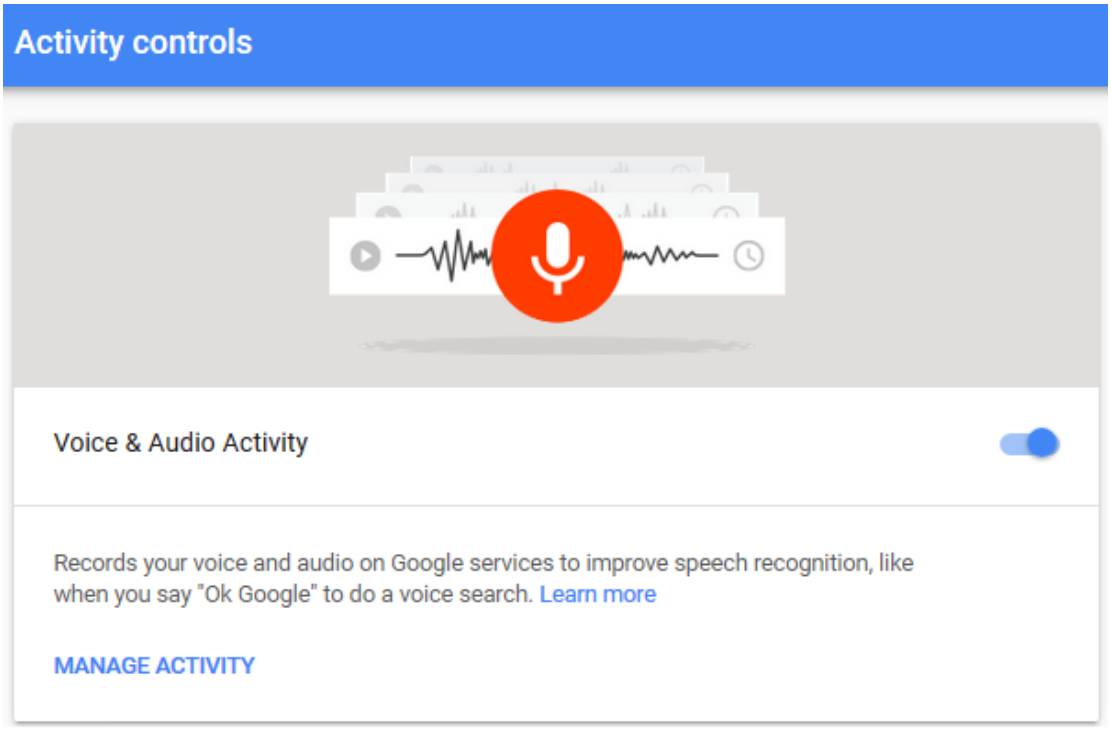
Screenshot no. 3 Device Information

Screenshot taken from personal Google account (24/03-2018)



Screenshot no. 4 Voice and Audio Activity

Screenshot taken from personal Google account (24/03-2018)



Screenshot no. 5 YouTube Search and Watch History

Screenshot taken from personal Google account (24/03-2018)