TECHNICKÁ UNIVERZITA V LIBERCI
Ekonomická fakulta

Využití digitálních dat v podnikání

Bakalářská práce

Studijní program: B6209 – Systémové inženýrství a informatika
Studijní obor: 6209R021 – Manažerská informatika

Autor práce: **Martin Havlišta**
Vedoucí práce: doc. Ing. Klára Antlová, Ph.D.

Liberec 2015

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

Digital data usage in a business environment

Bachelor thesis

Study programme: B6209 – System Engineering and Informatics
Study branch: 6209R021 – Managerial Informatics

Author: **Martin Havlišta**
Supervisor: doc. Ing. Klára Antlová, Ph.D.

Liberec 2015

TECHNICKÁ UNIVERZITA V LIBERCI
Ekonomická fakulta
Akademický rok: **2014/2015**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Martin Havlišta**
Osobní číslo:	**E12000728**
Studijní program:	**B6209 Systémové inženýrství a informatika**
Studijní obor:	**Manažerská informatika**
Název tématu:	**Využití digitálních dat v podnikání**
Zadávající katedra:	**Katedra informatiky**

Zásady pro vypracování:

1. Digitální obchodní data - jejich význam a použití
2. Popis výhod a nevýhod využívání digitálních obchodních dat
3. Analýza povědomí spotřebitelů o využívání jejich digitálních dat pro komerční účely
4. Vyhodnoceni míry znalostí spotřebitelů o způsobu nakládání firem s digitálními daty

Rozsah grafických prací:

Rozsah pracovní zprávy: **30 normostran**

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

PROVOST, Foster and Tom FAWCETT. Data Science for Business. 1st ed. Beijing: O'REILLY, 2013. ISBN 1449361323.
STRAUSS, Judy, Adel EI ANSARY and Raymond FROST. E-Marketing. 4th ed. Upper Saddle River, N. J.: PEARSON/PRENTICE HALL, 2006. ISBN 0131969021.
WALDO, James, Herbert LIN and Lynette I. MILLETT. Engaging Privacy and Information Technology in a Digital Age. 1st ed. Washington, D. C.: NATIONAL ACADEMIES PRESS, 2007. ISBN 0309103924.
WINDLEY, Phillip J. Digital Identity. 1st ed. Beijing; Farnham: O'REILLY, 2005. ISBN 0596008783.
Elektronická databáze článků ProQuest (knihovna.tul.cz).

Vedoucí bakalářské práce: **doc. Ing. Klára Antlová, Ph.D.**
Katedra informatiky

Konzultant bakalářské práce: **Ing. Petr Rozmajzl**
Katedra informatiky

Datum zadání bakalářské práce: **31. října 2014**

Termín odevzdání bakalářské práce: **7. května 2015**

doc. Ing. Miroslav Žižka, Ph.D.
děkan

Technická univerzita v Liberci
Ekonomická fakulta
L.S.

doc. Ing. Jan Skrbek, Dr.
vedoucí katedry

V Liberci dne 31. října 2014

Prohlášení

Byl jsem seznámen s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu TUL.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Bakalářskou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Anotace

Mezi obchodními společnostmi a jejich zákazníky se občas vyskytují vztahy, které nejsou vždy podle etických pravidel a to zvláště v souvislosti s vlastnictvím digitálních dat nebo s ochranou soukromí spotřebitelů. V dnešní době firmy usilují o zákazníky pomocí různých nástrojů, včetně nejnovějších technologií jako například big data. Ochrana osobních digitálních údajů je dnes velmi diskutované téma, které má jistý dopad jak na společnost, v podobě důvěry v e-komerci, tak i na fungování ekonomiky jako takové. Tato bakalářská práce zkoumá podstatu obav o digitální soukromí a také studuje obecné mechanismy spojené s ochranou soukromých dat spotřebitelů a speciálně se zaměřuje na oblast označovanou jako privacy paradox. Studie byla provedena pomocí on-line dotazníku, kde se potvrdilo, že lidé mají obecně obavy o vlastní soukromí, kde hlavním důvodem všech obav je prodávání osobních informací třetím firmám.

Jako většina studií, i tato práce potvrzuje existenci privacy paradoxu, ačkoli nevyvozuje závěry, že spotřebitelé jsou ochotni zveřejnit své osobní údaje výměnou za služby či produkty zdarma.

Odkazy na původní zdroje byly v této bakalářské práci vytvořeny podle stylu APA 6.

Klíčová slova

big data, digitální důvěra, digitální soukromí, obavy o soukromí, privacy paradox

Annotation

Relationships between companies and customers are not always completely ethical, especially when considering digital data ownership and privacy matters. Corporations would like to leverage the potential from customers using advanced analytical methods and techniques such as big data. Digital data privacy is quite discussed topic nowadays that has an impact on society in terms of trust in e-commerce and also has some implications into the future of economy. This dissertation explores digital privacy concerns and privacy in general terms inside the society in particular a privacy paradox. Regarding the research method, self-distributed online questionnaire was used in this research concluding that respondents are concerned about digital data privacy and the main cause for these concerns was selling personal information to third party companies.

As a quite big number of researchers claimed that there exists a privacy paradox, this study acknowledges the fact that the privacy paradox exists although the study does not suggest that customers are willing to exchange their personal information for goods.

This undergraduate dissertation was referenced according to APA 6th.

Key words

big data, digital privacy, digital trust, privacy concerns, privacy paradox

Acknowledgements

Firstly I would like to thank my parents, Jitka and Daniel, for their enormous amount of support they have always had for me, because without them this process would not be possible. Then I would like to thank my supervisor David Colley for the faith he had in me and last but not least I would like to thank my study mentor Jane Crookes for her enthusiasm and drive to set me in the right direction.

Table of contents

List of charts

List of tables

List of abbreviations

CEO	Chief Executive Officer
EC	European Commission
GfK	Gesellschaft für Konsumforschung (German research institute)
MIT	Massachusetts Institute of Technology
NSA	National Security Agency
PCAST	President's Council of Advisors on Science and Technology (USA)
PRISM	Planning Tool for Resource Integration, Synchronization, and Management (surveillance program, USA)
PwC	PricewaterhouseCoopers (Professional services network)

Introduction

Digital technology is becoming more embedded into lives of majority of individuals. Sometimes it is so present that people rely on digital outcomes and believe that digital responses are true, for example traffic lights, information systems or social networks. Within the digital environment, people create and share huge amounts of data every day, for instance using credit cards or just by browsing in an e-shop, and those entries are recorded all the time (Rosen, 2010).

Early at the beginning, companies quickly recognised the opportunity and embraced the idea of analysing customer data as a necessary business activity (Saxena, 2014). Later, customers became conscious that not only functionality of technological devices is important, but also aspects of ethics should be used. Nevertheless, technology advances and intense research brought the concept of big data analytics; the technology that can collect, store, analyse and also forecast customer behaviour in almost real-time speed (McAfee & Brynjolfsson, 2012). This concept was fully demonstrated by Target, a company that used big data techniques to predict customer behaviour to the extent that they offered baby clothes to a woman several months before she was pregnant (Duhigg, 2012). Continuing on data ethics, large scale surveillance programmes were revealed in the United States (The Guardian, 2013) increasing privacy concerns to a considerable level. In this case, the main problem for individuals is that the Internet (digital environment) remembers every entry and people need to be aware of this fact and carefully choose their steps and behaviour on the Internet (Rosen, 2010).

Despite all customer concerns about methods that companies use to invade privacy it was argued that ”[people] freely submit personal information and accept being monitored, both by businesses and government” (Dinev, 2014, p.97). These totally opposite actions were researched in last years and researchers call this the privacy paradox (Blank et al., 2014; Norberg et al., 2007). Regardless of the paradox, concerns about privacy are increasing and the issue has to be addressed otherwise this problem could have an impact on economy (Mutz, 2009).

This study sought to explore privacy issues firstly because it is very recent topic; secondly because privacy concerns have an impact on most of the population and finally because the topic combines several scientific disciplines for example mathematics, computer technology or psychology.

This dissertation aims to achieve five main objectives. Firstly it focuses on the underlying technology that companies use to collect, process and use data for leveraging value from customers. This information is mainly obtained from the literature review and the purpose is to explore important processes and compare them with knowledge that people have about those practices. The second aim is to estimate the extent of awareness that people have within the digital environment using primary and secondary data. The extent is simplified to a social network involvement as a factor for determining privacy issues, which significantly relates to the next aim of exploring the privacy paradox. In this case, the purpose is to examine primary data and compare them with previous research and outline possible conditions when this situation occurs. The last aim is to establish a practical solution from the theoretical base and the research in order to find a solution for collecting and storing data safely in regard to consumers.

1. Literature review

The purpose of this literature review is to explore and identify previous research in the field of digital privacy in order to gain an understanding of main concerns and search for underlying causes. Digital privacy is nowadays quite discussed topic within society regarding both customers and companies. One of the main reasons for this is because in 2013 electronic government surveillance programs (PRISM) became public through several revelations from an intelligence analyst (The Guardian, 2013).

This review will assess key concepts of society related to privacy such as, transparency, anonymity, control or protection and the review will also explore each aspect. In this sense, privacy concerns are often connected with a question of a control whereas privacy paradox exists due to lack of transparency. Moreover, historical or technological background that increased privacy concerns will also be explained in relation to big data concept.

1.1 Digital society

"We are in information society, [where] information [is] the basic driving force for social development, with information technology" (Zhen & Jilan, 2011, p.663). In other words, information powered by technological advances is getting into the centre of our social interactions and also influences other areas of interest such as economics or culture.

> *[Information society is a] concept that responds to the expansion and ubiquity of information. The term has been in use since the 1970s, but has gained in popularity and is now widely used […]. Sustained and accelerated growth of media, of education provision and participation, as well as computer communications technologies has led many to posit that the attendant information explosion distinguishes a new epoch. The information society is one in which information is the defining feature (Information society, 2011).*

With the concept of information as a central and connecting aspect of a society there have been society shifts and Solove (2006) says that our lives are changing from physical dimensions into more digital views of the world.

> *In a sense, the digital environment creates a more pure version of the physical world; pure in that everything in the digital world is information, and there are no physical boundaries to limit and shape how and when information is obtained (Kerr, 2005).*

Kerr's interpretation is valid for both legal procedures and also for every-day practise. Chayko (2014) says that "digital (online) and face-to-face (offline) spaces become fully integrated and experienced as a single, enmeshed reality" (Chayko, 2014, p.976).

As the environments sometimes collide together in a person's mind, there are still certain differences especially in terms of technology. Solove (2006) says that information about individuals in digital form are easily stored than physical ones and that contemporary technology allow us to track records of every individual. Even whole human generations could be stored in, what he calls, 'digital dossiers', which is "a collection of detailed data about an individual" (Solove, 2006, p.1). Moreover, Mathiesen (2013) points out that digital dossier has to be physically stored in a specific place, which suggests that those dossiers (data) could be possibly stolen or revealed to different groups of people in different sizes. He claims that "information systems are also surveillance systems" (Mathiesen, 2013, p.17) and their purpose is to gather and track information about individuals.

As a response for huge surveillance revelations in year 2013 (The Guardian, 2013), American government released a report (PCAST report, 2014) considering technology perspectives about big data and privacy. According to the report, a digital environment should "promote the free flow of information" (PCAST report, 2014, p.1) and use particular information only to determine the identity of an individual. The report also states elements of a fair communication between a customer and a company. The most discussed aspects are control and transparency, which are both necessary factors included in the underlying mechanism of customer-company interaction – trust.

"One of the most discussed and worried-about aspects of today's information age is the subject of privacy" (Waldo et al., 2007). Although there is no doubt about it, some researchers still incorrectly use the terms 'privacy' and 'anonymity' interchangeably. Yanes (2014, p.1) says that "anonymity is intrinsically present in the concept of privacy",

however refers to different matter. For example Skopek (2013) explains the difference between privacy and anonymity.

> *[U]nder the condition of privacy, we have knowledge of a person's identity, but not of an associated personal fact, whereas under the condition of anonymity, we have knowledge of a personal fact, but not of the associated person's identity. In this sense, privacy and anonymity are flip sides of each other. And for this reason, they can often function in opposite ways: whereas privacy often hides facts about someone whose identity is known by removing information and other goods associated with the person from public circulation, anonymity often hides the identity of someone about whom facts are known (Skopek, 2014, p.1755).*

Privacy is a quite complicated topic because different people perceive it differently and privacy is also a complex topic which creates a space for a possible misinterpretation. According to the TRUSTe (2014) index 92 % of U.S. customers are concerned about their online privacy. The concerns are raised because information became important not only for individuals but even for companies. Businesses focus on information about former and potential customers by analysing their behaviour and predicting future choices (Stein, 2011). A technology technique used by businesses called data mining became a standard for companies in order to leverage customers' potential (Saxena, 2014) and ultimately there are companies such as Acxiom, which is "one of the largest data-brokering firms in the world [which] recorded $1.1 billion in sales last year offering 'analytical services' on 144 million households [in USA]" (Morris & Lavandera, 2012).

Accessing competitive advantage is not purely about technology, firms are trying to get the most from their customers using several psychological and sociological techniques. One of the techniques mostly discussed in relation to privacy is an idea of exchanging free products or services for personal information. This phenomenon is widely known as a privacy paradox (see chapter Privacy paradox) and the whole problem is nicely summarised by Rob Livingstone, University of Technology in Sydney, who says that "[w]e have become the product" (Porter, 2014).

In the digital age, companies would like to maintain their competitive advantage by knowing more about their customers using the latest technology whereas at the same time on the other side customers would like to receive fair treatment from the side of a company. In order to satisfy the need from the side of a company and also assure a customer, the exchange of a product/service and a personal information has to be based on a "good faith, and honesty of another party, with respect to a transaction that involves some risk" (Windley, 2005, p. 16). It is because economy stands on a simple principle of a trust of each party involved in a particular transaction. This principle is amplified especially within the digital environment where the transactions are not exactly clear considering their origins or hidden mechanisms inside and in this case transparency is highly discussed as prevention (Kelton et al., 2008). This issue significantly influences an economy, because when people realise what they give out and how they are treated they could stop to believe and stop buying products or services from companies, therefore, companies will not profit and will not survive. The economy without trust could not exist (Mutz, 2009).

1.2 Technology

Digital technology is getting more powerful and complex. For example Moore's law says that a "processing power for computers will double every two years" (Moore's Law, 2015) and, moreover, another very important technological prediction, Kryder's law, which seems to correlate with Moore's law, argues that a data storage of hard drives is expanding but on a smaller area with decreasing tendency of cost for each bit of data (Walter, 2005). In other words, there is more storage space inside smaller devices which become more powerful and cheaper every year. In summary, the future of technology is in huge data stores, which are smaller and cheaper than before.

Considering the previous point, there were already enormous data sets to store in 2012 and as McAfee & Brynjolfsson (2012, p.62) point out, "about 2.5 exabytes of data are created each day and that number is doubling every 40 month or so". The data are from videos, health monitoring gadgets, bank transactions and many similar digital communicating channels. For the most part, 95 % of this – is unstructured information. Today it is possible

to collect, store, analyse and use most of this unstructured huge sets of data. A process or technique that emerged in last few years is commonly known as Big Data. However, big data has not emerged from nowhere, it can be dated to mainframe era and beginnings of data warehousing in mid-1990s, where a simple data collection has its roots (Poulin, 2014). The difference is that todays' technology is more encompassing. In the past, a lot of big companies possessed large databases but with no analytical use or predicting mechanisms and according to Minelli et al. (2012) the break point was when big corporations started to cooperate and sell different kinds of data to other big companies in order to create immense databases.

With technology and new analytical models, employing statisticians and mathematicians, "[t]he era of Big Data has begun" (Boyd & Crawford, 2011, p.1). Gordon (2013) explains that there is not a precise definition of the term 'big data', it could refer to a technology for storing data or to analytical tools that make sense of data. However, there are several major characteristics of big data. The most obvious include volume, velocity and variety. Volume is related to an amount of data that could be stored and technology that can store all data. Next, big data are processed almost in a real-time way. This fact allows managers to instantly see, with minimal delays, the results or responses from environment. Finally, variety refers to different sources of data such as internet data, tracking devices or card transactions – different data are combined together in order to identify a unique person (McAfee & Brynjolfsson, 2012; Franks, 2012). Gordon (2013) adds two more characteristics, veracity and value, because big data is not just about collecting data; the goal for organisations, companies or governments is to make sense of all data. Therefore it is necessary to extract and sort only trustworthy and accurate data. In summary, all the characteristics described are important, because gathering, loading and displaying huge amounts of data do not necessarily aid companies, and in some cases it could do more harm than just focusing on the most reliable information in small amounts (Ross et al., 2013). Sorting the data and deliberately picking the most trusted sources of data is the new era that many refer to. The value of big data comes from all five characteristics and according to Zwitter (2014), the purpose of big data analysis is to find less transparent relationships in data sets that reveal patterns of human behaviour.

Alex Pentland, MIT professor, explains that there is still a problem with a representation of the data. The analysis systems are not completely autonomous and human interaction is still required to decide and make final decision. Although sophisticated systems could analyse and reduce the complexity of a situation, human evaluation, and in some cases even intuition, is still necessary (Pentland, 2012). However, as Ross et al., (2013) point out the most important is whether a company can make a use of data, because if a company cannot manage small amounts of information then there is the reason for investment into 'big data' infrastructure, which is expensive, and rely on the hype and a promise of big profits.

1.3 Digital privacy

Privacy is a topic of a tremendous scale and relates to each individual. Laudon and Traver (2014) define that "[p]rivacy is the moral right of individuals to be left alone, free from surveillance or interference from other individuals or organisations, including the state" (Laudon and Traver, 2014, p.533). At first, it is very important to understand that privacy norms are created by humans, it is a perceived right, not a strict law. Therefore, different people could view it differently, because of their beliefs or cultural traditions. In other words, a human would like to feel safe and secure as an individual and his or hers believes shape their perception of privacy. Thus privacy is more like a comfort zone of each person and if any of unwritten rules are broken then a person feels unsecure.

A specific part of privacy is an information (digital) privacy. Laudon and Traver (2014) explain that information privacy has all the aspects of general privacy and, moreover, "includes both the claim that certain information should not be collected at all by governments or business firms, and the claim of individuals to control the use of whatever information that is collected about them" (Laudon & Traver, 2014, p.533). Camenisch (2012) says that because of the nature of the Internet, making references to the openness of the web and its complexity, transactions are often open and unsecure. He argues that privacy should be a part of a design process in development of new goods. Technology is already available and also legislations are in place with increasing awareness about privacy issues, but privacy techniques are still not initially implemented into digital interactions.

As previously discussed about the need to implement privacy mechanisms, year 2013 could serve as an example of increasing concerns and controversial revelations. In that year Edward Snowden, National Security Agency (NSA) intelligence analyst, revealed current programmes and methods that NSA used in order to get information from U. S. or international citizens (Guardian, 2013). One of the biggest revelations was that NSA monitored almost every telephone call of American customers, which could be interpreted as a surveillance of a huge scale. Other revelations disclosed secret program called PRISM, where technological companies were obliged to provide their data about their customers from their data servers to the U.S. government. Other issues include specific hacking techniques or spying other countries and their authorities (Franceschi-Bicchierai, 2014).

Looking at the Snowden case from the point of privacy, it is a question of transparency versus privacy or surveillance versus trust. In particular, a government should defend privacy; instead, majority of people see authorities and their approach as a tool for manipulation and control rather than protection. Some may argue that governments are actually looking for possible terrorists in all those databases and that authorities do it in the spirit of transparency. On the other hand most of surveillance is done without consent, because there should not be any consent for stalking. As Snowden points out, "[e]ven if you're not doing anything wrong you're being watched and recorded" (Starr & Yan, 2013). Questions about privacy has always been here, but practises are becoming more intrusive, deceiving and, very often, unknown (Rainie et al., 2013).

Microsoft, one of the biggest technological companies in the world, presents their concepts of data privacy and anonymization. They say that, “[p]rivacy concerns arise from fears that data will be used to discriminate against or embarrass individuals. When data policies are based on anonymity, transparency, and fair value for consumers, objections diminish” (Salkowitz, 2014, p.20). On top of that, White house report (PCAST report, 2014, p.2) says that, “[a]nonymity overlaps with privacy”. This is very strong statement, even if a company says that it is their priority to respect and ensure each customer, it does not consider trust as a key element. The problem is that almost a half of customers, 45 %, do not trust companies with their personal information online (TRUSTe, 2014).

1.4 Privacy concerns

> *Imagine a world in which consumers' preferences can be so precisely estimated by observing their online behavior that firms are able to anticipate consumers' needs, offering the right product at exactly the right time. Imagine the same world, but now consider that extensive knowledge of consumers' preferences also allows precise inferences about their reservation prices (the maximum price each consumer is willing to pay for a good), so that firms can charge different prices for the same product to each of their buyers and absorb the entire surplus arising from an economic transaction (Lane, 2014, p.76).*

The imaginations from Lane (2014) are not completely distant, because knowledge about the subject and technology is already available. However, reactions to first attempts to implement new processes were not quite positive and raised many serious concerns about privacy. The underlying concept behind an exchange or a transaction is trust and there are several simple attributes of trustworthiness in a transaction – "competence, positive intentions, ethics, and predictability" (Kelton et al., 2008, p.367). From these listed, ethics are most probably related to privacy issues, it is because ethics includes aspects such as honesty, integrity or fairness. The argument is that people would like to be treated with a positive attitude without any risks whereas companies need to take a risk in order to gain certain advantage and push boundaries toward new areas with including new technologies. Therefore it is no longer a secret that customers are tracked and monitored via card transactions, web search, online purchases, mobile applications or health tracking devices (Zwitter, 2014).

> *It is enough for marketers to know that 'Customer X,' part of a group of males between the ages of 18 and 24, living in a specific set of ZIP codes, with a household income between $25,000 and $40,000, with a specific set of interests (music liked and streamed, events attended, celebrities 'liked,' etc.), and a social circle that includes more than five others with similar interests, is most likely to respond to an offer delivered on his Xbox after a third successful attempt to level up, between 7:00 and 10:00 P.M., featuring a product discount of 20 percent (Salkowitz, 2014, p.18).*

One of the greatest examples of precise profiling and forecasting on customers was demonstrated by Target. This company wanted to attract more customers into their shops,

in other words persuade them to do all the shopping in their stores. In order to do this, marketers needed to find a moment when a person or a family changes their behavioural patterns. They found that this moment occurs when there are strong emotional feelings such as a birth of a child. Therefore Target started to seek patterns in data sets in order to discover who is going to have a family. The company developed quite precise model and was able to predict 'the right' moment. Finally they send a leaflet to a girl encouraging her to buy a children's clothes, however her parents were not familiar with the fact that their daughter was pregnant (Duhigg, 2012). This is a great example what companies can do with data harnessing the value in benefit for the company and gaining competitive advantage. However, it is also a great example of a privacy breach. The company legally got a consent for using personal data, but the problem is that they did not consider ethics and probably lacked positive intentions.

The previous research is crucial for comprehensive understanding of the topic. TRUSTe report (2014) says that 92 % of U. S. customers are concerned about their privacy and 89 % of all customers avoid firms that have low privacy protections. The biggest causes of concerns include selling private information to third party companies, 58 % of participants, or government surveillance, 38 % of participants. Activities that raise concerns are online banking and shopping or using social networks and mobile apps. These findings are supported by another study from GfK (2014) saying that 88 % of respondents are concerned and 80 % of participants call for regulations for stopping to sell information to third party organisations.

1.5 The privacy paradox

Nowadays many newly established companies are based on a business model that concerns only digital environment, for instance technology companies such as Google or social network companies such as Facebook. These companies are quite often in the centre of criticism for incorrect data management and incorrect practices because they operate with personal information in a way that is sometimes on the edge of an ethic code (Dwyer, 2011). In particular, it is free to register for using Facebook, thus it could seem that the service is also free. However, the company profits from the input from customers because

customers do not exchange any tangible goods, they exchange their privacy for the service. The problem is that networking sites are already deeply embedded into everyday lives of individuals and people could perceive them as something natural not something computerized, therefore people might be less cautious about the content they share (The Economist, 2010).

Making a connection with the statement of Rob Livingstone, in web-based network services people become customers and products for themselves at the same time (Porter, 2014). Laudon and Traver (2014) say that such network sites "invade the personal privacy of millions of users on a scale unprecedented in history" (Laudon and Traver, 2014, p.534). It is similar to a description of a product. As a customer you want to know what you buy and as a seller you also want to know what you sell and in case of social network sites, it is a description of a person, knowing location, previous or current employment, education, preferences, appearance and friends within a network. People voluntarily reveal and share their personal information, their life, for a 'free'. In Johnson (2010), CEO of Facebook, Mark Zuckerberg, states that, "privacy was no longer a social norm" which supports the argument of Chayko (2014) that the offline and online experience is formed as an enmeshed reality.

The problem is that, for example, TRUSTe report (2014) says that 92 % of U. S. customers are concerned about their privacy and 89 % of respondents avoid businesses that do not protect privacy of customers however, in contrast to that, PricewaterhouseCoopers (PwC) (2012, p.1) argue that "consumers are eager for companies to deliver exciting, personalized services … [and] they [consumers] are willing to share personal information to get it." This statement goes in exactly opposite direction and is justified in the survey which explains that 73 % of participants are willing to share their personal data providing they get some benefit in return. Finally, this is the essential point of privacy study, no matter if people are increasingly concerned, because to certain extent, when companies offer some benefits, customers fall into psychological traps and eventually provide their personal data, because they firstly see a benefit and do not realise an exchange behind it. This is what many call a privacy paradox (Dinev, 2014; Blank et al., 2014; Norberg et al., 2007). On one side

consumers are concerned about privacy whereas on the other side customers willingly provide their private information.

It could be said that the true origin of all the concerns begin with a notion that "[we] live our lives in a world where the Internet records everything and forgets nothing" (Rosen, 2010). Again, the issue goes back to the difference between two environments; in the natural (face-to-face) communication people tend to forget and forgive whereas as Brin and Page (1998) stated, the Internet is a huge and uncontrolled collection of data from past and presence and people need to have this fact on their minds when evaluating any digital-based information, because some information could be visible even if it is not real.

1.6 Digital data privacy in law

The latest law, Data protection directive, which regulates data protection was created in 1995. Since then technology has rapidly changed, although European Commission (EC) still sees it as a solid basis. Fortunately, with increased technology and customer concerns, in 2012 EC proposed a reform that will adjust and extend latest legislation. One of the proposed and already applied principles is the 'right to be forgotten'. This principle should help to manage personal digital data and, for example, delete or edit old invalid entries on internet search engines. The next core principle that EC proposed is an explicit consent for using services or generally consent for participating in any kind of private information exchange. Moreover, the legislation says that protection should ensure better "responsibility and accountability for these processing personal data". The reform also lists certain benefits for businesses such as singular rules, net savings, enhanced global competition and simplified processes (European Commission, 2012a).

First of all, the 'right to be forgotten', already applied by EU, was significantly challenged in 2010 when a Spanish citizen raised complaints against a Spanish newspaper, Google Spain and Google Inc. (Google is an American company maintaining internet search engine) about an old search result which invaded his privacy rights and at that time the record was no longer relevant.

After a long lawsuit, the court said that if a search engine operates within an EU state, the company has to comply with European rules. This is because the EC directive says that people can ask for deletion of their personal information when information is "inaccurate, inadequate, irrelevant or excessive" (European Commission, 2012b). In conclusion, Google Inc. was fined by a certain amount of money and has erased personal information from EU search results. However, according to Porter (2014) it is not easy to control or restrict the Internet at one point, because it is spread across multiple jurisdictions and that is what actually happened. The results remained the same for the most used American based search engine google.com (Powles & Chaparro, 2015). Secondly, terms and conditions in digital agreements are designed as a Yes/No option, where consent usually means using a service or product and indirectly provide some personal information. In case of a disagreement, potential consumers are usually not allowed to use a service or product (Minelli et al., 2012). Thirdly, responsibility of data is frequently debated topic. According to the Microsoft Trustworthy Computing (2013) survey, between the EU citizens, 40 % of participants wanted to see responsibility for online privacy and protection in hands of individuals whereas 30 % wanted the responsibility in hands of companies and another 30 % in hands of government. The results were similar with results from PwC (2012) study, although this study showed more significant result regarding the ownership for individuals, because 87 % of respondents agreed that they would like to manage and control the data they provide.

In terms of protection, the control and management of digital dossiers is becoming more important and it could be considered as a sign of inequalities between digital and physical environments. Sometimes it could be quite easy and quick to identify a person on the Internet with intention to find as much information as possible, for instance in terms of employment, friends, hobbies or opinions. A first impression about a certain individual could also be made through digital data, a collection of almost every action made online, and this first impression could mean for example false prejudice even though the people have not met yet (Microsoft official blog, 2012). Alex Pentland, MIT professor, says that the best way is to "own your own data" (Pentland, 2009, p.79). His concept is to put an individual above all the data and let him/her to manage their data through Personal Data Store. He says that in this case people know where their information is stored and who they

share the information with. Another possible way is to let a special company to manage personal data of individuals or let a government to manage all data. According to Lewis (2015) there are also specialised companies that provide a “data reputation” management on demand meaning that publicly accepted information or positive information are displayed first in a search engine, which could also be another approach to the personal data management.

1.7 Summary

This literature review discussed increased concerns that society has about digital privacy. Technology, as one aspect of human environment, is changing and people need to understand it and be able to manage own data and own legal rights. Nowadays, most of digital devices allow companies to track and analyse digital data of their customers and there should be a discussion about transparency and anonymity in the digital world that records and stores every single piece of data. The problem with technology is that current big data analysis, based on mathematical and statistical models, can make huge data sets meaningful for humans and find patterns of consumer behaviour that are not easily visible, as Target did.

In regard to digital privacy, the biggest concerns were discussed, such as government surveillance and selling information to third party organisations. Although all these concerns are present, it appears that individuals “freely submit [their] personal information” (Dinev, 2014, p.97), which results to a privacy paradox.

Keeping in mind customer as a priority, there is a possible way how to reduce consumer concerns and that is to create a sustainable controlling system based on trust. There are three possible controlling methods, which derived from the literature review. The first method lets a customer manage their own data in form of a data store. The second option proposes to let a specific company manage digital data for individuals and, thirdly, a government could take place as a superior entity and manage all data.

2. Methodology

This chapter will briefly explain the whole research process including adopted philosophy, research methodology, sampling method as well as limitations of the study. The purpose of this research is summarized in two main aims below.

- To identify the extent to which people are aware of their presence and privacy in the digital environment.
- To analyse conditions when consumers are willing to provide personal data.

The main objectives for the research were to identify how people perceive their digital privacy rights and what are their preferences regarding digital environment. The second objective was to analyse the conditions in which people are prepared to provide personal data.

2.1 Philosophy of the research

The research proposal suggested a philosophy of constructivism as an ontological view of reality. The justification for constructivism was that each person perceives privacy differently in their own way, which matches with the view of constructivism that "reality is unique to each individual" (Quinlan, 2011, p.105). However, mainly due to time limitations the philosophy of this study was changed to positivism. Within this philosophical view, the researcher studied a phenomenon; privacy issues, which were considered the same for each person and observable separately. In the positivistic view, the reality is always observable and participants cannot change it with their behaviour or interactions (Robson, 2011; Saunders et al., 2009). According to Quinlan (2011, p.105) a positivist researcher sees the reality as "singular, objective and apart from participants". In summary, the main characteristics of a positivist approach are that there is one reality and aspects of a reality are presented as facts and these facts are precisely quantifiable according to the mathematical principles (Robson, 2011). Another important characteristic of a positivist approach, which is again very similar to natural sciences, is that the

researcher himself is absolutely objective – thus creating "value-free" research (Saunders et al., 2009).

In balance, the main criticism of positivistic philosophy is that social research cannot use same methods as research for natural sciences. The critics say that there is not only one truth (or reality) that can be observed but many others as they are influenced by social interactions and social behaviour. In this opposite case, the reality is not consistent and easily quantifiable with precise mathematical principles. Also, another argument against positivism is that social sciences outputs, presented by positivists as facts, are actually opinions, thus it is impossible to separate facts and values (Robson, 2011).

2.2 Research methodology

Privacy issues have been studied profoundly in recent years (Dinev, 2014; Blank et al., 2014; PwC, 2012; GfK, 2014). This study used the theory devised from those studies, which is summarised in the literature review, as a basis for a research. This approach is commonly known as deductive where the research is based on a previous theory and then tests the theory. Building up on a philosophy of positivism, the research is purely quantitative, meaning that all data were collected, processed and analysed in a numerical form (Quinlan, 2011).

The main reason for adopting quantitative approach for this research was to focus on behaviour and objectivity of the research, supported by the statistical analysis. Whereas, qualitative researches tend to focus on meanings in relation to the research questions and also to the relation between participants itself (Robson, 2011).

As a social research, the study used survey as a methodological concept for a whole research, which means that "[i]nformation is gathered primarily by asking people questions" (Groves, 2009, p.3). This direct approach is efficient, considering the topic of privacy issues, because for example observation would not be sufficient in deciding on feelings or attitudes towards privacy concerns.

2.3 Data collection method

Following survey methodology, self-administered an online questionnaire was used as a method of a data collection. Online questionnaires are advantageous because of the easy way of distribution, collection and analysis of final data together with short collection time in comparison for example with interviews. Moreover, online questionnaires are especially effective in reaching large number of a population (Sekaran & Bougie, 2010; Hair et al., 2011). On the other hand, Gray (2009) points out that questionnaires do not allow direct interaction or feedback with participants and results may be sometimes based on "common-sense reasoning or even speculations" (Gray, 2009, p.165), because relationships between variables are chosen according to the researcher's judgement. Another argument says that some phenomena cannot be simply described using a scientific approach.

The questionnaire was designed mostly in type of scales in particular Likert and frequency scales (13 questions out of 21) and the rest (8 questions out of 21) was designed as a multiple choice question answering system. The scales for quantitative research were chosen largely because Quinlan (2011) says that scales measure attitudes toward certain subject, which expresses at least different points of view or feelings about the subject of digital privacy.

2.4 Sample and sampling method

Privacy concerns of digital data are related to large populations of people. Therefore the sample for the study was consisted of people of all age categories with access to the Internet. Regarding the method, the sample was based on mixed sampling techniques in particular convenience and then snowball sampling. Convenience sampling was chosen because of simplicity of execution and also because of time limitations. With this sampling method, the researcher has specifically chosen people that he has known (covering most of the age categories) and asked them to participate in the study. Then applying snowball sampling techniques, the participants were asked to distribute the questionnaire between their peers. In total, there were 107 participants involved in the research.

2.5 Data analysis

The data analysis was divided into two parts. In the first part there are basic characteristics that describe the data set such as sum or mode whereas the second part focuses on dependencies between variables using correlations. As the tested variables were only Likert scales, which is an ordinal variable, the Spearman's rho correlation coefficient was chosen for the analysis. This coefficient has values from 0 to 1 to indicate the strength of the relationship and positive or negative values which show the direction of a relationship (Bryman, 2012).

2.6 Pilot study

A pilot study was conducted prior to the research, primarily with two purposes – to examine coherence and correct understanding of the questionnaire and secondly to verify any language mistakes because the English language in not the first language of the researcher. In general there were found several minor language mistakes and a possible misinterpretation in the question number two because of the complexity of the question.

2.7 Survey questions and their relation to the research

Table 1: Questionnaire justification

#	Question
1	In terms of privacy, I feel a difference between person-to-person communication and communication through an electronic device.
	This sentence tries to identify whether people make difference between privacy in general terms and digital data privacy.
2	Imagine a situation: One day you search for a television online and few days later you go to a store and a shop assistant asks you which of those televisions you searched for online you would like.
	I prefer this kind of connection between digital and non-digital environments.
	This hypothetical situation deepens the 1st sentence and shows an example of how to perceive privacy from a different angle.

3	I am active daily on at least one social network.
	Statement 3 identifies advanced internet users.
4	I am satisfied with data privacy policies of social network sites (e. g. Facebook, Snapchat, Twitter etc.).
5	I am familiar with terms and regulations of Facebook.
	This set of sentences examines the extent to which are internet users aware of policies on social networking sites.
7	Who do you think owns the data of your digital profiles?
10	What is your main concern regarding digital privacy?
11	Who do you think should have a control over your data?
	This group of questions asks about ownership and control of data profiles.
8	I am concerned about digital data privacy.
	This statement is key for measuring general attitude towards privacy issues.
9	In terms of digital privacy, which kind of personal information do you consider as the most private?
	Ninth question tries to identify the most private information.
12	In terms of digital privacy, what is more valuable for you?
13	In terms of digital privacy, which information is more valuable for companies about their customers?
	These questions try to recognise customers' behaviour on the internet.
15	I have abandoned a service or company because of possible privacy issues.
16	I have left an internet site because of a lot of personal information I would have to provide.
17	I think that personalised services can limit my choice by showing me only a limited range of products/services.
	These statements aim to determine whether customers are able to spot any privacy issues on the internet.
6	I would like to be recognised (by my name and face) every time I am present on the web (for example, people can see which sites I visit or which products I buy).
14	I am willing to provide personal information in exchange for a free service.

18	Anonymity in digital environment is very important for me.
19	I am willing to pay for anonymity on internet.
	The statement number six outlines possible risks of privacy, followed by a statement exploring privacy paradox and statement questioning anonymity as an important counterpart to privacy.
20	What is your gender?
21	Which age category do you fit in?
	Basic demographic data

Own source

2.8 Issues of validity and reliability

The questionnaire followed a highly structured scheme, which means that the study could be replicated with high reliability. Regarding internal validity, the questions asked in the study matched the research aims. On the other hand, it was not possible to ensure external validity because of non-probability sampling methods. Thus no explicit measures were used to test validity or reliability.

2.9 Ethics

Saunders et al. (2009, p.201) state that "ethics are critical aspects for the conduct of research" consisting mainly of four areas to ensure against; harm to participants, lack of informed consent, invasion of privacy and deception (Bryman, 2008). Regarding the distribution of the questionnaire, there was no pressure for participants to take part in the study and the researcher assumed that even completing the questionnaire have not had any mental health consequences for participants. Additionally, Bryman (2008, p.125) says that "it is rarely feasible or desirable to provide participants with a totally complete account of what your research is about", given this fact, participants were given core information about the study and about their rights. Moreover, participants had to give their consent with information provided about this study. Also, the participation in this study was anonymous.

2.10 Limitations

There were two main limitation of this study. From the statistical point of view, a non-probability sampling does not allow represent and generalise the results to the whole population. Also, given the sample size of 107 participants, it is even more impossible to generalise the results. Therefore, the results are only applicable to the sample.

2.11 Summary

This chapter briefly explained the main principles and research instruments that were used for this study. The primary data collection was based on a quantitative survey using an online questionnaire as a data gathering method.

This study used non-probability sampling method with overall sample size of 107 participants, which means that from quantitative perspective, small sample size and chosen sampling method are indicators of limitations such as inability to generalise results to a whole population and representational bias, which means that same groups of people are encouraged to participate in the study according to the snowball sampling method.

3. Research results and analysis

This chapter shows results from the research firstly in a form of basic characteristics of the sample and then in a correlation table showing dependencies between variables.

Two main research aims were of particular interest in the primary data research.

- To identify the extent to which people are aware of their presence and privacy in the digital environment.
- To analyse conditions when consumers are willing to provide personal data.

3.1 Characteristics of the sample

In the whole there were 107 participants in this research. As the table below shows, each age category was covered. However the age category of ‘less than 18’ was represented only by two examples which is definitely not sufficient number for a statistically significant analysis. Nevertheless, the results are biased in favour of the category ‘18 – 28 years’, which is represented by 58.9 % of the whole size.

Table 2: Which age category do you fit in?

	Frequency	**Percent**
Less than 18	2	1.9
18 – 28	63	58.9
29 – 38	13	12.1
39 – 48	7	6.5
49 – 58	15	14
More than 59	7	6.5
Total	**107**	**100**

Own source

The next pie chart shows a gender ratio where male responses slightly dominate at level of 55.1 % (59 responses out of 107). The research sample is also characterised by social network activity because 73.8 % of participants acknowledged that they frequently use social networking sited during the day.

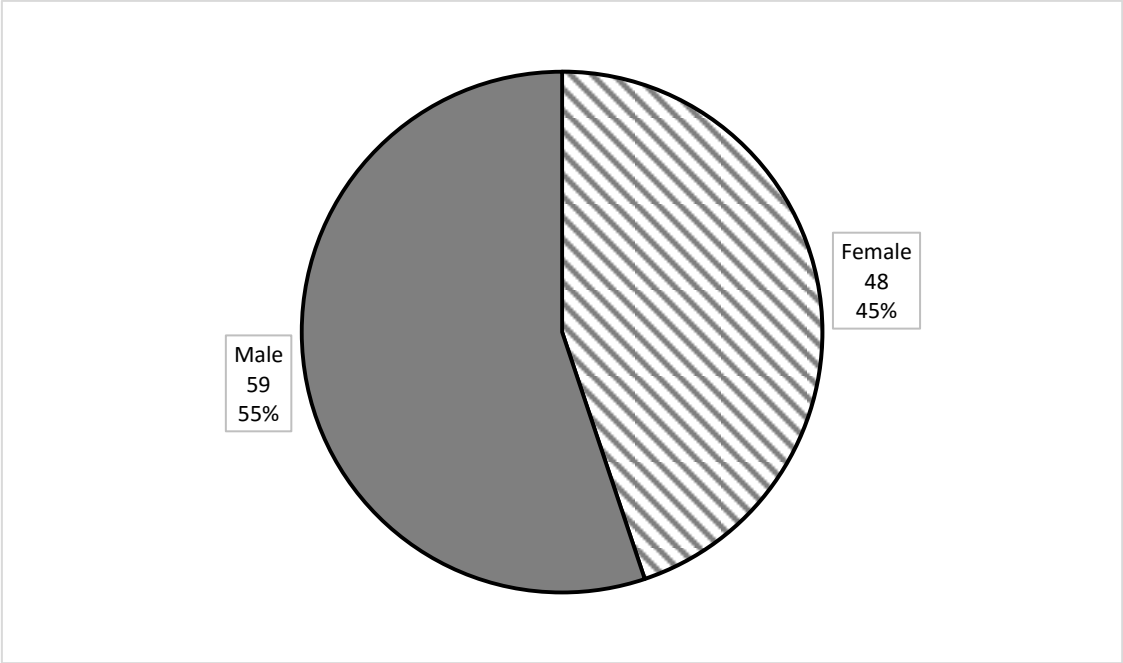

Chart 1: What is your gender?
Own source

Most importantly for the research, 85 out of 107 respondents said that they were in a way concerned about digital privacy. As expected, the main reason for increased concerns was identified as the act of selling personal information to third party companies (76 responses) followed by concerns about government surveillance (only 15 responses).

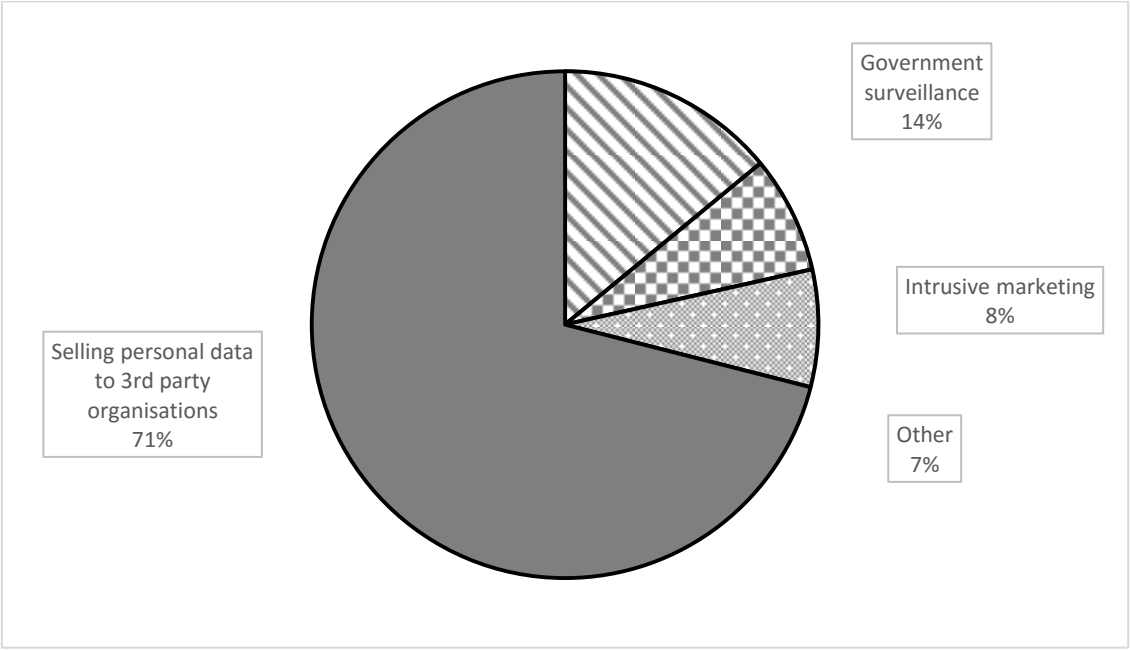

Chart 2: What is your main concern regarding digital privacy?
Own source

The question regarding the most private personal information where participants had an option to choose from seven items namely date of birth, gender, income, location, mobile number, name, online behaviour (patterns and preferences) – 50.5 % of participants selected all of the items. As individual responses, location and mobile number were highlighted as the most private information. These answers were expected in view that all items on the list could be considered as private information which could determine important characteristics about an individual and all of them show certain level of vulnerability.

In another question, 63.6 % of respondents said that they think that particular companies own digital data of individuals. However, on the other hand in the following question, 83 participants answered that they, as individuals, would like to have a control over their own data. This means that respondents acknowledged that contemporary businesses have a control over data of individuals, but more importantly, people would like to maintain their data on their own.

Private information in hands of companies could cause some problems to users and customers. Statements 15 and 16 tested whether customers are cautious about their digital privacy and whether they are able to spot some deceptive services. As the research showed 83 respondents have frequently left internet websites because of an amount of personal information they had to provide. Additionally, the same number or participants (83) said that they occasionally abandoned a service or a company because of possible privacy issues. These results were expected in light that 79.4 % of participants expressed concerns about their privacy in digital environment.

It seems from the results that people from the sample are still not sure how to include digital privacy into their every-day life, because only slightly over a half of the people (52.3 %) had positive attitudes towards integrated sharing of private information, which examined question number two. This is partly justified by the first question from the survey where 81.3 % of participants said that they still feel a difference between person-to-person communication and communication with a machine. Extending this point more in depth, question number six asked about profound recognition on the web and, not surprisingly, only 10 people expressed positive attitudes towards web recognition using

information such as face elements, name or other identifying factors. The robust majority, 90.7 %, from the whole sample indicated negative attitudes and 67.3 % people stated that they strongly disagree (the most negative point on the scale) with the recognition.

As the research has shown, 91 participants indicated positive attitudes to the fact that anonymity in the digital environment is very important for them. Also, majority of participants – 75.7 %, showed positive levels of agreement with a statement that personal services could limit their range of choices.

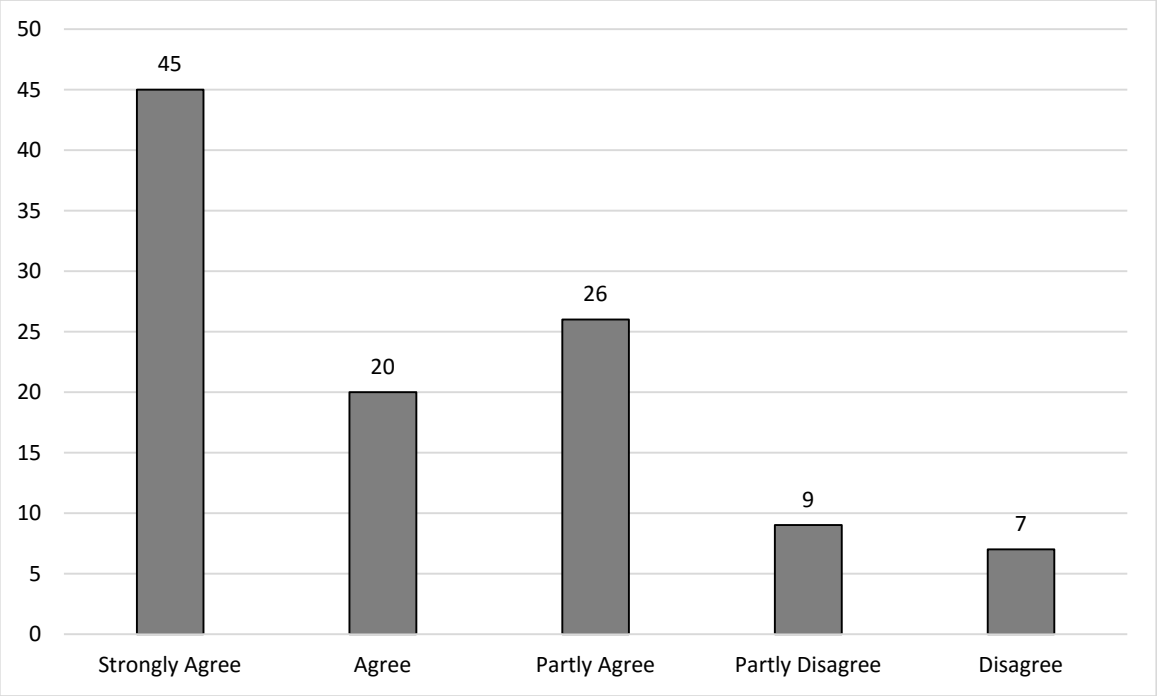

Chart 1: Anonymity in digital environment is very important for me.
Own source

The research results suggested that anonymity is important for the sample population of 107 participants. However, according to the question about willingness to pay for anonymity only one third of respondents, 36.4 %, indicated positive attitudes for paying for anonymity. The most frequent answer, 27, was that people strongly disagree with paying for anonymity inside the digital environment.

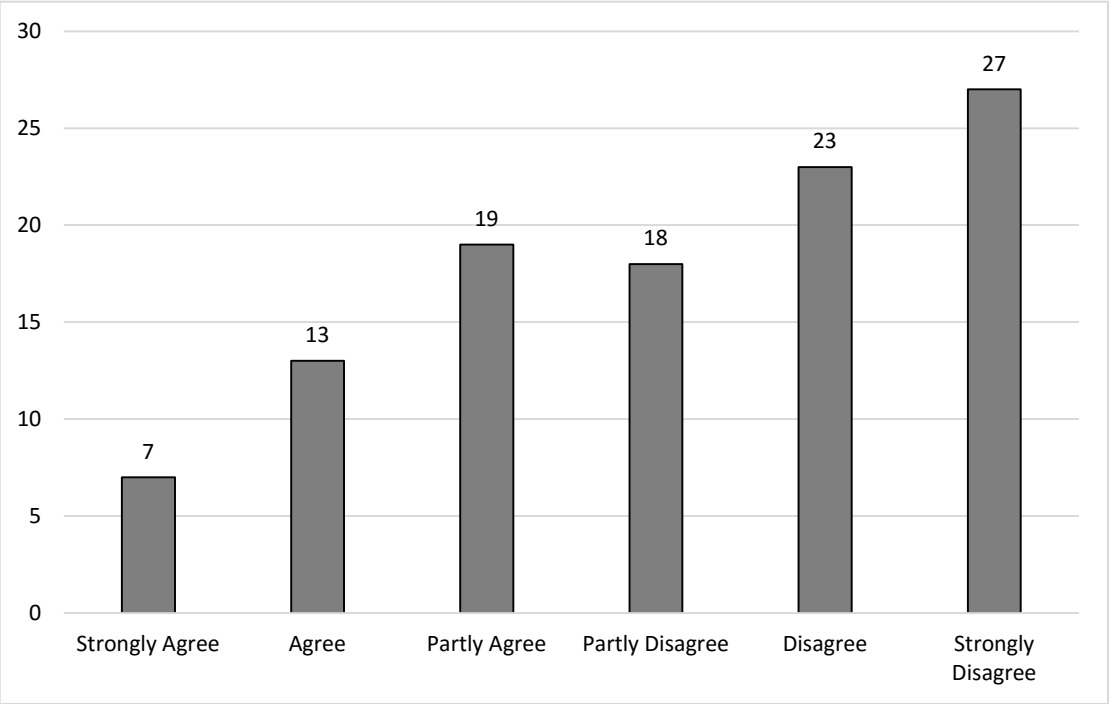

Chart 2: I am willing to pay for anonymity on the Internet.
Own source

Examining the privacy paradox, only 33.6 % of people showed positive attitudes for exchanging their personal information for a free service. The rest 71 respondents indicated negative opinions about the statement where the answer strongly disagree, 34 responses, was the most frequent answer. As it could be seen from the table below, exactly the same number of participants (34) is the highest amount of people who agreed that they are concerned about digital data privacy. It is also clearly visible that there is no connection between concerns and willingness to provide personal information.

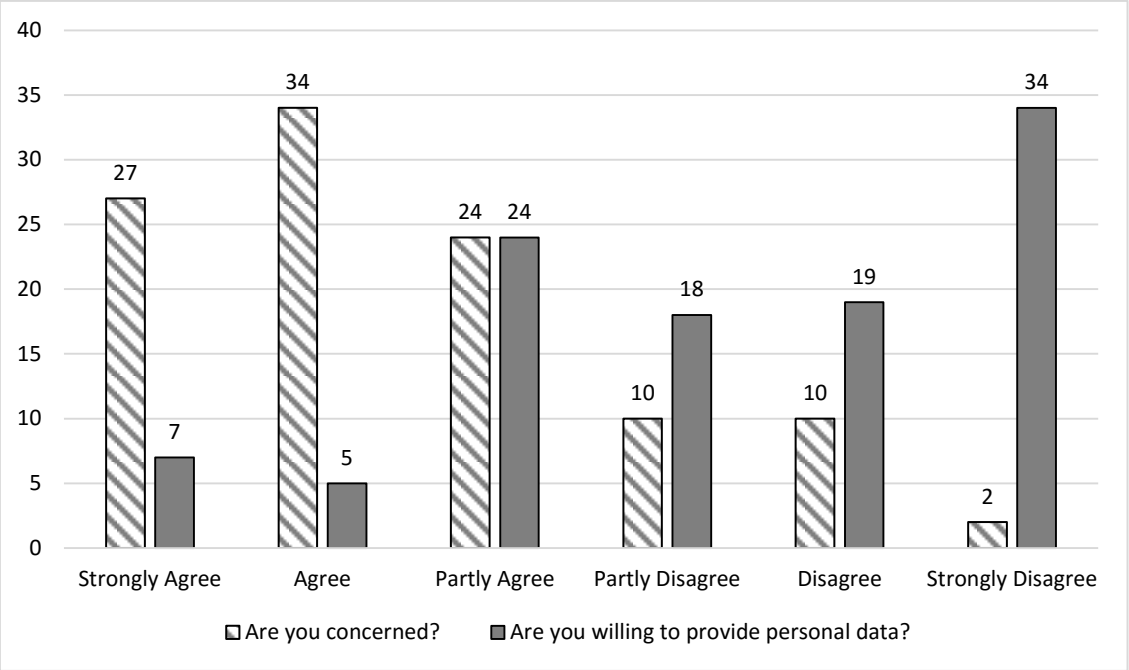

Chart 3: Examining privacy paradox.
Own source

3.2 Correlation dependencies

Looking at the relationships from the table of correlations (please see apendix), question number 18 has three moderate relationships. In the first one, it seems that people who feel difference between digital and physical environment think that anonymity is very important. Secondly, people who think that anonymity is important do not want to be recognised on the web and lastly people who are concerned about privacy issues think that anonymity is important for them.

There is a strong relationship between variables from questions 15 and 16 which has also indirect implications to the question number 8. It could be said that people who have concerns are more cautious about the amount of personal information they provide, therefore they are more likely to recognise bad practice from companies and consequently leave the service or abandon the brand (company).

Interestingly, according to the responses from questions 3, 4 and 5, people who are active on social sites tent to be familiar with terms and regulations on those social websites.

3.3 Summary of important findings

There were 107 participants altogether in the research sample. Regarding the privacy concerns, 79.4 % of respondents indicated concerns about their privacy and the reason for these concerns was primarily selling personal information to third party companies.

Exploring the topic of data management, 83 indicated that individuals should control and manage their own data, not specialized companies nor government. Results also showed that anonymity is important for the participants in the digital environment indicating agreement of 85 % with the response ‘Strongly agree’ as the most frequent answer. Moreover, only a third of participants said that they would exchange their private information for a free service.

4. Discussion and recommendations

This chapter discusses research findings and also offers some recommendations for a further research in the area.

At the beginning of this chapter it is also important to repeat the fact that analysed results are only valid for the research sample of 107 participants. It is not possible to generalise the results for the whole population. It is firstly because the sample is too small and also because non-probability sampling method was used in the research, which means that the sample cannot be statistically interchanged for the population. Therefore all data interpretations relate only to the sample.

4.1 Research sample

The research sample of 107 participants includes some notable flaws. It is significantly biased in favour of the age category of '18 – 28 years', which is represented by 58.9 % of the whole sample. It is because the questionnaire has mostly spread (according to the snowball technique) between students at the University. Secondly, age category 'Less than 18' includes only two responses, which is definitely not of statistical significance. The best option would be to ideally have the same portion of participants in each category. On the other side, it could be said that nothing in social sciences is ideally proportional. Therefore, majority of results is possibly deformed or distorted, for instance social network activity could be the most apparent.

4.2 First objective

To explore the privacy paradox in relation to digital data management.

Most interestingly, results from the research about exchanging of personal information for free services or products were absolutely opposite than the literature review have suggested. One way to looking at this is that most studies were conducted by corporate companies such as PwC, where 73 % of respondents were happy to exchange personal

information for some kind of benefit in return whereas individual researchers focused more on the general fact that the privacy paradox exists rather that examining customer willingness in detail. This could possibly mean that companies want to promote their personalised services and justify why those free services collect personal data.

In conclusion, this study states that people are definitely concerned about their digital data privacy and also that the paradox is presents to the certain extent. However, this research does not conclude that people are willing to provide their personal data in exchange for some kind of benefit.

4.3 Second objective

To describe the way businesses collect, process and use data in regard to customers along with legal restrictions.

Now it is clear that data are one of the most important aspects of a success for many businesses. Companies use many ways of how to obtain customer data and one of the ways is to use the newest technology such as big data. It would be strong to say that customers know all techniques that companies use, because sometimes the methods are not simple. For example Facebook, company operating mostly inside the digital environment, faces many criticism or sometimes lawsuits because of their privacy policies or regulations. Additionally, it was found from the research that the majority of participants are not satisfied with data policies that social network sites use. It is safe to say that those policies are sometimes very complex and lengthy.

Although, the aim was to examine more than one legal restriction, the literature review described and put into context the most important and discussed directive the ‘right to be forgotten’. It was because this right has been discussed most and brings another perspective to the problem of old search results.

4.4 Third objective

To identify the extent to which people are aware of their presence and privacy in the digital world.

In the research, participants acknowledged that they feel a difference between digital and non-digital communication. The difference in communication could be interpreted quite easily by saying that it is not natural for people to talk to a machine. It is because person-to-person communication has existed many years before people started to interact with machines. Participants also said that they do not want to be recognised (identified) on the Internet. It is mostly because "[o]ne of the attractive features of the Internet is its freedom" (Andrew, 2010, p.1098) and many illegal activities are happening inside the digital environment such as illegal reselling of movies, games or music with different themes that possibly break the law.

Regarding the Internet, people would like to possess the power of freedom and sustain in their anonymity. It is because anonymity assures certain level of freedom on the Internet, however many companies and especially e-shops, for instance amazon.co.uk or ebay.co.uk and many others, require login details and verifiable personal information about customers in order to make purchases on those websites. Although log in users lose their anonymity, and part of their freedom, companies get in return personal data about their customers and also preferences and patterns about products that customers buy or review for a potential purchase, which means that marketers are able to tailor an offer exactly for specific needs of customers using for example remarketing techniques. This theoretically means that a potential offer could be narrowed in terms of products or services.

In summary, it could be said that people recognise the difference between environments (online and offline) and they would like to embrace technology as a supporting tool for every-day life. On the other hand, people have increased concerns which means that they are more cautious on the Internet.

4.5 Fourth objective

To analyse conditions when consumers are willing to provide personal data.

In general, customers are willing to provide personal data when they trust the other side of a transaction. The trusted side has usually a good reputation which is based on a previous experience or a recommendation. Apart from reputation, complete knowledge about a transaction also increases trust between interested parties. Another condition for voluntarily providing personal data is when customers have a control over their data. This means that for example customers know who operates with their data, how are the data used and what will happen after the transaction or relationship (company-client) ends. These functions are absolutely vital because customers are cautious about their data management and, moreover, customers are not afraid to quickly abandon deceptive services or companies. This implies that companies should focus on trust building especially in the digital environment and for instance ensure that behind a machine there is a human aspect that customers can relate to.

4.6 Fifth objective

To establish possible solutions of how to gather and use data safely.

It was observed that most people would like to have a control over their own data. From this finding, it seems that the most probable way of how to gather and use data safely is that people will manage their own data storage centres. Other possible solutions suggest that government or specialised companies could administer personal databases of individuals. However, these other options are unlikely because for example trust in a government has been decreasing over recent years especially because of surveillance programmes and it is the same for data management in hands of companies. In particular, businesses were in many cases accused of selling personal information to third party companies, which is generally not considered as an ethical solution. Therefore this study concludes that personal data stores in hands of individuals could be the safest option available.

4.7 Recommendations

As it was remarked several times in this dissertation, the topic about digital data privacy is complex and includes many possibilities for further research. One of which could be of interest in particular is anonymity on the Internet and related issues. This is mainly because there has been several attempts to charge the usage of the Internet and also because anonymity (related to freedom) appears to benefit the digital environment and its users. There seems to be a pattern on the Internet to identify each user and make him or her pay for internet services. The assumptions for a further study could be that opposite sides could swap and users might pay for anonymity online.

Secondly, what could be the cause of all privacy concerns is that people do not know exactly what is happening with their data. Customers (users) would like to have a control over their data and a research into this field could bring interesting results. Of course, the Personal Data Store developed at MIT promises potential solution to the problem.

Conclusion

The main purpose of this chapter is to give a full summary of the dissertation research, state aims of the research and show how each point was studied and evaluated.

Overall there were five aims of this research.

- To describe the way businesses collect, process and use data in regard to customers along with legal restrictions.
- To identify the extent to which people are aware of their presence and privacy in the digital environment.
- To explore the privacy paradox in relation to digital data management.
- To analyse conditions when consumers are willing to provide personal data.
- To establish possible solutions of how to gather and use data safely.

The first objective was purely literature based and the purpose was to describe ways how businesses collect, process and use data in regard to customers along with legal restrictions. It was shown that firms use many techniques to obtain data from customers. One section discussed technology – big data analytics – whereas other section discussed purchasing personal information from data brokering companies. Additionally, businesses use other methods from a field of sociology or psychology in order to obtain private information about customers. One huge area of interest from the point of privacy is known as a privacy paradox, which was also examined in this dissertation. Finally the literature review focused on a basic legislative directive called the right to be forgotten, which was briefly discussed on an example of a dispute between Google and a Spanish citizen. In total this objective was sufficiently covered for purposes of an undergraduate dissertation.

The second objective tried to identify the extent to which people are aware of their presence and privacy in the digital environment using primary research data. At first, it was a bit complex objective because the extent could be measured by many ways, for example by a computer literacy or according to the knowledge of internet protocols. In this study the extent of awareness was measured using questions based on a social network involvement and the knowledge about policies, terms and regulations on social network

sites. Although the aim had a broad span of possible interpretations, the basic extent of awareness was identified and the aim was achieved.

The third objective focused on exploring a privacy paradox in data management. The topic of the privacy paradox was firstly studied in the literature review and then explored in the primary research. Although the literature suggested that the paradox is present at least in two thirds of transactions, the results from the primary research have shown that this argument is not true. Nevertheless, it is fair to say that the paradox was tested on a small sample size, therefore the results could be subject to misinterpretation. Overall, the paradox was explored in the literature review and then compared with the results of the study, therefore the aim was achieved.

The fourth objective was to analyse conditions when consumers are willing to provide personal data. This objective was, unfortunately, too broad and the literature review focused solely on the general aspect which is trust between participants in a transaction. This issues would be more for a qualitative data method rather that for a quantitative research. Nonetheless, in general terms the objective was reached.

The last objective of the research was to outline possible safe solutions for gathering and maintaining digital data. The literature review indicated that people think that the safest solution is to maintain data individually. These statements were later confirmed with the results of the research. It was concluded that digital data in hands of individuals is the safest way, acknowledging the opinion of Alex Pentland, who promoted personal data stores from MIT production. It could be said that this objective was completed because the other ways of gathering such as company or government data maintenance were not entirely supported by participants and therefore not thoroughly examined.

References

Andrew, A. M. (2010). *Internet freedom*. Kybernetes, 39(7), 1097-1099. doi:10.1108/03684921011062719

Blank, G., Bolsover, G., & Dubois, E. (2014). *A New Privacy Paradox: Young people and privacy on social network sites.* Prepared for the Annual Meeting of the American Sociological Association (Vol. 17).

Boyd, D., & Crawford, K. (2011). *Six Provocations for Big Data*. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. Retrieved from http://dx.doi.org/10.2139/ssrn.1926431

Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine*. Computer Networks and ISDN Systems, 30(1), 107-117. doi:10.1016/S0169-7552(98)00110-X

Bryman, A. (2008). *Social research methods* (pp.341-344). Oxford: Oxford University Press.

Bryman, A. (2012). *Social research methods*. Oxford: Oxford University Press.

Camenisch, J. (2012). *Information privacy?* Computer Networks, 56(18), 3834-3848. doi:10.1016/j.comnet.2012.10.012

Chayko, M. (2014). *Techno-social life: The internet, digital technology, and social connectedness*. Sociology Compass, 8(7), 976-991. doi:10.1111/soc4.12190

David, J. (2014). *Big data is big business*. City A.M.

Dinev, T. (2014). Why would we care about privacy? *European Journal of Information Systems* doi:10.1057/ejis.2014.1

Duhigg, C. (2012). How Companies Learn Your Secrets. The New York Times. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=1&

Dwyer, C. (2011). Privacy in the age of google and facebook. *IEEE Technology and Society Magazine*, 30(3), 58-63. doi:10.1109/MTS.2011.942309

European Commission. (2012a). Data protection reform: Frequently asked questions. Retrieved April 4, 2015, from europa.eu/rapid/press-release_MEMO-12-41_en.htm?locale=en

European Commission. (2012b). Factsheet on the "Right to be Forgotten" ruling. Retrieved April 4, 2015, from http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

Franceschi-Bicchierai, L. (2014). The 10 Biggest Revelations From Edward Snowden's Leaks. Mashable. Retrieved from http://mashable.com/2014/06/05/edward-snowden-revelations/

Franks, B. (2012). Taming the big data tidal wave: *Finding opportunities in huge data streams with advanced analytics*. Hoboken, N.J: Wiley

GfK. (2014). GfK: Survey on data privacy adn trust. Retrieved April 1, 2015, from http://www.gfk.com/Documents/GfK-Privacy-Survey.pdf

Gordon, K. (2013). What is big data? *ITNOW*, 55(3), 12-13. doi:10.1093/itnow/bwt037

Gray, D. E. (2009). *Doing research in the real world*. Los Angeles, [Calif.]; London: SAGE.

Groves, R. M. (2009). *Survey methodology*. US: John Wiley & Sons Inc.

Hair, J. F., Celsi, M. W., Money, A. H., Samouel, P., & Page, M. J. (2011). *Essentials of business research methods*. Armonk, N.Y; London: M.E. Sharpe.

Information society (2011). In Dictionary of Sociology (3th ed.). Oxford University Press.

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. Journal of the American *Society for Information Science and Technology*, 59(3), 363-374. doi:10.1002/asi.20722

Kerr, O. S. (2005). Searches and seizures in a digital world. Paper for Harvard Law Review, The George Washington University Law school Public law and legal theory, Working paper no. 135.

Lane, J. I. (2014). *Privacy, big data, and the public good*: Frameworks for engagement. New York, NY: Cambridge University Press.

Laudon, K. C., & Traver, C. G. (2014). *E-commerce: Business, technology, society*. Boston: Pearson. (pp.533-543)

Lewis, T. (2015). Michael Fertik: online reputation is becoming more valuable than money or power. The Guardian. Retrieved April 1, 2015, from http://www.theguardian.com/technology/2015/jan/18/michael-fertik-online-reputation-economy-interview-valuable-money-power

Mathiesen, T. (2013). *Towards a surveillant society: The rise of surveillance systems in europe.* Hook, Hampshire: Waterside Press.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, 90(10), 61-68.

Microsoft official blog. (2012). Microsoft & Data Privacy Day: Put Your Best Digital Foot Forward - The Official Microsoft Blog. Retrieved April 1, 2015, from http://blogs.microsoft.com/blog/2012/01/24/microsoft-data-privacy-day-put-your-best-digital-foot-forward/

Microsoft Trustworthy Computing. (2013). Privacy Survey Results. Retrieved April 4, 2015, from http://download.microsoft.com/download/A/A/9/AA96E580-E0F6-4015-B5BB-ECF9A85368A3/Microsoft-Trustworthy-Computing-2013-Privacy-Survey-Results.pdf

Minelli, M., Chambers, M., & Dhiraj, A. (2012). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. John Wiley & Sons.

Moore's Law. (2015). Moore's Law. Retrieved April 1, 2015, from http://www.mooreslaw.org/

Morris, J. & Lavandera, E. (2012). *Why big companies buy, sell your data*. CNN. Retrieved May 30, 2015, from http://edition.cnn.com/2012/08/23/tech/web/big-data-acxiom/

Mutz, D. C. (2009). Effects of internet commerce on social trust. *The Public Opinion Quarterly*, 73(3), 439-461. doi:10.1093/poq/nfp042

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1), 100-126. doi:10.1111/j.1745-6606.2006.00070.x

PCAST report. (2014). *Big data and privacy*: *A technological perspective*. Retrieved March 12, 2015, from https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

Pentland, A. (2009). Reality mining of mobile communications: Toward a new deal on data. *The Global Information Technology Report* 2008–2009, 75-80.

Porter, C. (2014). Little privacy in the age of big data. *The Guardian*. Retrieved April 1, 2015, from http://www.theguardian.com/technology/2014/jun/20/little-privacy-in-the-age-of-big-data

Poulin, C. (2014). Big data custodianship in a global society. *SAIS Review of International Affairs*, 34(1), 109-116.

Powles, J., & Chaparro, E. (2015). *How Google determined our right to be forgotten*. Retrieved April 4, 2015, from http://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search

PwC. (2012). *Consumer privacy: What are consumers willing to share?*: The speed of life: Consumer intelligence series. Retrieved from http://www.pwc.com/sg/en/tice/assets/ticenews201208/consumerintelligence201208.pdf

Quinlan, C. (2011). *Business research methods*. Andover: South-Western Cengage Learning.

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. (2013). Anonymity, Privacy, and Security Online. PewResearchCenter. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_AnonymityOnline_090513.pdf

Robson, C. (2011). *Real world research: A resource for users of social research methods in applied settings*. Chichester: Wiley.

Rosen, J. (2010). The Web Means the End of Forgetting. *The New York Times.* Retrieved April 1, 2015, from http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?pagewanted=all&_r=0

Ross, J. W., Beath, C. M. & Quaadgras, A. (2013). You May Not Need Big Data After All. *Harvard Business Review*, 91(12), 90-98.

Salkowitz, R. (2014). *From Big Data to smart data: Using data to drive personalized brand experiences*. Retrieved from http://download.microsoft.com/download/E/5/F/E5FCBC1F-84FE-4A04-A526-213973B12435/From_Big_Data_to_Smart_Data_White_Paper.pdf

Saunders, M., Lewis, P., & Thornhill, A. (2009*). Research methods for business students. Harlow*: Financial Times Prentice Hall.

Saxena, A. (2014). Mine Big Data's Promise. Selling Info You Gather Is a Profitable Sideline. Entrepreneur Media. Retrieved May 30, 2015, from http://www.entrepreneur.com/article/239871

Sekaran, U., & Bougie, R. (2010). Research methods for business: A skill-building approach. Chichester: Wiley.

Skopek, J. M. (2013). *Anonymity, the production of goods, and institutional design*. Fordham L. Rev., 82(4), 1751-1809.

Solove, D. J. (2006). *Digital person: Technology and privacy in the information age New York* University Press (NYU Press).

Starr, B., & Yan, H. (2013). *Man behind NSA leaks says he did it to safeguard privacy,* liberty. CNN. Retrieved April 1, 2015, from http://edition.cnn.com/2013/06/10/politics/edward-snowden-profile/

Stein, J. (2011). *Data Mining: How Companies Know Your Personal Information*. TIME magazine. Retrieved May 30, 2015, from http://content.time.com/time/magazine/article/0,9171,2058205,00.html

The Economist. (2010). *Dicing with data; facebook, google and privacy*. 395(8683), p.16.

The Guardian. (2013). *NSA files decoded: Edward Snowden's surveillance revelations explained*. Retrieved March 12, 2015, from http://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/7

TRUSTe. (2014). 2014 TRUSTe US Consumer Confidence Index. Retrieved April 1, 2015, from http://www.truste.com/us-consumer-confidence-index-2014/

Waldo, J., Lin, H., & Millett, L. I. (2007). *Engaging privacy and information technology in a digital age.* Washington, D.C: National Academies Press.

Walter, C. (2005). *Kryder's law*. United States: Scientific American, Incorporated. doi:10.1038/scientificamerican0805-32

Windley, P. J. (2005). Digital identity. Beijing; Farnham: O'Reilly.

Yanes, A. (2014). Privacy and Anonymity. arXiv preprint arXiv:1407.0423.

Zhen, J., & Jilan, Z. (2011). On interaction between network communication and society. *Paper presented at the International Conference on Business Management and Electronic Information*, 662-664. doi:10.1109/ICBMEI.2011.5917999

Zwitter, A. (2014). *Big data ethics. Big Data & Society*, 1(2) doi:10.1177/2053951714559253

Appendices

Item A: Questionnaire

1
In terms of privacy, I feel a difference between person-to-person communication and communication through an electronic device. (AG answer)

2
Imagine a situation: One day you search for a television online and few days later you go to a store and a shop assistant asks you which of those televisions you searched for online you would like.

I prefer this kind of connection between digital and non-digital environments. (AG answer)

3
I am active daily on at least one social network. (FR answer)

4
I am satisfied with data privacy policies of social network sites (e. g. Facebook, Snapchat, Twitter etc.). (AG answer)

5
I am familiar with terms and regulations of Facebook. (AG answer)

6
I would like to be recognised (by my name and face) every time I am present on the web (for example, people can see which sites I visit or which products I buy). (AG answer)

7
Who do you think owns the data of your digital profiles?

- Me
- Government authority
- Particular companies that run services
- Neither

8
I am concerned about digital data privacy. (AG answer)

9
In terms of digital privacy, which kind of personal information do you consider as the most private?

- gender
- sexual status
- name and surname
- location
- mobile number
- date of birth
- income
- online behaviour patterns and preferences
- all
- none

10
What is your main concern regarding digital privacy?

- Government surveillance
- Intrusive marketing
- Selling personal data to 3rd party organisations
- Other

11
Who do you think should have a control over your data?

- Government
- Special agencies for maintaining information
- Me
- Someone else
- No one

12
In terms of digital privacy, what is more valuable for you?

- Identification characteristics (name, address, salary, gender …)
- Behaviour patterns (what do you do and when)
- Both are equal

13
In terms of digital privacy, which information is more valuable for companies about their customers?

- Identification characteristics (name, address, salary, gender …)
- Behaviour patterns (what do you do and when)
- Both are equal

14
I am willing to provide personal information in exchange for a free service. (AG answer)

15
I have abandoned a service or company because of possible privacy issues. (FR answer)

16
I have left an internet site because of a lot of personal information I would have to provide. (FR answer)

17
I think that personalised services can limit my choice by showing me only a limited range of products/services. (AG answer)

18
Anonymity in digital environment is very important for me. (AG answer)

19
I am willing to pay for anonymity on the Internet. (AG answer)

20
What is your gender?

- Female
- Male

21
Which age category do you fit in?

- Less than 18
- 18 – 28
- 29 – 38
- 39 – 48
- 49 – 58
- More than 58

Additional information

(AG answer) = AGREEMENT LIKERT SCALES

- Strongly agree
- Agree
- Partly agree
- Partly disagree
- Disagree
- Strongly disagree

(FR answer) = FREQUENCY SCALES

- Always
- Frequently
- Occasionally
- Rarely
- Never

Item B: Correlation table

		21	1	3	4	5	6	8	14	15	16	17	18
21	CC	1	-0.052	.276**	.282**	0.159	0.093	-0.016	0.16	-0.15	-0.137	-.191*	-.279**
	S	.	0.596	0.004	0.003	0.102	0.338	0.872	0.101	0.123	0.161	0.049	0.004
1	CC	-0.052	1	0.058	-.280**	-0.006	-.218*	.256**	-0.187	.213*	.208*	0.189	.435**
	S	0.596	.	0.553	0.004	0.949	0.024	0.008	0.054	0.027	0.032	0.051	0
3	CC	.276**	0.058	1	.346**	.349**	0.009	.244*	.267**	0.111	.305**	0.041	0.034
	S	0.004	0.553	.	0	0	0.93	0.011	0.005	0.253	0.001	0.676	0.724
4	CC	.282**	-.280**	.346**	1	.331**	0.183	-0.128	.236*	-.209*	-0.125	-0.033	-.269**
	S	0.003	0.004	0	.	0.001	0.059	0.19	0.014	0.031	0.199	0.737	0.005
5	CC	0.159	-0.006	.349**	.331**	1	0.151	0.008	0.189	-0.018	0.023	0.149	-0.075
	S	0.102	0.949	0	0.001	.	0.121	0.933	0.051	0.852	0.815	0.125	0.443
6	CC	0.093	-.218*	0.009	0.183	0.151	1	-0.188	.194*	-0.177	-.269**	-0.171	-.440**
	S	0.338	0.024	0.93	0.059	0.121	.	0.052	0.046	0.068	0.005	0.079	0
8	CC	-0.016	.256**	.244*	-0.128	0.008	-0.188	1	-0.079	.337**	.378**	.287**	.429**
	S	0.872	0.008	0.011	0.19	0.933	0.052	.	0.419	0	0	0.003	0
14	CC	0.16	-0.187	.267**	.236*	0.189	.194*	-0.079	1	-0.025	-0.082	0.09	-.285**
	S	0.101	0.054	0.005	0.014	0.051	0.046	0.419	.	0.801	0.403	0.354	0.003
15	CC	-0.15	.213*	0.111	-.209*	-0.018	-0.177	.337**	-0.025	1	.605**	0.121	.288**
	S	0.123	0.027	0.253	0.031	0.852	0.068	0	0.801	.	0	0.215	0.003
16	CC	-0.137	.208*	.305**	-0.125	0.023	-.269**	.378**	-0.082	.605**	1	0.156	.300**
	S	0.161	0.032	0.001	0.199	0.815	0.005	0	0.403	0	.	0.107	0.002
17	CC	-.191*	0.189	0.041	-0.033	0.149	-0.171	.287**	0.09	0.121	0.156	1	.287**
	S	0.049	0.051	0.676	0.737	0.125	0.079	0.003	0.354	0.215	0.107	.	0.003
18	CC	-.279**	.435**	0.034	-.269**	-0.075	-.440**	.429**	-.285**	.288**	.300**	.287**	1
	S	0.004	0	0.724	0.005	0.443	0	0	0.003	0.003	0.002	0.003	.

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

CC = correlation coeficient
S = significance ratio