



LUND UNIVERSITY

School of Economics and Management

Department of Informatics

Finding the potential privacy gap in the Big Data Supply Chain

Master Thesis, 15 ETCS, INFM10 in Informatics

Presented in June 2014

Authors: Flogerta Banaj
Klaudia Dardha

Supervisor: Markus Lahtinen

Examiners: Björn Johansson
Benjamin Weaver

FINDING THE POTENTIAL PRIVACY GAP IN THE BIG DATA SUPPLY CHAIN

Master Thesis

Flogerta Banaj

int13fb2@student.lu.se

Klaudia Dardha

int13kda@student.lu.se

Presented in June 2014

Supervised by Markus Lahtinen

Examined by Björn Johansson
Benjamin Weaver

Abstract

We live in a digitalized society. All the abundant data we produce, today called “Big Data” it changing our lives, and will soon disrupt it. Different studies and analysis argue about the advantages that Big Data comes in, not only as a competitive advantages for the data holders, but also in health, government, for the citizens and society as a whole. Nevertheless, Big Data comes with significant questions and poses challenges toward the privacy concern. So the path to Big Data gains is risky and also rocky. The decision we take over that data have a real human consequences such as ethical issues. Any data on social subjects raise privacy issues, and when the risk of misuse, intentionally or not, is huge it becomes an issue for the entire information society.

In this research, we explore potential gaps among the participants and deduct various reasons of these breaches reaching thus to reasons for improving the interplay among them. The study reflects on the interplay between government, business and consumer in a Big Data Supply Chain. It shows an existing inconsistency partly because of the lack of enforcement government legacy that is also attributed to lack of educated public.

Data holders lack transparency and consumers retain their trust toward them. The communication, barriers and legal rights between their interplay are vague, leading so to an important question toward ownership. When data sets are available to be gathered and used in analysis, there is a mist about its usage rights and requirements.

Keywords: Big Data; Big Data Supply Chain; Gap; Privacy; Data Subject; Data Holder; Stakeholders; Ethics; Ownership; Transparency;

Acknowledgments

We would like to express our gratitude to Trevor Peirce, Cees Boon and Jon Page for the immense contribution and help with their participation in our research. Their availability, patience and good humor made our communication easier with valuable suggestions and assistance.

We would also like to thank our supervisor Markus Lahtinen for his tireless supervision, inspiration and guidance in this study.

Special thanks

Three intensive months of working spent together from morning till night at study areas, library and also home, made sometimes discussion exhaustive. And so a special thank goes to each of us, Flogerta and Klaudia for supporting and never lose the good spirit in our team.

Flogerta Banaj and Klaudia Dardha
Lund, June 2014

Table of Contents

| | | |
|----------|--|----|
| 1 | Introduction | 9 |
| 1.1 | Background..... | 9 |
| 1.2 | Problem area | 11 |
| 1.3 | Motivation of proposed study and Research question | 11 |
| 1.4 | Purpose | 12 |
| 1.5 | Delimitation..... | 13 |
| 1.6 | Terms and definition | 13 |
| 1.7 | Structure of the thesis | 14 |
| 2 | Privacy in the context of Big Data | 16 |
| 2.1 | Introduction..... | 16 |
| 2.2 | The changing landscape towards Big Data..... | 16 |
| 2.2.1 | <i>Big Data Supply Chain</i> | 18 |
| 2.2.2 | <i>Big Data Supply Chain in the Macro-environment</i> | 21 |
| 2.2 | Privacy - a Big Data challenge?! | 22 |
| 2.2.1 | <i>Privacy ethical dimension</i> | 22 |
| 2.2.2 | <i>Privacy transition from physical to information one</i> | 22 |
| 2.2.3 | <i>Privacy concerns</i> | 24 |
| 2.2.4 | <i>Privacy calculus theory</i> | 26 |
| 2.3 | Theoretical framework | 26 |
| 2.4 | Concept-centric theoretical framework | 28 |
| 3 | Methodological Approach | 29 |
| 3.1 | Research Strategy | 29 |
| 3.2 | Research Approach | 30 |
| 3.3 | Data collection | 30 |
| 3.4 | Thematizing of study..... | 31 |

| | | |
|----------|---|-----------|
| 3.4.1 | <i>Big Data Supply Chain</i> | 32 |
| 3.4.2 | <i>Big Data Supply Chain – Participants</i> | 33 |
| 3.4.3 | <i>Big Data Supply Chain – Individual</i> | 34 |
| 3.4.4 | <i>Extended research framework</i> | 34 |
| 3.5 | Design of collection techniques..... | 35 |
| 3.5.1 | <i>Design of interview guides</i> | 35 |
| 3.5.2 | <i>Design of diary instructions</i> | 36 |
| 3.6 | Selection respondents | 37 |
| 3.6.1 | <i>Selection of interviewee</i> | 37 |
| 3.6.2 | <i>Contact with diarists</i> | 38 |
| 3.7 | Transcribing..... | 39 |
| 3.8 | Analyzing | 40 |
| 3.9 | Ensuring research quality..... | 41 |
| 3.9.1 | <i>Reliability</i> | 41 |
| 3.9.2 | <i>Validity</i> | 42 |
| 3.9.3 | <i>Bias</i> | 42 |
| 3.9.4 | <i>Ethics</i> | 42 |
| 3.10 | Reporting | 43 |
| 4 | Empirical findings and data analysis | 44 |
| 4.1 | The declaration of empirical findings..... | 44 |
| 4.2 | Big Data Supply Chain..... | 44 |
| 4.2.1 | <i>Data Subject</i> | 44 |
| 4.2.2 | <i>Data Holder</i> | 45 |
| 4.2.3 | <i>Ownership</i> | 46 |
| 4.2.4 | <i>Summarizing Big Data Supply Chain</i> | 47 |
| 4.3 | Big Data Supply Chain – Government..... | 47 |
| 4.3.1 | <i>Legal Assurance</i> | 47 |

| | | |
|----------|--|-----------|
| 4.3.2 | <i>Policy Assurance</i> | 48 |
| 4.3.3 | <i>Invasion</i> | 48 |
| 4.3.4 | <i>Summarizing BDSC-Government</i> | 49 |
| 4.4 | Big Data Supply Chain – Business..... | 49 |
| 4.4.1 | <i>Management</i> | 49 |
| 4.4.2 | <i>Responsibility</i> | 50 |
| 4.4.3 | <i>Secondary use</i> | 51 |
| 4.4.4 | <i>Summarizing BDSC-Business</i> | 51 |
| 4.5 | Big Data Supply Chain – Consumer..... | 51 |
| 4.5.1 | <i>Privacy benefits</i> | 51 |
| 4.5.2 | <i>Transparency</i> | 52 |
| 4.5.3 | <i>Privacy concerns</i> | 52 |
| 4.5.4 | <i>Summarizing BDSC-Consumer</i> | 53 |
| 4.6 | Big Data Supply Chain-Individual..... | 53 |
| 4.6.1 | <i>Trust</i> | 53 |
| 4.6.2 | <i>Privacy Awareness</i> | 54 |
| 4.6.3 | <i>Educated Public</i> | 55 |
| 4.6.4 | <i>Summarizing BDSC-Individual</i> | 55 |
| 4.7 | From the diaries..... | 56 |
| 4.7.1 | <i>Diarists interview</i> | 58 |
| 4.8 | Summary of empirical findings | 60 |
| 5 | Discussion | 62 |
| 5.1 | Big Data Supply Chain..... | 62 |
| 5.2 | Big Data Supply Chain - Government | 63 |
| 5.3 | Big Data Supply Chain – Business..... | 64 |
| 5.4 | Big Data Supply Chain – Consumer..... | 65 |
| 5.5 | Big Data Supply Chain - Individual..... | 66 |

| | | |
|----------|--|-----|
| 6 | Conclusions and further research | 68 |
| 6.1 | Research question | 68 |
| 6.2 | Implications..... | 69 |
| 6.3 | Contribution | 70 |
| 6.4 | Limitations and further research..... | 70 |
| 7 | Appendix | 72 |
| 7.1 | Appendix 1 - Concept-centric theoretical framework | 72 |
| 7.2 | Appendix 2 - Interview guide for experts | 74 |
| 7.2.1 | <i>Appendix 2.1 – Table of themes in accordance to interviews and diary.....</i> | 76 |
| 7.3 | Appendix 3 - Diary guide for consumers | 78 |
| 7.4 | Appendix 4 – Interview transcript with Cees Boon | 80 |
| 7.5 | Appendix 5 – Interview transcript with Jon Page | 88 |
| 7.6 | Appendix 6 – Interview transcript with Trevor Peirce..... | 97 |
| 7.7 | Appendix 7 – Interview transcript with Interviewer 4..... | 105 |
| 7.8 | Appendix 8 – Personal Data Submission Diary – Diarist 1 | 110 |
| 7.9 | Appendix 9 – Personal Data Submission Diary – Diarist2 | 118 |
| 7.10 | Appendix 10 – Personal Data Submission Diary – Diarist3 | 126 |
| 7.11 | Appendix 11 – Personal Data Submission Diary – Diarist4 | 134 |
| 7.12 | Appendix 12 – Personal Data Submission Diary – Diarist5 | 142 |
| 8 | References | 151 |

Table of Figures

| | |
|--|----|
| Figure 2-1 Hype Cycle for Emerging Technologies, 2013 (Gartner, 2013) | 18 |
| Figure 2-2 Big Data Supply Chain (Croll, 2014)..... | 19 |
| Figure 2-3 Complex Data Infrastructure/Ecosystem (WorldEconomicForum, 2012) | 20 |
| Figure 2-4 Taxonomy of Privacy in a Big Data Ecosystem (Based on Solove (2006))..... | 27 |
| Figure 3-1 Seven stages of an Interview Investigation (Kvale &Brinkmann (2009))..... | 31 |
| Figure 3-2 Diary Investigation (Based on: Alaszewski (2006))..... | 32 |

Table of Tables

| | |
|---|----|
| Table 2-1 Information Privacy Taxonomy | 23 |
| Table 2-2 Evolution of information privacy concept following the evolution of IT (Adapted from Smith et al. (2011)) | 24 |
| Table 2-3 Measuring scale of privacy concerns..... | 25 |
| Table 2-4 The concept-centric framework | 28 |
| Table 3-1 Extended research framework..... | 35 |
| Table 3-2 Data collection (Interviews) summarize | 38 |
| Table 3-3 Data collection (Diarists) summarize | 39 |
| Table 3-4 Coding structure for data analysis Interviews | 41 |
| Table 3-5 Coding structure for data analysis Diaries | 41 |
| Table 4-1 Data Subject | 45 |
| Table 4-2 Data Holder..... | 45 |
| Table 4-3 Ownership..... | 46 |
| Table 4-4 Legal Assurance | 47 |
| Table 4-5 Policy Assurance..... | 48 |
| Table 4-6 Invasion | 48 |
| Table 4-7 Management..... | 49 |
| Table 4-8 Responsibility..... | 50 |
| Table 4-9 Secondary use | 51 |
| Table 4-10 Privacy benefits..... | 52 |
| Table 4-11 Transparency | 52 |
| Table 4-12 Privacy concerns..... | 53 |
| Table 4-13 Trust..... | 54 |
| Table 4-14 Privacy awareness | 54 |
| Table 4-15 Educated Public | 55 |
| Table 4-16 Diaries behavior | 60 |

1 Introduction

This introductory chapter seeks to provide a general outline of the social- technical dimension of Big Data, introducing though to the research question of the study. The Purpose and motivation of our research will follow up the background and problem area of the study. Consequently, we will discuss delimitation of it and some key concepts of definitions.

1.1 Background

Marion Coseja Gonzalez case was one of more than 200 similar cases in Spain (EuropeanVoice, 2014) but its judgment brought crucial consequences for all European Union (EU) citizens. He complained that Google's search engine linked his name to a 1998 announcement of an auction of his house to cover unpaid social security contributions, relating that information undesirable (EuropeanVoice, 2014). On 13 May 2014 the European Union's highest court, according to European Commission data protection proposal in 2012, made available to EU citizens "the right to be forgotten" making thus exceptions for politicians or other historical data purposes (EuropeanVoice, 2014).

Currently, most of the countries in the world are being part of a social, technological revolution. Their way of living, communicating and doing business is moving on the digital world (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013). Citizens in these developed countries are all the time surrounded by Internet: in phones, homes, cities, and cars. All this abundant produced data, today called "Big Data", will disrupt our life, and change the relationships between government, business, consumers and citizens (Needham, 2013).

According to Russom (2011) companies have gathered more data in the last two years than they did in all previous 2,000 years. In addition, Epic (2014) predicts until 2020 data will be doubled every two years. This data is generated from different sources as sensors, digital pictures or videos, social media posts, as well as purchase transactions (IBM, 2012). Every day, 24 Petabytes data are being processed by Google and 400 million tweets are being sent per day (Mayer-Schönberger & Cukier, 2013). Information and communication systems secretly merge within our environment. Thus, not only many objects around us will be on the network in one way, or another through the Internet of Things (IoT), but Radio Frequency Identification (RFID) as well as sensor network technologies will also help to meet these requirements (Gubbi, Buyya, Marusic, & Palaniswami, 2013).

Surely, many companies were “doing” Big Data long before the term became worthy of market capitalization, but many factors have brought the practice into the mainstream (Murphy & Barton, 2014). According to Mayer-Schönberger and Cukier (2013), Big Data is revealing as a revolution and is disrupting the way of living, working and thinking. It is not the big numbers in big data that stick out, but the big culture changes that it imply (Mayer-Schönberger & Cukier, 2013).

Various definitions of Big Data have emerged lately. Most of them focus on the size of data in storage. Firstly, Big Data was defined by Laney (2001) as “the rapidly increasing volume of data, velocity as the data go in and out, and the variety as the dispersed type and sources of data”. Chen, Chiang, and Storey (2012) in their paper, describe significant data as large and complex data sets in applications that need advanced and unique data storage, management, analysis and visualization technologies. The three Vs of big data that Laney determine, bust the myth that Big Data is only about data volume. Size is substantial, but there are also other important attributes of significant data, named data variety and data velocity (Russom, 2011). In contrast to previous approaches, where a sampling technique was used, now it is possible to use all the observations without needing a hypothesis in advance before collecting the data (Mayer-Schönberger & Cukier, 2013). Consequently, hypothesis can be changed after being run different regression one by one, arising so to one of the significant challenges to data information, the changing of ownership (Pavolotsky, 2013). Big Data offers a competitive advantage through utilizing the vast amount of data, and radically improve the company’s performance (McAfee & Brynjolfsson, 2012), but it can also imply issues in terms of ethics and privacy (Davis & Patterson, 2012). One of the crucial ethical issues of the information age is information privacy (Stone, Gueutal, Gardner, & McClure, 1983). Stone et al. (1983) defined this information privacy as the ability that any individual has to control their information by and about themselves. Meanwhile in information ethics, according to Mason (1986) arise four critical areas of concern: privacy, as what personal information should individuals be required to disclose and under what conditions; accuracy, as what policies and standards are needed to protect individuals from errors; property, as who owns data and how should ownership be determined; and access, as who can have access to what information.

Confusingly the terms "ownership" and "use" usually substitute each other although they have a different meaning (Van Alstyne, Brynjolfsson, & Madnick, 1995). While the usage rights involve the capability to access, make, normalize, and change data, ownership means the right to establish these privileges for others (Van Alstyne et al., 1995). Mason (1986) related the information ethics concerns with ethical dilemmas arising from the collection, use, and management of information.

Base on the fact the information in a supply chain mechanism flows among the participants (Lee, Padmanabhan, & Whang, 2004) we investigate for the data management in the Big Data Supply Chain. The Data Supply Chain as a methodology to manage company’s data assets (EvanLevy, 2001), would lead into a conceptual framework of the collecting- analyze-share-feedback of the information (Groth, 2013). Further on, the data supply chain expanded the traditional corporate information life cycle in order to entourage several data sourcing, provisioning, and logistical activities that are needed to manage successfully the company's data (EvanLevy, 2001). In the context of Big Data, the participants within this chain are the data generator and the data users that

interchangeable are the individuals, the public/ development sector, and the private sector (WorldEconomicForum, 2012). For the purpose of this study we will use the taxonomy of privacy of Solove (2006) where the stakeholder grouped into data subject such as the individual and data holder such as business, government and other individuals. We will generalize these stakeholders respectively as the consumer, the government, and the business.

1.2 Problem area

Notice and consent are problematic, particularly so for Big Data as many consumers do not read privacy policies (McDonald & Faith Cranor, 2008). Privacy policies displayed on devices, such as cell phones, with small-form factors are at best difficult to read, and layered notices, providing a brief summary of the privacy terms with a link to the full privacy policy, are often unavailable (McDonald & Faith Cranor, 2008). Again, even if an individual read and understood the privacy policy, that policy may be wrong or at least incomplete (Pavolotsky, 2013). Thus, Big Data, the value of which lies in identifying secondary (and so unimagined) uses of data, stretches the practical limits of meaningful consent (Pavolotsky, 2013). Enabling individuals to exert some control over how these data are used will be an important aspect of an overall solution (McKinseyGlobalInstitute, 2011).

Big data poses other challenges to existing approaches in data privacy, pointing to the need for dialogues among stakeholders on a policy framework that can help create a sustainable data ecosystem that will strongly protect individual rights (McKinseyGlobalInstitute, 2011). It may be unreasonable to pretend to have a consent from every person who deliberately posts information for their self or others, but it is awkward to justify actions of organizations that analyze that information as ethical just because the data are accessible (Boyd & Crawford, 2012).

Although the content is publicly available does not imply that it is intended to be consumed by just anyone, leading through to serious issues involved in the ethics of online data collection and analysis (Boyd & Crawford, 2012). Users are not aware of the multiplicity of agents and algorithms currently gathering and storing their data for future use (Boyd & Crawford, 2012). But they are not necessarily conscious of all the multiple uses, profits, and other gains that come from information they have posted (Boyd & Crawford, 2012). Therefore, Big Data leads to the question that how the information will be used.

1.3 Motivation of proposed study and Research question

The critical evaluation of Big Data phenomenon is vital important as a consequence of its data driven approach. Thus, the decision we take over that data have a real and significant human

consequences (Mayer-Schönberger & Cukier, 2013) such as ethical issues of online data collection and analysis (Boyd & Crawford, 2012). Understanding of the impact of this development is crucial, and broad social discussion is needed to ensure sustainable social and economic development. Studies have clearly found Internet users' information privacy concerns to be a main originator of their disposition to disclose personal information to online companies (Son & Kim, 2008). Hence, if we lack in the balancing of social values that we are sensitive, like privacy, identity, reputation, transparency, then our Big Data society is in risk to lose these values because of the innovation and efficiency (Richards & King, 2014).

As we already presented, Big Data phenomenon is quite new approach, leading so to the emerging of new and not previously faced problems. There have been various researches for the Big Data technology approach or the advantages and opportunities that Big Data can enhance. Also, there have been several studies of the privacy in IS research, but few investigation of privacy in the Big Data Supply Chain, which is a new dimension within the individual and organization social context. Hence, we will conduct the research based on the below raised up research question:

Research question: What is the potential privacy gap in the Big Data Supply Chain?

1.4 Purpose

The purpose of our study is to identify the potential gap between privacy and openness among the stakeholders acting in a Big Data Supply Chain. The study will explore the interplay between privacy and openness in the context of information society driven by Big Data.

Any data on human subjects certainly raise privacy issues, and the real risks of abuse of such data are difficult to measure (Boyd & Crawford, 2012). The difficulty is that privacy gaps are hard to make specific if there is damage done at the time, after 20 years and so on. In the coming decade, the whole integration of our online and offline lives will become increasingly important issue facing society (Bus & Nguyen, 2013). This research is, therefore, important as the information privacy of many individuals seems to have been seriously threatened, if not compromised (Son & Kim, 2008). It urges for individuals to be able to manage information related to them, meeting with their preferences, contexts, and values, within existing social and legal boundaries (Bus & Nguyen, 2013). We will try to investigate into these potential privacy gaps in a Big Data Supply Chain in order to bridge these weaknesses and improve the interplay between government, business and consumer.

1.5 Delimitation

According to the main mission of IS field as providing an extensive approach of IT within the individual, organizational, and social contexts (Bakos & Treacy, 1986), the delimitation of our research as a critical evaluation of Big Data would be in the social perspective of it. As Big Data is becoming more common tool in business decision-making, a number of new social jeopardizes arise. The most evident is the risk of privacy violations (Bollier & Firestone, 2010). Although there are physical aspects of privacy, our study will be delimited in the information privacy concerns. People disclose personal information to obtain the benefits of a close relationship; the benefits of disclosure are balanced with an evaluation of the risks (Mary J. Culnan & Armstrong, 1999). Additionally, our study will endeavor to gain a better understanding of the risks in an individual's intention to disclose personal information, and the influence of other stakeholders.

Being unable to gather information from National Security Agency or China legislation, we delimited our study only to European Commission (EC) regulation and also the interviews from the legacy part were from EC.

The research will be delimited further on the privacy concerns measurement done towards the individual context rather than group, or the organizational level. Besides, the business delimitation is done towards the one participating in the Big Data Supply Chain.

Finally, as a delimitation within this research we considered the different feature of online and offline privacy within the limited time to study in details. While, in the offline setting, there is no clear way visually to assess a consumer's personal data when dealing with, in online firms, the information transparency is easier to be tracked (Awad & Krishnan, 2006). For this reason we chose to concentrate only in online privacy issues

1.6 Terms and definition

Gap

The term of gap is used in our research question as well as throughout the study. We used gap in the context of representing a non-desirable break in continuity (OxfordDictionaries, 2014) between the participants of the Big Data Supply Chain, allowing thus for infringements of privacy and ethical issues.

Stakeholders

We use the concept of the stakeholder through our study as a person, group or organization are interested or concerned within an organization (BusinessDictionary, 2014), for us the Big Data

Supply Chain. For the purpose of simplicity, we use these interested actors in three large groups such as the business, government and consumers.

Data Supply Chain

The Data Supply Chain seeks to broaden the traditional corporate information lifecycle in including various data sourcing, storing and logistical activities required to manage data (EvanLevy, 2014). For this purpose, we will look through Big Data within the Data Supply Chain naming though Big Data Supply Chain (EvanLevy, 2014).

Privacy

General privacy is considered as the right of freedom from secret surveillance and to determine whether, when, how and to whom one's personal information is to be revealed (BusinessDictionary, 2014). Although privacy might be categorized into four areas: physical, informational, decisional, dispositional (BusinessDictionary, 2014) we will focus our study only in the Information privacy.

Information Privacy

Information privacy is a multidimensional notion that is reliant on context and also varies with a person's life occurrences (Xu, Dinev, Smith, & Hart, 2008), and it is a limitation on revealing data or searching for them unknown or unknowable to others (BusinessDictionary, 2014).

Big Data Analytics

Advanced analytic techniques that process on Big Data (Russom, 2011).

Ethics

Ethics is called moral principles that lead the behavior or the manners of doing activity in a person (OxfordDictionaries, 2014).

1.7 Structure of the thesis

The rest of the thesis will continue as follows:

Chapter 2: Big Data socio technical phenomenon explains the theoretical guideline on which our research framework is constructed. Afterwards, we evaluate critically the implications of individual information privacy in the new Big Data ecosystem, as a socio technical phenomenon.

Chapter 3: Methodological Approach shows the aim and motivation we went throughout our research structure and methodology. We start with the research strategy and approach and further

on with data collection and analyzing considering through it the reliability, validity and generalizability of the study. Finally, we end up with thematizing and an extended research framework.

Chapter 4: Empirical results and data analysis illustrates our empirical findings as a result from the interviews and the diaries conducted. A short analysis follows each of the main themes we based our research.

Chapter 5: Discussion and findings include the observations we extracted from the empirical findings combined with the literature review. We point out and discuss the main findings from our study.

Chapter 6: Conclusion will summarize the findings inducing for further researches and limitations we went through. In other words, we will give an answer to our research question.

2 Privacy in the context of Big Data

This chapter explains the theoretical guideline on which our research framework is constructed. Firstly, we will conduct a top-down analysis of the macro environment where Big Data operates. Further on, analyzing it as Big Data Supply Chain we will find the potential privacy gap among its stakeholders. In conclusion, it terminates with the selection of five main themes where our research framework is based: Big Data Supply Chain; Individual; Government; Business; and Consumer.

2.1 Introduction

According to Recker (2013) a literature review is important in conducting research because of its enormous contribution in findings of specific problem domains, the theories that are available to analyze the raised up problem, and the methodologies appropriate for the research.

We organized our literature review as top-down analyses. We start from a macro level analysis where our research is based on, such as Big Data, and afterwards we went down to the micro analysis like the privacy. First, we started with a description of Big Data associated with the new opportunities and challenges that emerge from it. Furthermore, we approached Big Data with the Data Supply Chain to make the description of data flow more understandable (McKinseyGlobalInstitute, 2011). Second, we conducted the analysis towards the impact of Big Data in terms of privacy. Based on that, we started with the transition of privacy from physical to information one related to the technology evolution (Westin, 2003). Consequently, we reflected the analyses of privacy concerns as impinged more in the presence of a new information technology phenomenon that is perceived as a threat from the public (Mary J. Culnan, 1993; Westin, 1967). Finally, we conclude with the theoretical framework adopted from (Solove, 2006) as the taxonomy of privacy. We used the privacy calculus theory to examine the position of Big Data towards consumer's privacy as it attempts to explain the costs and the benefits towards the individual behavior of the consumer (Mary J. Culnan, 1993).

2.2 The changing landscape towards Big Data

The term "Big Data" firstly appeared in 2001(Laney, 2001). Although many of the technologies it bases on have been used in this domain for many years such as clustering, parallel computing, and

network file systems (Needham, 2013). Furthermore, the increase of storage information capacities in the last twenty years and the exponential rates increasing technological information processing capacities (Hilbert & López, 2011), lead many organizations towards the use of Big Data. The Internet and the Web, starting from the early 2000s, started to offer new data collection and opportunities within the analytical research (Chen et al., 2012). While, the Internet of Things, firstly referred from Ashton (2009), lead toward the extend of data captured nowadays as the data are generated. Because of the Internet connectivity among the things, the information systems can collect up-to-date information on physical objects and processes (Mattern & Floerkemeier, 2010), resulting thus to increase of the amount of data continuously. Nowadays, Big Data overcomes the social networking and machine-generated web logs, leading through the procession of the tremendous amount of data (Epic, 2014) that are done in parallel and quickly (Needham, 2013).

When minicomputers firstly appeared in the early 1970s, corporations started to take advantage of the distributed computer sources that became the essential ground for the “decision support system” and recognizing DSSs from the MIS (Hosack, Hall, Paradise, & Courtney, 2012). It was not until the 1990s, when the business intelligence term became popular in IT communities and business (Chen et al., 2012). However, the analytical techniques used in these systems, commonly in the 1990s, are mainly based on the statistical methods emerged from the 1970s as well as data mining techniques that evolved in the 1980s (Chen et al., 2012). Also, Big Data techniques mostly rely on elaborate commercial technologies of relational DBMS, data warehousing, OLAP, ETL and BPM (Chaudhuri, 2011).

Nowadays, as the amount of data available is emerging increasingly, and the knowledge hides within that abundant data, the processing of it is not designed for traditional business intelligence tools (AppliedDataLabs, 2014) which make Big Data stand on the top. In support of this, Big Data as shown in Figure 2.1 from Gartner’s IT Hype Cycle (Gartner, 2013), has been identified as one of the emerging technologies in IT that will take 5-10 years for market adoption.

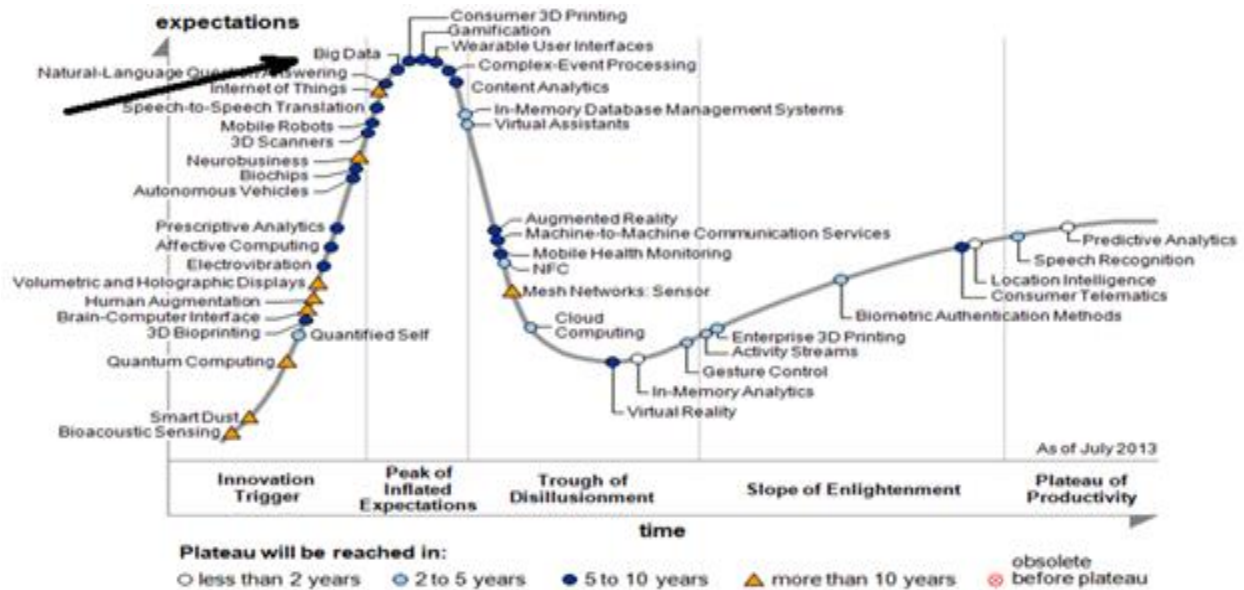


Figure 2-1 Hype Cycle for Emerging Technologies, 2013 (Gartner, 2013)

Boyd and Crawford (2012) generalized Big Data as a social-technical phenomenon that is associated with both opportunities, problems and big responsibilities (Filippi, 2014). We can find Big Data in different aspects of everyday life, from governments, e-commerce to health (Chen et al., 2012) making it a source of strategic use for both the business and the government (Bollier & Firestone, 2010). Big Data is also known as a new potential class of economic asset (WorldEconomicForum, 2012). But now, in terms of Big Data social-technical phenomenon, possible uncertainty arises as the consequence of the new data ecosystem that the technology has built up (Chen et al., 2012).

So, the correlation techniques and data analysis can lead towards both behavior and event prediction that potentially can raise new questions and challenges to deal with (Bollier & Firestone, 2010). Hence, we have to analyze ethical implication of the government and the business that get value from the consumers' (Bollier & Firestone, 2010) and the consumer's knowledge about the value they generate. In addition, the (WorldEconomicForum, 2012) emphasize that in parallel with the individual, as a data generator, there is a need for the development of instruments to assure their privacy and security. Also, the stakeholders responsible for data generations and data use can control the ethical dimension with their behaviors as the privacy policy agreement process (McDonald & Faith Cranor, 2008).

2.1.1 Big Data Supply Chain

Firstly, mention by EvanLevy (2001), the Data Supply Chain was seen as a new approach to manage the company's data assets, expanding the traditional corporate information lifecycle

management. DataSociety (2014) claims that as data moves between participants what prescribes the whole phenomenon is a Data Supply Chain. In the context of Big Data the processes that the data goes through are as follow:

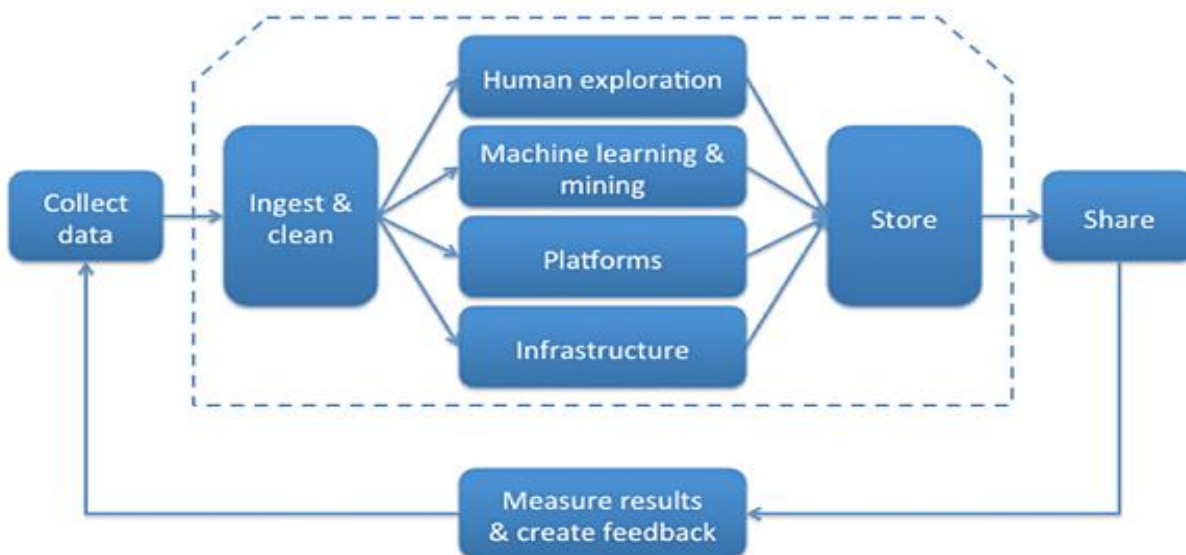


Figure 2-2 Big Data Supply Chain (Croll, 2014)

These activities helped us defining the participants that take place in a Big Data Supply Chain:

Data subjects, the one who generate the data and information (Solove, 2006). Furthermore, according to the European Directive 95-46, the data subject is any individually identified subject, directly or not, through personal number to his physical, physiological, cultural or social identity(EuropeanCommission, 2014).

Data holders, the one who collect, analyze, acting and sharing, and measure results and create feedback of the information gathered or produced (Solove, 2006). EuropeanCommission (2014) reflects data holders as the recipient who represents a natural or legal person, public authority, agency or any other entity to whom data are disclosed, whether to a third party or not.

1-Collect: The collection process within Big Data Supply Chain is based on several types and sources of data emerging nowadays (Chen et al., 2012). It can be traditional collection, as through transaction systems purchase transactions (IBM, 2012) but also, through data brokers companies (US.GOA, 2006). Data brokers, or third party (EuropeanCommission, 2014) base their business model upon the collection of information on consumers and afterwards resell it to their customers, in a private or public sector (US.GOA, 2006).

2-Big Data Analytics: The companies that compete on analytics differentiate themselves among the competitors (Davenport, 2006). Once the data is collected, it is being ingested, and manipulated (Chen et al., 2012). But, to achieve the competitive advantage from the analytics, we have to focus

on analytics effort, establish an analytical culture, hire the right people, and use the right technology (Davenport, 2006).

3-Sharing and acting: Davenport (2006) argues that the use of data has disrupted the decision making process. All the analysis made to the data gathered are not valuable if we are not able to act over them (Chen et al., 2012). But, the collection is not only a technical matter it involves even legislation, and organizational politics (Croll, 2014).

4-Measure results and create feedback: Big Data is mostly about feedback (Croll, 2014). Just the collection process or analysis will not add value without the feedback (Croll, 2014).

The data generated within the Big Data Supply Chain, is composed by three main participants such as individuals, public/development and the private sector (WorldEconomicForum, 2012) (Figure 2.3).

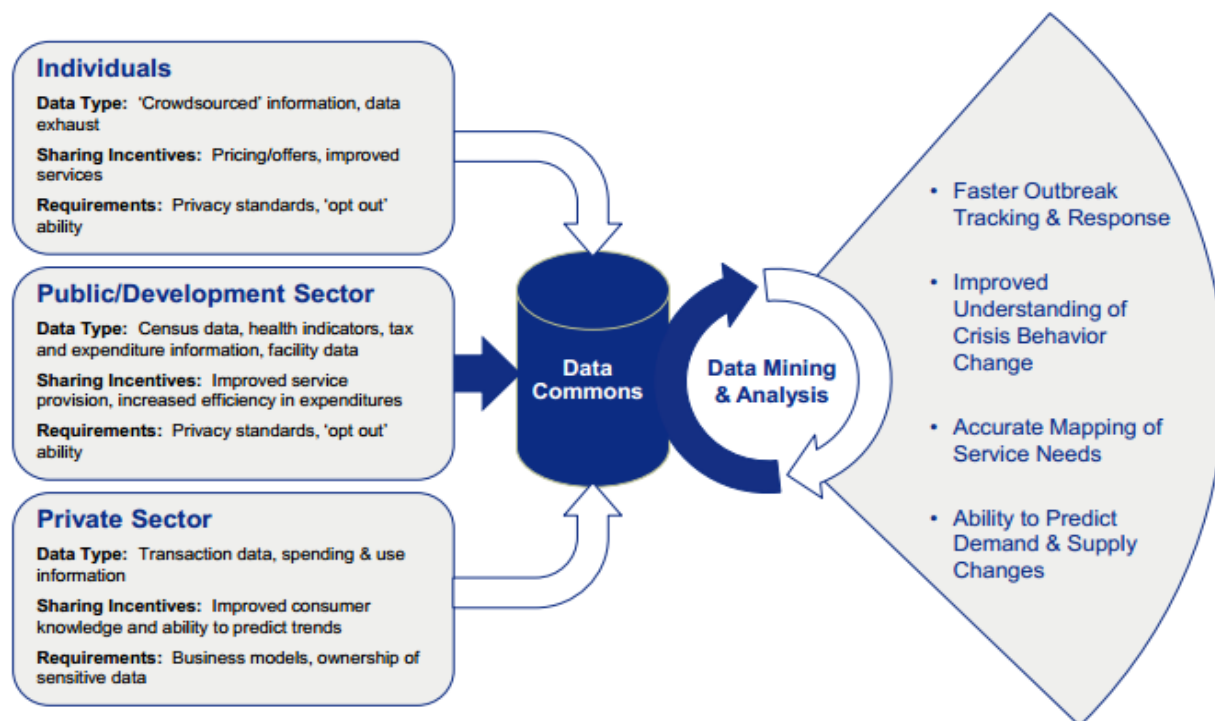


Figure 2-3 Complex Data Infrastructure/Ecosystem (WorldEconomicForum, 2012)

Further on, from the McKinseyGlobalInstitute (2011) was claimed that the data in a Data Supply Chain was roughly duplicated, leading through to disruption and eventually to a changed ownership of data. In addition consumer's privacy is a centric ongoing debate between business representatives, governments, and consumers' representatives (Mary J Culnan & Bies, 2003). Based on this approach we established to make an investigation of the stakeholders of the Big Data Supply Chain in order to characterize any potential privacy gap among them.

2.1.2 *Big Data Supply Chain in the Macro-environment*

Treating Big Data in the context of the social-technical phenomena (Boyd & Crawford, 2012), we looked through the business and social implications that affects the participants of Big Data Supply Chain (McAfee & Brynjolfsson, 2012). Although Big Data enhances the decision process within a business entity, several uncertainty social related rise up, where the most risky one is the challenge of privacy (McAfee & Brynjolfsson, 2012). After revealing the stakeholders in a Big Data Supply Chain from the first theme, we investigate into their relation with the Big Data Supply Chain to characterize any potential privacy gap among them. Accordingly, we will divide this theme into three ones:

- Big Data Supply Chain - Government
- Big Data Supply Chain – Business
- Big Data Supply Chain – Consumers

In addition, when seeing this environment in the context of a Macro - Environment four factors take place: Political, Economic, Social and Technological (Recklies, 2006). The political factor, through the government is revealed in the data regulation process e.g. the “General Data Protection Regulation” (EuropeanCommission, 2014). Moreover, privacy policies attempt to assure the privacy issues related to the data movement. But, they are complicated, long and difficult to read (McDonald & Faith Cranor, 2008).

Furthermore, some companies do not provide appropriate privacy protection for consumers because they either do not have privacy policies, or they do not comply with the Federal Information Processing (FIP) standards (Blanger & Crossler, 2011). However, there are several political initiatives to assure the privacy within the Big Data Supply Chain e.g. “Privacy of Big Data” (WhiteHouse, 2014). The economical factor and the competitive strategies are based on the vast amount of consumers’ data they possess (Mary J. Culnan & Armstrong, 1999). Because, it is hard to differentiate yourself based only on products, the achievement of differentiating is done using the data analytics (Davenport, 2006). The social factor, Big Data as a social-technical phenomenon is associated with several social opportunities (Chen et al., 2012). Several challenges also come across (Boyd & Crawford, 2012) e.g. the privacy in health care data, the prejudice because of the information ingested (Mary J. Culnan & Armstrong, 1999) from this Big Data etc. The technological factor, the low cost of the hardware and the evolution of cloud computing make it easy to store data compared with previously (Agrawal, Das, & El Abbadi, 2011).

2.2 Privacy - a Big Data challenge?!

Privacy as a concept is studied from several contexts including even the Information Systems (Pavlou, 2011). Hence, the research within its domain is quite broad. In order, to narrow the objective of the research, according to Smith, Dinev, and Xu (2011), the research in the privacy domain is based in three main aspects such as: 1) the privacy definition, 2) the privacy relationship with the other constructs 3) the privacy importance in a specific context.

For the purpose of our study the review is organized as follows: 1) the privacy definition analyzed as a transition from the physical to the information one, 2) the privacy relationship expressed as the privacy concerns but supported also with the privacy calculus, 3) the privacy importance in the context of Big Data Supply Chain.

Context is outlined as the catalyst and the phenomena that surround the external environment of the individual (Mowday & Sutton, 1993). So, Big Data perceived as the stimuli phenomena will be the environment where the potential privacy gap will be evaluated.

Moreover, the previous experience of the Internet technology that disrupted completely the privacy landscape leads the scholars towards the assumption that another disruption force like the first one we are facing with (Smith et al., 2011). Thus, Big Data as the social-technical phenomena is a potential context that will reshape the privacy debate within the participant and mostly the evaluation of the privacy concerns that rises within the enormous. In order to evaluate despite the macro environment factors we have to take in consideration even the microenvironment one that contains both internal and external factors (Jackson, Joshi, & Erhardt, 2003).

2.2.1 *Privacy ethical dimension*

Typically, general privacy is assumed under the ethics theme (Smith et al., 2011). Indeed, it is found mostly affiliated with “ethical issues” in several sociology readings (Bynum, 2008; Pearson & Saunders, 2009). General privacy has been investigated based on many ethical theories such as social contract theory, duty-based theory, stakeholder theory, virtue ethics theory, and the power–responsibility equilibrium model (Caudill & Murphy, 2000). Even though privacy is not the same with ethics and it can be analyzed without dealing with the ethical perspective (Smith et al., 2011).

2.2.2 *Privacy transition from physical to information one*

Firstly, the privacy was perceived as the physical privacy (Smith et al., 2011). Today, the communication is based on digitization and storing it as information (Blanger & Crossler, 2011). Furthermore, the information collected about individuals and groups evolved into a phenomenon,

converting information privacy into general privacy (Smith et al., 2011). Thereupon, we will use the term privacy to refer to the information privacy that we are analyzing in the context of Big Data Supply Chain.

But, what is the definition of information privacy is hard to define it clearly. "Nobody can articulate what it means" (Solove, 2006). Even if there are many definitions for it, there is not much difference on the essential elements describing it, that usually consists of any potential secondary use of individual's personal information (Belanger, Hiller, & Smith, 2002). Where secondary use indicates the data usage with the purpose outside the primary purpose it was firstly collected (Smith, Milberg, & Burke, 1996). According to a literature review of information privacy in IS the main taxonomies of the information privacy are as follow (Belanger et al., 2002):

Table 2-1 Information Privacy Taxonomy

| Author | Taxonomy | Constructs |
|-------------------------------|--|--|
| (Smith et al., 1996) | Information privacy consists of four dimensions: collection, unauthorized secondary use, improper access, and errors. | <ul style="list-style-type: none"> ▪ Collection ▪ Unauthorized secondary use ▪ Improper access ▪ Errors |
| (Solove, 2006) | Information privacy includes information collection, information processing, information dissemination, and invasion. | <ul style="list-style-type: none"> ▪ Information collection ▪ Information processing ▪ Information dissemination ▪ Invasion |
| (Skinner, Han, & Chang, 2006) | Information privacy in a joint environment is centered on time, matter, and space dimensions. Where the space dimension demonstrates the information privacy related to individual, group, and organization. | <ul style="list-style-type: none"> ▪ Time ▪ Matter ▪ Space dimensions ▪ <i>Individual</i> ▪ <i>Group</i> ▪ <i>Organization</i> |
| (Smith et al., 2011) | Information privacy research is related to individual, group, organizational, and societal. | <ul style="list-style-type: none"> ▪ Individual privacy ▪ Group privacy ▪ Organizational privacy ▪ Societal privacy |
| (Clarke, 1999) | He defines information privacy as the concern that an individual has in controlling, significantly changing, or in the administration of data about themselves. | <ul style="list-style-type: none"> ▪ Controlling |

Furthermore, the concept of information privacy existed long before information and communication technologies changed its occurrences, impacts, and management (Blanger & Crossler, 2011). With the IT development, the boundaries started to be reduced and eventually rises up the privacy paradox that associated privacy with the commodity (Bennett, 1995). Thus, a cost-benefit analysis can be embedded to it (Smith et al., 2011). We revised the model of the

information privacy evolution as a function of IT development (Westin, 2003), adapted from (Smith et al., 2011) in the context of Big Data as shown in Table 2.2.

Table 2-2 Evolution of information privacy concept following the evolution of IT (Adapted from Smith et al. (2011))

| Evolution of the Information Privacy Concept Following of IT | |
|--|---|
| Period | Characteristics |
| Privacy Baseline <i>1945-1960</i> | Limited information technology developments, high public trust in government and business sector, and general comfort with the information collection. |
| First Era of Contemporary Privacy Development <i>1961-1979</i> | Rise of information privacy and explicit social, political, and legal issue. Early recognition of political dark sides of the new technologies (Brenton 1964), formulation of the Fair Information Practices (FIP) Framework and establishing government regulatory mechanisms established such as the Privacy Act of 1974. |
| Second Era of Privacy Development <i>1980-1989</i> | Rise of the computer and network systems, database capabilities, federal legislation designed to channel the new technologies into FIP, including the Privacy Protection Act of 1984. European nations move to national data protection laws for both the private and public sectors. |
| Third Era of Privacy Development <i>1990-2000</i> | Rise of the Internet, Web 2.0 and the terrorist attack of 09/11/2001 dramatically changed the landscape of information exchange. Reported privacy concerns rose to new highs. |
| Upcoming Era <i>2001-Present</i> | Raise of Big Data , gets another insight toward the privacy issues related to the volume, variety, veracity, and value they are generated. The Data Supply Chain disrupting the ownership definition, and the <u>anonymization</u> of the information related to several concerns of privacy. |

Even though, in relationship with that commodity, privacy is still an individual and social value (Westin, 2003).

2.2.3 Privacy concerns

According to the literature review related to privacy concerns we found it as follows: Mainly privacy concerns are defined as an individual subjective judge of legitimacy from the perspective of information privacy (Malhotra, Kim, & Agarwal, 2004). More particularly the privacy concerns are associated with the organizations' information privacy policies (Smith et al., 1996). Furthermore, Internet privacy concerns represent individuals' perceptions of what happens with the information they provide via the Internet (Dinev & Hart, 2006). So, there are consumers claiming that their personal information is collected and analyzed not in a transparent way (Mary J Culnan & Bies, 2003). Privacy concerns are not new, but they emerge even more in the presence of a new information technology phenomenon, that the public perceive as a threat (Mary J. Culnan,

1993; Westin, 1967). According to Mary J. Culnan (1993) the use of technology for a strategic purpose can raise privacy concerns if they are not based on some common values. Thus, it is of a growing concern to multiple stakeholders including business leaders, privacy activists, scholars, government regulators, and individual consumers (Smith et al., 2011).

But, how are the privacy concerns measured?! There are scales using the ‘concerns for information privacy’ (CFIP) such as the collection, the errors, the secondary use, and the unauthorized access to information (Smith et al., 1996). Furthermore, Internet user information privacy concerns (IUIPC) is related to Internet users that have three first order components: collection, control, and awareness (Malhotra et al., 2004). Moreover, Smith et al. (2011) stated that the privacy concerns need a macro model to be evaluated from the perspective of their source, and the outcomes resulted from them. So, they introduced the macro model “APCO” compound of three components such as, "Antecedents => Privacy Concerns => Outcomes". This model is delimited into the personal borders of privacy, excluding the group and the organizational one (Smith et al., 2011). So, the macro-environment analyses previously leads to the antecedents for these privacy concerns in the area of political, economic, social, and technological factors (PEST). Summarizing, the measuring scale of privacy concerns associated with the respective constructs are shown on the table below.

Table 2-3 Measuring scale of privacy concerns

| Author | Measure Scale of Privacy Concerns | Constructs |
|-------------------------|---|---|
| (Smith et al., 1996) | CFIP (Concerns for Information Privacy) | <ul style="list-style-type: none"> ▪ Collection of data ▪ Unauthorized secondary use of data ▪ Improper access to data ▪ Errors in data |
| (Malhotra et al., 2004) | IUIPC (Internet User Information Privacy Concerns) | <ul style="list-style-type: none"> ▪ Collection ▪ Control ▪ Awareness |
| (Smith et al., 1996) | APCO Macro Model (Antecedents, Privacy Concerns, Outcomes) | <ul style="list-style-type: none"> ▪ Privacy experiences ▪ Privacy awareness ▪ Personality differences ▪ Demographics differences ▪ Culture/Climate ▪ Regulation ▪ Behavior reactions ▪ Trust ▪ Privacy notice ▪ Privacy benefits ▪ Privacy risk |

Nevertheless, the outcomes of privacy concerns generate a taxonomy of information privacy-protective responses known as (IPPR) (Son & Kim, 2008). Furthermore, these concerns have impact on the individual attitudes towards the regulatory environment preferences and also for the profilization willingness (Milberg, Smith, & Burke, 2000; Van Slyke, Shim, Johnson, & Jiang, 2006). Also, the concerns have impact on the trust that is related with the mobility to disclosure information apart the perceived risk (Mayer-Schönberger & Cukier, 2013; Mayer, Davis, & Schoorman, 1995) where the privacy risk is perceived as any potential loss of control related to personal information (Dinev & Hart, 2006).

Summarizing, these concerns influence the technology acceptance such as the online purchases (Malhotra et al., 2004; Smith et al., 1996). As a result, the privacy concerns have to be addressed in order to find the potential privacy gap in the Big Data Supply Chain. Furthermore, there is a privacy paradox that privacy calculus explains (Mary J. Culnan, 1993).

2.2.4 Privacy calculus theory

Privacy Calculus is the theory used to evaluate the tradeoff of costs and benefits towards the individual behavior (Mary J. Culnan, 1993) by assuming that a consequentiality tradeoff of costs and benefits is salient in determining an individual's behavioral reactions. This view is found in various works (Eddy, Stone, & Stone-Romero, 1999; Klopfer & Rubenstein, 1977) where the privacy is perceived not as a static concept but as a debate generator in terms of costs and benefits (Klopfer & Rubenstein, 1977). Thus, privacy calculus leads towards the risk-benefit approach when it is requested to disclose personal information (Mary J. Culnan, 1993). Where privacy risk is perceived as a person believes towards a high potential loss of personal information to an organization (Pavlou, Liang, & Xue, 2007). And, privacy benefit related with privacy calculus perspective is associated with the personal assumption that a particular behavior leads to the most level of outcomes (Stone & Stone, 1990) such as, financial rewards, personalization, and social adjustment benefits.

2.3 Theoretical framework

After the literature review of information privacy and Big Data we have chosen the theoretical framework where the study is based on. We will use the taxonomy of privacy (Solove, 2006) for the purpose of our study. As we have claimed previously, what prescribes the data movement is a Big Data Supply Chain. That one is composed of four main activities such as collection, processing, share, and measuring and creating feedback. According to the taxonomy of Solove (2006) information privacy includes information collection, information processing, information

dissemination, and invasion. They are the similar activities that reside in the “Big Data Supply Chain”. Thus, the potential privacy gap will be searched among the stakeholders within the Big Data Supply Chain related to the privacy taxonomy of Solove (2006). We adopted a new framework to characterize the potential privacy gap among the stakeholders in Big Data Supply Chain. Even though the privacy concerns are related to consumers, business entity, and the government (Mary J Culnan & Bies, 2003) we delimited the research within the individual level privacy concern. Moreover, Big Data in itself has not any value framework, but by analyzing the privacy implications within it as a Data Supply Chain we assure that Big Data is aligned with the values the participants inspire (Davis & Patterson, 2012). Thus, as an outcome of this analyses we came up with a framework of information privacy adapted from Solove (2006). Furthermore, we added the government intervention through its mechanisms such as regulation assurance and policy assurance within the processes of the data holder. Also, the framework is accomplished with the education impact towards the data subject.

The adopted theoretical framework is shown in the Figure 2.4.

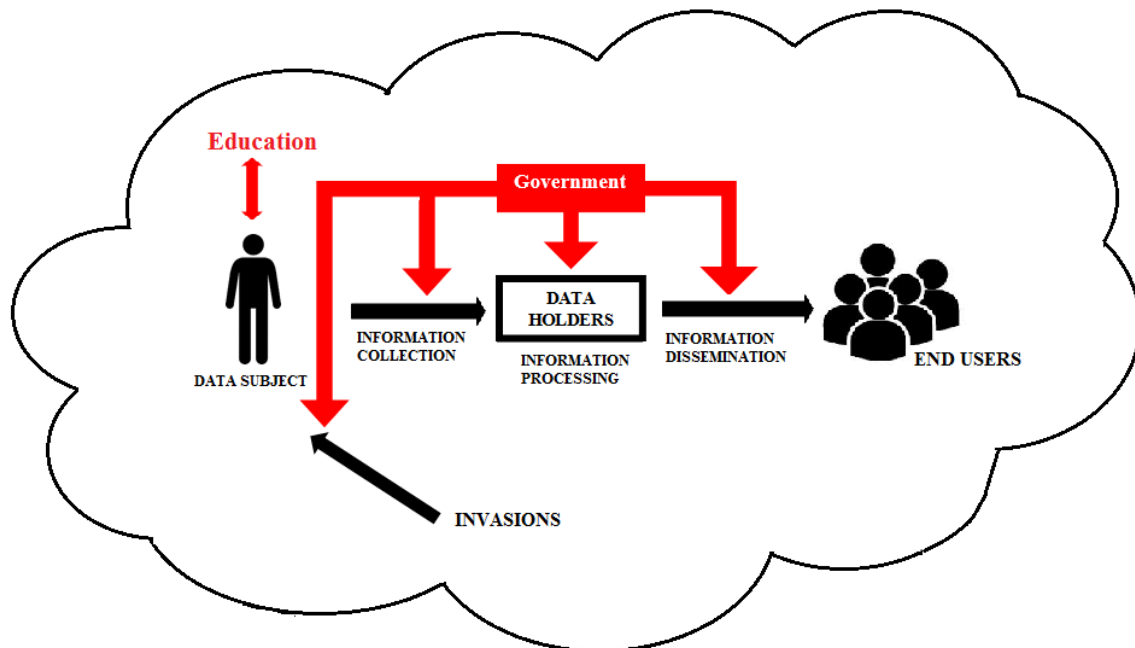


Figure 2-4 Taxonomy of Privacy in a Big Data Ecosystem (Based on Solove (2006))

2.4 Concept-centric theoretical framework

The approach we choose for the reviewed literature was the concept - centric one because of the more effective insights that can be achieved based on the key concepts of our research and to find any potential privacy gap among them (Webster & Watson, 2002). It would be a baseline for our framework. Thus, the concepts we have established our study are arranged into two main groups. The first one, macro-environmental as a composition of Big Data Supply Chain and Stakeholders participating in the Big Data Supply Chain. Further on, the micro-environmental as the individual information privacy. Also, the key concepts we described above are the one we will use in the interviews conducted for the purpose of data gathering (Kvale & Brinkmann, 2009) that will be elaborated in the next chapter. Summarizing, the concept-centric framework is illustrated in the table 2.4.

(See appendix 1 for the explanation of each of the subthemes)

Table 2-4 The concept-centric framework

| Literature | Theme |
|---|--|
| Big Data Supply Chain (BDSC) | |
| <ul style="list-style-type: none"> ▪ (EvanLevy, 2001) ▪ (Solove, 2006) ▪ (Solove, 2006) ▪ (Mary J. Culnan & Williams, 2009) | <ul style="list-style-type: none"> ▪ Data Subject (DS) ▪ Data Holder (DH) ▪ Ownership (OW) |
| BDSC - Government | |
| <ul style="list-style-type: none"> ▪ (Smith et al., 1996) ▪ (Smith et al., 1996) ▪ (Solove, 2006) | <ul style="list-style-type: none"> ▪ Legal Assurance (LA) ▪ Policy Assurance (POA) ▪ Invasion (IN) |
| BDSC - Business | |
| <ul style="list-style-type: none"> ▪ (EvanLevy, 2001) ▪ (Filippi, 2014) ▪ (Smith et al., 1996) | <ul style="list-style-type: none"> ▪ Management (MA) ▪ Responsibility (RE) ▪ Secondary Use (SU) |
| BDSC - Consumer | |
| <ul style="list-style-type: none"> ▪ (Stone & Stone, 1990) ▪ (Mary J Culnan & Bies, 2003) ▪ (Dinev & Hart, 2006) | <ul style="list-style-type: none"> ▪ Privacy Benefits (PB) ▪ Transparency (TP) ▪ Privacy Concerns (PC) |
| BDSC - Individual | |
| <ul style="list-style-type: none"> ▪ (Malhotra et al., 2004) ▪ (Dinev & Hart, 2006; Malhotra et al., 2004) ▪ (Malhotra et al., 2004) | <ul style="list-style-type: none"> ▪ Trust (TR) ▪ Privacy Awareness (PA) ▪ Educated Public (EP) |

3 Methodological Approach

In this chapter we offer the aim and motivation we went through out our research structure and methodology. As part of this, we extend our research framework by conducting a review of Privacy and Big Data Supply Chain literature, in order to dig into a potential gap between privacy and openness among stakeholders in relation to each of the theoretical themes established in our framework. The Micro and Macro Models factors are used to help lead our data collection through interview and diary process. Finally, we describe and analyze in details the results in our research as well as the steps we took to provide the quality of our research.

3.1 Research Strategy

Our research aims to identify the potential gap between privacy and openness among stakeholders acting in a Big Data Supply Chain. Thus, our focus when selecting the appropriate approach for our research was choosing the best methodology for our data collection requirement and analysis. We chose to discuss directly with different sector experts within the Big Data ecosystem and found it most appropriate to observe the daily habits of several individuals for one week through a diary written by them. Indeed, diverse interpretations to what counts as good quality work in our research will exist (Seale, 1999). The conduct of a qualitative approach for our research will help us to study the three stakeholders as well as the social frameworks in which they live, work and behave (Recker, 2013). In our study we assume that social reality takes form based on the social contexts and individual experiences that allow us for subjective interpretations (epistemology) (Bhattacharjee, 2012). The qualitative approach is used to inquiry the collection of data in a normal setting tangible to the participants of the study, while data analysis determine patterns or themes (Creswell, 2012). Brinkmann and Kvale (2005) evaluate the qualitative research interview as the one which enquiries individuals' life in detail. Further on they say that it allows researchers to describe personal aspects of people's life. Diary methods estimates observational research and are mainly beneficial in situations where first-hand observations are not possible (Czarniawska-Joerges, 2007). Latham (2003) sees the diary as a way of performance or reportage of the week of the participants.

Initially, we thought of conducting a mixed method, but then again we found unsuitable to ask questions regarding privacy and ethical issues to consumers. Eventually we would get in the end only their importance perceivment of personal information and nothing about our assumption of having a neglected behavior. As Sapsford (2006) argues, when measuring traits of personal behavior the direct question would probably be unsuitable in any circumstance, because there is no motivation to assume that the respondents are able to place themselves in terms of the researcher's theoretical constructs.

In conclusion our study, even because of the deficiency of considerably number of observation, cannot be largely generalizable. According to Kvale and Brinkmann (2009), the lack of generalizable is a feature that characterize the qualitative research. Nevertheless, we expect our study to be a good start up for future research and further quantitatively validation in the context of individual privacy within the Big Data Supply Chain.

3.2 Research Approach

The descriptive research, according to Sandelowski (2000) is the lower form of inquiry, while the qualitative descriptive studies designed in sampling, data collection and analysis, summarize the everyday life events. Therefore we reflected this approach as most suitable for our study and the fact that we want to show the potential gap between privacy and openness among the participants acting in a Big Data Supply Chain. Our main goal was to infer the findings from the data describing to a proposition that best describe the potential gap among stakeholders (Josephson & Josephson, 1996). In our study we aim not to determine the three stakeholders as the only one taking part in a Big Data Supply Chain, but rather digging in to discover where this gap lays on and which are the reasons for that, generating so enough variations to yield desirable results (McGrath, 2001). Further on, deductive theory derived from more general principles (Ezzy, 2013) to determine other implied stakeholders and causes for gaps as stated in our study may strengthen our research.

3.3 Data collection

Initially, we had two important institutions agreements in different time, in Sweden, for assisting us during our thesis with their data, documents, help and expertise. But meanwhile they resigned, the first one found limited the time for the thesis, and the second company was not comfortable with our research toward the privacy.

After that, we decided to conduct a mix research strategy, interviews for experts (business side and legal side) and a quantitative study for the individuals. But quantitative research can be quite challenging as a high amount of sampling is needed in order to make it generalizable and avoid bias (Kvale & Brinkmann, 2009). Thus, we changed direction into diaries as a qualitative research. Further on, we would have preferred a face-to-face method for data collection, as they give a rich qualitative response rates (Bhattacharjee, 2012), due to the time limit and the dispersed location, we used Skype interview. Through this mean, we were also able of recording and afterward transcribing it. Though it can be disadvantaged because of the missing image and body gestures

rather than just words (Kvale & Brinkmann, 2009), the distance that is imposed through phone, made our interviewees free and comfortable. For diaries we chose convenience sample (Marshall, 1996) as a technique that allows us in selecting most accessible subjects with fewer costs in terms of time, reachable and money. Although it might lead for poor quality data and credibility (Marshall, 1996), the restriction in time justifies this selection. For the purpose of generalizability, we chose diarists in different age and residence. We asked them to keep a diary for the specified issue we were interested in to study. When a diary is well managed, it can be used to gather data that are not obtainable through interviews or other data collection (Alaszewski, 2006). Despite, the reserved confession of their activities that we can assume, the use of time in a typical day life of an individual can tell a great amount about a social pattern (Corti, 1993). We were focused in a structured diary for our study, and also provided instructions and questions on how to fill and keep it. Further on we explained the purpose of our study to the diarists when we had the first interview, and tried to persuade them to be part of this study by keeping an everyday diary.

3.4 Thematizing of study

Our research phenomena will be investigate into three categories of actors, part of the three-dimensional area: Consumers, Government and Business. The collection of data will be made in two different ways. We chose to conduct interviews for data collection of Government and Business and diary for the Consumers. In order to structure our study we used Kvale and Brinkmann (2009) and Alaszewski (2006) books, respectively for qualitative research interview and diaries for social research. Kvale and Brinkmann (2009) suggests to use the linear progression of seven stages, as shown in Figure 3.1, from the original purpose to the final report. Although they prognoses that due to adoption of new conditions that might occur, the interviewer's initial scope and purpose might evolve. Alaszewski (2006) rises some key issues in designing, choosing the population or analyzing the diary, which a researcher should consider when choosing diary as their qualitative research. Figure 3.2 visualize the steps carried out in a diary investigation, adapted from the steps Kvale and Brinkmann (2009) used.

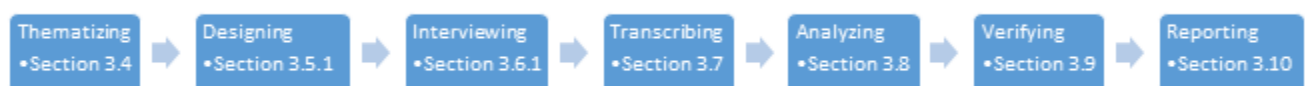


Figure 3-1 Seven stages of an Interview Investigation (Kvale &Brinkmann (2009))



Figure 3-2 Diary Investigation (Based on: Alaszewski (2006))

During the first stage, thematizing, we will express the purpose of our study and the outset of our research question previously the interview begins (Kvale & Brinkmann, 2009). Further on, we have to clarify the theme of our study (Kvale & Brinkmann, 2009). Which means developing a theoretical understanding of the phenomena we wish to investigate. In addition a base from the literature review, the hypothesis derived from social theory, could be tested against the observations and then enriched and with our findings and knowledge throughout the way (Alaszewski, 2006).

Our study was initially inspired for the use of our information from third parties. We have been reading a number of forums and articles from New York Times, IBM, Gartner, Forbes and other press release from EU, NSA etc., which raised up the Big Data ethical issue, and we decided to investigate some of the questions that previous research let as unsolved. Several thoughts and hypothesis answer regarding our research question were assumed from the beginning of the study. Although they were initially based on the literature review, later investigation through interviews and diary changed somehow the initial answer we had in mind. This was a crucial component in thematizing our study through interview (Kvale & Brinkmann, 2009) and dairy research as it constructed a knowledge base where to develop our further findings. We performed a PEST analysis previously the interview and the release of the diary, in order to make a strategic plan (Recklies, 2006) of the Big Data ecosystem.

Finally we concentrated in the interview responses as well as in the diaries, and provide for both of them coding based on the data gathered.

3.4.1 *Big Data Supply Chain*

According to the aim of our study where we were interested to find the interplay between privacy and openness driven by Big Data, we wanted to investigate within the Big Data Supply Chain as the information in a supply chain mechanism flows among the participants (Lee et al., 2004).

Based on the fact that data in a Data Supply Chain is being duplicated leads into changing ownership (DataSociety, 2014), we decided to make an investigation of the participants of the Data Supply Chain. We chose this theme initially to define the participants, such as individuals, public/ development and the private sector (WorldEconomicForum, 2012) that act as data subject and data holder. Additionally we explored on the roles that consumer, government, and business entity have in this Data Supply Chain.

1. Data subjects as the one who generate the data and information.
2. Data holders, the one who collect, analyze, store, and share the information gathered or produced.
3. Their relationships can interchange sometime, meaning a data holder can become a data subject and vice versa. For this reason we also looked to the data ownership, and its management within the Data Supply Chain.

3.4.2 *Big Data Supply Chain – Participants*

After revealing the participants in a Big Data Supply Chain from the first theme, in reliance to their individual interplay within the Big Data Supply Chain we chose to adopt a theme for each relationship. Hence, we split this theme into three ones:

- Big Data Supply Chain - Government
- Big Data Supply Chain - Business
- Big Data Supply Chain - Consumers

According to that we chose subthemes as:

1. Legal assurance and its effectiveness in the Internet age we are living.
For this purpose we also looked through the invasion of the government within the individual data privacy.
2. Policy assurance, as it is complicated, long and difficult to read (McDonald & Faith Cranor, 2008).
3. Management for data protection.
For this component we also investigated through the responsibility among the participants, how will data be managed, and who should be responsible for its integrity in a case of secondary use.
4. Further on we used secondary use of data as a consequence either from lack of legal assurance or flaw from policy assurance.
5. We used transparency as a subtheme, because the importance it has as a component in putting light in the manageability of personal information.
The privacy benefit of the consumer from one side and his concerns toward privacy issues were factors that explained this component.

3.4.3 *Big Data Supply Chain – Individual*

According to the fact that there are several factors such as culture, regulatory laws, previous experiences, and personal characteristics that affect an individual in showing various level of information privacy concerns (Malhotra et al., 2004), we chose to study this theme apart of the previous one BDSC- Consumer. Individuals from different countries and different cultures, values, and laws, might generate different perceptions of information privacy and even their impacts toward them might be diverse (Blanger & Crossler, 2011).

From this theme we issued two potential gap components that both relies on their privacy awareness:

1. Educated public as it is only through it that we can have any form of effective policing.
2. Trust as there is increasingly people sharing information, which are misrepresentation of the real world and disruption for the Big Data validity analytics.

3.4.4 *Extended research framework*

Based on the literature review, European Commission press, and PEST analysis we came into conclusion firstly suggesting the stakeholders in a Big Data Supply Chain with compounding data subject and data holders, and afterwards we extended the theoretical framework by connecting the theme-based framework with potential gap components as shown in Table 3.1.

The theme of Big Data Supply Chain as the general ecosystem where the phenomena lies, we associated it with the ownership issue. Further on from each of three themes in Big Data Supply Chain - Participants we associated them with potential gap components such as legal/policy assurance, management, responsibility, secondary use and transparency. Finally we added trust educated public and privacy awareness under the BDSC- Individual theme as the last potential gap components in this privacy gap.

The division of the information into macro and micro environmental framework and the analysis we previously made based on the literature, but also the potential gap components we provided, supported us in addressing the accurate questions for interviews and diary as well as when analyzing the content.

Table 3-1 Extended research framework

| Themes | Theories | Potential Gap Components |
|--|-------------------------|--|
| Big Data Supply Chain | PEST analysis | <ul style="list-style-type: none"> ▪ Ownership |
| BDSC – Participants <ul style="list-style-type: none"> • BDSC - Government • BDSC - Business • BDSC - Consumer | Privacy Calculus Theory | <ul style="list-style-type: none"> ▪ Legal Assurance ▪ Policy Assurance ▪ Management ▪ Responsibility ▪ Secondary Use ▪ Transparency |
| BDSC - Individual | Privacy Calculus Theory | <ul style="list-style-type: none"> ▪ Trust ▪ Educated Public ▪ Privacy Awareness |

3.5 Design of collection techniques

3.5.1 Design of interview guides

We used a semi-structured type of interview for the experts where we outlined the topics and questions to be covered, but also allowed the interviewee to be open for his judgment and point of direction (Kvale & Brinkmann, 2009). We explained the purpose of our study from the beginning and in addition set direct questions as ‘why’, ‘what’ and ‘how’ questions (Kvale & Brinkmann, 2009). Thus, we developed two interview guides, based on our research framework. The first with study’s thematic research question and the second with interview questions to be directed (Kvale & Brinkmann, 2009) as shown in appendix 2 and 2.1 Our prior literature findings helped us in knowing what to ask, and simplify the meaning to the answers during the interview to our study (Kvale & Brinkmann, 2009). The interviews, in consequence, were designed to be flexible based on the type of the experts we were going to interview, whether from Big Data experts or regulatory researchers within the Big Data ecosystem. We were attentive in listening to the answers as we encouraged them with second question to keep the direction of our research investigation. The interview guide was based accordingly to the conduct of interviews of Kvale and Brinkmann (2009) keeping a semi-structured form and including an outline of the subjects to be covered through the interview.

The interview guide we provided was mainly separated into 7 Parts:

Part 1: Introduction and general questions. We mainly introduced and thanked the interviewee and asked him for his background and some general questions about Big Data as a concept, its opportunities and challenges.

Part 2: Management This part provided questions related to the management of Big Data within a company, which department makes more use of it, who is the one keeping responsibilities within the organization and how are the providers of data from where companies do get data to analyze.

Part 3: Authority This following part covers questions regarding the authority. It covers the questions about legal assurance, policies of data transfer, data management due to ethical issues, misinterpretation, and finally responsibilities toward data analytics.

Part 4: Ownership We addressed some important questions about the ownership information, changing ownership, data collection, integrity and permanent privacy choices.

Part 5: Big Data Privacy This is where we aimed to know the point of view of the interviewee toward privacy. We also dig into the privacy issues and obligations and use of technology in the context of privacy.

Part 6: Future vision This was the pre- final part, where we wanted to investigate through the future of privacy and the interviewee concerns about privacy.

Part 7: Closing and debrief. In the end, we asked the interviewee if there were any other issue that he would wanted to add and that we did not covered. Next to that we thanked the interviewees for their collaboration and let them know when we would send the interview transcriptions.

A full summary of the interview questions related to our research framework is in appendix 2.

3.5.2 *Design of diary instructions*

We used a structured diary for collecting the data from consumers (individuals) (Alaszewski, 2006). The diary was asked to be kept for one week (7 days), and for each day there was a timeline to follow and questions to be considered when writing it (Alaszewski, 2006). We also provided a set of instructions for diarists for how were they expected to keep it and fill the diary recording system (Alaszewski, 2006). In addition the set of question following the instructions were in accordance to the purpose of the study and the structure of it (Corti, 1993). We checked after 1-2 days during that we hand in the diary, to verify for accuracy and if the diarist had any questions regarding the use of it (Alaszewski, 2006). The diary's formal structure imposed the diarists to have regular entries in order to insure for relevant data (Alaszewski, 2006). For the format and design of the diary we followed the guidelines of (Corti, 1993). Although the instructions in the first page of the diary, we also provided a sample of how the diary could be kept (Corti, 1993).

According to (Corti, 1993) our diary was designed as:

1. A 7 structured pages for each day.
2. The first page contains a clear set of instructions on how to keep and fill the diary, emphasizing the importance of writing it as soon as the events occur.

3. A set of questions for the respondent were also placed in the first page of the diary to be considered accordingly while filling it, or used as a questionnaire in case of aborting the keeping of it.
 4. The second page contained a sample of model filled correctly for one day.
 5. The subsequent days had denoted the day and date of the diary and the time grouped in 4 time frames.
 6. One final page for any further comment as it might helped us in our coding and analyzing.
- A full summary of the diary design can be found in appendix 3.
- Finally after getting back the diary from the respondents we made a call interview asking only one final question which would contribute in the analysis of the diary (Corti, 1993).

3.6 Selection respondents

For the purpose of our study we chose different set of respondents for interviews and diarist. In order to give our study generalizability, we selected two experts to interview for each category (Big Data experts, and regulatory experts in European Commission). Further on for diarist, we did a convenience sample (Marshall, 1996), selecting 5 respondents from different ages, living in different regions but with residence in Europe for at least the last 10 months.

3.6.1 Selection of interviewee

Cees Boon is the director at Scenarius, which is a management consulting company in Netherland, for almost 4 years. He has been working for more than 19 years as a retail intelligence consultant at Newway. We found Cees appropriate to our research to interview as an expert in Big Data ethics, as he was the founder of many professional network groups but especially of Big Data Ethics. Cees has also experience in Business intelligence, Science and IT.

Jon Page, now a freelance consultant motivated in defining BI Strategy and ensuring linkage with Corporate Strategy. He is the 'thought leader' for EMEA in The Big Data Institute (Stockholm). As he is working recently with HP and EMC driving Big Data solution development and sales, we found him adequate to interview as an expert in the Big Data field. The previous background in Business Intelligence and Data Warehousing in Oracle EMEA, and also working for seven years and being a founder member of Teradata Europe, added him expertise for the questions we planned to have. Has directly assisted in the design of some of Europe's largest Data Warehouses.

Trevor Peirce is the Activity Chain Leader of governance, security, and privacy in European Commission - Internet of Things, European Research Cluster (Brussels). And also, a Public Policy Activity Leader and CFO in RFID, Brussels, particularly orientated toward the European Parliament, European Commission and European Member State Governments. The engagement

in European Commission toward privacy and security but also working as a Supply Chain Technology Consultant made him relevant for our investigation. We were interested mostly on the reaction, role and contribute of the regulatory intermediate role into the Big Data Supply Chain, which Trevor had more than 8 years' experience in that area.

Interviewer 4 is an affiliated senior researcher at Institute of European Law at K.U. Leuven. His experience also as a member legal service at European Commission and at European Data Protection Supervisor made him a valuable respondent for our study. Added his study within the field and his recent PhD thesis about data protection and personal data, as well as some other publications related to it, despite the appreciated knowledge for our study also give us finely insights.

Table 3.2 outlines the interview selections and respective interview details.

Table 3-2 Data collection (Interviews) summarize

| Interviewee | | Interview | | |
|---------------------------|---|---|------------------|---------------|
| Name | Job Title | Date (Duration) | Method | Transcription |
| Experts | | | | |
| Cees Boon | Director at Scenarios Business to Consumer Intelligence | 25 th April 2014 (59 minutes) | Skype voice call | Appendix 4 |
| Jon Page | Thought leader EMEA at Big Data Institute | 30 th April 2014 (56 minutes) | Skype video call | Appendix 5 |
| Regulatory Experts | | | | |
| Trevor Peirce | Public Policy Activity Chain Leader - Governance, Security and Privacy at EC | 29 th April 2014 (45 minutes) | Skype voice call | Appendix 6 |
| Interviewer 4 | Affiliated senior researcher at Institute of European Law European Law-K.U Leuven | 07 th May 2014 (40 minutes) | Skype voice call | Appendix 7 |

3.6.2 Contact with diarists

The approach for recruiting diarists was related to the purpose of our research and the data which we were seeking to get (Alaszewski, 2006). Additionally the selection of the cases in the study were guided by the nature of the design (Alaszewski, 2006). Through the diary research methodology we tried to make generalizations for a large population (Alaszewski, 2006), by choosing only a set of respondents. Thus we used a convenience sample (Marshall, 1996) as a technique that allows us in selecting most accessible subjects, choosing thus different people from

18-30 years old, male-female, residing in Europe for at least 10 month . We were interested in our study not in a specific group of population but rather as randomized as possible. Although, we would have prefer to choose a larger number of respondents, to have more generalizable results (Alaszewski, 2006), that was conditional to the time required for coding. Lacking that, a random selection process was used to reduce the probability of bias undermining the generalizability of the findings (Alaszewski, 2006).

Table 3.3 outlines the diarist selections and respective diary details.

Table 3-3 Data collection (Diarists) summarize

| Diarist | | | | Diary | | |
|-----------|-----|--------|--------------------|-----------------------------------|---------|---------------|
| Name | Age | Gender | Residence | Date (Duration) | Method | Transcription |
| Diarist 1 | 21 | Female | Italy | 4/23/2014 – 4/29/2014 (7 days) | Written | Appendix 8 |
| Diarist 2 | 23 | Female | Albania | 4/23/2014 – 4/29/2014 (7 days) | Written | Appendix 9 |
| Diarist 3 | 22 | Female | Germany | 5/03/2014 – 5/09/2014 (7 days) | Written | Appendix 10 |
| Diarist 4 | 28 | Male | India ¹ | 4/23/2014 – 4/29/2014 (7 days) | Written | Appendix 11 |
| Diarist 5 | 22 | Male | Sweden | 4/24/2014 – 4/30/2014 (7 days) | Written | Appendix 12 |

1

3.7 Transcribing

The transcribe of the interview recordings, and diaries will enable the produce of a neat, typed copy (Kvale & Brinkmann, 2009). Kvale and Brinkmann (2009) suggests to be cautious when transcribing from interview to typed copy as the spoken context might change or transform due to other factors as body language rather than only spoken words. Although Kvale and Brinkmann (2009) say that there is no only way to transcribe as it depends from the purpose and the analysis of it. We were quite attentive and wanted to capture as much as possible from the interviewee, transcribing thus, in a narrative way rather than artistic. Nevertheless, we avoided pauses, conjunctions between sentences, or other clarifications between talking etc. beyond the general answer to the question itself. While for the diaries, it was easier because the diarist chose by themselves the writing style they wanted. Interviews, as well as diaries, were conducted in English. In consequence no further ambiguity in translating them was shown. For the interviews we made, after transcription, we sent it back to the interviewees asking if we did any unintended

¹ For the last 10 months has been living in Sweden.

misinterpretation to be corrected. Kvale and Brinkmann (2009) advice that the transcribe should represent a good, careful attempt to reproduce the interview, rather than just being accurate. After coding the interview, we sent the interview a copy so they would also know how their words were interpreted, but we didn't do the same as per diaries due to the difference in expertise between them. We did the transcribe as soon as the interview ended so our memories regarding it were fresh in case of not understanding the voice from recording. We also cross checked between us the transcribe of the interview to be sure that nothing was missed or misinterpreted.

3.8 Analyzing

Our sources for collecting data were interviews and diaries. Kvale and Brinkmann (2009) recommends researchers to think in advance of how they will do the analyze of the transcription of interviews, before they start doing interviews, as during that point principally it is too late to start thinking about analyzing. The method of analyses decided will guide the preparation of the interview guide, its process and in the end the transcription. (Kvale & Brinkmann, 2009). Qualitative content analysis can be reflexive and interactive as the researcher continuously modify the treatment of data to assist new data and new insights about them (Sandelowski, 2000). The written text created by diaries does not have the inbuilt structure of diaries used in experimental or survey research, so the researcher needs to decide how to manage this text. (Alaszewski, 2006). We used coding which categorizes the text into segments and categories (Bhattacharjee, 2012). For diaries, the process of involving coding of precise entries can be very intense, almost in the same way as it is for treating qualitative interview transcripts (Corti, 1993). Coding according to Kvale and Brinkmann (2009) includes assigning one or more keywords to a document segment which allowed us to uncover potential gaps within the Big Data Supply Chain. The coding of the text's meaning into categories made it possible to quantify how often specific themes were addressed in the text (Kvale & Brinkmann, 2009) themes that we used for our research framework.





Content analysis in diaries involves taking a number of written texts such as diaries, breaking them into their constituent parts and re-assembling these parts into a new scientific text (Alaszewski, 2006). The process of identifying themes on diary, indeed begins after the content of it have been transcribed (Alaszewski, 2006), that in our case was no need as the diary was kept in written form. The coding process in both forms required observing the text, rereading it many times, in order to identify and confirm themes and to organize and summarize all the evidence relevant to each theme. We used methods such as investigator triangulation when coding in order to ensure inter-reliability within the process (Golafshani, 2003). We code the interviews and diaries transcription separately and then exchange it between us to establish thus the existing and new codes we might add during coding, and compered coding with each other in order to conclude with a final established coding structure (Table 3.4 and Table 3.5). Additionally we used text coding for each

theme and subheading of it for interviews, and color coding for three main themes we were interested in diaries.

Table 3-4 Coding structure for data analysis Interviews

| Big Data Supply Chain | BDSC – Government | BDSC – Business | BDSC - Consumer | BDSC – Individual |
|-----------------------|---------------------------|------------------------|--------------------------|---------------------------|
| Data Subject (DS) | Legal Assurance (LA) | Management (MA) | Privacy Benefits (PB) | Privacy Awareness (PA) |
| Data Holder (DH) | Policy Assurance (POA) | Responsibility (RE) | Transparency (TP) | Educated Public (EP) |
| Ownership (OW) | Invasion (IN) | Secondary Use (SU) | Privacy Concerns (PC) | Trust (TR) |

Table 3-5 Coding structure for data analysis Diaries

| | |
|--------------------------|---|
| Usability |  |
| Privacy Concerns |  |
| Privacy Benefits |  |
| Privacy Awareness |  |

3.9 Ensuring research quality

3.9.1 Reliability

When using qualitative approaches, Seale (1999) says that there need to be reassured into a skeptical audience. Qualitative studies are different in those of quantitative as they use credibility and dependence, and as the validity in such qualitative research do not just provide measure of the studied phenomena but rather show whether the data reflects the interested phenomenon (Kvale & Brinkmann, 2009). We tried to obtain this objective by selecting interviewees from different sector and different background and employing investigator triangulation in coding (Seale, 1999) and by providing a more reliable method of research for the consumer such as an observation method. Finally we used semi-structured interviews so that respondents could feel free to overpass the formal question, and diaries as a narrative way for an individual to get express as he wishes to. Although, structuring the diaries including written instructions and sampling how to be used are relatively impersonal ways of influencing diarists (Alaszewski, 2006).

3.9.2 *Validity*

In our study we try to achieve validity by providing as much rich description as possible, selecting interviewees with diverse backgrounds as well as diarists from different ages and different countries, and further on employing methods such as investigator triangulation when coding to establish inter-reliability within the process (Golafshani, 2003). As we separately coded interviews and diaries using though an investigator triangulation (Seale, 1999). Further on, we approached other supporting methods to make our findings agile and credible (Seale, 1999). Thus we send only the coded interview to our interviewees and have feedback according to the interpretation we made whether correct or not. While for our diarist, as Corti (1993) recommends, we checked during the keeping diary, whether they had questions, they were inserting occurrences continuously and reminding them to add any small detail as it might be valuable to us.

3.9.3 *Bias*

Unbiased sampling processes leads into producing representative outcomes which is causal and supports explanation and prediction (Josephson & Josephson, 1996). Seale (1999) recommends using several methods at once so that the biases of any one method might be canceled out by those of others. Thus we used both interviews and diaries to limit the potentiality of biases. The sampling process, when it is unbiased, it is explained also as part of the sample frequency, and any independent data for or against it bears on its reliability as part of the explanation (Josephson & Josephson, 1996). Although Corti (1993), says that diaries are more prone to errors derived from respondent conditioning, not properly keeping diaries, inadequate recall of events, insufficient cooperation and sample selection bias. Respondents, also, might change their behavior due to the fact of keeping a diary (Corti, 1993), so the period of which it is being kept has to be long enough to depict the behavior of diarists (Corti, 1993).

3.9.4 *Ethics*

The researcher should follow a group of common practices that scientific researchers use in order to avoid the ethical issues (Bhattacharjee, 2012). Kvale and Brinkmann (2009) suggest moreover, the consideration of ethical implications at every stage of research. In addition, we asked the interviewees and diarist permission to be part of our study and whether they wished to stay anonymous or not (Bhattacharjee, 2012). Moreover, for took for interviews the consent to record their interview (Creswell, 2012). For both interviewees and diarists make them know the possibility to withdraw the continuously of being part of the study at any time (Creswell, 2012). Finally, using Kvale and Brinkmann (2009) we send to interviewees the transcribe of their recordings asking whether they approve how we interpreted and whether it was no conflict of interest for them to publish their names. For three interviewers it was alright proceeding while we kept anonymous due to his request the last interviewer. Whereas for diarist, as we try to use a convenience sample, there was no need for informed consent, as we kept their names anonymized.

3.10 Reporting

The writing of the report is the last stage of the seven method stages of an interview study. As such Kvale and Brinkmann (2009) recommend to take the final report in consideration from the beginning of the inquiry. Thus working toward the final report from the start contribute to a readable report of methodologically well-substantiated and interesting findings (Kvale & Brinkmann, 2009). Moreover, Recker (2013) recommends using a structure that audience is familiar with which allows them to navigate easily. In order to evaluate the quality, validity and transferability of the findings, information concerning the methodological steps of an investigation is required (Kvale & Brinkmann, 2009). The reader of a report was introduced to the methodical procedures in order to evaluate the trustworthiness of the results (Kvale & Brinkmann, 2009). Then we used quotes as the usual style of presenting the findings (Kvale & Brinkmann, 2009), giving the reader the impression of the interview content. Additionally we used short contextualized quotes related to the general text. That meant that the quotes were fragmented helping the reader to get more into the context of the study (Kvale & Brinkmann, 2009). Finally transferability as a mean to replace applicability and external validity (Seale, 1999) was applied in our research.

4 Empirical findings and data analysis

In this chapter, we declare our empirical findings. According to the methodology, defined in the previous chapter, both interview and diary are built up in the theoretical framework presented in Chapter 2. Thus, this chapter is the display of our empirical findings as a result from the interviews and the diaries conducted previously. The empirical results are based on the thematizing done on the previous chapter, on the five themes that we have based our research.

4.1 The declaration of empirical findings

The empirical findings of our research are grouped in five major themes, based on the issues we revealed from the interview and diary data. For both of them, we used coding, structured as it was shown in the Methodology Chapter (Figure 3.4 and Figure 3.5 for diaries). Additionally under each theme we extracted crucial observations that helped us define the research framework. All of them were supported with citations taken from parts of interviews or diaries. Further on, for each theme and corresponding observation, a short summary of them will follow up. Finally we conclude our empirical findings with an analysis of the themes we based our study. Our citations in the empirical findings are referenced to the according interview or diary available in the appendix. In the analogy, referencing (7:15) will refer to a statement found in appendix 7, line (in the right of the table) 15 for interviews, while for diaries we used referencing e.g. Diarist (2, 9) for diarist number 2 found in appendix 9.

4.2 Big Data Supply Chain

According to one of the interviewed experts, referring to Big Data as Big Data Supply Chain is an appropriate approach. “I like the fact looking the data in supply chain. Because I think data is putted into information and information into insights and insights into usable form. Data in itself is not usable only if it is provided.” (4:36)

4.2.1 Data Subject

The empirical findings towards the data subject lead into the reveal of the data sources within Big Data Supply Chain. We found different data sources. Data is coming from traditional sources (5:20) and social media (6:10; 5:20). Furthermore, on the Internet of Things we basically see the

virtualization, which means data being generated by the Internet itself (6:10). However, the useful data resides in the transactional bases in the system, for an airline or a retailer or bank because of the analytics (5:20). “Because we have data at its very basic level, we also have information (or ‘modified’ data) created from processing of information.” (6:10) Thus, “without actually knowing you or having a profile which is linked to your name, so you are not necessarily identifiable but you are part of a group which is targeted by the person using the Big Data. So you’re a victim requirement not so obviously” (7:6)

Table 4-1 Data Subject

| Jon Page | Trevor Peirce | Interviewer 4 |
|--|---|--|
| “Well traditionally, data comes from their own operational systems. So a billing system in Telco, the system that records ordinary transactions in a bank, they have been the Big Data sources for 20 years. Now, you are starting to see, obviously social media...” (5:20) | “We also see in Internet of Things, real and virtual things that will publish things directly and we see all kinds of processing of these information leading to yet more big data.” (6:10) | “Big Data is about general information from different sources indeed, which can single out proofs of people or even people. With single out I mean reading or combining all kind of sources of information. Single out is of course an issue.” (7:4) |

Summarizing, “Big Data is a topic that is in the center of all of these, in terms of collection and processing of data with impacts on privacy and data protection.”(6:8) Moreover, Big Data “in need to the privacy context somewhere should be saying that it is for personal data of persons, or data which tells something about people.”(7:4)

4.2.2 Data Holder

Table 4-2 Data Holder

| Cees Boon | Jon Page | Trevor Peirce | Interviewer 4 |
|--|--|---|--|
| “Big Data is a big business. A lot of companies are selling information, and information of all kind.”(4:36) | “Do they (government) buy data, I am absolutely sure they do. So it is probably two way data for them.” (5:24) | “I see quite, increasing business, (commercial activity but also government related activity) associated with big data.” (6:12) | “I think from the privacy point of view, it (Big Data) breaks with the traditional approach of privacy.” (7:6) |

We found that the data collection is not based only in the traditional data sources, data that comes from businesses' operational systems (5:20). Accordingly to (6:12)'s professor "data is in abundance and we have a lot of it now, and most of it we cannot rely on". Thus, "we have to be very careful about validating the conclusions we make from the big data analytics."(6:12). Also, Big data is not only about new data, it is associated with reviewing old one (4:68). "Because everything that you are collecting now, data is a potential violation of privacy risk." (4:50). Thus, "the collection of information, I think there are severe limitations to what can be done." (6:32). "Because if there are no good rules you don't know if you are collecting data that you shouldn't have or that you should be collecting." (4:48). "Because the aggregation of information often bypasses data protection." (6:20). However, "the evolution, the Internet of Things is something which is driving forward. It is the ability to captures information about beyond our capacity to possibly understand even."(6:38).

4.2.3 Ownership

Table 4-3 Ownership

| Cees Boon | Jon Page | Trevor Peirce | Interviewer 4 |
|--|--|--|--|
| "The problem is also buying and selling information. It is an issue in itself. You don't know where they are from." (4:36) | "I am not sure the concept of ownership actually exists, people either have permission or no permission to use the data. Who actually owns it is quite difficult to say." (5:44) | "This is an awful addition of Big Data that is not owned by anybody..." (6:34) | "...the regulations of EU, the current one or the future ones, and they are all about responsibility, ownership and rights." (7:14). |

Actually, when you use Big Data, it is difficult to identify the owner (6:28; 5:40). "So not so much changing ownership, just the relations are more complicated nowadays." (7:18). Although we own our data in the first place, that might not be exclusive (7:14). Furthermore, data duplicates itself. (4:54). That, most data providers do not really own their data they just cannot be competitive without them (4:28). Also, " it is not that every data processing is subject to your consent that might be any other reason that can follow from the contract, there can be legal information to answer for information , which is not everything based on your consent. "(7:14).

Summarizing, the challenge of the data protection should take a leadership (6:32) because, if we operate into a world where data governance is correct, then we can identify the owner of the data (5:40).

4.2.4 Summarizing Big Data Supply Chain

- Big Data is about general information from different sources indeed that can identify people. Eventually, creating and an issue.
- Big Data has disrupted the traditional approach of privacy
- It is difficult to define the ownership in Big Data Supply Chain

4.3 Big Data Supply Chain – Government

4.3.1 Legal Assurance

Table 4-4 Legal Assurance

| Cees Boon | Trevor Peirce | Interviewer 4 |
|---|--|--|
| “The rules are not good enough in EU that is really true. All the things that come the last years were not already updated in the moment they put in act.” (4:28) | “I do not think there can be an effective protection for Internet use. There cannot be. It is a never ending race to get ahead.” (6:18). | “In January 2012 the European Commission came with two proposals... In those proposals the commission tries to tackle a little bit the mobile challenging for privacy in data protection.” (7:8) |

The actual legal assurance is interpretable in different ways (4:26; 6:32). There is no pure law but, it is pure ethics among the data scientists (4:26). “Clearly, law enforcement and intelligence services want to have access to information to assist them in fighting against organized crimes. But their influence over legally permissible information security created a weaknesses. Means that organized crime and other people can easily gain and access to “secured” Internet services and information.” (6:16). For example, issues like the SSL security (5:28; 6:16). However, even if now it is too late to adopt the legal assurance we have to discuss and take the chances (4:64). Firstly, to know it as a problem (4:64). Thus, “some of the initiatives around the European cloud are found good but I would be skeptical about how they are going to operate.”(6:32). But, the rules are not well developed (7:8). Furthermore, the introduced rules are oriented towards the anonymization in order to not allow to retrace the person (7:8) as nowadays there is protection only for some of the core elements (6:32). Furthermore, comparing Big Data Supply Chain with the food supply chain there is no control activity over it (4:38) because of the missing good rules there is no clear understanding of the data that should or shouldn’t be collected (4:48). Even the aggregation from data to information is one of the problems (4:42). But, it is important to consider the discrimination

occurrence because “data protection is not purely an elaboration of detail rules of our right of privacy but also there are to afford discriminations on the basis of the classification of persons due to certain characteristics they join, they share” (7:10).

Summarizing, there is a low level of legal assurance even customs and culture to regulate the data operation process (5:34) and there is a need of vision for the law to protect the fundamental rights (6:50).

4.3.2 Policy Assurance

Table 4-5 Policy Assurance

| Jon Page | Trevor Peirce | Interviewer 4 |
|--|---|---|
| “I think that the whole area of transmitting data between companies especially in international boundaries and is a very complex area.”(5:26). | “The privacy agreement is too long and difficult to read in terms. And even if you disagree with that, it is very hard to participate in society without ticking the box, saying: Yes you do!” (6:52) | “...you just have to say yes if you have to purchase the product or to use the service, that is where the new rules try to change...”(7:14) |

Summarizing, policy agreement are not simple to read (6:52). Furthermore, it does not allow the consumer to disagree with it without affecting the value generated from the product or the service (6:52; 7:14). However, there are some attempts to strengthen these rules (7:14). But, different organizations are lobbying into the process of simplification for policy privacies, making it a complicated one (6:54).

4.3.3 Invasion

Table 4-6 Invasion

| Cees Boon | Trevor Peirce | Interviewer 4 |
|---|---|---|
| “...most government are really involved trying to get data to strengthen their power...” (4:26) | “...political institutions will seek to obtain more control over the world we have. Today, already we cannot live in a completely free world.” (6:46) | “...nobody may be subject to a decision with legal effects which is purely based or performed by automated means so there should always be a human being involved...” (7:8) |

Political institutions will seek to obtain more control over the world we have (6:46; 4:26) through the concentrated power of Big Data (4:18). E.g. NSA (4:18). Even if there are several application to exploit Big Data for security reasons (6:12), there is also a lot of encryption to avoid the surveillance (6:14). Also, there is a concern about the decision taken over the use of information:”I concern for groups of individuals who want to take actions or take initiatives where information will undermine fundamental rights and organizations...” (6:46). But, according to (7) nobody can be subject to a decision with legal consequences, as a result of relevant profiling Big Data, without human being involvement (7:8).

4.3.4 Summarizing BDSC-Government

- Legal assurance are not relevant (misinterpretable, not updated, not well developed).
- Lobbying against the simplification of privacy policies.
- Invasion is related to concentrated power.

4.4 Big Data Supply Chain – Business

4.4.1 Management

Table 4-7 Management

| Cees Boon | Jon Page | Trevor Peirce |
|--|--|---|
| “...the electronic intelligence ... that’s something completely the different should be regarded totally different from the human intelligence. This means yes. (The lack of data management)”. (4:24) | “If you look at data governance, it is very immature. Hardly anybody, especially in Europe has an organization in roles and responsibilities in the world of governance.” (5:18) | “Like in real world there are virtual objects...These things are done way beyond the political reach of governments.”(6:30) |

Big Data is not about data, but the way business can be improved using these data (5:10) and the success is sourced from the collaboration of business and IT people (5:14). Traditionally Big Data is used in accounting, but further on it will be especially part of the tactical issues rather than strategic one (4:20; 5:16). And responsible for the ownership should be the same even for the integrity (5:42).

So, the management of sensitive information should be made all over the company (4:44). But, there is no differentiation among the data processed because the aggregation overcomes the data protection (6:20).

4.4.2 Responsibility

Table 4-8 Responsibility

| Cees Boon | Jon Page | Trevor Peirce | Interviewer 4 |
|---|--|--|--|
| “But, in the situation we are now we have not only human intelligence...And who is responsible for that? It is very hard to point that.” (4:22) | “Well, that person (the responsible person) does not exist in most of the companies.” (5:38) | “Nobody. I have not seen anybody responsible (data integrity).” (6:30) | “In the new rules we try to improve the clarity of defining who’s responsible...” (7:18) |

There are some attempts to clarify the responsibility within Big Data and the article 29 as a data protection authority for all the countries is one of them where even the consumers are mentioned as responsible for certain kind of data (7:18). However, “there are many levels of responsibility. Firstly, you might need to set that person responsible of the collection of data, privacy, and security of data and the management of it. I think that is a business concern.” (5:38). On the other hand, too many companies violate the privacy because the information is not protect well and is used only to get competitive advantage (4:18; 6:24; 4:44) so the owner of responsibility has to be the CEO (5:38).

However, when we try to find the responsibility within this situation we have to consider the fact that now we have electronic intelligence in high level which is missing the social background (4:22). Also, Big Data is trying to predict something that was not firstly intended (4:26). So, the aggregation of data into information is the problem (4:42). Furthermore, “most people don’t really matter what they are doing with us. If they, most people don’t understand what they are doing. If they knew they wouldn’t understand, so it’s a problem in itself.” (4:34). The companies are responsible for the data integrity and they operate according certain rules to assure that the data are secured (7:14). “The controllers are the one responsible on processing the data cannot certainly do something different and that is the whole idea.” (7:16).

Nevertheless, people who are really responsible are hardly going to take responsibilities among the organizations, e.g. social media does not put the responsibility to any central organization but mostly delegate the responsibility to the lowest level of individual (6:42).

Summarizing, no one is clearly responsible in the event of privacy break. “No one, I think if you look at 98% of companies, they would not be able to point the finger at one person and say you are responsible for big data and information management in this company.” (5:18).

4.4.3 Secondary use

Table 4-9 Secondary use

| Jon Page | Interviewer 4 |
|--|---|
| “There are other companies, I am sure that work in the dark world here in the web collecting data.” (5:62) | “So, as obliged to consumer act of commercialization of data, in principle you have to give your consent first before they can use it for other purposes.” (7:12) |

We do not find convenient that certain our information is found on the web (5:54). Also, governments buy and sell data (5:24). Furthermore, as obliged to consumer act of commercialization of data, the data holder should have the permission of the data subject before they can be used for any other purpose (7:14).

Summarizing, the important fact towards the data protection is that the data are used on the purpose they were firstly collected (7:16). So if you want to use them for another unrelated problem you have to have a new legal which mainly is the consent or legal obligation.” (7:16).

4.4.4 Summarizing BDSC-Business

- Data governance is immature especially in Europe
- There are many levels of responsibility (*the need for the person responsible of the collection of data, privacy, and security of data and the management of it*)
- The important fact towards the data protection is that the data are used on the purpose they were firstly collected.

4.5 Big Data Supply Chain – Consumer

4.5.1 Privacy benefits

Even if we generate a lot of profitability from Big Data usage such as good information and its use for good purposes we have to discuss about the societally problems emerging from it (7:22). Generally, the individual as a data subject has the right to ask the company to disclose their data

at one point in time (5:44). But, e.g. the location of a person used as a marketing idea from the business is not liked from 70% of the people but they still accept it for benefits of the product or the service offered (5:68). Also, there is a discussion if there is a balance in terms of profitability business-consumer (5:68).

Table 4-10 Privacy benefits

| Cees Boon | Jon Page | Interviewer 4 |
|--|---|---|
| “Most people they don’t know that they are at risk it is still very profitable to take the advantages.” (4:30) | “So the main point is what the consumer gets to what the provider benefits. And that is not balanced, that means to be addressed.” (5:68) | “Again, the existence of big data is development which we cannot deny...we should consequently think about societal levels...” (7:22) |

4.5.2 Transparency

Table 4-11 Transparency

| Cees Boon | Trevor Peirce | Anonymous |
|---|--|--|
| “Most people don’t know what they are doing but even if they will know they don’t know how it will affect them.” (4:32) | “...there are all kind of big data applications today exploit Big Data. And all kind of ways that are not transparent to majority of us.” (6:12) “I think transparency aspect is one that we don’t do enough.” (6:24) | “...being informed in advance... but at least in the law should be clear, in what circumstances you can be subject of a police search of this kind of information.” (7:12) |

Thus, there is the need of more pressure from the political mechanisms to assure the transparency because it is an aspect that has not taken the necessary importance (6:24). “There are challenges, major political challenges to overcome at national level, European level and global level” (6:32). In conclusion, the simplification is the main requirement to achieve the data protection in a transparent way (6:32).

4.5.3 Privacy concerns

Table 4-12 Privacy concerns

| Cees Boon | Jon Page | Trevor Peirce | Interviewer 4 |
|--|---|--|---|
| “The problem is that I don’t really know how the things that I do in this moment, using data, using information, I don’t really know in about let’s say five years we are looking at that.” (4:68) | “What I am worried is about data that i am not aware of, some data that is generated by me.” (5:64) | “I concern for groups of individuals who want to take actions or take initiatives where information will undermine fundamental rights and organizations...” (6:46) | “So discrimination is important in all this data, again it leads more to a sociable debate you should have than a debate on the basis of the right invoked by an individual when the court gasps.” (7:10) |

Because of the time limitless of data, we don’t know if the data that we are using now how will be treated in the future related to privacy (4:68). Also, the concern is sourced from the data that is generated without the individual awareness (5:64).

Summarizing, the discrimination that arise from the use of this data leads to a social debate (7:10). Big Data it is really a societally problem discussion we should have on where we want to go with this. (7:22).

4.5.4 Summarizing BDSC-Consumer

- People mostly do not know the worth and the consequences of the data they generate.
- The majority of the consumers are not transparent about their data.
- The concerns from the individuals are related with initiatives that can potentially ruin the fundamental rights.

4.6 Big Data Supply Chain-Individual

4.6.1 Trust

We might not agree for example Google to hold certain information about us, but on the other hand the people we share our everyday life might have different limits toward the privacy attitude on the web (5:54). Also, people have started not to trust in handling of their information. “More increasingly since the revelations of data manipulation and capture in the US, more and more

people are sharing information, especially the younger generation, creating electronic profiles which are completely misrepresentation of the real world.” (6:14)

Table 4-13 Trust

| Cees Boon | Trevor Peirce |
|---|--|
| “Retailers have large transactional databases that are the DNA of consumer. Privacy is an issue for ages, even before the Big Data hype.” (4:8) | “One other thing established by popular European Union is level of trust, sufficient level of trust of people participating.” (6:36) |

4.6.2 Privacy Awareness

Table 4-14 Privacy awareness

| Cees Boon | Trevor Peirce | Interviewer 4 |
|---|--|--|
| “I don’t think that they (consumers) are aware of the data what is happening some people I don’t think they are interested.” (4:30) | “So, I think when it comes to Big Data, we need to be very wary about of the analytics, the sources of information, evolution of all of these.” (6:50) | “...if there is a privacy break you have to notify about certain casual.” (7:18) |

Based on the imbalance power in management, in government it is difficult to generate value for the world we are living in through Big Data without being aware of its ethical issues (4:18). “Because people today think of very, very short term view of the immediate environment and they do not think very far at all.” (6:50). Furthermore, the biggest challenge is that the individual, the consumer, relies in the work of the companies but there is misinformation there (6:14) that is done deliberately or by accident (6:38). So, most people rely only on the benefits perspective (4:30) without knowing the consequences of their behavior within Big Data (4:32). “It is the ability to captures information about beyond our capacity to possibly understand even. So, I mean, most people still have a very simple idea of this.”(6:38). However, the people become more sensitive to the information they generate (6:40). But, the Big Data world developed from the younger people is not the one representing themselves, consequently we have to be aware that there is a problem in trusting it (6:40). Thus, when we discuss about Big Data we have to be wary about the data source, the analytics and their evolution because people think very in short term without taking in consideration the consequences (6:50).

In conclusion, “...in new rules there are still proposals, if there is a privacy break you have to notify about certain casual. You have to inform those people and the data protection authority. This

should create awareness of what happens, and put pressure on the companies to prevent this kind of breaks.” (7:18)

4.6.3 Educated Public

Table 4-15 Educated Public

| Jon Page | Trevor Peirce | Interviewer 4 |
|---|--|---|
| “Incredible (lack of education). There should be some adverts on TV or documentaries or in the newspaper. It is amazing what people don’t know.” (5:60) | “But I think it is only through the educated public that we can have any form of effective policing.” (6:24) | “So in order to make data protection privacy still relevant and realistic, in the future the focus will be not so much on what is personal data but what do we consider as a privacy.” (7:20) |

“People understands only after something bad happens.”(5:60). Thus, through the media could be improved the education and the awareness of the people (5:60; 6:18). “But there need to be more guidance given to and more information shared about where the information is going and how we really are not protected, more awareness of logic into things that could be threatening in some way to our fundamental freedom.” (6:18). Thus, because of the abundance of data and the results coming from their analytics (6:12). Furthermore, the consumer have to be educated how to act in any case of privacy break (7:14). “So in order to make data protection privacy still relevant and realistic, in the future the focus will be not so much on what is personal data but what do we consider as a privacy.” (7:20).

In conclusion, it is a challenge to have an effective protection for Internet use not only because of the technological reasons but also in the education of the public (6:18). “It is only through the educated public that we can have any form of effective policing.” (6:24).

4.6.4 Summarizing BDSC-Individual

- People understands only after something bad happens
- The discrimination that arise from the use of this data leads to a social debate
- When generating value from Big Data we have to be aware about its ethical issues
- Only through the educated public we can have any form of effective policing

4.7 From the diaries

All of our 5 diarist used social network every day. Although they find social media most of the time not interesting nothing or useful, they looked quite obsessed with being aware each moment of the day what was happening around there. Each of them had at least two email addresses that they used daily, for work, school or personal purposes. Four to five of them use different applications on the smartphone, but they admitted that they never went through their privacy agreement. Sometimes annoyed by too much information asked when in different sites, or found it suspicious they decide to abandon the service. Almost all of them were annoyed, in different grades, with email, ads, and advertising when browsing or cookies in general. Two of them used ads blocker and cookies blocker while others just were not aware. All of them found in a way useful and profitable online services, even online banking, purchasing or just communication mediums, as well for work or studies their data is strictly related to cloud computing. Only one of them was aware that he was trading its privacy for the products and services offered from the data holders.

Further on we will look on details each of them:

Diarist (1, 8):

- **Usability:** Facebook; Twitter; Instagram; Spotify; online streaming; online banking; different apps for smartphones; YouTube; Skype; Weather Channel;
- **Privacy Benefits:** Google scholar, Google maps; other search engines; communication with school mates.
- **Privacy Concerns:** Advertises on mail; find Facebook nothing important or funny; hates ads on YouTube. Diarist 1 admits that might be a little over obsessed with social media. Although she is concern about what others might think when she posts or like or comment in different posts she is most of the time attentive. Even though diarist 1 sometimes would choose personally not to get involved in posts, she does not do anything, and ultimately she thinks as she does not care.

Diarist (2, 9):

- **Usability:** Facebook; WhatsApp; LinkedIn; YouTube; Skype; online streaming; Weather Channel; different apps for smartphones; Wikipedia; Google translate
- **Privacy Benefits:** Google scholar; online news; communication; job tasks
- **Privacy Concerns:** Finding social network not safe as people who are not within your network can still see and share your information.

Diarist (3, 10):

- **Usability:** Facebook; WhatsApp; LinkedIn; Skype; online streaming; customer cards; Amazon; Shopify; Spotify; Google Drive; Google +, Apple store

- **Privacy Benefits:** Communication; purchasing product and services; discounts from the customer's cards, but it is not always valuable compared with the personal information to disclosure. E.g. *"I realized I had to give some of my main personal data such as email address, telephone number etc. As I still wasn't sure if I want it to take advantage of that discount, I decided not to fill out that form and save myself from the annoying emails."*
- **Privacy Concerns:** Annoyed by ads, email ads, or newsletters, which her attitude is to unsubscribe from them continuously.

Generally, the diarist 3, accepted and agreed to term and conditions without reading anything, but when it is required to give her basic personal data such as, email address and the phone number she found it really annoying. Also, when Facebook, for games, asked to access her public profile and friends list she skipped it and closed. So, she didn't feel comfortable when other sites used Facebook profile to make them registered. She is cautious of every move she does on Internet, as well as annoyed with e-commerce ads through email.

Diarist (4, 11):

- **Usability:** Facebook; WhatsApp; LinkedIn; online streaming; online banking; Twitter; YouTube; eBay
- **Privacy Benefits:** Google scholar, LinkedIn to connect with experts for master thesis purposes
- **Privacy Concerns:** Email in Gmail most of the time is not useful, so he hates Ads he deletes them and mark them as spam. He also deletes previous wall posts on Facebook; and founds annoying google AdSense; He cancelled premium account in LinkedIn, as he was skeptical about the website keeping his card number. As a matter of fact, most of the ad mails are due to chrome using autocomplete while typing the name in forms. Now he wishes to cancel in which all websites that have his email ID. Found it annoying that truecaller.com website had his phone number and address listed. After that data.Scribd.com asked him to login with LinkedIn/Facebook/google+ for free documents, he become a bit cynical about websites accessing Facebook, Twitter or LinkedIn. He is more careful about using those details to log into website, so he skipped the search.

Diarist (5, 12):

- **Usability:** Facebook; Google drive; Google Drawing; Doodle; Smartphone Apps
- **Privacy Benefits:** The diarist is aware of the deal between him as a consumer with the data holders such as Facebook, Google etc. of the data generated from him. He knows that everything is stated in their term of use and also it is a fair deal because the profitability he obtains from these data holders is immense, because of the tremendous networking value containing. Also, he is aware of the fact that he surrenders his privacy to Facebook / Google etc. because of the great services they provide to him. Thus, if he is not satisfied with this relation he thinks he is free to back away.

- **Privacy Awareness:** Diarist 5 is an aware consumer for his consequences He states: *“The moment I press “upload” Facebook owns my content, they are free to use it in whatever advertising they want, and they are free to use it to target me with advertising. This is the “price” that I pay for getting the social experience that Facebook offers.”* Also, the diarist is aware about the fact that he has to take the responsibility even for the data generated from the others about him such as the photo that the diarist is tagged even if there is the possibility to limit the audience of the tags the diarist hadn’t choose to take that action. The diarist is aware that this is the risk to take in order to profit from the benefits that are offered within in this environment in the context of data generator and data holder. He claims that *“When I post something on e.g. Facebook I do so with the mindset that anybody on the internet may see it. Facebook is only so good at keeping people outside of my friends circle out. I think it is naïve to view Facebook as some sort of secluded forum where you can discuss topics, shielded from outside scrutiny.”* Furthermore, the diarist suggest the “healthy” behavior within these social networks. He declares that: *“I think that if you view these social networks as the corporate entities that they are, and realize that their core business is exploiting your data, you can have a healthy relationship to them.”* Moreover, the diarist argues that *“I think that’s to some degree correct, but that the ethics problem of entities like Facebook “watching us” is awareness and not demonizing.”*

Summarizing, all the five diaries we compiled the table 4.16:

4.7.1 Diarists interview

At the end of the diary week, after they handed in their diaries, they were asked the following question in order to understand what was their opinion about the information privacy in the context of their daily behavior:

If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 1: Yes, it is my life and I do not want to share it with the internet world.

Diarist 2: I don’t want that my activity be known from my employer and they know every single activity I do during the day. Maybe it affects my employer opinion about me. But, also I don’t like that the other people know in detail what I do during the day even I have described common things that people generally do. I don’t like it. It seems like someone is spying to me.

Diarist 3: I don’t see it comfortable that my diary to be accessible somewhere in the Internet. Everything that you have done, online in the middle of the “jungle” because I perceive Internet in this way. Also, even with my photo and my name.

Diarist 4: I don't want my mail ID to be published online. Although all the material I provided is fine, as I didn't do anything bad, I don't like it. And more over I don't like my picture to be found in google.

Diarist 5: I surrender my privacy to FB/Google etc. because they provide great services in return. If someone would take this info and slap it on a blog I get nothing back. I feel that I'm knowingly trading my privacy (as much as I want to give them anyway) in exchange for a service.

Summarizing, the main privacy concerns that the diarists said to have are:

- It is my life and I do not want to share it with the Internet world.
- I don't want that my activity is known from the people surrounding me.
- I don't see it comfortable.
- I don't want my mail ID to be published online.
- I surrender my privacy to FB/Google etc. because they provide great services in return. But, in this case I get nothing back.

In conclusion, the diarists from the view of the consumer even if they have several concerns related to the disclosure of the information because of the privacy they still continue to get value from these services and products that make the source of their concerns. Thus, they can be divided into three subgroups.

Firstly, the one that have concerns about their privacy but do not take any action (diarist 1, diarist 2). We examined their behavior from the perspective of an individual, and they are not aware and educated to conduct the privacy calculus for their behavior and to prevent or minimize their concerns such as the annoying towards ads that can be simply eliminated by ads blocker tools.

Second, diarist 3 and diarist 4 are the consumer that make a privacy calculus for their behavior. Even if they take action to eliminate the privacy concerns they still are not fully aware and educated towards the privacy issues within this Big Data Supply Chain.

In the end, the diarist 5 is the consumer that is aware for the consequences of his behavior related to privacy benefits and privacy concerns. He has a "healthy" behavior that is characterized from the privacy awareness and well educated level within this topic.

In conclusion, the consumers should operate based on the privacy calculus theory when they decide to disclose information. But, in order to conduct the privacy calculus they have to be informed and educated about the privacy issues that emerge because of the disclosure of their information accurately.

The table 4.16 is to depict an overview of the consumers' behavior, related to the benefits, the concerns, and the awareness towards privacy. But, also through the usability of the online services it is clear that they are the data subject of the Big Data Supply Chain even if they are not aware of it. They do not express any particular concern to the data not specified as personal one or the data generated without their consent.

Table 4-16 Diaries behavior

| Diarist | Privacy Benefits | Privacy Concerns | Privacy Awareness |
|------------------|---|---|--|
| Diarist 1 | Google scholar Google maps Search engines Communication | Mail advertising. Ads on YouTube. Others' opinion on social media activity. Facebook not import or funny | No awareness |
| Diarist 2 | Google scholar Information Communication Job tasks | Social media unsafe. People outside your network can see and share your information. | No awareness |
| Diarist 3 | Communication Purchasing product and services Discounts from the customer's cards valuable compared with the personal information to disclosure | Annoyed by ads, Email ads, and Newsletters | Unsubscribe & Privacy calculus |
| Diarist 4 | Google scholar Professional connection (LinkedIn) | E-mails, Ads, Facebook posts, Bank transaction, Security | Delete & Unsubscribe |
| Diarist 5 | Online community Tremendous networking effect | N/A | Surrenders privacy for the products and services gained. |

4.8 Summary of empirical findings

In this chapter, we presented and analyzed the findings of our study based on the five main themes of the theoretical framework. We analyzed Big Data Supply Chain to characterize the potential privacy gap among its stakeholders. Then, we analyzed the stakeholders of it such as, government, business, consumer, and individual. Furthermore, we developed three components for each of the

themes. Thus, Big Data is about general information from different sources, which indeed can identify people, creating thus and an issue. Big Data has disrupted the traditional approach of privacy. Furthermore, it is difficult to define the ownership in Big Data Supply Chain.

According to the government stakeholder we found that the legal assurance is not relevant. It is mostly misinterpreted, not updated, and not well developed. Also, the privacy policies are complicated to read. But, even if there are attempts to simplify them exists lobbying against the simplification of privacy policies. Additionally, the invasion is related to the concentrated power. Furthermore, the business stakeholder is faced with the data governance in an immature level, especially in Europe. Additionally, the business needs a person responsible of the collection of data, privacy, and security of data and the management of it within the business entity. Also, the secondary use of data appears to be a problem towards the privacy.

But, when we come to the consumer we found that most of them do not know the worth or the consequences of the data they generate. Also, they are not transparent about their data. Thus, they understand only after something bad happens to them. So, only through the educated public we can have any form of effective policing.

5 Discussion

This chapter is the discussion of the empirical findings in the context of the extended theoretical framework. The discussion is based on the five themes we conducted our research previously, introducing for each of them the potential privacy gap components.

5.1 Big Data Supply Chain

As we previously discussed in the literature review, what describes the data flow in the Big Data context is a Data Supply Chain (DataSociety, 2014). Our empirical findings also supported this approach. Moreover, EvanLevy (2001) introduced Data Supply Chain as a new way to handle the data's business. Again, our empirical data supports that concept, because now the data is aggregated into information and further on into a usable form. The purpose of our research was to identify the potential privacy gap among the stakeholders in Big Data Supply Chain. As a result, firstly we defined the stakeholder based on the activities that reside within the Big Data Supply Chain such as collecting, analyzing, sharing and acting, and generate feedback (Croll, 2014) and the taxonomy of privacy (Solove, 2006).

According to Solove (2006) they are grouped into data subject and the data holder. Where, data subject is analyzed as an individual who generates data and information (Solove, 2006). Also, in the empirical data we found that the data subject within the Big Data Supply Chain is a person who feeds it with data. They can be generated through the traditional way such as the data from the transactions within the providers of their services and products like telco, banking, etc, in general the data kept in the operational systems of any company. But, Big Data Supply Chain is different because the other data sources (Chen et al., 2012). Actually, the individual analyzed as the data subject in the context of Big Data Supply Chain is living in a digitized world. Thus, the Internet and the web lead towards new data sources for collection (Chen et al., 2012), making, the social media and the Internet of Things are possible data sources. Still, the empirical findings confirm it. As Ashton (2009) claimed, the Internet of Things extends the capture of data. Obviously, Big Data analyzed in a Data Supply Chain is about information gathered from several sources as we described previously. Also, it is capable through analytical techniques to convert these data into identified data subject. That is an issue itself with this phenomenon, accompanied from both opportunities and responsibilities like the privacy issues (Filippi, 2014). Again, the data confirmed the literature. Big Data has raised a social debate towards the privacy issues in terms of collection, processing, and acting over it.

So, we went further with the examination of each of the stakeholders within Big Data Supply Chain. Hence, we go through the analyze of the second group of stakeholders, the data holders. According to Solove (2006) data holder is the one who collect, analyze, act, and create feedback over the outcomes the analyze and action. Previously, we described the data sources available for collection. But, how is proceeded with the whole process and who collects the data generated from the individual, analyzed as a data subject?! Davenport (2006) argued that analytics based on Big Data are a competitive advantage for the companies. As findings suggest, “Big Data is a big business.”(4:36)” thus, the data are seen as an asset for the companies. Because of this several companies, and government too, are selling and buying information. According to Solove (2006) the data holder could be, the government, the business, or any other individual. Thus, the empirical findings cited previously support the literature review.

Obviously, Big Data is disrupting the way we live (Needham, 2013), but, what happens with privacy? According to Davis and Patterson (2012) privacy is one of the issues that emerge because of Big Data. Even the empirical findings claim that Big Data Supply Chain is disrupting the traditional approach of privacy.

Further on, we have to be aware that data is duplicated within the Big Data Supply Chain (DataSociety, 2014). Thus, as we underlined above the data are sold and bought broadly. As such, it is an issue because it is not clear where that data is coming from. So, it is a challenge to define the privacy and even more to find the owner in the Big Data Supply Chain. Ownership is one of the perils of Big Data Supply Chain. For this purpose, the EU is emphasizing it in the current and future regulation.

Summarizing, one of the issues Big Data imposes is the privacy. When analyzed as a Big Data Supply Chain we found the stakeholders and also the data ownership uncertainty. Further on, we will discuss more in detail each of the stakeholders in order to identify the potential privacy gap among them in the context of the Big Data Supply Chain.

5.2 Big Data Supply Chain - Government

As we discussed previously, government is one of the stakeholders in Big Data Supply Chain. According to the political factor that Big Data operates, it is represented through the government. Thus, the government presence is assured through the legal assurance such as data regulation process (EuropeanCommission, 2014). Hence, we found that the rules are not updated or well developed to Big Data Supply Chain. Also, they are interpretable in different ways. Therefore, the rules within EU are not good enough even if there are several attempts to challenge the privacy in the data protection. Moreover, it is quite impossible to have an effective protection for the Internet use. We are emphasizing the Internet use because the Internet of things is an enormous data source (Ashton, 2009) as such brings even privacy issues because of the activities after the collection

phase within the Big Data Supply Chain. But, why is the achievement of this effective protection such a challenge?! We have to look further on the stakeholders.

Additionally, the government presence is associated with the policy assurance. How is the privacy assurance positioned within this discussion?! According to McDonald and Faith Cranor (2008) the privacy policies are complicated, long and not simple to read. Also, it is strongly supported from our empirical findings. Also, McDonald and Faith Cranor (2008) interpreted the time spent as included in the price as the opportunity cost for the privacy. Moreover, they associated time and economic value to the privacy policies reading. Furthermore, you have to accept the privacy policies for gaining value from the services or the products you aim to get. “And even if you disagree with that, it is very hard to participate in society without ticking the box, saying: Yes you do!” (5:52). Hence, only one privacy policies requires from 8 to 12 minutes to read (McDonald & Faith Cranor, 2008), and, your opportunity cost is any other activity related to you it means that you paid and did not gain the benefits you were looking for. It is an issue. Hence, the new rules are trying to change in relation to the above discussion. But, still there is the confront with the business part. So, there is the resistance of different organizations through lobbying to inhibit the privacy policies simplification process.

In the end, we found the involvement of the government in the overall discussion also related with the invasion. According to Solove (2006) invasion as is seen from both intrusion and interference. Where intrusion is associated with the involvement of the government in the data subject decisions from the privacy issues. We found that governments are using the opportunities that Big Data offers to strengthen their power to obtaining more control. But, the most important fact is that, nobody can be subject of an automated result without the involvement of any human being.

Summarizing, the discussion of government we are confident that the investigation of the other stakeholders will create the complete tableau of the phenomenon. Until now, we found that the legal assurance and the policy assurance are facing a challenge to operate effectively towards privacy in the context of Big Data Supply Chain. It is only through the government intervenient that this challenge can be succeeded through an effective legal and policy assurance. In the next section, we are going to discuss the implication of the business as the stakeholder that operates within this legal and policy assurance.

5.3 Big Data Supply Chain – Business

As stated above the companies competing on analytics differentiate themselves creating a competitive advantage (Davenport, 2006). Also, the data gathered from the consumers create the competitive advantage (Mary J. Culnan & Armstrong, 1999). Furthermore, we have clarified at the beginning of this discussion that Big Data Supply Chain is a new approach to manage the data. According to the empirical data we found that the data management is facing several problems

within Big Data Supply Chain as follows: Firstly, the data governance it is immature. “Hardly anybody, especially in Europe has an organization in roles and responsibilities in the world of governance.” (4:18). Further on, the involvement of electronic intelligence that is obviously different from the human intelligence leads to the lack of data management in any company. Along this line, the obstacle of the management resides in the fact that Big Data Supply Chain is associated with the virtual objects.

As we approached Big Data in a Data Supply Chain, comparing it with the traditional one we found as follow: In the last one there are responsible and control points along the processes (Cai, Liu, Xiao, & Liu, 2009). But, our empirical data reveals that there is not any responsible person within the Big Data Supply Chain for the data integrity and other issues during the processes the data goes through. It is not only difficult to assign the responsibility to any person but it does not even exists within the companies. However, it is promising that the new rules are trying to clarify the responsibility role associated to the data.

According to Smith et al. (1996) secondary use, as the usage of data not for the first purpose it was collected, leads to possible privacy concerns. Thus, as long as there are companies operating not in trustworthy way in the data collection it means that they are using them not for the initial purpose gathered. We can mention that the consumer have to give the consent for any possible secondary use. But, we return back to the policy assurance discussion. Thereupon, as far as the privacy policy reading has a considerable opportunity cost we (as consumer) would not be aware of the secondary use of the data they generate.

Summarizing, the business is getting value from the opportunities that offers Big Data Supply Chain. But, there are some important issues to regulate and control their operation. Thus, the lack of management, the uncertainty of the responsibility, and the secondary use have to be regulated. Where the regulation process involves not only the government, the business, but even the consumer in order to be an effective one. Further on, we are expanding the discussion in the individual level.

5.4 Big Data Supply Chain – Consumer

Big Data is defined as the revolution that will transform our life (Mayer-Schönberger & Cukier, 2013). Even if it offers several opportunities we have to discuss it even in the societal level, where information privacy is defined as the importance associated with the controlling process that the individual has about themselves data (Clarke, 1999). According to Klopfer and Rubenstein (1977) privacy generates debate in terms of costs and benefits. But, empirical data reveals that most of the people do not know about the value and the consequences that emerge from the data they are generating. It is still convenient to gain the benefits. Thus, they are trading their privacy in order to get value from the services and the products offered from the provider that could be both

business and government. Although, we have to discuss about the balance if it really exists in the value generated for both data subject and data holder in the Big Data Supply Chain.

Through the privacy calculus theory can be examined the tradeoff of costs and benefits related to individual behavior (Mary J. Culnan, 1993). But, there are consumers claiming that their personal information is collected and analyzed not in a transparent way (Mary J Culnan & Bies, 2003). Even the empirical data reveal that majority of the consumers are not transparent about their data then confronting costs and benefits is not reliable. It does not lead towards the real tradeoff of costs and benefits making the consumer to have privacy concerns. Privacy concerns are not new. They emerge even more in the presence of a new information technology phenomenon that the public perceive as a threat (Mary J. Culnan, 1993; Westin, 1967). Also, the empirical findings reveals other concerns not only in the context of gaining benefit from a product or a service. Thus, people are worried about the data generated without their consent. There are concerns related to action taking in relation to their fundamental rights. Finally, discrimination is another concern.

Summarizing, the consumer as the third stakeholder in the Big Data Supply Chain trades the privacy in order to get the benefits offered from the provider of the product or the service. But, are they aware of this trade relation? Because of the lack of transparency even if they behave based on the privacy calculus theory it will not be a reliable evaluation. For the sake of this investigation the next and the last section of the discussion chapter will be used.

5.5 Big Data Supply Chain - Individual

According to Solove (2006) it is a challenge to define the privacy clearly. Apart of this, there are several definitions for it and usually they are related with any possible secondary use of individual's data (Belanger et al., 2002). But, trust as the readiness to undertake the risk of data disclosure (Malhotra et al., 2004) controls the level of data generated from the data subject in the Big Data Supply Chain. Nowadays, the individual has started not to trust in the publication of the data because of the revelations of data management and data capture in the US. As a result they are creating several ways to protect themselves from the surveillance or any other intrusion because of their data. Because of the low level of trust, the young generation is disclosing false information. It means that their data is not facing the real world. Hence, we have to discuss the consequences of this behavior and recognize it as a problem before it becomes a phenomenon.

But, the trust level is not only affected from any previous experience. We have to consider even the privacy awareness, as the extent to which an individual is informed about the privacy practices within the organization (Malhotra et al., 2004). But then again, most of the people are not aware what happens with their data in terms. Additionally, in the Big Data context it is important to be wary about the analytics and the source of information that the analytics are based on. Although, an individual should be aware of the practices to go through in case of any possible privacy break. Nevertheless, if we discuss the awareness it is also as a consequence that people think in a short

term. Therefore, we have to raise the discussion even in this direction. The individual have to be aware that the Big Data is not only about new data but it is also based in the data generated previously. It's true that there is such a debate for the right to be forgotten, as the example discussed in the introduction, but then again, it is thank to the awareness that we are moving now towards the "right to be forgotten".

But, where is the source of the awareness level? Why some individual are more aware and take action in any possible privacy break? According to Awad and Krishnan (2006) the educated public is related with the privacy awareness. Furthermore, the education is defined as the information or the training within a specific subject (OxfordDictionaries, 2014). Thus, in the view of our study the education is related with the privacy in the Big Data Supply Chain context. But, there is a lack of education within the area. The media, the government and the education institutes have to discuss how to improve the public education towards the privacy awareness increase. Furthermore, the individual has to be educated how to act in case of privacy break. As a result, the privacy awareness will be increased, and also it will contribute in the improvement of the interplay between the stakeholders in Big Data Supply Chain in terms of privacy.

Summarizing, we discussed the interplay among the stakeholders in the Big Data Supply Chain to identify any potential privacy gap. Eventually, the privacy discussion is enhanced because of the Big Data and the chances for exploitation it offers towards the stakeholders in the Big Data Supply Chain. Thus, we discussed that the legal and policy assurance need to be updated according to the evolution of technology. Also, further on, we discussed the business stakeholder. It is faced with the lack of management and the uncertainty of responsibility towards the data. Moreover, the secondary use of data is a source for privacy concerns. But, as long as the legal and privacy policies are not effective there is place for discussion. Then, the consumer gain privacy benefits and generate privacy concerns through the participation in the Big Data Supply Chain. But, the low level of transparency leads to more concerns among the consumers. In the end, we identified that the trust and the privacy awareness are low among the individuals.

Concluding, only through the educated public we can make the legal and policy assurance effective. Even more, to make this discussion realistic and relevant we have to be concentrated not only in what is personal data, but rather what is the privacy as altered within the IT evolution (Smith et al., 2011).

6 Conclusions and further research

In accordance to the empirical findings and the discussion we had about them, this chapter will summarize the main findings attempting to answer to the research questions we posed from the beginning of our study. Further on, a guideline to implications to this study, our contribution to the IS sphere will follow up. Concluding thus with short outlines of limitations in our study and proposed further research.

6.1 Research question

The purpose of our research was to identify the potential gap between privacy and openness among the stakeholders acting in a Big Data Supply Chain. In accordance to this aim, we extended a research framework and used a Big Data Supply Chain helping us therefore in the identification of the potential gaps into this ecosystem. Through our empirical findings and analysis we verified that there is a gap (if not more than one) and moreover, we indicate components that might affect this gap. We provide an answer to the privacy issue seen in different perspectives, such as the role of government, the responsibility of the organizations and the informed consumers regarding privacy. Accordingly to the interviews and diaries responses that we analyzed through coding and the findings we presented in Chapter 5 we answered our research questions as follows:

Research question:

What is the potential privacy gap in the Big Data Supply Chain?

Although Big Data is known for the competitive advantages it brings to the parties using it, it also comes with a number of challenges. The challenges of Big Data as a phenomena not only are associated to its return of investments and lack in proficient employees, but it also brings important concerns to consumers and the society in general, in terms of privacy and infringements of rights. For this reason we derive from our study that the potential privacy gap in a Big Data Supply Chain lies in or because the following paths:

Volume vs. Value: Data holders are putting significant effort in collecting every kind of data and detail generated but less in estimating and understanding the real value of it. All this hasty analysis carries space for misinterpretation.

Data Overconfidence: Overreliances in the analytics produced by Big Data can lead to blind spots when making decisions. These might affect forever an individual or company and their discrimination toward it. Although, data analytics and its predictive models are a reflection of the real world, it is still not the real world itself and as such there should be constraints toward its confidence.

Accuracy: It is essential and though there are spaces in which marketing analytics fails to support confidence. Mining text and other information is a good example of this gap. Based on our findings, a part of the consumers, when on Internet world, they used encryption as a way to avoid surveillance and to protect themselves from spam and other kind of generated information that they were not pleased to receive.

The Wrong Data: In all data analysis, big and small, the quality of the output is based on the quality of the inputs. Garbage in. Garbage out. From our study we perceived that consumers deliberately generate different identification profiles online, which many of them are false as a way of evading to surveillance. So the path to Big Data gains is risky and also rocky. Public should be educated with the value of data they generate. Their data is not only used for marketing but also improved health services or enhanced safety from government.

Legal/Policy assurance: As with most novel technologies, legal frameworks and policy assurance are not able to catch up the speed of their emerging. Nowadays it is a challenge to resolve any legal/policy issues that might carry barriers in using the data for analytics. Clearly, law enforcement and intelligence services want to have access to information for fighting against organized crimes, but consumers also want anonymity. They also want transparency that is most of the time missing. Consumers are diminishing their trust in companies they deal with, in their governments, in entities such as the European Union, and in the law.

Data Ownership: Consumers are owner of their data in the first place, but that is not exclusive. One of the legal/ policy restriction on sharing and using data is when the data belongs to different agencies. The communication, barriers and legal rights between this interchange are still vague, leading therefore to an important question toward ownership. Once data sets are available, data ownership becomes questionable. There is a mist about usage rights and requirements when data sets are available for analysis.

Summarizing our conclusions, we emphasize that there should be more effort, pressure and responsibilities put upon on governmental mechanism in order to build higher requirements for any subject handling data or information. In order to achieve this, data holders have to be more transparent in order for consumers and citizens can increase trust in companies and government. But any form of effective policing cannot be obtained without enhancing an educated public.

6.2 Implications

One of the first implication of our study is providing a simplified and generalizable policy assurance for consumers. A common data protection should be established in order to be understood from any citizen. Second, a legal assurance should have stronger implementation of data protection. Instead of a privacy policy, where different companies design it accordingly to their purpose of data usage and then data from different sources is being gathered and re-

disseminated, we suggest a privacy plan. Big Data should make organizations to refocus their attention on the flow of data in the supply data chain, and if the collected data is to be used for a context different from, the initially purpose it was collected, the consumer should be required to consent to that new use. As the notice-and-consent model is based on control, and a consumer cannot control what she/he does not know. Finally, responsibilities for data protections and use should be kept in the high levels of operations within organizations.

6.3 Contribution

By finding the gaps among the participants in the Big Data Supply Chain toward privacy, our study aims to give importance to privacy concerns. Any data on human subjects certainly raise privacy issues. Observing at how fast technology capabilities and computing power continue developing it is an open question whether computing will return privacy and confidentiality protections back into the hands of consumers. Exploring the interplay between privacy and openness in the context of information society will contribute the designers of system, designers of policy makers and CIO's of management.

6.4 Limitations and further research

Due to practical and conditional limitations the study we conducted has space for improvements. One of the limitations of our study findings is that we used a small number of consumers to keep a diary and additionally we could not analyze all the factors that might affect them in their perceivment of privacy. Further on, in our study, we looked at the privacy concern in an individual context rather than group, or in the organizational level. As it was impossible for us get all these data in a short time. Moreover, we collect data only in Europe and also relying on the European regulation.

Finally, due to simplification and the ease of the information to be tracked online, we limited our research in online disclosure of information.

However, we tried to diversify the collection of data from different participants acting in a Big Data Supply Chain so that our findings could be generalizable form a conceptual significance.

Although we achieved our purpose of this study, we believe that there is space in our research to be further developed. Based on the practical limitation or the time condition we suggest that this study can be further improved. Thus, a further research can involve a mixed research, substituting the qualitative research of consumers with quantitative one. A larger and randomized number of respondents worldwide can be included in the sampling.

A comparison between China, with restricted privacy standards toward personal information disclosed on Internet, and other countries might also give important insights. Finally, privacy concern in the context of the group and organizational level should also be considered in further researches. In reliance to these our findings would have reach greater generalizability.

7 Appendix

7.1 Appendix 1 - Concept-centric theoretical framework

| Literature | Theme |
|---|--|
| Big Data Supply Chain (BDSC) | |
| <ul style="list-style-type: none"> ▪ (EvanLevy, 2001) ▪ (Solove, 2006) ▪ (Solove, 2006) ▪ (Mary J. Culnan & Williams, 2009) | <ul style="list-style-type: none"> ▪ Data Subject (DS), an individual who generate the data and information. ▪ Data Holder (DH) <i>the</i> one who collect, analyze acting and sharing, and measure results and create feedback of the information gathered or produced such as government, business, or any other individual. ▪ Ownership (OW), the act, state, or right of possessing something (OxfordDictionaries, 2014) |
| BDSC-Government | |
| <ul style="list-style-type: none"> ▪ (Smith et al., 1996) ▪ (Smith et al., 1996) ▪ (Solove, 2006) | <ul style="list-style-type: none"> ▪ Legal Assurance (LEA), a promise or guarantee, an act that inspires confidence (YourDictionary, 2014) related to the law. ▪ Policy Assurance (POA), a promise or guarantee, an act that inspires confidence (YourDictionary, 2014) related to privacy policies. ▪ Invasion (IN), <i>Intrusion</i> concerns Invasive acts that disturb one's tranquility or solitude. <i>Decisional Interference</i> involves the government's incursion into the data subject's decisions regarding her private affairs. |
| BDSC-Business | |
| <ul style="list-style-type: none"> ▪ (EvanLevy, 2001) ▪ (Filippi, 2014) ▪ (Smith et al., 1996) | <ul style="list-style-type: none"> ▪ Management (MA), the process of dealing with or controlling things or people (OxfordDictionaries, 2014). ▪ Responsibility (RE), the state or fact of having a duty to deal with something or of having control over someone (OxfordDictionaries, 2014). ▪ Secondary Use (SU) indicates the data usage with the purpose outside the primary purpose it was firstly collected. |
| BDSC-Consumer | |
| <ul style="list-style-type: none"> ▪ (Stone & Stone, 1990) ▪ (Mary J Culnan & Bies, 2003) | <ul style="list-style-type: none"> ▪ Privacy Benefits (PB), related with privacy calculus perspective is associated with the |

| | |
|---|--|
| <ul style="list-style-type: none"> ▪ (Dinev & Hart, 2006) | <p>personal assumption that a particular behavior leads to the most level of outcomes</p> <ul style="list-style-type: none"> ▪ Transparency (TP), consumers claiming that their personal information is collected and analyzed not in a transparent way ▪ Privacy Concerns (PC) are defined as an individual subjective judge of legitimacy from the perspective of information privacy. Furthermore, Internet privacy concerns represent individuals' perceptions of what happens with the information they provide via the Internet. |
| BDSC-Individual | |
| <ul style="list-style-type: none"> ▪ (Malhotra et al., 2004) ▪ (Dinev & Hart, 2006; Malhotra et al., 2004) ▪ (Malhotra et al., 2004) | <ul style="list-style-type: none"> ▪ Trust (TR) reflects a willingness to assume the risks of disclosure ▪ Privacy Awareness (PA), reflects the extent to which an individual is informed about organizational privacy practices. ▪ Educated Public (EP), information about or training in a particular subject (OxfordDictionaries, 2014) related to information privacy. |

7.2 Appendix 2 - Interview guide for experts

I-Introduction

Introduce to the research question of our study and the purpose of it.

Confirmation for recording the interview.

Introduction of interviewers and interviewees

Warm up question

1. Can you tell us little about your professional background?
2. Do you remember what lead you to this particularly field? The motivations of it?

II-Main session

Big Data Background

3. According to Gartner (2001), where the term was initially used, Big Data is described as the rapidly increasing volume of data, velocity as the data go in and out, and variety as the dispersed type and sources of data. Do you think is something missing in this term of definition?
4. What are some of the opportunities of Big Data analytics? Can you give a concrete example from your experience?
5. Which are the main challenges Big Data faces within different fields?

Management

6. Which department uses more Big Data analytics?
7. Who is responsible for the data management in your company?
8. Who are the providers of data from where companies get the data to analyze?

Authority

9. What are the policies in place to secure data transfers between companies? How is this data protected from being hacked or leaked?
10. Do you think that existing legal assurances present offer an effective protection for Internet use?
11. Are any data aggregation being treated differently because when analyzed prediction might violate or prejudice ethical issues?
12. When and where should access of information be treated differently because of the potential for misinterpretation or abuse?
13. Who should be responsible for addressing different issues that emerge because of predictive analytics?

Ownership

14. Who is the owner of the information produced?
15. Who is responsible for its integrity?
16. What rights, if any, do consumers (individuals or organizations) have to data collected by and about them?

17. What rights do they have over the commercialization use of their data? Does the reuse (reselling to other companies) lead toward the changing ownership?
18. Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete.

Big Data Privacy

19. How would you define Privacy in the era of IT evolution, specifically related to with Big Data?
20. What privacy issues and obligations arise from changing ownership?
21. In the event of a major privacy break, who is ultimately responsible?
22. When the use of the technology is an ethical one towards the power that the Big Data offers?

Future vision

23. What privacy concern do you have regarding the information that you as an individual generate?
24. How do you assume the future of information privacy? (Any concrete example)

III-Closing questions

Other useful information

25. Based on your experience and point of view, is there any other issue you would like to add in terms of privacy as a challenge of Big Data, before closing the interview?

IV-Debrief

Thank you very much for your time and collaboration, we appreciate it a lot. Once we finish the transcription, we would like to share it with you, so if you have any comments or interpretations we would be happy to get them.

7.2.1 Appendix 2.1 – Table of themes in accordance to interviews and diary

| Theme | Subthemes | Interview / Diary | Interview questions |
|------------------------------|---|-------------------|--|
| Big Data Supply Chain | <ul style="list-style-type: none"> ▪ Data Subject ▪ Data Holder ▪ Ownership | Interview | <ul style="list-style-type: none"> ▪ Who is the owner of the information produced? ▪ Who is responsible for its integrity? ▪ What rights, if any, do consumers (individuals or organizations) have to data collected by and about them? ▪ What rights do they have over the commercialization use of their data? Does the reuse (reselling to other companies) lead toward the changing ownership? ▪ Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete. |
| BDSC – Government | <ul style="list-style-type: none"> ▪ Legal Assurance ▪ Policy Assurance ▪ Invasion | Interview | <ul style="list-style-type: none"> ▪ What are the policies in place to secure data transfers between companies? How is this data protected from being hacked or leaked? ▪ Do you think that existing legal assurances present offer an effective protection for Internet use? ▪ Are any data aggregation being treated differently because when analyzed prediction might violate or prejudice ethical issues? ▪ When and where should access of information be treated differently because of the potential for misinterpretation or abuse? ▪ Who should be responsible for addressing different issues that emerge because of predictive analytics? |

| | | | |
|-------------------------|---|----------------------|--|
| BDSC- Business | <ul style="list-style-type: none"> ▪ Management ▪ Responsibility ▪ Secondary Use | Interview | <ul style="list-style-type: none"> ▪ Which department uses more Big Data analytics? ▪ Who is responsible for the data management in your company? ▪ Who are the providers of data from where companies get the data to analyze? |
| BDSC - Consumer | <ul style="list-style-type: none"> ▪ Privacy Benefits ▪ Usability ▪ Transparency ▪ Privacy Concerns | Interview + Diary | <ul style="list-style-type: none"> ▪ How would you define Privacy in the era of IT evolution, specifically related to with Big Data? ▪ What privacy issues and obligations arise from changing ownership? |
| BDSC- Individual | <ul style="list-style-type: none"> ▪ Trust ▪ Privacy Awareness ▪ Educated Public | Interview + Diary | <ul style="list-style-type: none"> ▪ In the event of a major privacy break, who is ultimately responsible? ▪ What privacy concern do you have regarding the information that you as an individual generate? |

7.3 Appendix 3 - Diary guide for consumers

Personal Data Submission Diary

Instructions: How to fill your Diary

Thank you for agreeing to help us with our research. Here are some points to consider while filling in your diary.

We are interested in all your personal data submissions - not just internet data.

Remember that this is your diary. We are interested in finding out as much as possible about your data submission and your experiences of using different offline and online services. So, please tell us as many detailed information as you can about your daily habits. For example, when doing shopping and using your loyalty supermarket card; online web pages usage; when acquiring an app for your phone; any activity within different social networks.

Please make every entry in it as soon as possible after it happens, possibly on the day it happens.

Please fill in the day and date in the space provided on each new diary page.

Please write as clearly as you can. If you are not sure whether to tell us about something or not, please include it! - We would rather have too much information than too little.

If you decide at one point you no longer want to proceed with keeping it, or find that you cannot complete the full week, *please return the short questionnaire at the end of these instructions, and whatever part you have completed.*

Questions:

Please tell us about any data subscriptions you did today – no matter how unimportant they might look at you. While filling the diary, please consider the following questions:

- How your privacy concerns affected your daily routine?
- How much information do you give and get both online and offline in your daily activity?
- Are you concerned of who might obtain and use the information you produce/ provide?
- Any information you did not give because you thought it was unnecessary for the service required?
- Did advertising (in web sites as Ads or through email) annoyed you?
- Any alternative way you found for achieving your objective, avoiding to give personal information?
- Any privacy consent you read before checking/ signing for giving personal data?
- Anything that changed your attitude for further continuity of a service/ facility because of personal data requirements/ publishing without consent/ using for marketing etc?
- Anything that happened today that forced your privacy issues?

Sample ModelF Age: 28M Nationality: Albanian**Personal Data Submission Diary***Day 1: Date April 21, 2014*

| Time | Activity |
|-------------|--|
| 07:00-12:00 | <p>At 8 AM i checked my Facebook account; liked 3 posts, two of them from pages and public share, one from my friend and shared with friends; Also commented on 1 status and 2 photos (for those I didn't see the privacy). Further on I opened a shoes site, Adidas, offered by Facebook as they have 50% discount. I love Adidas sneakers.</p> <p>I then went to Telia (Telecommunication Shop) refilled my card and went to school.</p> <p>Around 10 PM I checked my Instagram profile, liked some pics of my friend and posted one from the last day i made in Malmo for Easter. My profile has only 52 followers.</p> |
| 12:01-14:00 | <p>As soon as I reached school, at 13:45, i checked my Gmail and school mail. I had a suggestion for a job related to my education and experience; I googled for the company and then applied through my LinkedIn account. The site said that they will keep my CV in their databases for further matches so that is a really nice! I re-checked Facebook and commented on a pic of my friend she did in Monte Carlo. And as I saw an Ad about new app 'Promillekoll' which measures the degree of alcohol one can drink without getting drunk, I took it in my HTC smartphone. As usual a privacy agreement, which lead me to an extended version, before installing it popped up, but i just press Accept without reading. After school, I stopped at ICA (supermarket) bought some daily grocery and used my Debit Card and ICA card so I can gain points from buying. I took the bus home. Here in Sweden the buses are with camera.</p> |
| 14:01-20:00 | <p>As I'm doing my master thesis I have to search a lot through internet. Once it was hard to read books only in library, but now with google is much simpler. From 5PM to 7.30 PM i was working on my thesis, trying to find good articles. I used Google Scholar and subscribed in New York Times Magazine; Oracle BI and some other pages so that news related to my topic will daily been sent to my email. After some time I checked my twitter account but i wasn't in the mood to write something.</p> |
| 20:01-00:00 | <p>At 10 PM I checked my email account; I had an email with seasonal offers in Spain, booked NH hotels before May were with 25% discount, but I have been in a NH hotel I Netherland and it was not worth the money so I just discarded the offer. Then I searched in google for the 'Frozen' movie to watch. I have an account in a movie stream so I used it. Re-checked my Facebook account; one of my crazy friend at school posted a quite embarrassed photo and I wanted so much to do a comment, make fun with him, but then I thought what if that shows up as news feed to my whole Facebook friends...so I reconsidered the idea and didn't comment.</p> |

7.4 Appendix 4 – Interview transcript with Cees Boon

Interview date: 10:00, 25th April 2014

Participants: Cees Boon (CB), Flogerta Banaj (FB), Klaudia Dardha (KD)

Interviewee: Cees Boon (CB)

Interview type: Skype voice call

Interview duration: 59 mins

Transcribed by: Flogerta Banaj (FB)

Transcribed date: 27th April 2014

| Line | Speaking | Text | Code |
|------|----------|---|------|
| 1 | FB | Thank you for the willingness to participate and in this interview. Thank you for your time. I am here with my writing partner of the master thesis, Klaudia. We will conduct this interview through Skype and for the purpose of the research we need to record it. Is it all okay for you? | |
| 2 | CB | Yes, it's ok. | |
| 3 | FB | We are making a study about Big Data towards Privacy Issues. And additionally we will look through any gap among the participants in this Data Supply Chain. We have prepared several questions and we can go straight to them in order not to take you a lot of time. | |
| 4 | CB | Yes, ok. | |
| 5 | FB | Can you tell us little about your professional background? | |
| 6 | CB | I work as a data scientist on consumer behavior for retailers Business Intelligence, Retail Management, Data mining, Information Analysis, Data Scientist. Founder: 'BIG DATA Ethics', Professional Network Group to stop Big Brother stealing value from Big Data. Founder: 'Retail Intelligence Initiative'. Network group for professionals in Retail, Business intelligence, Science and IT Founder: 'Meteo Sales Research Group'. Joint insight on the impact of weather condition and forecast on consumer behavior. Founder: Business to Consumer Analytics Group. Professional Group for marketers, business analysts, retailers and e-tailers Founder: 'Prospectant.com', the Innovation Sales Network for Knowledge based Selling Professionals Expertise: Business to Consumer Intelligence, Intelligence Cycle, Consumer Mood Research, Business Process Simulation and Design. B2C Model | |
| 7 | FB | Do you remember what lead you to this particularly field? The motivations of it? | |
| 8 | CB | Retailers have large transactional databases that are the DNA of consumer. Privacy is an issue for ages, even before the big data hype | TR |
| 9 | FB | You are the founder: 'BIG DATA Ethics' group on LinkedIn, what was the main motivation for this initiative? | |
| 10 | CB | Data scientists like me often see data that cannot be transformed into information or insight without violating laws. I remember attending a convention on big data issues about a year ago where we discussed the ethics. We decided to extend the discussion on LinkedIn. | |

| | | | |
|----|----|--|------------------------|
| 11 | FB | According to Gartner (2001), where the term was initially used, Big Data is described as the rapidly increasing volume of data, velocity as the data go in and out, and variety as the dispersed type and sources of data. Do you think is something missing in this term of definition? | |
| 12 | CB | Volume and velocity are obvious. For me Variety and veracity are even more important in the big data definition. The increased power of analytics provides us with tools that can associate data of great variety (i.e. consumer behavior and weather conditions) and the result is information with superior veracity (i.e. predictions with 99% accuracy) | |
| 13 | FB | Can you compare Big Data with traditional Data Mining? | |
| 14 | CB | In fact they are two different subjects. | |
| 15 | FB | What are some of the opportunities of Big Data analytics? Can you give a concrete example from your experience? | |
| 16 | CB | Big data analysis is not a strictly rational science anymore. It is open for more intuitive, creative thinking minds. I think this will lead to surprising benefits for better or worse. | |
| 17 | FB | Which are the main challenges Big Data faces within different fields? | |
| 18 | CB | Privacy, Crime, Concentrated power. Privacy is a challenge because of the ethics. 1-I think too much companies violate the privacy and it will get worse in the line of getting bigger data and copying it everywhere and loosing this information and not protecting good enough, trying to use it for competition. All of this use will violate privacy and Big Data will not be maybe even easier to use but it will also make more profitable. 2-Concentrated power. You see that the governments, we have the example of NSA, we have the example of Google they have so much data, they know so much what I and you are doing and it is so easy to use them nowadays for marketing purposes or enforcing people, government, organization, well whatever as a power organization. So I think it is concentrated because just like money because if you have some money it is easier to make more and if you have some data it is easier to make more. I am in this business too, because I think I can make this a better world. But, if you ask me from an ethical point of view a better or whatever it is that world does not agree what a better world is. My better world could be very different from your better world. So this has an imbalance power in management, in government and so I am afraid that is making this world better is only knowing the sense of Big Data having this ethical issues. | RE IN PA |
| 19 | FB | Which department uses more Big Data analytics? | |
| 20 | CB | Let me say that traditionally the part of the company is accounting but in the future Big Data will be used all over there, all over the organizations and especially in the more tactical issues not that strategic anymore So I think it will be used on the lower level in any organization even on even on the executional level. | MA |
| 21 | FB | Who is responsible for the data management in a company or in your company? | |

| | | | |
|----|----|---|------------------------------------|
| 22 | CB | I think this is a key question. Who is responsible of the problem is that if we look at in the humans, and human intelligence we have social background in not violating Big Data and using information we have control. But, in the situation we are now we have not only human intelligence. But even the electronic intelligence not like the artificial intelligence but I mean the data that is stored everywhere. And who is responsible for that? It is very hard to point that. You are responsible for what you know and I am responsible for what I know. But who is responsible for what is in our databases. That is very hard to pinpoint. And I think that s the key question. So, who is responsible for your PC. Who is responsible for your systems? Very hard to say, that s the key question. You can pinpoint the woman or the men being responsible for having data that they don't even know to have. | RE |
| 23 | FB | Would you call this, like the lack of data management being the source of these privacy problems? Top management? | |
| 24 | CB | If you pinpoint that as the electronic intelligence what it is that's something completely the different should be regarded totally different from the human intelligence. This means yes. | MA |
| 25 | FB | Who should be responsible for addressing different issues that emerge because of predictive analytics? What is the role of the government? Of data providers? Of existing tools? Of educational institutions? Of mass media? How should people and organizations be held accountable? | |
| 26 | CB | Well, I am not really sure. Think, it is not possible to predict what are you going to predict something, I have to put like that. Because if you take Big Data and trying to predict very often you don't know what are you going to come across. You might find things you want to find what you are going. You can find things that will interpretation without knowing that. If you say who is responsible, well, it's me like a data scientist I would like to have legislation. These rules are very interpretable in a thousand of ways and in practice that means that a lot of data sciences take advantage that they don't know that shouldn't do other data science may be the more ethical people. Warrens that there is no pure law. It is a pure ethics and not legislation. Government: well the only think they should do is to make laws and most government are really involved trying to get data to strengthen their power and that is something that should be as soon as we can. They should not do anything in that place only get rules for us. | RE LEA LEA IN |
| 27 | FB | Are the actual rules are suitable for the data ecosystem we are living? | |
| 28 | CB | The rules are not good enough in EU, that s really true. All the things that come the last years were not already updated in the moment they put in act. At this moment Data provider, they are responsible of course. It is not like the issues you have on owning the data. Most data providers do not really own the data they can't live without the data that the people violate the laws. | LEA OW |
| 29 | FB | And from the consumer perspective, do you think they are aware? | |
| 30 | CB | I don't think that they are aware of the data what is happening some people I don't think they are interested. Most people they don't know that they are at risk it is still very profitable to take the advantages. | PA PB |
| 31 | FB | About the education institutions? | |

| | | | |
|----|----|---|--------------|
| 38 | CB | Well, if you take for instance the food supply chain. We have all the kind of the checking where the food is good and where the products are made and where the base wrong material is coming from. Something like that should be in place for data. Data transformation in information, information transformation into insights. There should be. If you talk about ethical legislation or looking at that this transformation. | LEA |
| 39 | FB | What are the policies in place to secure data transfers between companies? How is this data protected from being hacked or leaked? | |
| 40 | CB | Technically I am not in this business. I have an idea but I am not an expert on that. | |
| 41 | FB | Is any data aggregation being treated differently when analyzed prediction might harm ethical issues? | |
| 42 | CB | The aggregation from data to information. I think it is there was one of the problems | LEA RE |
| 43 | FB | When and where should access to information be curtailed because of the potential for misinterpretation or abuse? Who should get to decide when information should be curtailed? | |
| 44 | CB | Well yeah that the key question. When and where should access to information...? On the point we know that that is necessary. The problem is that who will be doing it? Well that is all what it is about and I can't really say that in any organization these decision will be made everywhere in any organization Often we cannot predict the occurrence of ethical issues in Big data, we just come across interesting insights that may beat the competition. Well. I think both. They are both because who's to blame?! For me this question maybe is too Key. This is what it is all about, and is very hard to get a straight answer from an ethical point of view. We are all to blame! It depends on the problem... Yeah, well, I anybody who is violating any of your privacy based on what you did, and that's the reason why I said that anybody has the rights of its own truth, so anybody that is violating is responsible for what he is doing, even for yourself. | MA RE |
| 45 | FB | What rights, if any, do individuals have to data collected by and about them? | |
| 46 | CB | I think you are referring at privacy data only or data in broad way section? | |
| 47 | FB | Yes, we can say data in broad way because some data are personal private the same data for another one aren't private. But from your perspective as an individual. | |
| 48 | CB | Well, I think that everybody has the right to collect data but the truth is to extract from anything as well as it is violating the rights and the obligations that everybody else has. So as long as there are not very good rules you cannot answer this question. Because if there are no good rules you don't know if you are collecting data that you shouldn't have or that you should be collecting. Let's go back to what we said about 'do-s'. You know that we shouldn't sell/talk about someone else. So if you know about data that you collect from other. Other people were not meant to be used by someone else, just don't sell it. Probably | DH LEA |

| | | | |
|----|----|--|----|
| | | in some cases it is though because of getting money from it, or in some cases even it is done because you think that you are doing good at them. Let me go to the cancer checking. That says it all because just for health care we are re-analyzing the picture that we're taking for cancer, and now we see a lot more than we saw in those pictures many years ago. So we know a lot of people that are going to get cancer within 2 or 3 years. What are we going to do with this information? It was of course meant to heal people. Now we have information of people but we cannot tell them because we are not allowed to this, this really ethical issue. Now we have information of people that we know they are going to die in 1-3 years but they don't even know themselves. And that can be a problem, not the only one. The people that will be reanalyzing this information are the people of the insurance companies. Now what happens when insurance companies have an information like that? That is really an ethical problem from doing-doing ? | |
| 49 | FB | So it is all about analysis and the way we chose to share these results that arises these privacy issues in Big Data? | |
| 50 | CB | Well, in this case it was collecting the data, its providing us now with the possibility to analyze and re-analyze. So you might say that collecting the data intrudingly it was Ok, but maybe it should have strongly curtailed guards? Because everything that you are collecting now, data is a potential violation of privacy risk. | DH |
| 51 | FB | But now everything has started, do you think that it is too late to re-design a so we have to think in this position that we are to find any possibility for improvement or government to stop this things? | |
| 52 | CB | I think it is a bit late but we will see the example that I gave... we can deal with that will be a lot of more examples we will not be able to deal I hope we will but I am not sure... | |
| 53 | FB | Do you think that data is often duplicated? If so, does that leads towards the changing ownership? | |
| 54 | CB | Yes it even duplicates itself. | OW |
| 55 | FB | What rights do data providers have over the commercialization use of their data? | |
| 56 | CB | Rights don't really matter | |
| 57 | FB | Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete. | |
| 58 | CB | Yes, even we have the tech to wipe up all the I think that we don't have the tech wipe out all the data some companies say they have it is not really possible wipe out all the data everything... especially the things that you wanted to wipe out I think that are not very ... Well, privacy it is not Wiping out the things to wipe out It is almost impossible. It is done without knowing that is knowing what is done. E.g. I worked for a lot of companies If I want to use the database as soon as I going to analyze it I going to copy it., cleaning it making it usable If my college want to go over my work he has to copy it again that is a problem The duplicated is far more trigger if you want to wipe out It is coming back.. That are a lot of situation ... It is a problem...It is very necessary this law but it is very difficult to make ... | |
| 59 | FB | How do you define Privacy? How is it related to the IT evolution, specifically with Big Data? | |

| | | | |
|----|----|---|-----|
| 60 | CB | I think that every person has the right to define for himself what is information about him or her and what is truth of information what is due to information what is the issue of that privacy. As soon as I have information for someone else I should always aware of that. Well, what information about me is the right info and everybody should get that in mind. Privacy is only about personal privacy, there isn't other privacy then personal privacy. It is already changed and adopted .The way that I think to it is that next human privacy we have electronic privacy and because we don't have personal way to interact with this, the electronic intelligence and well as long as there is no an artificial at it is not very yet that has the social human background maybe it will be in 10 or 20 years and the foundation. It is very difficult to make really ethical question. | |
| 61 | FB | In the event of a major privacy break, who is ultimately responsible? | |
| 62 | CB | Very good question, very hard to answer. I think everybody involved in the process of collecting, reporting, analyzing and consuming the data. | |
| 63 | FB | Do you think that existing legal assurances present in Europe offer an effective protection for Internet use? | |
| 64 | CB | No. It is too late, there should be a lot of discussion about it and thoughts to take chances It is very good that we do this well it only one part and it is very important ? of it a fact and the fact that you know there is a problem is a good thing. So, the government does not recognize it as a problem. | LEA |
| 65 | FB | 21. What are the limits on how public records data can be used and, vice versa, on how commercially collected data is used by public institutions? | |
| 66 | CB | I think if you have the really due to find the... ethical legislation... the problem of using the data everywhere ? I don't really see the difference between public data and private data and commercial use. Yes, I see the difference between them but there is not difference about an ethical point of view. | |
| 67 | FB | What privacy concern do you have regarding the information that you as an individual generate? | |
| 68 | CB | The problem is that I don't really know how the things that I do in this moment, using data, using information, I don't really know in about let's say five years we are looking at that. I know that's data science in all over the world are working with data and doing things with this data that might be ok now, but not in five years So, I really often use the case of LanceArmstrong, ten years ago everybody was saying that he was doing the right thing. Ok, he is using it I think its fine using that. But, now ten years later, now we are blaming him for doing things that he should not done ten years before for things we didn't blame him before. What is going to happen with data science. The things that now we are doing could be quite different in five or ten years but could be traced back to anybody who give things. I could be make things now for my client Big Data maybe not violating personal privacy but in five, ten years maybe we can look at that in a quite different perspective. I am worried about that. | PC |
| 69 | FB | How do you assume the future of privacy? (any concrete example) | |

| | | | |
|----|----|---|----|
| 70 | CB | Big data is not only about new data, its merits and pitfalls are also in associating and reviewing old data. Take for instance cancer-scan-data: reviewing old data now gives us far better predictions on the chances of getting cancer. This leads to personal health information owned by insurance companies. That is an ethical issue. | DH |
| 71 | FB | Based on your experience and point of view, is there any other issue you would like to add in terms of privacy as a challenge of Big Data, before closing the interview? | |
| 72 | CB | Not really. I think the questions were addressing the right things that I had the possibility to put in there what I think it was an issue. | |
| 73 | FB | Okay, thank you for the interview. We appreciate it. As soon as we will finish the transcription of the interview we will send it to you. If you have any comment to it, it would be very helpful for us. | |
| 74 | CB | I appreciate it. Thank you. | |

7.5 Appendix 5 – Interview transcript with Jon Page

Interview number: 2

Interview Date: 10:00, 30thApril 2014

Participants: Jon Page (JP), Klaudia Dardha (KD), Flogerta Banaj (FB)

Interviewee: Jon Page (JP)

Interview type: Skype video call

Interview duration: 56 mins

Transcribed by: Klaudia Dardha (KD)

Transcribed date: 30thApril 2014

| Line | Speaking | Text | Code |
|------|----------|---|------|
| 1 | KD | First of all, thank you for the willingness to participate in this interview. We hope it's ok to conduct the interview through Skype as we need to record the call. | |
| 2 | JP | Yes it is Ok | |
| 3 | KD | I'm here with my thesis writing partner Flogerta. We are making a study about Big Data toward Privacy Issues. And additionally we will look through any gap among the participants in this Data Supply Chain. We have got a number of questions to go through. We do not want to take you a lot of time, so let go straight to them. | |
| 4 | JP | Ok | |
| 5 | KD | Can you tell us little about your professional background? | |
| 6 | JP | I worked about 35 years, all of the years in IT and all of it around the subject called data warehousing which is the precursor of big data. As a senior consultant in sales in Teradata. We started a company in 1997. I was working for Oracle, in charge with data warehousing, BI in Europe for 5 years. I did freelancing consulting in the area of Big Data with EMC2 and HP. And i have written a number of books for different subjects related on Databases, BI architecture, and I wok now for the European Big Data Institute. | |
| 7 | KD | Do you remember what lead you to this particularly field? The motivations of it? | |
| 8 | JP | Well, Big data is 4 or 5 years old term. What actually brought me into the world of DW etc., I can't remember. I was trained very early as Cobol programmer. I was good at it and I loved programming. And just went on from that. | |
| 9 | KD | According to Gartner (2001), where the term was initially used, Big Data is described as the rapidly increasing volume of data, velocity as the data go in and out, and variety as the dispersed type and sources of data. Do you think is something missing in this term of definition? | |
| 10 | JP | I think the definition is not very helpful. Anyway, it comes from one guy in Gartner several years ago, who was probably put on the spot and asked what Big Data is. And he came up with a decent definition, don't get me wrong, but not very helpful. It misses out a huge amount of the traditional data that you find in BI data | MA |

| | | | |
|----|----|---|----------|
| | | warehouses. And it gives an impression that Big Data is something about data. But I don't think it has anything to do with data really. It is much more about the way of thinking, businesses actually can improve the way they do business by using data. Whether it is big or little, high velocity structured and unstructured make no difference. It is the willingness of the company to use data. That is what is changed in last couple of years. | |
| 11 | KD | What are some of the opportunities of Big Data analytics? Can you give a concrete example from your experience? | |
| 12 | JP | Oh yes, I mean basically, the real opportunity I think can be collectively thought of as being predict, what is going to happen in future and then use data to create business strategy, which enable you to take advantage of what you think is going to happen. The simple one that is being used for years is the chain in Telco. It is so simple, what is going to happen and if some people are going to turn profitable to you, you have a chance to do something about it. Big Data might enable certain companies to identify and change the future. Lots of money can be made, opportunities, if you can change the future in the way you want it to be. | |
| 13 | KD | Which are the main challenges Big Data faces within different fields? | |
| 14 | JP | Oh yes, I am trying to be clear that, the problem is not technical, I don't see any real issues with the technology today. The issue that separates the companies that fail and companies that succeed in this area are simply how well the IT environment people work with the business people. Because all the advantages of the Big Data comes from a business perspective. If the business cannot identify what that perspective is, then nothing the IT guys can do can come. So it is what I call, i spent a lot of time in this subject, 'Business Alignment'. Who does the business work closely with the IT to take advantages of initiatives such as Big Data?! Exactly the same in what they called in Data Warehousing. | MA |
| 15 | KD | Which department uses more Big Data analytics? | |
| 16 | JP | It is clear, marketing. Absolutely clear. This idea of Big Data, which is actually ridiculous comes from the analytics and social media. And that is why it is hyped at the marketing department more than anywhere else. It won't change, other operations will gradually understand the benefits, which is manufacturing, may be Supply Chain Management, or whatever, But it is clear that today it is marketing. | MA |
| 17 | KD | Who is responsible for the data management in your company? | |
| 18 | JP | No one, I think if you look at 98% of companies, they would not be able to point the finger at one person and say you are responsible for big data and information management in this company. And is a huge failure in our business. I am sure you guys would have heard about the term governance. If you look at data governance, it is very immature. Hardly anybody, especially in Europe has an organization in roles and responsibilities in the world of governance. Only when you have done that you will be able to answer your questions and | RE MA |

| | | | |
|----|----|--|--------------|
| | | point your fingers and say that person, is responsible for transformation in information. | |
| 19 | KD | Who are the providers of data from where companies get the data to analyze? | |
| 20 | JP | Well traditionally, data comes from their own operational systems. So a billing system in Telco, the system that records ordinary transactions in a bank, they have been the Big Data sources for 20 years. Now, you are starting to see, obviously social media, (people talking in Twitter and Facebook). And that helped, people are starting now to figure out is there value in what they call non structured data, such as pictures or scanned images or documents etc. And some people talk about the ability to capture, process, and control streaming data basically. Such as, thousands of media in the nuclear site, So those are the kinds of the sources from where the data is coming from. But still the useful data lies in the transactional bases in the system, for an airline or a retailer or bank. That is still where the money lies in analytics. | DS DS |
| 21 | KD | So you think that companies are able to capture data by themselves or there are other companies or brokers who sell data? | |
| 22 | JP | Oh no, the basic facts if you like for a CEO to run a company, lies in that companies operating system today. That is how, 90% of the real world BI data comes from. To argument that you might need all sources of information such as demographic information competitor information and you can buy that, companies do buy that stuff. In England or Europe - Neilson. But in terms of big amounts of data sourced from a different company that in my experience is not a huge priority at the moment. It comes with a lot of issues, especially privacy and legal issues, And i have not seen that term much. It would be interesting. Like I said there is one thing storing your own data and something completely different storing somebody else's. | OW |
| 23 | KD | Based on the large amount of data, you will see the government as a seller or as a buyer? | |
| 24 | JP | I don't think government should sell data anyway. They should give it away free and to a larger extent they do. And you can find a lot of statistical information now, at least in UK on the web. Do they buy data, I am absolutely sure they do. So it is probably two way data for them. Buying and selling, and the law of giving away information free to the general public, which is new. I have no idea. | DH,DS,SE |
| 25 | KD | What are the policies in place to secure data transfers between companies? How is this data protected from being hacked or leaked? | |
| 26 | JP | I don't really know how to answer that question. I think that the whole area of transmitting data between companies especially in international boundaries and is a very complex area. And when it comes down to governments it should be a group of people whose job is to understand and how they move within the company and between companies they deal with, making sure that the only data they really need gets moved making sure that they understand how and where the data is stored who is using it and making sure that they understand various international law, that is a complex area. I am sure you need specialist people. Because it is complicated, I | POA |

| | | | |
|----|----|--|-----|
| | | mean if you do not do and control these things then you might as well open the doors for every hacker in the world and say here is my data, you have it. That is just the way we see things at the moment. Responsibility has to be taken by some people and I suspect many organizations do this well and i know many of them don't. | |
| 27 | KD | Do you think that existing legal assurances present offer an effective protection for Internet use? | |
| 28 | JP | No, Clearly you see it every day. Major internet issues, like the SSL security stuff. And then you find the biggest world's companies depending on some security that is written by some open source chap, are around somewhere. You can't expect it can be honest big data, transmitting data that are supposed to be public network. These are new things. 6-10 years ago, people used to lease clients to do this step that were private, and very difficult to hack in to with actual physical devices. Now anyone with brain can get anything they want from internet. Personally it do not scare. But as a professional it would worry me a great deal. | LEA |
| 29 | KD | Are any data aggregation being treated differently because when analyzed prediction might violate or prejudice ethical issues? | |
| 30 | JP | Not sure whether I got the question, I will paraphrase it. Whenever you deal with detailed data that actually has in depth data, a way of identifying the individual person, like your name, then there are real issues. In every country. However, quite a few companies are really interested in your own personal detail. They want to know how many people are there like you. That is much more useful for them. They just have techniques to change your names into some meaningless, aggregate by your different attributes that is the stage when they anonymize and can't link it back to the individual person. So it is dealt somehow safe and non-intrusive. | |
| 31 | KD | So you mean that companies are not interested in personal prediction analytics? | |
| 32 | JP | No, there are very few big companies who are interested in you as an individual because it is too expensive. Let's say in UK, let's say they got 20 million subscribers. There is no way they are interested in individuals. For marketing purpose or anything like that. They want to know whether there are other many people who look like you and behave like you and that kind of advertising. But the big guys with millions of customers they are not going to look at you as an individual. They are looking at same as , look and behave the same way | |
| 33 | KD | When and where should access of information be treated differently because of the potential for misinterpretation or abuse? | |
| 34 | JP | As soon as you can identify a person individually there is an issue really. I don't need your name to look at you, but all sources of data that you do every day, where you are, whom you are calling. If you are clever, you can start to understand who you are without knowing your name. I think that is kind of disturbing, it is someone who came up with this idea of Big Data, and now it is matching your everyday transactions like a bank or shop whatever, with the way you communicate, on social media and bring in that data together. I just | |

| | | | |
|----|----|---|----------------------|
| | | think at the moment there are few laws or regulations or even customs and culture that stop that happening or to regulate the way that it is done. But i think it is just matter of time before people wake up and say ‘no we can’t allow this. This should stop’. Because on Facebook and other social media, you pretty much give your name away all the time. CDR (Call Detail Record) never has your name on it. But it is quite there up in social media. Remember, social media goes on the same line as the telephone call. And it can be interpreted and trapped by exactly the same technology and company. So when you look at the telephone guys, mobile guys when they are looking, calling and probing the network they are not just seeing mobile calls, they are seeing your SMSes, all your social media chat, in text. That sort of thing is deeply disturbing, and government should make sure that the telephone companies actually store the content of the call, then i think are deeply worrying. | LEA EP LEA |
| 35 | KD | But what happens for example when an insurance company captures my data from social media or other sources, then this data analysis will be personal not only general as in marketing?! | |
| 36 | JP | No: Again it is kind of difficult to generalize because, different industries are very different in the way they react in the way they hold your information. I think insurance companies could be a problem. That would be exactly what you are saying. I think banks obviously could be a problem, Insurance in particular, because there is so much fraud in insurance. There is so many ways of suggesting, fraudulent ways are over the social media. Posting about what you go away with. There is a lot of stuff on social media. That is very interesting to an insurance company. Even on legal basis. Unless they start using that stuff and prosecuting people. Based on social media comment, to perhaps the time between someone took an insurance policy and made a claim. All the time that is they can give away some bodies can make a fraud. Another problem is that the fraud is so extensive in insurance. That is difficult to know. We would want them to stop fraud. But in order for them to stop fraud, we may need to give them a lot of the data we let them have. | |
| 37 | KD | Who should be responsible for addressing different issues that emerge because of predictive analytics? | |
| 38 | JP | Well, That person does not exist in most of the companies, There are many levels of responsibility. Firstly, you might need to set that person responsible of the collection of data, privacy, and security of data and the management of it. I think that is a business concern. I have not seen individuals in any companies, Now, for example, within a marketing department, i would point fingers at vice president of marketing. If there are people in marketing doing things unethically with analytics, to me, the perk stops with the top guy. Which is vp of marketing. I think if a company is using data unethically, the owner of responsibility has to be the CEO. I have no doubt on that. He carries the responsibility of everybody in his company. They are using data unethically, in correctly or madly is CEOs responsibility. | RE RE |
| 39 | KD | Who is the owner of the information produced? | |

| | | | |
|----|----|--|----------|
| 40 | JP | No Idea, It is a good question. if you get into a world where you actually deployed, data governance correctly unspoiled by an organization, then you probably are going to have someone like a data steward in marketing, another one in sales, another one in operations and if you can get to that way, then they can become owners of the data. If you can't, then i got no idea. That indeed is a scary thing. Today if i am the CEO of a company, and i did not have those roles and responsibilities in position and i was bringing social media data into my company, and i am seeing my employees online then i would be very concerned. | OW |
| 41 | KD | Who is responsible for its integrity? | |
| 42 | JP | Same person who is in charge of its ownership. Which we figured out that who really does not exists. If i have DW system and Big Data system, the integrity of the data is looked down on a system by system basis. So that is not good enough for a company. So normally in an operational system the integrity of data is good/ high quality because it is companies customers. For example, let see some purchasing data, is very high quality. Obviously, because people pay and tests are going out of business. But that data is not necessarily running their stock control system. Their stock control system also might be very high quality in terms of integrity of data. But it is measuring something different from the sales system. So you can't compare them, at any time for example you cannot make a single calculation says what is the value of 'tooth paste' i have in stock? Because the two systems are fine on their own, but they don't understand the same meaning of the word stock or price or customer. Those prones have been there for 20 or 30 years, big data is going to make them worse, when i talk to companies i tried to make them put the basics right. Before you carry on and make things worse. Basics are getting into the feed into the company in words of master data management, quality and governance. | MA MA |
| 43 | KD | What rights, if any, do consumers (individuals or organizations) have to data collected by and about them? | |
| 44 | JP | There are more changes every now and then, generally i think the picture is that the person owns data and has every right to ask the company to disclose all the data they have about them in a certain time period. I think they are protected by the laws in UK. So for example, under certain circumstances i can demand customs tell me within 20 or 30 days of all the data they have about me. I don't know how far that rule goes. But at least somebody thought about it, which is good. Generally who actually owns or aligns collection of data is determined by whether you take a vote. Like it says, I don't mind the company using my data or I do mind company using my data. Unless people don't really understand that box exists, I think that is game changing and so people tend to opt in. They tick the box on the webpage to allow companies to use their data. I am not sure the concept of ownership actually exists, people either have permission or no permission to use the data. Who actually owns it is quite difficult to say. | PB OW |

| | | | |
|----|----|---|-------|
| 45 | KD | Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete. | |
| 46 | JP | I think for that person it becomes almost impossible. Logically there is certainly laws and things allow people to force that the data to be deleted from a computer. I am sure of that. But delete the data from web is an entirely different thing. Because it is not possible to find where your data is in web. We just don't know where to go. It can be anywhere. I am not talking about in the cloud, i mean physically once you put the data on the web, you are never going to convince yourself it has been deleted everywhere. One copy left somewhere in the cloud, people can reconstruct the original data. | |
| 47 | KD | So the inability to delete this data makes the consumer for sure not the owner of the data | |
| 48 | JP | I think your ownership is purely conceptual, it is not a physical thing, at the end of the day RBS owns physical disk that have your data that does not mean they really have it. Or they are going to keep it, or they will use it. Some countries/companies recommend how to keep it and use it whatever you say and store it. And in many cases there is nothing you can do about that. | |
| 49 | KD | How would you define Privacy in the era of IT evolution, specifically related to with Big Data? | |
| 50 | JP | Well, I think a lot of the big data stuff we have talked about is not such personal data as the stuff you do every day, in a bank, retail, or on the web. So you look at the big data and common stuff. None of that has any personal issue , where your personal data is, what is in Facebook, i don't use any of them myself as i don't trust them, on twitter , whatever bank you use, that is where your personal data is. ICA, CITY GROSS. They know wherever i go, they know what i eat, they know my credit card details, and they know everything. That is a retailer, so i don't worry too much about big data. Certainly some really nice elegant solution for a guy with a mechanical heart sending out senses to some place to tell when his heart machine would break. There are fantastic applications. That man's data can be picked up by anyone with a Wi-Fi connection closed to it. But there applications are rare in the real world. Everybody eats, everybody banks, and everybody travels in airplane. | DS |
| 51 | KD | In the event of a major privacy break, who is ultimately responsible? | |
| 52 | JP | The CEO is ultimately responsible. No doubt whatsoever. He is the one going to jail. | RE |
| 53 | KD | You mean that google or Facebook might be the responsible? | |
| 54 | JP | This is typical. I think FB & Google are classic examples of these guys by putting data, you are agreeing to them, there is a tick box somewhere. I think if you check that box and say I do not want you to share my data. It is difficult. The idea of using Fb is to spread personal information. So, you might not want google to hold certain information of you, but can you be sure that all your friends post about you, or they don't mind being on the web. I don't know, i don't use twitter and Facebook. Because i think it is totally pointless. | SU,TR |
| 55 | KD | Does the consumer have any possibilities to win over Google in a privacy issue break? | |

| | | | |
|----|----|--|----|
| 56 | JP | <p>No chance. This is the problem. But again you know in context if google come along and steal all your patents and your company and by doing so put you out of business, you got no chance to win against google. They are so big, you can't even imagine it. When you say big, 5 or 10 times bigger. It is not that scary. We will see how they will turn out in the end. The individual taking a privacy case against one of these companies, i don't think we should kill ourselves. It is not worth doing it. I don't want to say just google. It is unfair. Any company which makes such difficult process causes much to do for the individual. I don't use these services, it is high risk to go on Facebook and twitter, I would be thinking, I am not going to set anything but on this, I don't mind the world see. I think that is twitter and Facebook aim. I don't think they are trying to get private data from you. And if you put private data it is tough. How FB is supposed to know what is private and what is not. It is a difficult issue. But again, i don't want to say this, you should ignore this, by talking about social media you are assuming that social media got something to do with big data. I am not sure. I think people who are talking about are, how many companies in Europe download terabytes of tweets everyday through analytics. I am telling you, none. It could be for example, British Airways, they are people very interested in social media , analyzing social media stream, because they are looking for tweets that have BA in it. May be 10 GB a day, Be careful when you associate fb with big data. If you are a retailer in your country Albania, in one year they would collect 3 GB of data. That is the size.</p> | |
| 57 | KD | <p>When the use of the technology is an ethical one towards the power that the Big Data offers?</p> | |
| 58 | JP | <p>Definitely it does. No question. Let's take any mobile company. You are in Sweden, so let's take "3". Those type of companies for example own the cable that your tweets and messages go on. At some stage depending on what they use when you subscribe etc. all of your stuff will go on one cable, in 1, 2 or 3 switches. The company has the capability of proving to find out everything you have done basically. Every sms, mms, tweet, your location, persons location etc., they know everything. The proven technology is hardware and software in the switch and absolutely, the data they can collect can be totally compromising. There are laws and regulation of stopping being used in a bad way. That is the same way the peace force, fire engine people, anti-terrorist people, in the same way the same methods save our lives, it is a bouncing game. The devices and software, they are all there. You walk into any place and log on to their hotspot, it is a terrible place to be. In terms of persons holding personal information of your pc.</p> | |
| 59 | KD | <p>Do you think there is lack of education among consumers? regarding their privacy and data they share</p> | |
| 60 | JP | <p>Incredible. There should be some adverts on TV or documentaries or in the newspaper. It is amazing what people don't know. For many years i have lectured to marketing people that you can get in telephone companies and banks about what information they already</p> | EP |

| | | | |
|----|----|---|-----|
| | | have from their network and the market people look at me like I am crazy. They don't even understand what is there in their own data or what they collect every day. So how the general public understands i really don't know. People understands only after something bad happens. | |
| 61 | KD | You think that these companies are using this lack of education and information consumers have, to collect data and reuse them? | |
| 62 | JP | I think most companies in the world today are thoroughly respectful and working in guidelines. That is most of the companies. There are other companies, I am sure that work in the dark world here in the web collecting data. And is bad to be a company that collects data but it is also bad to be a company which buys it. I think there are companies like that out there for sure. It takes long time for big company to change policies change roles and responsibilities. We need to be patient. We can't have expectations too high... Don't put anything on the web, you don't want everyone to see. | SU |
| 63 | KD | What privacy concern do you have regarding the information that you as an individual generate? | |
| 64 | JP | I would answer in two ways, The data i put deliberately or knowingly on the web, i have no worries about it. What I am worried is about data that i am not aware of, some data that is generated by me. For example when cameras in the street, supposed to capture my speed, capture photos of me, a part of the basic data they would collect, my speed, they also know where I'm going, which direction etc. the generation of that data i cannot control. | PC |
| 65 | KD | How do you assume the future of information privacy? (Any concrete example) | |
| 66 | JP | I think we will see more regulation about usage of the personal data. In what form i don't know. And i think we are going to have a more user education. In a more rational way i would say don't say anything that you don't want other to know on social media. Generally we are in lot better position that we might have been but where it goes from now i don't know. | LEA |
| 67 | KD | Based on your experience and point of view, is there any other issue you would like to add in terms of privacy as a challenge of Big Data, before closing the interview? | |
| 68 | JP | I think i will say one thing that is important. At the moment people are wary of this idea of Big Data. So you can take a simple example, the location of a person and the idea to market it by sms. i see that 70% of people don't want that to happen. And when they ask them why, they say that as i buy the service i want to get some real benefits of it. So the main point is what the consumer gets to what the provider benefits. And that is not balanced, that means to be addressed. | PB |
| 68 | KD | Thank you for your time we appreciate it a lot. Once we finish the transcription, we would like to share it with you, so if you have any comments or interpretations we would be happy to get them. | |

7.6 Appendix 6 – Interview transcript with Trevor Peirce

Interview number: 3

Interview Date: 10:00, 29th April 2014

Participants: Trevor Peirce (TP), Klaudia Dardha (KD), Flogerta Banaj (FB)

Interviewee: Trevor Peirce (TP)

Interview type: Skype voice call

Interview duration: 45 mins

Transcribed by: Klaudia Dardha (KD)

Transcribed date: 29th April 2014

| Line | Speaking | Text | Code |
|------|----------|--|------|
| 1 | KD | First of all, thank you for the willingness to participate in this interview. We hope it's ok to conduct the interview through Skype as we need to record the call. | |
| 2 | TP | No problem, Ok | |
| 3 | KD | I'm here with my thesis writing partner Flogerta. We are making a study about Big Data toward Privacy Issues. And additionally we will look through any gap among the participants in this Data Supply Chain. We have got a number of questions to go through. We do not want to take you a lot of time, so let go straight to them. | |
| 4 | TP | No, problem. | |
| 5 | KD | Can you tell us little about your professional background? | |
| 6 | TP | I am currently coordinating research for EU funded work on the Internet of Things in the areas of governance, security, and privacy. I have fulfilled that role since 2010. I have been involved in European data protection; legislation, proposed changes which is still on the table now. I have been involved in the EC RFID recommendation and the mandate. So I have been involved in a number of EC level Internet of Things related topics since about 2008, 2009. The EC RFID recommendation was published in 2009. Further background about me: I have been working for myself since 2006. I work as a consultant in ICT space, supply chain, technologies and encompassing all kinds of related topics. I became involved in EC initiatives since 2009, because they touched upon many of the things which I have been working on in this ICT space. Prior to that I was the Standards Director for EPC Global (a GS1 organization) focused on Walmart and retail standards, but also on health care such as the serialization that is going on at the moment all over the world. And prior to that I worked at DHL, as a program director for RFID and prior to that as an engineer manager at their global office for 12 years. My qualifications are in engineering. | |
| 7 | KD | Do you remember what lead you to this particularly field? The motivations of it? | |
| 8 | TP | Big Data is a topic that is in the center of all of these, in terms of collection and processing of data with impacts on privacy and data protection. Which are two topics I see differently. Often interchanged by politicians and others in the same sentence in a way that would lead people to believe that one thing represents that other. But I do not see it that way. Indeed, the reason I got involved with big data and the future | DS |

| | | | |
|----|----|---|------------------------------|
| | | internet related things, is because I see the Internet of Things as a sizable niche part of the future internet environment where Big Data is a significant element. | |
| 9 | KD | According to Gartner (2001), where the term was initially used, Big Data is described as the rapidly increasing volume of data, velocity as the data go in and out, and variety as the dispersed type and sources of data. Do you think is something missing in this term of definition? | |
| 10 | TP | Definitely! Certainly, that definition is ideal for the supply chain of data. I think there is a better formula. Because we have data at its very basic level, we also have information (or 'modified' data) created from processing of information. And indeed, we can see different origins of data. Things that are coming from social media that people publish. We also see in Internet of Things, real and virtual things that will publish things directly and we see all kinds of processing of these information leading to yet more big data. Internet of things illustrates basically the virtualization here so we have data being generated by the internet itself. And indeed, it is rather too simple to talk about it as volume data velocity alone and considering solely the acts of data creation, retrieval and processing. I think we need to be more precise when considering the definition of what Big Data is. | DS |
| 11 | KD | What are some of the opportunities of Big Data analytics? Can you give a concrete example from your experience? | |
| 12 | TP | Sure. I work on the security areas, so, a lot of organizations chase Big Data to authentication services related to security reasons, for intelligence reasons, there are all kind of big data applications today exploit Big Data. And all kind of ways that are not transparent to majority of us. I see quite, increasing business, (commercial activity but also government related activity) associated with big data. I think we need to take care with how we use Big Data. I remember my professor of Birmingham University, way back when he was giving his lectures, saying that, you know data is in abundance and we have a lot of it now, and most of it we cannot rely on. I mean a lot of it is just rubbish. So we have to be very careful about validating the conclusions we make from the big data analytics. | IN TP DH, DS DH, EP |
| 13 | KD | Which are the main challenges Big Data faces within different fields? | |
| 14 | TP | I think the biggest challenge is that individuals and increasingly move towards creating a digital environment where there is more and more misinformation deliberately created as a defensive measure. For example we have different identification profiles online, many of them are false. We see people who deliberately do that because they are trying to evade some kind of surveillance or cyber threats while engaging in legal activity but these same measures are the same or very similar to those deliberately trying to do things that are perhaps illegal. Increasingly since the revelations of data manipulation and capture in the US, more and more people, especially the younger generation are creating electronic profiles which are completely misrepresentation of the real world. So I can see that shown in my own family when my daughter connected with me in Skype. I use this example in public and presentations too. She was 11 at the time, she went home on the computer and connect with me when I was travelling around the world. As she cannot have a Skype | PA TR |

| | | | |
|----|----|---|-----------------|
| | | address under 18 years old, she connected with me using a profile as Russian 26 year old blonde. I declined it. She asked me why you keep declining, and you know the answer is: ask your mother...Such sources of big data you have to be very cautious about. I also see many people turning to encryption and the dark internet as way to protect themselves from all kinds of threats (real and imaginary). Some of it is just to avoid surveillance, others are doing it for all kind of spam and other kinds of generated information which they do not want to receive. Some appear to seek some kind of anonymity, there are all kind of reasons driving people to exploit Internet tools and techniques much of which are also used by organized crime to escape surveillance. This will also impact upon Big Data's future | TR, IN |
| 15 | KD | What are the policies in place to secure data transfers between companies? How is this data protected from being hacked or leaked? | |
| 16 | TP | Poorly! We still have a huge number of conflicts. It is not clear exactly in which direction (information) security protection will go. Clearly, law enforcement and intelligence services want to have access to information to assist them in fighting against organized crimes. But their influence over legally permissible information security created a weaknesses. Means that organized crime and other people can easily gain and access to "secured" Internet services and information. We have seen a lot of issues picked up. Two weeks ago there was a public announcement of the SSL (secure socket layer) level in internet. There is a major flaw in it and two years have been spent working trying to solve it. It is not a surprise. You only have to look into the encryption methods to know that what we have is fairly basic level of protection today. | SU LEA |
| 17 | KD | Do you think that existing legal assurances present offer an effective protection for Internet use? | |
| 18 | TP | I do not think there can be an effective protection for Internet use. There cannot be. It is a never ending race to get ahead. We cannot win not only for technological reasons but also for the topic I mention above, about LI and Intelligence service access requirements. And so I think some of the answers must lie in education of the public. Public become increasingly made aware of Internet challenges through the media anyway. But there need to be more guidance given to and more information shared about where the information is going and how we really are not protected, more awareness of logic into things that could be threatening in some way to our fundamental freedom. There are all kinds of initiatives out there generating more information and we should be aware of things what Internet of Things will deliver. And we have seen more examples: Google Glass, and more and more of invasive methods of data capture that are monitoring our behavior that have potential returns in terms of wellbeing but also huge risks. | LEA EP DH |
| 19 | KD | Are any data aggregation being treated differently because when analyzed prediction might violate or prejudice ethical issues? | |
| 20 | TP | I am not really aware of any differentiation of the data out there. Sure, enough data protection focuses very much on specific elements of data to prioritize some of them. I think data protection is more of a political tool than an effective mechanism to protect information. Because the aggregation of information often bypasses data protection. | MA DH,MA |

| | | | |
|----|----|--|--------------------|
| 21 | KD | When and where should access of information be treated differently because of the potential for misinterpretation or abuse? | |
| 22 | TP | That is a very, very challenging topic because of what I mentioned is by pulling information around the edge of traditional protection. We have it, the legal protections, you can get to describe or to get the answer you are looking for or discover the world you are looking for to uncover. So I think that there is no way I can think of that we could effectively offer that kind of protection. | |
| 23 | KD | Who should be responsible for addressing different issues that emerge because of predictive analytics? | |
| 24 | TP | I think the responsibilities imposed by legislations today are on the operators. I think there should be some efforts or some more responsibilities or pressure put up on political mechanisms to put up on or greater requirements on anybody handling information or data, to be more transparent about what they are doing. I think transparency aspect is one that we don't do enough. What challenge in putting the whole thing is putting up a control mechanism. You might be able to have a form of office that is busy identifying the potential threats, taking effective action against them. I cannot see it happening without this transparency. But I think it is only through the educated public that we can have any form of effective policing. | RE TP EP |
| 25 | KD | But you also said that there is no effective law or regulation to force the organization to be transparent?! | |
| 26 | TP | The work I have been involved with was rather soft law, so it was rather encouraging industry. I have seen some positive results from this approach. I think we have to exploit these mechanisms to a larger extent as they are nowhere near fully utilized. I worked for European retailers who have taken onboard the recommendation on RFID to talk to the public. But have not seen a follow through from European Commission. We still have from the political institutions at European level, because I think the member states are not officially on board. Anyway if you talk about Europe, some of the challenges of the political mechanism are the member state relationship with the central European institutions. But you see some kind of challenges in different ways in the US and other places. Well, in US it seems to be more effective. While many people think there is no such thing as privacy in China. There are many examples of information security though. Security and privacy are not too far apart. Certainly there are lot of people in China who absolutely assure their privacy. So there are mechanisms there, and there are people interested in these topics and in fact people who listen to the presentations I've made, on issues of privacy, security and internet of things in other countries are representatives of governments in Asia. | LEA LEA |
| 27 | KD | Who is the owner of the information produced? | |
| 28 | TP | I think when you get big data, the owner is very difficult to identify | OW |
| 29 | KD | Who is responsible for its integrity? | |
| 30 | TP | Nobody. I have not seen anybody responsible. The whole trend for the internet is that it is a virtual world. A virtual representation of something often completely different from the real world. Like in real world there are virtual objects, there are virtual people, virtual currencies, there are virtual everything. These things are done way beyond the political reach | RE MA |

| | | | |
|----|----|---|------------------------------------|
| | | of governments. In a global world we live in. So, it is not something I can see we are going to find an easy answer to data integrity and is one of the governance challenges. United Nations is not fully global. So I do not see any political mechanism yet. | LEA |
| 31 | KD | What rights, if any, do consumers (individuals or organizations) have to data collected by and about them? | |
| 32 | TP | I think, if you look at protection individuals, I do not see protection is there in part for some fundamental core elements. But, as I said, for Big Data you can bypass most of those anyway. So information that is not directly related to an individual, like name or something such as specific details that are personally identifiable. Data beyond those core criteria that give access to pinpointing an individual or group of individuals. These type of things you can easily achieve. If not outside, legislation will be a grey area that would be very difficult to interpret it. And I think some of the initiatives around the European cloud are found good but I would be skeptical about how they are going to operate. I mean, we have to look at the challenges facing their visions, to data protection and number of obstacles to that. The collection of information, I think there are severe limitations to what can be done. If you look pragmatically, we can do lot of things to try and help the situation. But, I think the complications are already introduced into the processes and trying to establish a new common data protection from Europe is challenging for any citizen to possible understand. So I think simplification is going to be essential requirement if you want to achieve really data protection. I think someone should take leadership and challenge some of the arguments we have of not having data protection. I think transparency is going to be another essential element, there is going to be greater visibility for European systems and those things can be done. I think there is a willingness to do some of these things. There are challenges, major political challenges to overcome at national level, European level and global level. | LEA LEA DH TP OW TP |
| 33 | KD | What rights do they have over the commercialization use of their data? Does the reuse (reselling to other companies) lead toward the changing ownership? | |
| 34 | TP | I do not think. There is only one element of big data. This is ownership they talk about. This is an awful addition of Big Data that is not owned by anybody or the data after being processed, may be in the far corner of the world somewhere published again. It is hard to know, how we would ever address that. | OW |
| 35 | KD | Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete. | |
| 36 | TP | I think that we have to be aware about deletion. The promise that things cannot be permanently erased, I think this is a myth. One other thing established by popular European Union is level of trust, sufficient level of trust of people participating. But I do not believe that information is truly deleted, it kind of deletes itself. Probably one of the best protections is velocity. The change of the world we have around, the change of online profiles, online identities will change things. So the data reflected 5-10 years ago will no longer be relevant. Or even possibly connected to current profile. | TR |

| | | | |
|----|----|---|---------------------------|
| 37 | KD | How would you define Privacy in the era of IT evolution, specifically related to with Big Data? | |
| 38 | TP | I think, the evolution, the Internet of Things is something which is driving forward. It is the ability to captures information about beyond our capacity to possibly understand even. So, I mean, most people still have a very simple idea of this. But in the future there will be all kinds of sensors and information capturing devices, and there could be all kinds of virtual devices creating information about us as well. So the information about any one individual would be made up of a whole list of sources of information that are real and correct and all other kinds that are not. We have to also understand that many of these devices have flaws in them. It could be easily subject to malware and information they publish could be entirely false and incorrect. All sources of information points to different directions and describe differently and that is a challenge of Big Data, really. I think people will move towards craziness of misinformation deliberately or by accident. There is a lot of information in Big Data world that we should rely on. There is danger in people thinking that they can rely on it, possibly when they cannot. | DH,PA RE PA |
| 39 | KD | What privacy issues and obligations arise from changing ownership? Or not having a defined proper ownership? | |
| 40 | TP | The challenge that we have it is going to be an extension of where we are today. Challenges will exists that there are no boundaries for this. We have also challenges that, I do not know which way it is going to evolve. But, I would imagine that people become more sensitive to the information that they publish and the information that they share. That does not mean to say that they share information that means that they completely move into this virtual world. They will think about it. I think younger people will start to develop big data world which is not really representation of themselves. I think we escaped some of these by treating the technology degree of contempt, understanding that we cannot trust it. | PA PA |
| 41 | KD | In the event of a major privacy break, who is ultimately responsible? | |
| 42 | TP | I mean, there will be legal responsibilities placed on organizations who have money, or having influence. That is the way things work. Manmade disasters are already in the world, like airplane crashes etc. It is always a big name, organizations with lots of cash will take it. There are always responsibilities rest upon local governments that are in opposition to the central government. These are all kinds of political and economic kind of trust, and motivation. So I do not see that perhaps people are really responsible will ever really get to or feel any heat from it. Look at the social media today. Social media pushes the responsibility not to any central organization. But the people who are publishing the information, we see that in all kind of industry sectors like critical infrastructure, where the responsibilities are bounced to the lowest level of individual. | RE RE |
| 43 | KD | When the use of the technology is an ethical one towards the power that the Big Data offers? | |
| 44 | TP | I do not think so. It has limitations, but it does not have limits. The technology will continue to evolve. The more rubbish we put on internet, the more challenges and threats we have , the more big organizations will | |

| | | | |
|----|----|--|------------|
| | | put in resources to avoid that it has significant disruption. Probably there is no real limit for that, but as long as there is some interest, political or commercial or a combination of the two there will always be. The limit would be inexistence of the challenges that are being faced. | |
| 45 | KD | What privacy concern do you have regarding the information that you as an individual generate? Personally | |
| 46 | TP | My consideration for privacy are really related to freedom. I concern for groups of individuals who want to take actions or take initiatives where information will undermine fundamental rights and organizations; political institutions will seek to obtain more control over the world we have. Today, already we cannot live in a completely free world. But I think we have really to worry about the number of people, the transparency that the people and their influence of control over will become greater and greater. I do not see we will be much different from China of today in the future. The rest of the world will be the same. I think we have a lot to learn from the Chinese in this respect. The Chinese attitude of an individual to the state and to the organization, the way they communicate is very measured. And I think rest of the world has a big lesson to learn from that. | PC IN |
| 47 | KD | How do you assume the future of information privacy? (Any concrete example) | |
| 48 | TP | I think you can only get information data protection. I do not think information privacy is really clear. People do not talk about information privacy, they talk about privacy of individuals, and they talk about data protection. So, I would prefer to think about data protection and cover the topics under data protection and think of privacy as something more than information. | |
| 49 | KD | Based on your experience and point of view, is there any other issue you would like to add in terms of privacy as a challenge of Big Data, before closing the interview? | |
| 50 | TP | I would like to add a degree of pragmatism that needs to be, and vision needs to be wrote to the debate about what we do. In policy and in law and in these type of things to protect the fundamental rights. We will make anybody around the world, describe their world, really. So, I think when it comes to Big Data, we need to be very wary about of the analytics, the sources of information, evolution of all of these. Because people today think of very, very short term view of the immediate environment and they do not think very far at all. | LEA PA |
| 51 | KD | Do you have any further comment regarding Privacy Agreement? | |
| 52 | TP | The privacy agreement is too long and difficult to read in terms. And even if you disagree with that, it is very hard to participate in society without ticking the box, saying: Yes you do! When I talked about simplification I meant that that is a key requirement for privacy and data protection. But also, for the whole mechanism in Europe, some forms of simplification require from people to be identified and understand. I think all these terms, conditions and things like this for sharing information, in some ways is good because somebody is taking some attention to comply for. But, they are far too complicated, unnecessary complicated. I mean most organizations around Europe do not understand data protection. And indeed I can tell you, I have been part of | POA POA |

| | | | |
|----|----|--|-----|
| | | European debates here, where organizations representatives have told me that we are here because we do not want any data protection authorities snooping around. Because we have all kinds of data protections flaws. And even companies that are known that have flaws, recognize it is too complicated and too costly to comply because of this complexity. | |
| 53 | KD | Do you know any law simpliflicated for policy privacy? | |
| 54 | TP | I know the ongoing initiative for data protection, directive supervisions, initially tried to simplify things, implementation of data protection, but there are just so many different organizations, lobbying left and right and putting a lot of money into this process making it frustrated and complicated. And we really want to see some improvement in this era but the whole thing is not so transparent and they are for all kinds of opportunities. | POA |
| 55 | KD | Thank you very much for your time and collaboration, we appreciate it a lot. Once we finish the transcription, we would like to share it with you, so if you have any comments or interpretations we would be happy to get them. | |
| 56 | TP | I wish you very success with your studies! | |

7.7 Appendix 7 – Interview transcript with Interviewer 4

Interview number: 4

Interview Date: 10:00, 07th May 2014

Participants: Interviewer 4 (I4), Klaudia Dardha (KD), Flogerta Banaj (FB)

Interviewee: Interviewer 4 (I4),

Interview type: Skype voice call

Interview duration: 44 mins

Transcribed by: Klaudia Dardha (KD)

Transcribed date: 08th May 2014

| Line | Speaking | Text | Code |
|------|----------|--|----------------|
| 1 | KD | Thank you for the willingness to participate and in this interview. Thank you for your time. I and my master thesis Flogerta are making a study about Big Data towards Privacy Issues. And additionally we will look through any gap among the participants in this Data Supply Chain. We have prepared several questions and we can go straight to them in order not to take you a lot of time. For the purpose of the research we need to record it. Is it all okay for you? | |
| 2 | I4 | Yes it is Ok, we can proceed | |
| 3 | KD | According to Gartner (2001), where the term was initially used, Big Data is described as the rapidly increasing volume of data, velocity as the data go in and out, and variety as the dispersed type and sources of data. Do you think is something missing in this term of definition? | |
| 4 | I4 | Well, you know, it is always difficult to snap definitions. I think it depends to the context. If it is in need to the privacy context somewhere it should be saying that it is for personal data of persons or data which tells something about people. That is the description of what big data is, but is not the present definition I would give. Big Data is about general information from different sources indeed, which can single out proofs of people or even people. With single out I mean reading or combining all kind of sources of information. Single out is of course an issue. | DS DS |
| 5 | KD | Which are the main challenges Big Data faces within different fields? | |
| 6 | I4 | I think from the privacy point of view, it breaks with the traditional approach of privacy. So the privacy normally was: the right to be left alone, which is the right of the individual. If the policy enters your home, of course it is an invasion of your privacy. But here it is less clear in terms of the use of Big Data, were data is actually a concrete victim and a victim triggers to right. As indeed you will be compensable of human rights, you have to be a victim of the measure. But Big Data it's just there, and by using Big Data people can make an analysis and again single out a group of persons. Without actually knowing you or having a profile which is linked to your name so you are not necessarily identifiable but you are part of a group which is targeted by the person using the Big Data. So you're a victim requirement not so obviously. So you have to develop something more like an action popularize, which means that you have to be able to challenge a measure which does not necessarily conclusively infringe upon | DH DH DS |

| | | | |
|----|----|---|-------------------------|
| | | your right but might potentially be. So probably Big Data from a privacy point of view is how people, individuals can stop up and do something against it. And original route of invoking your rights it is maybe not completely fixed; it is more a sociable change, more or less, where these are decisions which have to be taken adequately. Police is going to use Big Data. That is a decision that has to be taken by the legislator fully. So it is a phenomenon which is not so easy to prosp and it proves also that it just happens so we're all creating sources of information and then the next step is to join the information and create profiles or again to single out groups or people. So it is not so easy to take a breach from a legal point of view. | |
| 7 | KD | Do you think that existing legal assurances present offer an effective protection for Internet use? | |
| 8 | I4 | In January 2012 the European Commission came with two proposals. One for new rules for the market regulation instead of directive and one on the police sector, which is directive, so it has to be implemented first in national law. They have not been adopted yet, it has been discussing after two years now. In those proposals the commission tries to tackle a little bit the mobile challenging for privacy in data protection. There is also a provision in automated decision making so nobody may be subject to a decision with legal effects which is purely based or performed by automated means so there should always be a human being involved, and that provision has been beefed up a little bit and is rehearsed to profiling and relevant for Big Data. But as I said if you look at those rules it is not very well developed, it is a challenging thing. There is also a provision on anonymized data, so once data can no longer be traced back to a person it is anonymized. And there are introducing rules on several anonymization to change the data with a key in order to make it impossible to retrace the person. So in different sites they are introducing some elements which may play role in the discussion with Big Data and profiling but they cannot really deal with what I just mention which is quite difficult if there is no directly to individual but only contextual like behavior advertising. | LEA IN LEA |
| 9 | KD | Is any data aggregation being treated differently because when analyzed prediction might violate or prejudice ethical issues? What about discrimination? | |
| 10 | I4 | I think discrimination is a very strong argument. Discrimination is something that should be taken into account here. Can only be a privacy notion, but it is about discrimination actually. If you are able to distinguish between different groups of people, and you treat them differently then you are in a contextual discrimination. I think discrimination is also one of the founding principles of data protection. Data protection is not purely an elaboration of detail rules of our right of privacy but also there are to afford discriminations on the basis of the classification of persons due to certain characteristics they join, they share. So discrimination is important in all this data, again it leads more to a sociable debate you should have than a debate on the basis of the right invoked by an individual when the court gasps. | RE LEA PC |
| 11 | KD | When and where should access of information be treated differently because of the potential for misinterpretation or abuse? Who should be responsible for addressing different issues that emerge because of predictive analytics? | |

| | | | |
|----|----|---|---|
| 12 | I4 | <p>I think in a police sector of course, NSA, SAGA, there should be safe guards against abuse. First of all you have to decide at a sociable level, or the legislator should decide whether this kind of information can be used or should be used for these purposes, then there should be assessment of what is affecting necessary so just not get out everything but what is necessary. And then the proportionality of the measure should be assessed which links to all safe cards that there should be in place. Safe ... like being informed in advance, of course not at the moment of the collection of the searching profile the data, but at least in the law should be clear, in what circumstances you can be subject of a police search of this kind of information, again I'm in the where the law has passed, but also you have to have access to the information once you ask for it, or information about the information they have about you that you are able to go to court etc. and the data protection authority are able to delete your data and to look your data in your place in order to ensure confidentiality.</p> | SE TP |
| 13 | KD | <p>And about the ownership?</p> | |
| 14 | I4 | <p>I think ownership is exactly what I was talking about safe cards, you can look at the regulations of EU, the current one or the future ones, and they are all about responsibility, ownership and rights. So, the telephone company is the ... you are the owner of your own data in the first place but is not exclusive. It is not that every data processing is subject to your consent that might be any other reason that can follow from the contract, there can be legal information to answer for information , which is not everything based on your consent. So if your telephone company keeps your data for billing purposes, there is a lot of information and they are responsible for that information, and they are responsible for making sure that all the rules are applied. They hold responsibility for the integrity of the data, so once you define a purpose and actually perform the processing of the data then you have to obey to certain rules and integrity certainty involve them, so you have to make sure that the data is accrued and secured. And according to that you have rights to individuals, this is data directive implemented in Sweden in 95 46, Swedish data protection act.</p> <p>So the consumer can ask the access for the information, and if they check the information, to have the data deleted if you think it is illegally or the possibility to correct the information, update it etc. and you can go to a court, to the data protection authority to complain if there is a discussion between controller and you. So, as obliged to consumer act of commercialization of data, in principle you have to give your consent first before they can use it for other purposes. That is what you always see on Internet, do you agree your data is being used for advertising. In principle you have to consent to give your agreement to that. Especially if it is reselling, if the company uses it for its own purposes that come close to what you legitimately can expect may be. But if is sold to other companies then for sure that is something which you should have been able to say yes or no. Of course in practice sometimes is done on purpose often you don't have a choice, you just have to say yes if you have to purchase the product or to use the service, that is where the new rules try to change these to make sure actually there is free consent and not pre-ticked boxes which you</p> | OW RE EP SU POA |

| | | | |
|----|----|--|----------------|
| | | cannot un-tick or which are something which see on the Internet. The new rules try to strengthen in this. | |
| 15 | KD | Are consumer's data privacy choices permanent? As once they are put online it can be impossible to delete. | |
| 16 | I4 | In principle Yes. You agree to what you agree to, if there is a change, back with Facebook, it should be clearly communicated to you. And normally it sort of obviously follows from what you have asked them to do but normally they should ask your consent to it. The problem with Facebook is that they are so huge, you cannot have data made individual privacy policy per person. So what they do in principle, what these guys do, is to give you possibilities to arrange your privacy. But you also see that they also can do something which is really wrong, is that people join together to make sure that Facebook does not do that anymore. So in principle the choice is permanent. Controllers are the one responsible on processing the data cannot certainly do something different and that is the whole idea, I think the core of the data protection is that data are only used on the purpose which you have got them, so if you want to use them for another unrelated problem you have to have a new legal which mainly is the consent or legal obligation. | RE SU |
| 17 | KD | How would you define Privacy in the era of IT evolution, specifically related to with Big Data? | |
| 18 | I4 | In the new rules we try to improve the clarity of defining who's responsible and if you have read about article 29 working party, it is data protection authority form all states come together including the European data protection supervisor , they issue advice , they also advised on social networks. And there they discuss also consumers to certain extended is responsible with the data of others. You can find there discussion on current rules and a bit reflective on the new rules of data protection. So not so much changing ownership, just the relations are more complicated nowadays. If there is a big privacy leak in WhatsApp or Facebook, then in principle the company is responsible, unless like with the NSA, a government body who actually force the company to handover the data. So depending on the situation who is ultimately responsible. Normally it should be the controller the one who is processing data, but if other people, if someone is hacked then you get to do the question what the controller did, so Facebook should do sufficient effort to make sure that there is no hack and they should use state of the art technology etc. By the way in new rules there are still proposals, if there is a privacy break you have to notify about certain casual. You have to inform those people and the data protection authority. This should create awareness of what happens, and put pressure on the companies to prevent this kind of breaks. | RE OW PA |
| 19 | KD | How do you assume the future of privacy? | |
| 20 | I4 | I think we already get, but there will be only more information from every one of us. That is something you can't avoid or try not to share much information but sooner or later you are sharing this information. I think that is something which we cannot control. It is important to concentrate on what is really important and so if you try to object all the process which access data, you might be seen as a data protection freak. In the public opinion you might be seen as a data protection fundamentalist who try to block developments and you make it so ridiculous. So in order to make data | EP |

| | | | |
|----|----|--|----------|
| | | protection privacy still relevant and realistic, in the future the focus will be not so much on what is personal data but what do we consider as a privacy. So we might develop toward on something, right now we have sense of big data about health, political preference etc, so we have to categorize sensitive data but may be in the future discussion will be thinking about what we consider privacy sensitive data and other data. So the important thing is to focus on the important data floating around. And focus will be on safe guards. It is not about denying the things happened but being able to challenge things once they go wrong. | |
| 21 | KD | Based on your experience and point of view, is there any other issue you would like to add in terms of privacy as a challenge of Big Data, before closing the interview? | |
| 22 | I4 | I think discrimination also should be mentioned in this context. Again, the existence of big data is development which we cannot deny. If it constitutes good information and is used for good purposes, we should consequently think about societal levels, so it is really a societally problem discussion we should have on where we want to go with this. | PC PR |
| 23 | KD | Ok, thank you for the interview. We appreciate it. As soon as we will finish the transcription of the interview we will send it to you. If you have any comment to it, it would be very helpful for us. | |
| 24 | I4 | Thank you. I wish you very success in your thesis. It was a pleasure. | |

7.8 Appendix 8 – Personal Data Submission Diary – Diarist 1

F Age: 21M Residence: Italy

Day 1: Date 4/23/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | At 7 A.M I checked my FB account. I saw some photos. 0 likes, I don't like people say "hey, she's on Facebook at 7 am". I went to school. During the lesson I checked my fb home a lot of time, even if there are no news, I also checked my Instagram account, 2 likes to my friends' photos. I opened my university site to control my school mail, no one wrote me. |
| 12:01-14:00 | At 12:20 I went home and I ate Caesar salad. While I was eating I checked my fb account, a friend of mine wrote me on Messenger, so I replied to him. Then I've to run away. |
| 14:01-20:00 | At 13:30 went to school cause at 14:00 o'clock I had to help my math professor during the math exam of the chemistry class. Students wants me to help them, so when the professor was gone I did it. During the test I opened "Il sole 24 ore" site. Further, at the end of the exam, I went to the nearest library, I studied econometrics, I used the program STATA. Then I checked Instagram. At 19:30 I went home. I checked my Gmail, there's a mail, I work for a consulting firm, and my boss sent me the works to do during this week. I opened the "Giallo Zafferano" site, it's full of delicious recipes, I wanted to cook something different. |
| 20:01-00:00 | After eating I opened Skype to talk to my mum, just 10 minutes then I made a shower, so I listened to the music from YouTube cause I don't have "pretty fly for a white guy" on my phone. My cousin wrote me on messenger so I replied to her. Of course I checked my fb account, nothing funny. I decided to stay at home that evening so I started to do my work for the society I work for. I opened sites like WolframeMathWord, Forbes, Il sole 24 ore, The economist and The financial times..ok now I'm tired so I'm gonna go to bed. Every night, before going to bed I check my fb and my Instagram accounts, and then I sleep. So I'll do it also tonight, but now I know that maybe I've a little problem with fb....night!! |

F Age: 21M Residence: Italy

Day 2: Date 4/24/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | <p>At 7 am I checked my fb account (my morning routine). I opened the meteo.it site cause I wanted to know how the weather would have been. I went to school and I could not use my telephone cause I had to listen to the lesson of econometrics. At 11 am I went to my boss office and there I opened the site of the society. At 13:00 I went to a fast food to eat something, of course I checked my fb account and I liked 2 photos.</p> |
| 12:01-14:00 | <p>I went to a library to study. At the break a checked my Instagram account, I liked 3 photos. I opened my university mail, a professor answered me, and I replied. Twitter sent me some mail on Hotmail, I really hate their useless mail!! Also pay pal sent me a mail, but I am not interested.</p> |
| 14:01-20:00 | <p>I studied all the day...I used internet just to opened the "sole 24 ore". At 19 30 I went to run at the park.</p> |
| 20:01-00:00 | <p>I came back home at 9 pm, I made a shower. I listened to the music on my roommate Spotify account. I checked my fb account cause my mom wrote me on messenger, I replied. At 00:00 I went to a club and I used google maps to reach it. I came back home at 4 am, I don't know why but as always I checked my fb account before sleeping.</p> |

F Age: 21M Residence: Italy

Day 3: Date 4/25/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | I woke up at 10 am. I checked my fb account, a guy sent me a request to be friend but I didn't accept it cause I didn't know who he is and my philosophy is to not accept on fb people I don't know. I opened the "sole 24 ore site". At 11 am a started to work, I had to prepare an article about Green-Tao theorem and to do a boring work about accounting.. for the article I used Google scholar, and a lot of sites about mathematics. I did this work till 14:00 |
| 12:01-14:00 | |
| 14:01-20:00 | I ate something at 14, and I checked my fb account. I started studying at 15:00. At 5 pm I went out with a friend of mine. We wanted to know for what society my boyfriend's uncle works for, so we opened google and we searched for him, we opened Linked-in. I came back home at 7 pm and I continued to study. |
| 20:01-00:00 | I finished to study at 9 pm. I checked my fb account a friend of mine liked my photo. I searched on internet a Chinese restaurant, I find out that there was one near my home, and it's also on trip advisor so I thought it was pretty good. I went there and I ate my Chinese food at home. I watched "pretty woman" on streaming then I went to bed and I checked my fb account before sleeping. |

F Age: 21M Residence: Italy

Day 4: Date 4/26/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | I woke up at 8 am I ate my cereals and I opened "Il sole 24 ore" site. As always I checked my fb account, a guy want me to become her friend on fb, I didn't accept it cause I don't know him. I started to study industrial economics. I finished at 13. |
| 12:01-14:00 | I opened the "Apple Store" site to looking for a pc, I really like the macbook pro. Then a started to cook pasta, meantime I listened a song on fb, god!! I really hate the publicity on YouTube..anyway.. I ate my pasta and I checked my fb profile. A friend of mine wrote me on messenger so I replied. I checked my Hotmail account. I had 4 mails, two of them from my boss. I had to go to work. |
| 14:01-20:00 | I went to the agency, my boss gave me some works to do, so I started to work. I didn't use internet. I checked my Gmail account, 3 mails (Ryanair, Vitasnella and a friend of mine) but I couldn't reply. I finished at 6 pm so I went with some friends for a happy hour. We were 5 females, so of course we talked about other girls and we went on fb to look their photos...funny. |
| 20:01-00:00 | I went back home at 11 pm, I opened asos sites, cause I was looking for a dress. i love asos!! Even if there are lots of cookies. I went on bed at 1:00 and I checked Facebook home..of course! |

F Age: 21M Residence: Italy

Day 5: Date 4/27/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | I woke up at 8 am, I drank a coffee and I started to study..oops during my study I realized that I didn't check my fb profile! So I did it. I continued to study. I took a break and I opened "il sole 24 ore" site. Then I started to study again. |
| 12:01-14:00 | I finished my study at 13:00, I opened the "giallo zafferano" site cause I wanted to cook something light.. I ate and I checked on Asos something....really, i don't know what.. it's like a drug. |
| 14:01-20:00 | I started to work at 3 pm..then I checked fb and my friends tagged me on a photo that we did at the club on Thursday night. I didn't like it so much but I didn't care. I continued my work..i worked till 20:00. I was really tired so I went on bed and I started to watch my fb home. I finished at 21.....my Gosh. |
| 20:01-00:00 | I ate something and then I opened a gossip site "the gossipers" , I didn't opened it since I was at high school, it's funny. Then I watched "the departed" on TV. During the break I checked my mail. Nothing important. After watching the film I checked my fb profile, my friends created a group for a dinner they are organizing, I comment a post saying that for me there was ok. I went on bed. |

F Age: 21M Residence: Italy

Day 6: Date 4/28/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | I woke up at 7:20 am I was really late!!! I ate something fast and I went on school. A friend of mine wrote me on messenger, he was in Belgium so I replied. During the lesson I checked my email, my boss wrote me about a work, twitter sent me a mail and a professor wrote me about an exam. I opened the "il sole 24 ore" site. I also opened the site "stata blog" and the Columbia university (department of statistics) site cause I found a good lesson of stata. |
| 12:01-14:00 | I finished school at 13, ate a sandwich and I went to the library to work for my boss. I opened a lot of sites about mathematics. |
| 14:01-20:00 | I worked till 5 pm then I went home to for my exams, but before going I stopped to the market, and I paid with the bancomat. I study till 19:00, my roommate took a photo of our computers and posted it on fb. I commented the photo with "stata irony" and she did the same. for us it was really funny but I don't think other people would understand they maybe think "they are really stupid" but I don't care ☺ I was invited by my ex roommates to eat at their home so I went there. |
| 20:01-00:00 | We ate and then we watched a film on "mediaset premium on demand" on the pc. I went on fb, some friend liked my roommate photo, even a guy she likes sooo much, so we thought that maybe he likes her too. I went home at 1:00 and, after checking fb I went to bed. |

F Age: 21M Residence: Italy

Day 7: Date 4/29/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | I woke up at 7 o'clock... I ate my chocolate biscuits and I checked fb, nothing funny. I went to school at 8 am but before I drank a coffee. I had 5 hours of lesson of course during these hours I checked a lot of time fb and Instagram. I also opened my university mail because I had to send a request for a stage that I have to do this summer. I gap the last hour because I had to go to the hairdresser's. |
| 12:01-14:00 | I finished to cut my hair at 13, I was desperate. I didn't want my hair to be short like this!! So I took a selfie and I sent it to some friends of mine using whatsapp and messenger. I went to home and I ate something. I checked the app of "il meteo.it" to check how the weather like cause I wanted to go to run. |
| 14:01-20:00 | I worked on STATA and on SAS for my exams and at 5 pm I went to the library with a friend of mine to work for a judgment of commercial law that we have to prepare. We used a lot of sites regarding commercial law and we also download the application of the civil code. At 7 pm I went to Zara because before I checked its web site and I really wanted a shirt. When I was there a man I didn't know wrote me on messenger, he is a photographer and he want me to do the model for a fashion website called fashionoutlet.it he contacted me because a friend of mine is a "fashion blogger" (I know, it's not a work) and he told him about me. Of course I opened the site and I did a search of the man on fb (really useful) and on google. He seemed ok. I went home. |
| 20:01-00:00 | I ate something and I checked my Instagram account. I went out for a beer with my ex roommate. We drank a delicious beer so I took a photo and I posted it on Instagram, it's really cool ☺ we talked a lot, but at 00:30 i was really tired so I went home. I went on bed and I checked Instagram, some friends liked my photo. I checked the "meteo.it" site, then the last look at fb and then I slept. |

F Age: 21

M Residence: Italy

Do you have any further comments that are not included in the day diaries?

Interview:

KD: If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 1: Hem..yes..it is my life and I do not wanna share it with the internet world...

7.9 Appendix 9 – Personal Data Submission Diary – Diarist2

F Age: 23

M Residence: Albania

Day 1: Date 4/23/2014(choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | <p>At 8:00 AM I checked my Facebook account; liked 5 posts, two of them form ads pages and three from my friend and shared with friends; Also I checked my mail account. Then I went to work, I opened my work's mail. I opened shutter stock, a page in which I can find some images that I need for my work.</p> <p>On 11:15 AM I opened a news site (Ballkanweb) ,to inform about the news and show-bizz. On 11:30 I opened google, to search some information about google ads and mobile ads.</p> |
| 12:01-14:00 | <p>On 12:20 I checked my hotmail (personal mail) and school mail. I had an mail from one Proffesor about the diploma thesis,so I googled (only 30 minutes) for some papers about taxes in Albania. Then I opened Whatsapp to chat with some friends. On 13:05 I opened the pages of UNDP,USAID and World Vision in Albania.</p> |
| 14:01-20:00 | <p>On 14:20 I opened youtube to watch some videos and to listen music. On 14:45 I was searching on google for some creative ideas on advertising. On 16:00 I checked my Linkedin account. On 19:10 I re-checked Facebook and commented on some posts and some pictures of my friends. I was chatting with a friend of mine, when she suggested me to watch a funny video on youtube, so I watched it.</p> <p>On 19:45 I opened google scholar and the official page of taxes in Albania to read some papers for my diploma thesis.</p> |
| 20:01-00:00 | <p>On 20:30 I opened Skype to talk with my sister who lives abroad. Then, on 20:45 I re-checked my facebook account, I followed some make-up pages and economics pages. On 21:40 I watched a movie on megashare.</p> |

F Age: 23M Residence: AlbaniaDay 2: Date 4/24/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | At 7:30 I watched some videos of my favorite singers on youtube. On 8:15 I opened the Weather Channel to see if it was a sunny day or a rainy one. On 9:10 I checked my Facebook account, I liked two posts of my friends : the first was a pic on Instagram and the other post was a check in that a friend of mine had been in an event. On 9:20 I checked my work mail, I sent two e-mails and I received three. On 11 :05 I checked my Hotmail account (personal mail). |
| 12:01-14:00 | On 12:10 I opened shutter stock and I was looking for some images that express taste. After I found the images that I needed, I opened google, to search about some creative ideas for different events. On 13:10 I searched on google some information about Account Executive (Job description) and it helped me a lot. |
| 14:01-20:00 | On 14:20 I re-checked my hotmail account and I had an email to join Referral Key. On 14:30 I signed in on Referral Key. On 15:10 I opened Assos.com, to look for some clothes, because I had heard that Assos had 20 % discount. On 19:30 I checked my Skype account. |
| 20:01-00:00 | At 20:50 I was looking for some information about taxes in public sector in Albania. (for my master thesis). So I opened google and google scholar to find some information. Then, on 21:40 I searched for some economics pages and I opened the page of Ministry of Finance in Albania to take some information. On 22:25 I re-checked my facebook account and I liked two posts. |

F Age: 23M Residence: AlbaniaDay 3: Date 4/25/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | At 8:10 I checked the weather on Weather Channel on my mobile phone. On 9:02 I checked my work mail, to see if I had any inbox and I sent two emails. On 9:20 I opened my skype account. On 9:40 I checked my personal mail (Hotmail account) and I hadn't any new mail, so I signed out. On 10:15 I opened google translate to translate some words from English to Albanian, and it helped me a lot. |
| 12:01-14:00 | On 12:20 I checked my facebook account and I liked only a post. On 12:50 I opened some newspaper's pages online, to look for some information and to check if some ads were published or not. On 13:15 I re-checked my personal mail, because I was waiting for an important mail. I had an email, for a meeting with Network of Excellent students. |
| 14:01-20:00 | On 14:20 I watched some funny videos on youtube, to have an idea about creativity on advertising. On 15:15 I searched for some information for LinkedIn users, Google + users and Twitter users in Albania. On 16:20 I opened google translate to help me to translate some words from Albanian to English. On 18:10 (after work) I re-checked my facebook account and I liked some pages (.....) |
| 20:01-00:00 | On 20:30 I opened my Skype account and I watched some videos on youtube. On 21:50 I checked my LinkedIn Account, I had an invitation to add someone on LinkedIn and I saw that a friend of mine had a new job. On 22:10 I was searching for some information about taxes, when I saw that a friend of mine mentioned me on a comment on facebook, so I commented back. |

F Age: 23M Residence: Albania

Day 4: Date 4/26/2014(choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | At 9:15 I opened the Weather Channel to inform about the weather on weekend. On 9:20 I downloaded some apps on my mobile phone. On 10:10 I checked my work mail, because I had to sent an email for an application. On 10:50 I opened google translate to translate some words from Albanian to English. |
| 12:01-14:00 | On 12:20 I checked my facebook account and I liked some posts. on 12:35 I opened my Skype account and I talked with my sister who lives abroad. On 13 :05 I checked my Linkedin Account (I saw that a friend of mine started a new job, so I congratulate her for her new job position). |
| 14:01-20:00 | As long as it was weekend, I only checked my personal mail account on 14:30 and I had an email for a meeting with my college's friends. On 16 :30, I opened youtube to watch some videos of my favorite singers, for two hours. Then, I checked my friend's account on Instagram on my Windows Phone. I liked some pics and I followed two people. |
| 20:01-00:00 | As long as it was weekend, I didn't use too much internet. I re-checked my facebook account on 20:20 (I had a message from a friend of mine to confirm a meeting) and after that, on 21:05 I watched Top-Channel (a national channel in Albania), to watch my favorite reality show (Master Chef). |

F Age: 23M Residence: AlbaniaDay 5: Date 4/27/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | On 9:50 I checked my Facebook account, I had a message in which my sister suggested me to follow some ads pages. So, I opened that pages and they had interesting stuff. Then I liked some posts on facebook and I wished happy birthday a friend of mine. On 10: 22 I checked my hotmail account. I had an email to complete a survey about “ Product life cycle “. On 11:20 I watched some videos on youtube. |
| 12:01-14:00 | On 12:50 I opened BalkanWeb to inform about the news in Albania and all over the world. On 13:25 I checked my LinkedIn Account and I had an add invitation. After I accepted that invitation, I added my profile on LikedIn. On 13:50 I re-checked my facebook account and I chatted with some friends. |
| 14:01-20:00 | On 14:20 I opened weather channel, to be informed about the weather, and I opened Whats app to chat with some friends for free. On 18:45 I searched on Google for some information about taxes and fiscal management in Albania. On 19:50 I used ipiccy to adjust some photos. |
| 20:01-00:00 | As long as it was Sunday, I thought that was a good idea to watch a film. So, I opened Megashare and I watched 21 Jump Street, a comedy with Johnny Depp (my favorite actor) and Holly Robinson . I have to mention, that I watched 21 Jump Street, after I checked on Wikipedia about it. |

F Age: 23M Residence: Albania

Day 6: Date 4/28/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | At 8:20 I checked weather channel to be sure for if it was a rainy day or not. On 9:00 i checked my work mail account , I had two mails , I reply to one, and in the other mail someone sent me a file via WeTransfer, so I opened WeTransfer and I downloaded the file. on 9: 50 I opened my skype account, because I have to use skype even for my work. On 10:35 I checked my facebook account and I had 3 notifications and one message.The first notification was about activity log, a friend of mine said that I had attended the same University as she. The other notifications were two likes, two friends of mine had liked my photos. |
| 12:01-14:00 | On 12 :30 I opened send space to upload some files that i needed for my work. I used send space as long as the files were too large. On 13:15 I checked my LinkedIn Account . I had two request for add and two suggestions to follow some groups for taxes and finance. |
| 14:01-20:00 | On 14:20 I used whats app to chat with some friends of mine. On 20 :15 I opened Viber to chat with some friends of mine and to make a call. On 15:45 I opened Wikipedia to search about some words and concepts. |
| 20:01-00:00 | On 20:15 I was searching on Google Scholar, when I found an interesting material about tax evasion. So, I read that paper and after that I searched on Google and Google Scholar about taxes in Albania and Balkan Countries. On 23:15 I opened Investopedia, to see the shares of some big companies on S&P 500. |

F Age: 23M Residence: AlbaniaDay 7: Date 4/29/2014(*choose from the calendar*)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | At 9:02 i checked my work mail, i had a new mail, so i answered to it. On 9:05 i opened skype, because I needed it for my work. On 10:10 i checked my hotmail account, and I had an email form facebook, that some of my friends had their birthday. On 11:15 I checked my facebook account, and I liked two posts (a check in and a pic in Turkey of a friend of mine) and i wished a friend of mine for her birthday. On 11 :50 I opened google translate to translate some words from albanian to english. On 13:50 I opened http://www.english-test.net to test my english skills. |
| 12:01-14:00 | On 12:50 I opened http://www.english-test.net to test my english skills. After that I searched in Google about Toefl Scores and how it works. On 13:35 I opened shutterstock to find some photos that express happy family and irresistible taste. On 13:55 I created an gmail account, because I needed it for my work. |
| 14:01-20:00 | On 14:20 I used sendspace because it helped me a lot with transferring some files that I needed for my work. On 15:15 I re-checked my facebook account to search for someone. After I found him, I looked some photos of him. On 16:05 I checked my Linkedin account, and I joined a group. |
| 20:01-00:00 | On 20:50 I re-checked my facebook account, because I needed to find a person on Facebook. After I found him, I was looking some of his photos, photos which was opened even for non-friend. (This is one of the reasons why I thing that social media is not too safe, because people who are not your friends, or who don't know you, can see your staff). On 22 :15 I watched some videos on Youtube. |

F Age: 23

M Residence: Albanian

Do you have any further comments that are not included in the day diaries?

Interview:

KD: If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 2: I care about the fact that these data can be published in any blog with my name and a photo of mine. I care about that because I don't want that my activity be known from my employer and they know every single activity I do during the day. Maybe it affects my employer opinion about me. But, also I don't like that the other people know in detail what I do during the day even I have described common things that people generally do. I don't like it. It seems like someone is spying to me.

7.10 Appendix 10 – Personal Data Submission Diary – Diarist3

F Age: 22

M Residence: Germany

Day 1: Date (03/05/2014)

| Time | Activity |
|-------------|--|
| 11:00-13:00 | <p>After having breakfast, I had a conversation with my friend about the idea of opening an online store for handmade knitwear, so I started to make some research about E-commerce platforms. I registered on three of them to try out their features, pricing and templates. They required me to give my basic personal data, email address and the phone number which I find really annoying. I accepted and agreed their term and conditions without reading anything. I started using one of them and let the two other ones for later to be checked.</p> |
| 13:01-16:00 | <p>At 12 o'clock we started reading and listening to some news at top-channel.tv mainly about our home country and Ukraine. As the weather was good we went out for a walk. After a while we entered a shopping center where we were given a chance to open a customer card and collect points. We thought that would be convenient to us so we filled out the form and made the application. They told us that the card would be sent to us by mail. As we continued walking from one shop to another we saw a magazine which you could take for free, so I grabbed it to see the latest offers and discounts. When we were leaving the shopping center we saw a notification that it would be open tomorrow on Sunday from 13:00 – 18:00. That was nice for us as we have more free time on Sundays to make some shopping.</p> |
| 16:01-20:30 | <p>Before going home we had to buy some food and groceries so we entered the market. We were distracted by some action price on some products that we hadn't planned to buy but we thought to take advantage from the discounts and buy them before we need them to save some money for later. After having dinner I checked out the magazine with the new trends and discounts but I didn't find anything convenient for me so I threw it in the trash bin.</p> |
| 20:31-02:00 | <p>At 20 pm I checked my Gmail account, I had some discount offers from Apple but as I don't have any plans to buy some tech supplies I deleted it. And there were a bunch of email from my online courses which I have been registered but never really attended, so I didn't open any of them.</p> <p>At 10pm we opened TED Talks and chose one of them and started watching. When it ended we decided to watch a film so went to IMDB.com</p> |

F Age: 22M Residence: Germany

Day 2: Date (04/05/2014)

| Time | Activity |
|-------------|---|
| 11:30-13:30 | After having breakfast we read and listened to some news on the recent happenings at top-channel.tv as usual. At 1 pm we remembered that the shopping centers were open so we got ready and went straight to them. |
| 13:31-18:30 | I focused on the clothes which were on sale, tried some of them but in the end bought nothing. My friend needed a case for his smart-phone, Google Nexus 5 so we went straight to that shop afterwards. After having a look at all the cases we didnt find to many option for that type of smart-phone so decided to search on Amazon , of course there will be more options, its a global market. Before leaving the shopping mall we reminded ourselves that we need a coffee machine so went to check out the prices, we selected two of them but as we weren't 100% sure, took one of the magazines about those coffee machines to read them more carefully at home. |
| 18:31-22:30 | After taking a break I went back to analyze the other two e commerce platforms. I filled out the form for the other twos. It took me a while to check out the themes, prices and user interface. So after two hours I decided that I'd go for Shopify e commerce platform . It offers more interesting themes, apps and plug ins. At 10 pm we had scheduled two Skype meetings with our families. |
| 22:31-01:00 | I had to check out for the most cheapest flight ticket as my cousin was coming on the other week so she needed me to help her with this issue. I went to skyscanner and searched for the options and I realized that the mos convenient, cheaper and more comfortable was the Austrian airlines. So I sent her an email with that option. Afterwards we started listening music from Spotify and talking, trying to stay away from the tech devices. |

F Age: 22M Residence: Germany

Day 3: Date (05/05/2014)

| Time | Activity |
|-------------|--|
| 10:30-12:00 | <p>At 11am I checked my Gmail account, the other language school which I had sent the form yesterday replied me, I had to leave an appointment to go there and take a short exam. I decided not to answer at the moment and to think more about that.</p> <p>According to my previous made search at Amazon site, was sent to me an email with some other offers from similar products. I clicked unsubscribed from their newsletter and deleted that email as I didn't want it anymore to take their offers. Afterwards I checked my Facebook account. I had one friend request which I deleted and gave the feedback to FB (I don't know this person) as I really appreciate this feature, it helps to get rid of requests from strange people.</p> |
| 12:01-15:30 | <p>I went to my course by train and on the way near to the building a girl gave an advertisement banner, it was about a fast food pretty close to our course. I was in a hurry so I just put it in my bag and went straight to the class. During the break I wanted to check my fb profile but then I decided not to cause I want to spent less time on it so I skipped to LinkedIn. I scanned the home page and read quickly one funny article about bosses.</p> |
| 15:31-20:00 | <p>Before going home I went to the market, bought some groceries and went straight to the cash desk as I was a bit tired. There I realized that the market was giving some cards to collect points so I decided that the other day I would take one. I checked my facebook profile and played my favorite game, farm hero saga but then I saw another game which seemed interesting to me. I tried to play it but it asked me to access my public profile and my friends list so I skipped it and closed.</p> |
| 20:01-01:20 | <p>Since weeks I decided to go to a concert as Berlin is the city of the concerts so I entered the Songkick site and checked out my favorite bands and musicians. I wanted to track them but without a profile I couldn't do that. The bad thing about Songkick was that I could only register with my fb profile which I don't find it comfortable at all but I had to cause I found not so good all the other sites about concerts and tickets. I went on with the registration and tracked my favorite artists. After that I entered my shopify account and checked the template about the online shop which I am planing to open soon. Some of them were for free but I picked up some other ones which offered me more options, memory and plug ins and of course they were not for free.</p> |

F Age: 22M Residence: Germany

Day 4: Date (06/05/2014)

| Time | Activity |
|-------------|--|
| 10:30-12:00 | After having breakfast I logged in my facebook account, I checked out some articles and liked three, two of them about news in my home country and one from Buzfeed. I clicked on the respective links which redirected me to their official websites, and then I read them all. At top-channel.tv I continued to read some other news articles about the situation in Ukraine and all the precautions taken from Europe and USA towards Russia. |
| 12:01-14:00 | At 12 o clock I entered the U bahn station to take the train to the language course. Here in Berlin all the train stations are surveyed by camera 24/7 and the trains inside as well. During the break I checked my Gmail account, two of my friends has added me to their circles in Google+, a newsletter about the discount for the online courses at Udemey and one notification from Google drive about a reply from my colleague in one of my comments there. |
| 14:01-20:00 | After finishing my course I took the U bahn to home as usually. Inside the U-bahn you can find a lot of advertisements, today I saw an advertisement about another language school more close to my house. So when I arrived home I took a rest and checked their website and filled out the application form as the price were convenient. It took me a while to fill out the form, as they asked me for all of my previous knowledge and experience in my German language but still I went till the end of it as it really mattered to me. I looged in Google+ and added back to my circles the friends who had previously added me. |
| 20:01-00:00 | Around 20:00 I rechecked my facebook profile, I saw some really funny pictures of my cousins and friends, I clicked like and I decided not to comment cause I don't want it to be annoyed by the notifications coming from those photos. I checked my LinkedIn profile and accepted tow invitations there, read one article and then I closed it. Before going to sleep I decided to watch a film with my friend. We checked the IMDB site, choose the film and then tried to find a link which could provide us watching it for free. The first one was asking to register as a user but we skipped that and continued on till we found the one for free. |

F Age: 22M Residence: Germany

Day 5: Date (07/05/2014)

| Time | Activity |
|-------------|---|
| 10:00-12:00 | <p>I logged in my LinkedIn account, I accepted one invitation from a friend of mine. Then I opened the Fb account, saw some funny photos but I decided not to react at all, like or comment. Only when some interesting articles appeared to me I opened them. One video was from 9gag and the other from top-channel.tv I spend some time reading them and got ready to go to the course.</p> |
| 12:01-15:30 | <p>On the way I searched for a brain training app at Apple store. The most rated one was Lumosity, I skimmed quickly the demo provided and decided to install it. Of course it required me to approve their terms and conditions so I just pressed I agree. During the break I checked my Gmail account, some emails from the courses and the other ones from the e commerce platforms with advices and ways "how to use" which I didn't need anymore cause I took a decision and ignored them.</p> |
| 15:31-20:00 | <p>I went to the market REWE and at the moment I had to pay I asked the lady to give me a parback card to collect points. I asked her what did I had to do next, the told me that I could register at the terminal at the entrance of the market or online. As I could get more free points when registering at the terminal I decided to go straight there. I was a bit confusing to me cause of the German language so I gave up and changed my mind, I will make it online so as soon as I arrived home I made the registration.</p> |
| 20:01-01:00 | <p>At 8 pm I went to the Amazon store and started searching for the case. It took me a while till I got the decision as I refereed to the reviews which distracted me a bit cause for every product you will get negative and positive feedback so in the end I took the decision by myself and considering the brand and price. Before going to check out,there was a offer from Amazon, to become e Prime member for one month as e free trial, after that I could decide to go on that plan and pay a certain amount in a year or leave that membership. As it was for free I decided to try it and take advantage of it on the free prime shipping on the other day. As I had to work a little bit for the online store I started searching for models and designs applied on knitwear. I found some forums, online shops and blogs where you can get some patterns for free. I registered to their website and subscribed to their newsletter.</p> |

F Age: 22
 M Residence: Germany

Day 6: Date (08/05/2014)

| Time | Activity |
|-------------|--|
| 10:20-12:00 | As usually I read the news , checked my fb account and saw that was one of my friends birthday so I sent him a message. Actually I don't like to use fb for such things bu that was the only way of contact. Then I checked the mail box, there were some offers from the food markets so I read them and put some products in mind. |
| 12:01-15:30 | During the break I checked my Gmail account, there were too much emails from the e commerce platforms, payback card, website about knitting and crocheting, amazon etc. So I decided to edit the settings in managing my emails, made some modifications in the way that I could only see primarily the emails market as important. I tried to unsubscribe from some of them and it went successfully. |
| 15:31-20:30 | This time at REWE market were issued some coupons which you can redeem when buying certain products so I grabbed one brochure. With the discounted products in mind and the coupons I filled my basket, gave the lady my payback card and the coupons and went home. A bit exited from the coupons I entered the website of the Payback card and activated the E coupons from there as it is more easy for me when I forget the paper version and its environmentally friendly as well. |
| 20:31-01:00 | After having dinner I went to songkick website to see what my fav artist were up to, any tour or concert near me. I was pretty exited as two of my fav bands are playing in Berlin in the next two months. I purchased some tickets for one of them so I had to give my personal data. As this process will be always like this I decided to be registered rather than check out as a guest, in this way my data will always be auto completed in my next purchase. On the LinkedIn home page I read an article which in my personal opinion it wasn't suitable for such type of social professional networks but I didn't react and went through the hole home page and logged out. |

F Age: 22M Residence: Germany

Day 7: Date (09/05/2014)

| Time | Activity |
|-------------|--|
| 10:00-12:00 | After having breakfast I checked my mail box and finally received my order from Amazon. Inside of it I found a voucher code for a holidays. I entered the website given and then I realized I had to give some of my main personal data such as email ad, tel no etc. As I still wasn't sure if I want it to take advantage of that discount, I decided not to fill out that form and save myself from the annoying emails. |
| 12:01-15:30 | During the break I checked my Gmail account and my cousin had responded me. She had decided to purchase the ticket which I've sent to her. At this moment I thought I should make some plans as I would be her guide, so as soon as I will have free time I would start the search and planing. The other email were from online courses, e commerce platforms, online stores and payback card. But they were under the unimportant category so I skimmed through them and ignored them. |
| 15:31-21:00 | When my friend arrived home, he opened the case exited but he wasn't satisfied, the case wasn't as we expected it to be. So I entered the Amazon site and read the return policy, it was quite easy to send it back so filled out the form with the reason why we want it to send it back. Printed out the papers needed so the packet was ready to be sent tomorrow. At this moment I had to search again for another one, it was a bit difficult as I was more skeptic this time but in the end I decided to buy one more expensive, hoping that it will be better processed with a higher quality than the previous one. |
| 21:01-02:00 | After the dinner I remembered to make a guide for my cousin so I made e bit of research. To visit the Dome here in Berlin, you have to make an online booking days in prior. I applied for it and chose three options, gave my personal data such as name, surname, nationality, tel no and email address. An email confirmation came to me asking to confirm my email address, I confirmed it and another one was sent to me with a reply that the final option for the time and date will be sent to me on the other day. Afterwards I logged into my payback account, saw my recent points collected and checked out if there were any new e coupons issued. I activated one that interested to me, it was one of the partners DM market some of their products I needed to buy in the continuing days. Before finishing the day we decided to watch a film like usual. |

F Age: 22

M Residence: Germany

Do you have any further comments that are not included in the day diaries?

Interview:

KD: If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 3: Indeed, I don't see it comfortable that my diary to be accessible somewhere in the Internet. Everything that you have done, online in the middle of the "jungle" because I perceive Internet in this way. Also, even with my photo and my name. Thus, it is better to not publish it.

7.11 Appendix 11 – Personal Data Submission Diary – Diarist4

F Age: 28

M Residence: Indian²

Day 1: Date 4/23/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | At 8 am i checked Facebook. Saw two private messages. Replied to that. Opened Gmail. Out of 10 emails 8 were advertisements and 2 were from subscribed communities. Deleted all. Deleted all spam mails, most of them were advertising emails marked as spam to be diverted to spam mail. Went to school. Deleted all spam mails every 30 mints. Had Chat with 3 friends in Facebook. Marked as spam two advertising in Facebook which were uninteresting to me. Checked twitter only once. Browsed google scholar and google several times for thesis purpose and downloaded many pdf files. Checked whatsapp frequently to see updates and replied to friends. |
| 12:01-14:00 | Browsed google scholar and google several times for thesis purpose and downloaded many pdf files. Watched 3 or 4 YouTube videos. Checked Facebook frequently to see messages and replied to 3 friends. Replied to friends in whatsapp. Had a look at zalando.se for jackets and shoes as it was prompted in a blog. Blocked google ads in that page. Deleted previous wall posts in Facebook. Verified LinkedIn updates. Send requests to Subject matter experts for research in LinkedIn. Checked Facebook, Gmail and LinkedIn frequently to distract from thesis work and check for replies from contacts regarding thesis. Unfollowed a page in Facebook as it posted a picture which was uninteresting. Checked Facebook, Gmail and LinkedIn frequently to distract from thesis work and check for replies from contacts regarding thesis. |
| 14:01-20:00 | Send messages in whatsapp to friends. Watched videos online and many advertising websites popped up, made sure each popup window is closed before watching video. Checked in eBay for buying dress and found google AdSense showing the same product later as suggestions. Found it annoying. Cancelled premium account in LinkedIn, but was skeptical about the website keeping my card number and checked in forums regarding the same. Checked Facebook, Gmail and LinkedIn frequently to distract from thesis work and check for replies from contacts regarding thesis. |
| 20:01-00:00 | Send messages in whatsapp to friends. Watched videos online and many advertising websites popped up, made sure each popup window is closed before watching video. Got message from LinkedIn connection in Gmail and went to verify the same in LinkedIn. |

² For the last 10 months has been living in Sweden.

F Age: 28M Residence: Indian

Day 2: Date 4/24/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Checked emails, Facebook and whatsapp at 7:00. 6 emails in gtalk, except 1 everything was online ads. 1 email was in Swedish, I assume they obtained my mail id based on location. Went to school at 8:00 and checked emails and Facebook. Texted a friend in Facebook. Replied to friend in LinkedIn. As a matter of fact, most of the ad mails are due to chrome using autocomplete while typing my name in forms. Now I wish to cancel in which all websites I gave my email id. Liked a page and a photo in Facebook. Recharged a mobile phone online using online banking. The site is almost trustworthy as i was using it for quite some time. Previously i used debit card for the payment, but now prefer online banking considering it being safer. |
| 12:01-14:00 | Watched YouTube videos in TMZ channel. Verified Gmail, twitter and Facebook frequently. |
| 14:01-20:00 | Had chat with friends in Facebook, whatsapp and Gmail. Found it annoying that truecaller.com website is having my phone number and address listed. After that a bit cynical about websites accessing fb, twitter or LinkedIn data. |
| 20:01-00:00 | Watched movies in free website. Messaged 3 friends in Facebook frequently. |

F Age: 28M Residence: Indian

Day 3: Date 4/25/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Checked emails, Facebook, twitter, whatsapp and LinkedIn. One message in Facebook, 8 mails, 2 updates in a LinkedIn. All emails were from communities. And couple of spam mails which were directed by me to be in spam. Send message to a friend in Facebook before leaving for school. By 9:00 I left for school. After reaching school verified Facebook and Gmail. Found again spam mails which were directed to be in spam folder. |
| 12:01-14:00 | Watched a movie online. Thanks to YouTube for blocking ads. Walked around in school with my tablet pc. While switching zones the Wi-Fi handover was asking for password frequently. Looked suspicious. Had chat with friends in FB. Sent message to professionals in LinkedIn for interview request for thesis. Sent messages to friends in whatsapp frequently. |
| 14:01-20:00 | Left school by 4:00 pm. Away from internet for 5 hours. It is good that I am not using smart phone. Using old school text messages and phone calls. |
| 20:01-00:00 | Watched movies online using free websites. Had chat with friend in Facebook and gtalk. Sent messages through whatsapp. |

F Age: 28M Residence: Indian

Day 4: Date 4/26/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | Checked mails at 7:00 am. Only 1 useful mail. Deleted all the other mails. Replied to the one email from LinkedIn. Left for school at 9:00. Tried to make online purchase in ikea.se. Both debit cards did not work. Was not sure about what happened. So had to go and check whether money was deducted. Texted 2 friends in Facebook. Texted many friends in Whatsapp. |
| 12:01-14:00 | Texted 1 friend in Facebook. Texted many friends in Whatsapp. Browsed in google scholar for articles for thesis. Scribd.com asked to login with LinkedIn/Facebook/google+ for free documents. I am more cynical about using those details to log into website. So skipped the search. |
| 14:01-20:00 | Had few chats in Whatsapp. Sent few messages in Facebook. Although I don't like them a lot, I try to avoid the online communications, because I'm afraid people using my details. |
| 20:01-00:00 | Watched movies in online free website. |

F Age: 28M Residence: IndianDay 5: Date 4/27/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | Checked e-mails, Whatsapp, twitter and Facebook. No updates were there in FB. Ignored updates in Twitter. Ignored all messages in Whatsapp in groups. Many ad mails in Gmail and i wonder how my email id reached these websites. Need to make a rule to move everything to spam. Had chat with a friend in Facebook. |
| 12:01-14:00 | Watched movies online in free websites. |
| 14:01-20:00 | Received a mail in Gmail from a LinkedIn contact. Typed a reply for him. Based on his response sent mail to another person. Had talk with a friend in Facebook. Played a new game in Facebook and the moment i opened the game it made a post in my Facebook wall. That is annoying. And after that I removed it, I don't want posts by it to be shown in my profile without my consent. |
| 20:01-00:00 | Received a message from LinkedIn. Had to reply for the mail from university mail id. It was for my thesis. From the university mail it looks more legitimate than gmail, that's why I sent it from it. All other mails in Gmail were advertising and irrelevant. Sent messages in 2 groups and one friend in Whatsapp. |

F Age: 28
 M Residence: Indian

Day 6: Date 4/28/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Woke up at 7:00. Went through the e-mails. No Important mails as usual. Deleted spam mails. Checked FB and Whatsapp. No useful messages, only advertisings. Sent a message to a friend in US to do an online purchase for me. I am worried of using my debit cards in EU as in my experience US bank services are better. Reached school at 8:00. Had an online interview for thesis. So prepared for it until 10:00. When the interview was one it was 12:00 |
| 12:01-14:00 | Sent message to friend in FB. As I got free text messages in phone, now prefer texting than Facebook and Whatsapp. It is more safe, no one is planning to hack it. |
| 14:01-20:00 | Left school by 3:00. No internet communication until 7:00. Then sent message to a friend in FB. Asked another friend to do online purchase for me. |
| 20:01-00:00 | Browsed around internet and watched YouTube videos. Had chat with 4 or 5 friends in Facebook. Sent text messages in Whatsapp. |

F Age: 28M Residence: Indian

Day 7: Date 4/29/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Woke up at 9:00 am and checked mails, Facebook and Whatsapp. 5 mails out of which none are useful in Gmail. No messages in Facebook. A lot of messages in Whatsapp community. Marked as read all messages and cleared history. Left for school at 10:00. Had chat with friend in Gmail. |
| 12:01-14:00 | Worked continuously for 3 hours without distractions. Stayed in school until 4:00. |
| 14:01-20:00 | Watched few YouTube videos and slept until 8:00 pm. |
| 20:01-00:00 | Checked FB and Gmail. Got two important mails regarding thesis interview. Sent replies. Worked on thesis for 3 hours. 11-12:00 am watched series in free website. |

F Age: 28

M Residence: Indian

Do you have any further comments that are not included in the day diaries?

Interview:

KD: If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 4: -That is bad. I don't want my mail ID to be published online. Although all the material I provided is fine, as I didn't do anything bad, I don't like it. And more over I don't like my picture to be found in google.

7.12 Appendix 12 – Personal Data Submission Diary – Diarist5

F Age: 22

M Residence: Sweden

Day 1: Date 4/24/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | <p>I will start this diary by giving you some technical specifications. At the time of writing this I am on my Asus Zenbook notebook on which I only use one browser: Google Chrome. My Chrome is fitted with the Chrome extension Adblock Plus which of course prevents me from seeing any static, popup and video ads on web sites. It even kills the ads on YouTube that plays before videos. I also have an un-rooted Samsung Galaxy S4 phone running Kitkat. This phone is not fitted with any ad blocker of any description. In fact, Google makes it difficult to block ads on the Galaxy by e.g. providing Youtube in the form of an app of their own.</p> <p>The time is now 8:47 on the 24th of April 2014. I have thus far checked my work e-mail for a reply regarding a meeting taking place this week. It should be noted that I use several e-mail accounts. The two important ones are my work e-mail tied to a university-specific domain, and my personal @gmail.com address.</p> <p>Now that we got that out of the way I am going to pay my phone bill for my Galaxy. My carrier Tele2 bills me every three months, and they will do so by sending me a physical letter containing the bill. I also however have a Tele2 app installed on my phone. I log into this app using my private e-mail and a secret password as credentials. From here I can see my phone number, when my contract expires, my monthly fee and so forth. I can also look at how much data I've used. At the moment, if you are interested, I've sent 64 text messages and used up 0,25GB of mobile data. These counters will reset every three months. The reason for the low data usage is that I spend a lot of time at home and at my place of work, the University. In both of those situations I am linked up to either my home network or Eduroam. Eduroam is also pretty easy to find throughout the city of Lund, as every university building has it, and my phone is set to connect to it automatically. By tapping the "bills" tab I can view all my paid and pending bills. I can use the app to instruct Tele2 to send me the bill as an e-mail instead, I had it sent to my private e-mail. I will now retrieve the bill from my e-mail and proceed to pay it using online banking.</p> <p>The procedure is quite simple. My bank has provided me with a physical card reader device that plugs into one of my laptop's USB ports. I insert a card into the card reader and log onto the bank's website. There I can log in using my physical device by punching in a PIN code. I like this way of paying because it means that I can copy/paste the OCR-number from the bill that I received in PDF-format from Tele2. The computer is a lot less susceptible to human error, a factor that would come into play if I was reading the OCR-number off of a paper as I was typing it into the text field of the web site. I punch in my PIN code again and the payment is made, fast and easy.</p> <p>As I finish the transaction and disconnect the physical device I receive a PM on Facebook from my thesis partner. He asks me to try to see if I can open a link which I cannot, I report this unfortunate state of affairs back to him. In the future I won't go into detail regarding my</p> |

| | |
|-------------|--|
| | <p>Facebook private messages. It's not that they contain compromising stuff, it's just that the documentation effort would be too time consuming as I get quite a lot of them. Anything I say or do on Facebook is essentially tapped, I have this deal with Facebook (it probably says somewhere in the terms that I didn't read) that essentially says that I get to use their services (such as messaging) and they get to listen in on everything that I do. I think that this is a fair deal because Facebook is a very good service. During the master's programme that I'm currently taking, all communication within work groups have been done on Facebook mainly, and text messaging as a secondary option.</p> <p>I am now going to close this document and save it in a location where I can get to it again. This location is of course Google Drive which is a cloud storage service. All of my documents regarding work are on there in a neatly organized folder structure. This content is probably (again, I didn't read the agreements) entirely owned by google, as I recall Dropbox pulling something similar to that. The documents contain my work for university, but no sensitive information such as passwords are on there. The best thing they can data-mine is my social security number which is on the front pages of some of the hand-ins. All sensitive documents are kept in a heavily encrypted file stored locally on my notebook. In theory I could upload that file to google drive as well, it would take them years to break the encryption even if they tried.</p> <p>Arrived at my place of work and continued working on a Business Concepts Model (Bajec & Krisper, 2005) =) for my master's thesis. The Business Concepts Model is built using the Google Drawing web app, as it allows for more flexibility in drawing e.g. ternary associations. Something that other tools such as Microsoft Visio and IBM Rational Software Architect, annoyingly enough, don't support. While working I checked my e-mails as well as facebook on and off.</p> |
| 12:01-14:00 | <p>After lunch I went on facebook on my work computer at LUSEM (Stationary, dual monitors blabla..). After replying I quickly scrolled through my news feed and came upon a link to a TED-talk by Michael Shermer posted by the official TED facebook page that I follow. The link of course re-directed me to the appropriate TED-page where I proceeded to watch Shermer's talk, it was from 2006 and I think I've seen it before but it was good fun.</p> |
| 14:01-20:00 | <p>This time period consisted mainly of working in Google Drive while simultaneously chatting with people over facebook and text. No data was entered into any site.</p> |
| 20:01-00:00 | <p>I happen to work for Lundakarnevalen's communication section (ComSec) and they like to do their scheduling using doodle. So this evening at 20:00 sharp I got an email encouraging me to pick a 4 hour time slot during which I would be working at Lundakarnevalen. I did so with haste and alacrity, as the best places (during the day) get taken quickly. Doodle means that you essentially enter your full name and check whatever time slots fit you, easy enough. Around 9:30 I was again checking my Facebook news feed. I found an interesting link that a friend shared. It takes you to a website that plays a full screen interactive video simulating a drowning accident. You are the victim and in order to stay alive you have to repeatedly and continuously scroll upwards, an exercise obviously designed to get you tired. I lasted 5m 3s which is not bad.</p> |

F Age: 22M Residence: SwedenDay 2: Date 4/25/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Today is the 25 th , I checked Facebook and both emails briefly in the morning. Nothing of interest was to be found. At 9:48 I opened Google Drive to once again start work on my master's thesis. |
| 12:01-14:00 | 12 to 14 of the 25 th was rather uneventful data entry wise. Most of my time was spent either in computer lab sessions or in meetings. The only mildly interesting part I guess would be me sitting in my office having a conversation over Facebook's chat service with my thesis partner, then heading down for a lab session while keeping the conversation running via the Facebook messenger app on my phone. The seamless switch between devices, still being able to follow the same conversation thread is indeed trivial by today's standards but it is nevertheless neat. In the same way, the 7-12 entry of this very diary was typed on my work computer which is stationary |
| 14:01-20:00 | 14-20 was also pretty uneventful. The seamless conversations with my thesis-partner continued throughout the day. At the time of writing I am sitting on a bench located on a square in Lund with my laptop out. I have no WiFi, but I can still work with this document that is located on my google drive. It kind of adds to the seamlessness of the experience that companies like Microsoft and Google try to create for users. You constantly hear about this stuff "access your files from any device!", "one platform!" and so forth. The time is currently 6:49. |
| 20:01-00:00 | N/A |

F Age: 22M Residence: SwedenDay 3: Date 4/26/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | N/A |
| 12:01-14:00 | First instance of internet activity occurred at ~1pm when I, as usual, opened google drive to resume work on my master's thesis. During my work sessions I did check BBC news once, and yes I am well aware of the BBC's cookie policies. They have been bragging for some time about being one of the few websites compliant to the new cookie law. When it comes to cookies and the tracking of your browsing it's a bit of a double-edged blade. On one hand it helps to make your browsing experience seamless by remembering what services you are logged onto so that you don't have to punch in your password every time you e.g. access a google site. Cookies is essential for doing this since HTTP is rather senile. Cookies can of course also be used to track all of my internet activity and use that for whatever purposes. I'm sure I have 20+ cookies tracking me as we speak, and if I would like to get rid of them I'd install a Chrome Extension such as "BetterPrivacy" that kills even everlasting cookies. |
| 14:01-20:00 | This time-block was spent in Google Drive with the occasional visit to FB to check my news feed. Nothing of significance to report. I did have to access the Lund University Library website to pull a document off of it. No information was submitted though. |
| 20:01-00:00 | N/A |

F Age: 22M Residence: Sweden

Day 4: Date 4/27/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | N/A |
| 12:01-14:00 | Checked Facebook on my smartphone, clicked no ads or any other items. Also checked BBC news through the widget that I have installed on my phone, read a couple of stories while waiting for a friend to get ready. |
| 14:01-20:00 | This time-block was entirely consumed by facebook chatting. I went from working in google drive to working locally, so no data trafficked there. I had two facebook chats up simultaneously discussing work stuff with the other TA and thesis stuff with my thesis partner. |
| 20:01-00:00 | N/A |

F Age: 22M Residence: SwedenDay 5: Date 4/28/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | I arrived at work early and checked both of my e-mail accounts before heading to a meeting for my master's thesis. There will be a lot less activity on google drive from now on as I and my thesis partner decided to move the work to Microsoft Word. The reason for this is simply that MS Word is a better word processor than Google Docs and it allows for extensions. |
| 12:01-14:00 | Nothing significant happened here, I went to get lunch and as I was waiting in line I was checking BBC Technology news stories (I have this as a widget on my Galaxy) and Facebook on my phone. As mentioned previously my phone is not fitted with an ad blocker, causing my browsers (and apps of course) to display ads. I never click any of these ads however. Reflecting on this it might be because of the sheer amount (according to my perceptions) of compromised ads online. For example, if I'm on my PC and I hover my mouse over an ad link it won't tell me where it's going to take me. Instead it points to the site/server in between that's going to record my click. It's an old paranoia that I think has stuck with me, I don't know how much of it is real and how much of it is imagined but I don't mind it. The reason is that I'd like to think that I have better things to do than getting distracted by ads while I'm on the internet. I think that my "low" activity on e.g. Facebook when it comes to letting them know what movies I've watched, what products I like and so forth limits their abilities to target me with targeted and accurate ads. |
| 14:01-20:00 | Minimal internet activity during this time, I checked and replied to an email on my work e-mail. |
| 20:01-00:00 | Pictures were posted from an event I attended a few weeks earlier. I went on facebook and liked a few of them. |

F Age: 22M Residence: Sweden

Day 6: Date 4/29/2014 (choose from the calendar)

| Time | Activity |
|-------------|---|
| 07:00-12:00 | Checked e-mails and the usual stuff, checked Facebook but found nothing of interest. 10-12 was spent in meetings and on lunch, nothing to report there except for the usual checking of BBC news. |
| 12:01-14:00 | 12-14 was spent on the run, no notable internet activity to report. |
| 14:01-20:00 | This time block consisted of the usual e-mailing, e-mail checking and Facebook checking. |
| 20:01-00:00 | N/A |

F Age: 22M Residence: SwedenDay 7: Date 4/30/2014 (choose from the calendar)

| Time | Activity |
|-------------|--|
| 07:00-12:00 | <p>Dear diary, today is Valborgsmässoafton. This is the day when we try to cram 25.000+ people into one park in Lund. Most of these 25.000 people have smartphones or at least cellphones which means that the networks will be heavily overloaded due to 25.000+ phones in one such small area. I spent the entire two first time blocks of the 30th in the park, surprisingly my carrier didn't let me down and I was able to receive and send texts as well as Facebook messages.</p> |
| 12:01-14:00 | <p>This time block was spent in the aforementioned park, the only activity worth noting was Facebook messaging and texting, no specific data was submitted.</p> |
| 14:01-20:00 | <p>Now, as it happened, while I was in the park I was taking photos with my Galaxy which happens to have a pretty good camera for a smartphone. I will now upload these pictures on Facebook, a move that has some implications.</p> <p>The moment I press "upload" Facebook owns my content, they are free to use it in whatever advertising they want, and they are free to use it to target me with advertising. This is the "price" that I pay for getting the social experience that Facebook offers.</p> <p>After scanning through the pictures and deleting redundant or unwanted ones, I was ready to upload. The only task remaining was tagging. Some of the pictures that I took included my friends whom I proceeded to tag. This means that, whether they like it or not, they are now linked to my pictures. This is also something that I put myself "at risk" for by joining Facebook. Any one of my friends can tag me in any photo, a feature that is ripe for abuse. I can however limit the degree to which people get to see what I'm tagged in, a limitation I've decided not to impose.</p> |
| 20:01-00:00 | N/A |

F Age: 22

M Residence: Sweden

Do you have any further comments that are not included in the day diaries?

The writing of this diary has granted me some self-awareness when it comes to my own internet activity. I hope that I've been able to convey an accurate window into my daily internet activities and I suppose I should end this diary with some sort of summarizing comment.

When reading this diary it might appear as if I don't care much for my privacy. It might seem that I knowingly throw my data at Facebook, fully aware that they can exploit it in any way they want. It might seem reckless to some, but it all depends on how you view internet services, especially social networks. When I post something on e.g. Facebook I do so with the mindset that anybody on the internet may see it. Facebook is only so good at keeping people outside of my friends circle out. I think it is naïve to view Facebook as some sort of secluded forum where you can discuss topics, shielded from outside scrutiny. I think that if you view these social networks as the corporate entities that they are, and realize that their core business is exploiting your data, you can have a healthy relationship to them. Nobody forces you to post to Facebook (arguably, see social/group pressure). Posting sensitive material to Facebook and then complaining about it when it spreads to unintended recipients or appears in an ad is, I think, a problem that stems from lack of knowledge rather than Facebook being evil. I would argue that today there's a contemporary sense of these giant social networks having some sort of responsibility for the privacy of the individual rather than to their share-holders. I think that if we recognize the social networks for what they are, corporate entities, and deal with them as such, we can have a much more healthy relationship to social networks. The deal I have with Facebook, which is written in their terms of use, is that I get to use their services in exchange for (essentially) my data. If that is not a deal I'm willing to make then I am free to back away. But the tremendous network value contained in e.g. Facebook outweighs the perceived cons of my data being manipulated.

When I was in City Park I saw a person with a t-shirt with a picture of big brother from George Orwell's 1984 on it with the added caption along the lines of: "Facebook is watching you". I think that's to some degree correct, but that the ethics problem of entities like Facebook "watching us" is awareness and not demonizing.

Interview:

KD: If we will put together all the information you provided within a week, and post it online with your photo, name and other details we find available online will it disturb you?

Diarist 5: Well, the answer is yes and no. The best way I can explain it is this: All opinions included in that diary are those that I hold at the moment, at my age of 22 and I stand for them. My daily routine that is described in there isn't that exciting and I wouldn't be devastated if it leaked. But I've put together this document by participating in a scientific study where my identity is to be kept confidential. What I've written down in there, I've written knowing that it can't easily be traced back to me. So for that reason I would object.

Secondly, I surrender my privacy to FB/Google etc. because they provide great services in return. If someone would take this info and slap it on a blog I get nothing back.

I tried to make that point in the final section of the diary that I feel that I'm knowingly trading my privacy (as much as I want to give them anyway) in exchange for a service.

8 References

- Agrawal, D., Das, S., & El Abbadi, A. (2011). *Big data and cloud computing: current state and future opportunities*. Paper presented at the Proceedings of the 14th International Conference on Extending Database Technology.
- Alaszewski, A. (2006). *Using diaries for social research*: Sage.
- AppliedDataLabs. (2014). Retrieved 04/05/2014, 2014, from <http://www.applieddatalabs.com/>
- Ashton, K. (2009). That 'internet of things' thing. *RFiD Journal*, 22, 97-114.
- Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization *MIS Quarterly*, 30(1), 13-28.
- Bakos, J. Y., & Treacy, M. E. (1986). Information technology and corporate strategy: a research perspective. *MIS Quarterly*, 107-119.
- Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The Journal of Strategic Information Systems*, 11(3), 245-270.
- Bennett, C. J. (1995). *The Political Economy of Privacy: A Review of the Literature*. NJ: Center for Social and Legal Research.
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). DIGITAL BUSINESS STRATEGY: TOWARD A NEXT GENERATION OF INSIGHTS. *MIS Quarterly*, 37(2).
- Bhattacharjee, A. (2012). *Social science research : principles, methods, and practices*: Tampa, Fla.
- Blanger, F., & Crossler, R. E. (2011). Privacy in the digital age: a review of information privacy research in information systems. *MIS Q.*, 35(4), 1017-1042.
- Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*: Aspen Institute, Communications and Society Program Washington, DC, USA.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Brinkmann, S., & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology*, 18(2), 157-181.
- Bus, J., & Nguyen, M.-H. C. (2013). Personal Data Management—A Structured Discussion. *Digital Enlightenment Yearbook 2013: The Value of Personal Data*, 270.
- BusinessDictionary. (2014). BusinessDictionary.com - Online Business Dictionary. Retrieved May 24, 2014, from <http://www.businessdictionary.com/>
- Bynum, T. (2008). Computer and information ethics.
- Cai, J., Liu, X., Xiao, Z., & Liu, J. (2009). Improving supply chain performance management: A systematic approach to analyzing iterative KPI accomplishment. *Decision Support Systems*, 46(2), 512-521.
- Caudill, E. M., & Murphy, P. E. (2000). Consumer online privacy: legal and ethical issues. *Journal of Public Policy & Marketing*, 19(1), 7-19.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*, 36(4), 1165-1188.
- Clarke, R. (1999). Introduction to dataveillance and information privacy, and definitions of terms. *Roger Clarke's Dataveillance and Information Privacy Pages*.
- Corti, L. (1993). Using diaries in social research. *Social Research Update*, 2(2).
- Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*: Sage.

- Croll, A. (2014). The feedback economy. Retrieved May 24, 2014, from <http://strata.oreilly.com/2012/01/the-feedback-economy.html>
- Culnan, M. J. (1993). "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use. *MIS Quarterly*, 17(3), 341-363.
- Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization science*, 10(1), 104-115.
- Culnan, M. J., & Bies, R. J. (2003). Consumer Privacy: Balancing Economic and Justice Considerations. *Journal of Social Issues*, 59(2), 323-342. doi: 10.1111/1540-4560.00067
- Culnan, M. J., & Williams, C. C. (2009). HOW ETHICS CAN ENHANCE ORGANIZATIONAL PRIVACY: LESSONS FROM THE CHOICEPOINT AND TJX DATA BREACHES. *MIS Quarterly*, 33(4), 673-687.
- Czarniawska-Joerges, B. (2007). *Shadowing: and other techniques for doing fieldwork in modern societies*: Copenhagen Business School Press DK.
- DataSociety. (2014). The Social, Cultural, & Ethical Dimensions of 'Big Data'. Retrieved 20/03, 2014, from <http://www.datasociety.net/initiatives/2014-0317/>
- Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, 84(1), 98.
- Davis, K., & Patterson, D. (2012). Ethics of Big Data. *Linux Journal*(223), 66-66.
- Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, 17(1), 61-80. doi: doi:10.1287/isre.1060.0080
- Eddy, E. R., Stone, D. L., & Stone-Romero, E. E. (1999). The effects of information management policies on reactions to human resource information systems: An integration of privacy and procedural justice perspectives. *Personnel Psychology*, 52(2), 335-358.
- Epic. (2014). Big Data and the Future of Privacy. Retrieved 28/04/2014, 2014, from <http://epic.org/privacy/big-data>
- EuropeanCommission. (2014). European Commission. Retrieved May 24, 2014, from http://ec.europa.eu/index_en.htm
- EuropeanVoice. (2014). European Voice. Retrieved May 26, 2014, from <http://www.europeanvoice.com/>
- EvanLevy. (2001, 20/04/2014). The Data Supply Chain. Retrieved 20/04/2014, 2014, from <http://www.rosebt.com/1/post/2013/02/the-data-supply-chain.html>
- EvanLevy. (2014). Inside IT. Retrieved April 24, 2014, from <http://evanjlevy.wordpress.com/>
- Ezzy, D. (2013). *Qualitative analysis*: Routledge.
- Filippi, P. D. (2014). Big Data Big Responsibilities. Retrieved 01/03/2014, from Institute for Internet and Privacy
- Gartner. (2013). Gartner's 2013 Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines. <http://www.gartner.com/newsroom/id/2575515>: Gartner.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report*, 8(4), 597-607.
- Groth, P. (2013). Transparency and Reliability in the Data Supply Chain. *Internet Computing, IEEE*, 17(2), 69-71.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60-65.
- Hosack, B., Hall, D., Paradice, D., & Courtney, J. F. (2012). A look toward the future: decision support systems research is alive and well. *Journal of the Association for Information Systems*, 13(5), 3.
- IBM. (2012). 2012 IBM Annual Report. http://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf.

- Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of management*, 29(6), 801-830.
- Josephson, J. R., & Josephson, S. G. (1996). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press.
- Klopfer, P. H., & Rubenstein, D. I. (1977). The concept privacy and its biological basis. *Journal of Social Issues*, 33(3), 52-65.
- Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.
- Laney, D. (2001, May 26). 3D Data Management: Controlling Data Volume, Velocity and Variety. Retrieved from <http://blogs.gartner.com/doug-laney/>
- Latham, A. (2003). Research, performance, and doing human geography: some reflections on the diary-photograph, diary-interview method. *Environment and Planning A*, 35(11), 1993-2018.
- Lee, H. L., Padmanabhan, V., & Whang, S. (2004). Information distortion in a supply chain: the bullwhip effect. *Management science*, 50(12_supplement), 1875-1886.
- Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. *Information Systems Research*, 15(4), 336-355.
- Marshall, M. N. (1996). Sampling for qualitative research. *Family practice*, 13(6), 522-526.
- Mason, R. O. (1986). Four ethical issues of the information age. *MIS Quarterly*, 10(1), 5-12.
- Mattern, F., & Floerkemeier, C. (2010). From the Internet of Computers to the Internet of Things *From active data management to event-based systems and more* (pp. 242-259): Springer.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709-734.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- McDonald, A. M., & Faith Cranor, L. (2008). Cost of Reading Privacy Policies, The [article]. *I/S: A Journal of Law and Policy for the Information Society*, 4(3), 543.
- McKinseyGlobalInstitute. (2011). Big data: The next frontier for innovation, competition, and productivity. www.mckinsey.com/~./.../Big%20Data/MGI_big_data_full_report.ashx: McKinsey Global Institute.
- Milberg, S. J., Smith, H. J., & Burke, S. J. (2000). Information privacy: Corporate management and national regulation. *Organization science*, 11(1), 35-57.
- Mowday, R. T., & Sutton, R. I. (1993). Organizational behavior: Linking individuals and groups to organizational contexts. *Annual review of psychology*, 44(1), 195-229.
- Murphy, M., & Barton, J. (2014). From a Sea of Data to Actionable Insights: Big Data and What It Means for Lawyers. *Intellectual Property & Technology Law Journal*, 26(3), 8-17.
- Needham, J. (2013). *Disruptive Possibilities: How Big Data Changes Everything*. O'Reilly Media, Inc.
- OxfordDictionaries. (2014). Oxford Dictionaries. Retrieved May 26, 2014, from <http://www.oxforddictionaries.com/>
- Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go? *MIS Quarterly*, 35(4).
- Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, 31(1), 105-136.
- Pavolotsky, J. (2013). Privacy in the Age of Big Data. *BUSINESS LAWYER*, 69(1), 217-225.
- Pearlson, K., & Saunders, C. S. (2009). Strategic management of information systems.

- Recker, J. (2013). *Scientific research in information systems : a beginner's guide*. Berlin ; New York Springer, cop. 2013.
- Recklies, D. (2006). PEST-Analysis.
- Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review*.
- Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*.
- Sandelowski, M. (2000). Focus on research methods-whatever happened to qualitative description? *Research in nursing and health*, 23(4), 334-340.
- Sapsford, R. (2006). *Survey research*: Sage.
- Seale, C. (1999). Quality in qualitative research. *Qualitative inquiry*, 5(4), 465-478.
- Skinner, G., Han, S., & Chang, E. (2006). An information privacy taxonomy for collaborative environments. *Information management & computer security*, 14(4), 382-394.
- Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly*, 35(4), 989-1016.
- Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly*, 20(2).
- Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 477-564.
- Son, J.-Y., & Kim, S. S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly*, 503-529.
- Stone, E. F., Gueutal, H. G., Gardner, D. G., & McClure, S. (1983). A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. *Journal of applied psychology*, 68(3), 459.
- Stone, E. F., & Stone, D. L. (1990). Privacy in organizations: Theoretical issues, research findings, and protection mechanisms. *Research in personnel and human resources management*, 8(3), 349-411.
- US.GOA. (2006). PERSONAL INFORMATION Agency and Reseller Adherence to Key Privacy Principles. <http://www.gao.gov/new.items/d06421.pdf>: United States Government Accountability Office.
- Van Alstyne, M., Brynjolfsson, E., & Madnick, S. (1995). Why not one big database? Principles for data ownership. *Decision Support Systems*, 15(4), 267-284.
- Van Slyke, C., Shim, J., Johnson, R., & Jiang, J. (2006). Concern for Information Privacy and Online Consumer Purchasing. *Journal of the Association for Information Systems*, 7(6).
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare. *MIS Quarterly*, 26(2), 13-23.
- Westin, A. F. (1967). Privacy and freedom. 1967. *Atheneum*, New York.
- Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues*, 59(2), 431-453.
- WhiteHouse. (2014). Privacy of Big Data. from <http://www.whitehouse.gov/blog/2014/05/01/findings-big-data-and-privacy-working-group-review>
- WorldEconomicForum. (2012, 23/04/2014). Big Data, Big Impact: New Possibilities for International Development. 2012. Retrieved April 23, 2014, from <http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>
- Xu, H., Dinev, T., Smith, H. J., & Hart, P. (2008). Examining the formation of individual's privacy concerns: toward an integrative view.
- YourDictionary. (2014). Dictionary definitions you can understand - YourDictionary. Retrieved May 26, 2014, from <http://www.yourdictionary.com/>