Aino Nybacka

PRIVACY CONCERNS OF CONSUMERS IN BIG DATA MANAGEMENT FOR MARKETING PURPOSES
AN INTEGRATIVE LITERATURE REVIEW

Bachelor’s Thesis
Department of Marketing
CONTENTS

1 INTRODUCTION............................................................................................................. 4
  1.1 Introduction to the research and research questions................................. 4
  1.2 Importance of the research ......................................................................... 5
  1.3 Methodology ................................................................................................. 6
  1.4 Structure ......................................................................................................... 7

2 INTEGRATIVE LITERATURE REVIEW AS A METHODOLOGY ........... 9
  2.1 Acquiring material ....................................................................................... 9
  2.2 Categorising the found material ................................................................. 10
  2.3 Searching relevant information .................................................................... 11
  2.4 Analysing the findings and searching for patterns ....................................... 11
  2.5 Synthesis ...................................................................................................... 11

3 ANALYSIS AND SYNTHESIS ............................................................................. 12
  3.1 Big data as a marketing tool........................................................................ 12
    3.1.1 The elusive big data ........................................................................... 12
    3.1.2 The research ....................................................................................... 13
    3.1.3 Research ............................................................................................. 14
    3.1.4 Collecting data .................................................................................... 15
    3.1.5 Storing data ......................................................................................... 17
    3.1.6 Analysing data .................................................................................... 18
    3.1.7 Using data ........................................................................................... 19
    3.1.8 After the data ...................................................................................... 23
  3.2 Privacy concerns in collecting and using big data....................................... 24
    3.2.1 The research ....................................................................................... 24
    3.2.2 Collecting data .................................................................................... 25
    3.2.3 Storing data ......................................................................................... 27
3.2.4 Analysing data ................................................................. 27
3.2.5 Using data ..................................................................... 28

3.3 Synthesis ........................................................................... 29
3.3.1 Collecting data ................................................................. 30
3.3.2 Storing data ................................................................... 31
3.3.3 Analyzing data ................................................................. 31
3.3.4 Using data ..................................................................... 32

4 CONCLUSION ........................................................................ 33
4.1 Theoretical Implications ..................................................... 33
4.2 Managerial Implications ...................................................... 34
4.3 Limitations and Suggestions ............................................... 34

5 REFERENCES ....................................................................... 36
1 INTRODUCTION

1.1 Introduction to the research and research questions

This bachelor’s thesis is a literature review about big data, privacy concerns for individual consumers and how these two overlap together in a way to possibly explain what are the privacy concerns that customers have, and the companies maybe should think about, when they utilize data as a marketing tool. In the picture below you can see a simplifying picture which brings the issues in my thesis together and show how the research questions have formed.

![Image of a simplifying picture]

The aim of this research is to offer an integrative framework for big data management for marketing purposes and privacy issues of consumers and bring clarification to the issue of privacy in the context of big data management for marketing purposes and search for answers and clarify the meanings of the research questions which are:

Minor research questions, which lead the way to the main question, are:

1. What are the main phases of the process of big data management for purposes of marketing?

2. What are the privacy concerns among customers in each of the big data management process phases?

The main research question is:
1. How privacy concerns of consumers relate to big data management for purposes of marketing

1.2 Importance of the research

There are several reasons for conducting this research. The rabid growth of the data provides an environment to the research that is ever changing. The data growth, explained in further detail later in this work, creates new business opportunities and innovations all the time. (Breur 2015)

This being said there has been and influx of information regarding the use of data as a marketing tool and of the privacy concerns in collecting and using this data, but the research of the specific relationship between these two is not conclusive, or a lot of the research surrounding these two does not seem to be collective as in that the research done the relationship between these two is referred in many ways but there is no collective understanding of the subject. Although some researches have listed potential uses of big data or uses of the big data in their articles (Van Auken 2015) the nature of the growth of the big data and the speed of the innovations that are made, make these lists “old” or just incomprehensive and the words the privacy aspect is rarely mentioned in these in nothing more but in passing sentence. E.g. the research has been looking at the problem from the companies point or view and from the point of view of the effect on the effectiveness of the online advertising but not from the point of view of an individual consumer and if there might be implications on a larger scale about the privacy, not just on an one individual aspect.

The subject is very actual in nature, because there has been a lot of talk about the big data revolutionising not only digital marketing but marketing in general. In spite of this there have been voices that are very worried about the privacy intrusions against individuals that collecting data creates. People are thinking how much data of then is out there and how it being used and how it could be used in the worst case scenario.

Furthermore the legislation around data privacy is changing. 15th of December 2015 the European Union passed legislation that gives consumers new rights on how their personal data could be used and what they themselves can do with it. The most
important implications from a consumers’ point of view include easier access to their own data, a right to portability, clarified “right to be forgotten and the right to know when your data has been hacked” (European Commission 2015) Although there is some research done from the impact that the new legislation has had on the advertising the use of big data as a marketing tool and the privacy implications of it are still there. The law comes into effect in two years’ time and the implications of it still remain to be seen on a full scale but considering the marketing and privacy there is some things that can be seen.

In summary the context of data and privacy is constantly changing and the collective research around these two is vague at its best. The implications of these two are more thoroughly researched than the actual subject and this research is conducted in order to gain a better understanding of the relationship between these two subjects. Furthermore, this research might give some managerial implications as in giving a collective knowledge about the consumers concerns about online marketing and privacy by integrating them together thus giving new insight to the consumer behaviour.

1.3 Methodology

This research is conducted as an integrative literature review. Cooper (1988) defines literature reviews as detailed independent works or existing data or as introductions for research for new data. The literature reviews that do not serve to purpose of being introductions for new data have multiple different purposes and focuses e.g. research outcomes, theories or research methods from different perspectives or for different goals.

In this work there is no new primary data and it focuses on integrating already existing information and on building bridges between the two areas and this is also listed in the Cooper’s purposes on what the literature reviews might want to achieve. Cooper explains that this is integration of two specific sets is a form of literature review that is most frequent in the scientific literature. This research synthesis or integrative research review focuses on drawing conclusions by summarizing former research and empirical studies from several different sources. This is being done so
that that the existing knowledge of the issues is explained and the important issues that have not been resolved are brought to common knowledge. The literature review can also try and inspire the way of the future research.

In the book Cooper handles the literature review in six stages: the problem formulation stage, the literature research stage, the data evaluation stage, the data analysis stage and the interpretation and presentation stage. These stages together form the structure of doing an accomplished literature review.

1.4 Structure

The structure of this work comes from the methodological framework and the requirements for the integrating literature review. The structure of this work is following: first part of this research delves deeper to the acquiring of the information in order to do the research: how and where the information was searched, what was found, what articles were deemed relevant and what were the principles behind the decision and then these findings were put into two categories relevant to the research (big data as a marketing tool and privacy concerns affecting consumers).

After this the finding are analysed; how is big data used as a marketing tool and what are the privacy concerns for consumers. In doing this the acquired material is used and similitudes and patterns first in both of the subjects separately are found and then the synthesis is drawn together. The synthesis part tries to explain possible dependence between the subjects and tries to offer new information derived from old. This basically means that in the synthesis the findings are taken, similitudes and patterns found in the analysis of the text, and things are found that overlap and could such explain the dependence between the two. In the synthesis part the gaps are also contemplated upon.

Last part of this work is the conclusion. In the conclusion part managerial and theoretical implications are given that the conducted research suggests. Generalisations of the findings are presented in such a manner which could be beneficial in doing further research or in marketing in general. The last part of the
conclusion chapter is suggestions for the further research where gaps that this research has left or created in the general research area are contemplated further.
2 INTEGRATIVE LITERATURE REVIEW AS A METHODOLOGY

In this part the process of conducting this research is explained, the findings and analysis of the found information and the synthesis between the two main subjects in this research. Below you can see the process simplified;

Acquiring the material → Categorising the found information → Searching relevant information → analysing the findings → searching for patterns → Synthesis

The following chapters are going to be concentrating on each of the parts in the process and describe what I did to the material in each of them. Chapter 2.1 contains the first three phases of the process, Chapters 3.1 and 3.2 contain the analysis and the search for patterns and lastly chapter 3.3 contains the synthesis.

2.1 Acquiring material

The process of this work started by searching internet for the literature about these two things that were discussed in this work: the use of big data as a marketing tool and privacy concerns in this from the consumer’s point of view. Two main databases were used when looking for the information: ESBCO-databases and ProQuest-databases and as a secondary database Scopus-databases was used. All the databases were used when searching information from EBSCO and ProQuest collections because some areas of this work concern not only the business are but also information technology and social ethics. This meaning that if the searches were confined the in solely business journals, some of the researched work that would be beneficial in doing this research could be excluded. The searches were all peer reviewed to maximise the validity of the articles chosen.

It has to be taken into consideration with these searches were that the results overlapped with each other extensively. We can also see that when you take into the consideration the inevitable occasion of articles that do not fit the research being conducted even though the subject is near to the ones needed.
After searching for relevant articles some articles were chosen that were deemed most relevant to the purposes of this research including both “big data” and “marketing” and “big data” and “privacy”. After reading these the next stop was to go to the references of the articles already chosen to find more information. By doing these and choosing more articles the amount of then used in this work was almost doubled. Use of big data is a reasonably widely researched subject when it is taken into consideration that the subject itself is relatively new. Big data has many uses not only in the marketing sector but also in health research and information technology. In this research the articles used come from both academic marketing and information technology journals, because some of the published work about applications of big data for marketing has been published in information technology journals.

In this literature review 24 academic articles, one neuromarketing project and six non-academic articles were used. The reason for the decision to use also non-academic articles was the need for material for the privacy section of this work due to shortage of academic research of the subject from the companies’ point of view. The articles that were used in conducting the analysis are marked with * in the list of references.

2.2 Categorising the found material

Categorising of the material was actually done in two different phases when doing this research. First the material was roughly divided in two groups: big data and privacy and big data and marketing. This however proved to be not wholly conclusive as some of the articles that handled the issue of marketing had information about the privacy aspects and vice versa. This in mind it was chosen not to draw attention simply to the issue of articles “belonging” to these two categories but search information of both in every article if there was any.

The second phase of categorizing happened after the material was searched and the points relevant to the research were put together under privacy and marketing and this was to categorize the findings to subcategories: collecting data, storing data, analysing the data, using the data and after using the data. This way of categorising
can be seen as the processual view to big data management and it was not found in any of the articles but the pattern was found naturally by trying to tie all of this information together.

2.3 Searching relevant information

The articles were handled on by one, searching for relevant information and patterns of research from which point of view or how these subjects had been handled in the past. The way this was done was simply by first excluding the information that was not relevant and then excluding some of the more technical points of the information that was relevant because explaining all these different points was not necessary for this research.

2.4 Analysing the findings and searching for patterns

The analysing phase began when the findings were categorized again into smaller subcategories and finding the patterns in these findings. This phase also made clear what information was important to include and what was not. For example at this phase the more technical aspects about using the big data were left out because it was not really that relevant for this research to actually know how the data is analysed but more important to know that how the data is actually being analysed in wider terms.

2.5 Synthesis

In the synthesis chapter the patterns found from both of the sides together were bought and more importantly the things that were missing from the literature were discussed and through this some generalizations and connections were bought to discussion to get these missing parts.

The managerial implications and limitations and suggestions of further research are then given from the basis of the synthesis. The synthesis part for example points out some things that are not handled in the research and this information is then directed to the last part of this work.
3 ANALYSIS AND SYNTHESIS

3.1 Big data as a marketing tool

In this chapter I am going to first explain the concept of big data through the articles gathered, for the purposes of clarifying the issue, and then I am moving to more precise point of view; explain how the concept, big data as a marketing tool, has been researched; viewpoints, methodologies etc. and lastly how the big data could be and is used as a marketing tool which is the main issue of this part.

3.1.1 The elusive big data

The term big data is a vague concept. There seems to be no consensus about the definition among the researchers but several of them have tried to define it at least to some extent. In the article by De Prato and Simon (2015) this problem has been addressed and the article tries to bring clarification to the term.

At its largest the term refers to the data that is generated when people use technology and the information that is given by people in various websites in internet, such as Facebook. As such the sources that the data comes from are not homogenous but the data originates from people, different kinds of machines that are used or sensors (De Prato & Simon 2015). Weinberg, Milne, Andonova & Hajjat (2015) list the sources of information in these two different sources. First, by human interactions in the internet such as text, audio, clicks or cookies, and second in the machine interactions of consumers using the products, such as behaviors of the user. The digital and analogue data there exists in the world has fused together to create the amount of the data that exist today. (De Prato & Simon 2015)

Breur (2015) delves a little deeper to the creation of all of this data in terms of history. In his article he states that the first wave of big data has come from man to machine interactions and the second one comes from machine-machine interactions. The data that comes from the machine to machine is new and the influx of data is great. The man to machine interactions can be seen as limited but the data from machine to machine interactions has growth possibilities that are virtually unlimited.
The amount of data there exists is however not conclusive. The amount of this data is so big that we can only guess its correct size. The best guess we have is that the amount of data has grown from 3 billion gigabytes to 300 billion gigabytes in twenty years, from 1987 to 2007. This, as it was said, is at best a guess of the size of the amount of the data there exist in the world (De Prato & Simon 2015). Alemany Oliver & Vayre (2015) state that big data is big because it observes constantly and gives data that gathers in a raw cloud that has potential to uncover underlying relationships between the data and the world around it.

Other definitions extend the big data beyond just the size of it. Boyd and Crawford (2012) on the other hand claim that big data is not about the size of it but the capacity to search and cross-reference the large amount of data and datasets. Pavolotsky (2013) defines big data as increasing amount of data volume and types and the increase of speed of collecting it and technology to analyze, store and draw conclusions about it.

Although some researches include the analogical data to their work, for the purposes of this work I intend to focus more on the digital data and its applications, newer innovations regarding the data mining and the term big data is used at its broader context, which includes also the data, collected from machine to machine interactions to an extent.

3.1.2 The research

The first point found from the marketing research containing the big data is that the information is scattered: there was almost no consistency between how the articles used in this work handled the issue. Some articles handle the work from the viewpoint of the different analytics, the sources of data and their applications in marketing or just offer a one new specific application derived from the integration of many. All the previous literature reviews about big data, or just data, in the use of marketing were either unavailable, old or only tackled one specific issue about big data.

The second point was that the definition of big data in the articles derived greatly from one another. Some researches handled big data as only digital data or only
digital data coming from human to machine interactions, some researches took the
traditional data into account but excluded the data from machine interactions, some
took digital data coming from both human and machine interactions as big data,
some used term Internet of things to differentiate between human to machine and
machine to machine interactions and some used all these forms of data under the
same name.

The third point was that the research is highly technical in nature; the articles that
handled the issue directly from the point of view that was most beneficial to this
research were from journals of information technology rather than marketing. The
marketing journals handled the issue mostly from how the big data affects the
marketing sector rather than how it could be used. The articles that chosen to this
research were these few from the marketing journals that actually handled the issue.

The problem of categorizing the findings in this work comes from the fact that the
academic journals that have researched this subject all handle the categorising from
different angles. Categorizing the findings on this work has been done with different
stages of data as a marketing tool in a chronological order, from collecting data to
using it.

3.1.3 Research

The literature that exists has focused largely on specific points of using big data in
marketing e.g. analysing the collected data in specific way or on the impacts it may
conclusive information of big data as a marketing tool comes from non-academic
articles or the information has “expired” because of the new usages that are invented
all the time.

When we look at data from the marketer’s point of view, we can instantly see that the
possibilities of usage in marketing are great. However the ever changing data
environment, due to the emerging regulations and rabid growth of the size on the
data, brings forth new innovations that are directed on utilization of the data and on
the other hand regulates the use of it.
Breur (2015) states in his article, that using analytics for the purposes of marketing research is not new by any chance. Data has been used for marketing purposes for decades, some companies use it and some don’t. Big Data just brings new possibilities due to the enormous amount of data and this can be translated to achieve competitive advantage. Van Auken (2015) opens the development of the data as a marketing tool and states that data collected from consumer panels was the big data of that era.

The articles list several different uses of the big data as a marketing tool and in this part I try to list and explain these uses for the purpose of integrating them with relevant privacy concerns in the synthesis chapter.

3.1.4 Collecting data

How and where the data is collected might not be so easy for people to understand. Almost every article used in this work that tries to explain big data, its aspects or uses tackles the issue of data collection. There is no data if no-one collects it.

As explained earlier data is collected from various different places. The collection of data can be divided in analogue and digital (De Prato & Simon 2015) or to traditional, digital and neurophysiological data (Kumar et al. 2013). And digital data can be divided to data coming from human to machine interactions and data coming from machine to machine interactions (Internet of things) (Breur 2015, Weinberg et al. 2015)

Traditional and analogue data can be collected from focus groups, structured interviews, unstructured interviews, surveys, experiments, transactions, observations and products reviews (Kumar et al. 2013) The traditional data is not on focus of this research but it’s important to recognize also the traditional data as a source because the data that is collected now from other sources builds on the data that has been collected in traditional ways. That is also true regarding the uses of the data: the use of data is marketing is not new, the amount of data and the ways it could be used has changed.
What should not be forgot when talking about digital data that it is also collected in traditional ways such as surveys, experiments, transactions, observations and product reviews (Kumar et al. 2013). The thing that is changed, however, is that the data is collected in numerous other ways from people in the internet. Data collected from the internet can be done in two ways. Some companies collect it on the side (Facebook and Amazon) and some collect it as a form of business to sell it forward to companies that can deem the information relevant (Christiansen 2011)

In the internet data is collected from online interactions of humans (Weinberg et al., 2015) Clickstream data, contents of social media, video data, search queries, blogs, (Davenport, Barth & Bean 2012, Kumar et al. 2013) are forms of digital data that consumers give of themselves when they operate in the internet. Call center voice data, community forums and incentivized referrals are examples of data that is collected by the company giving the possibility out to consumers ask about things, discuss etc. (Davenport et al. 2012, Kumar et al. 2013)

Data is collected while consumer tries to collects information about products they are thinking of buying, compares the choices and makes the decision to make the purchase (Leeflang, Verhoef, Dahlström & Freundt 2014). Christiansen (2011) also gives insights on how this data is collected in the first place; the sharing of information from the customers side can be done voluntarily (what information you give in Facebook) or involuntarily (cookies). Cookies, tracking pixels, flash cookies are methods of collecting digital data that have risen in the turn of new millennia and in the late 2000’s (Couldry & Turow 2014)

Data from sensors, radio frequency identification, devices (Davenport at al. 2012) are forms of digital data that come from machine to machine interactions. Data has been collected from machines as far back as in 1982, but the amount of data been relatively small and it has been more difficult to put into action as it is today (Breur 2015) Machine to machine interactions mean that person doesn’t have to participate or enter the data anywhere for the data to be collected (Weinberg et al. 2015)

Neurophysiological data can be collected from several different sources such as eye-tracking, facial electromyography, skin conductance response, electrocardiography,
electroencephalography magnetoencephalography and functional magnetic resonance imaging (Kumar et al. 2013). In this instance it’s very important that the people who collect this data are trained professionals (Barocci 2011) It’s not the purpose of this research to get into the practicalities of these highly complicated things but as an example the eye-tracking collects information on where, when and how long the person looks at each elements for example in advertisement (Maughan, Gutnikov & Stevens 2007).

3.1.5 Storing data

Storing this huge amount of data that is collected from these different sources is actually a problem for companies. There has been claims about that the data storing is not arranged in companies in a good way. (Leeflang et al. 2014). Davenport et al. (2012) on the other hand claim that the big data should be not only seen as data in stock, storage in data warehouses as in the past, but as a flow that is constantly changing and moving, and further than that storing data is also expensive (Breur 2015)

Websites can use personal data they collect in many different ways. Christiansen (2011) goes deeper into this question and divides it into three different categories. First they collect personal data and sell it to other companies or use it themselves after making in anonymous. Second they keep personal data within their firm but give other companies opportunity to target their advertisements to people that are in some kind of range (age, location or something more specific like preference of horror movies). Third way of using the collected data is with “the intention of selling the information, sometimes including specific profiles or names, to third parties”. Cloud databases have more storage capacity and processing power than traditional databanks but also virtual data marts can help data scientists to use big data by helping with sharing it (Davenport et al. 2012). Ad networks collect and connect the information form several different websites and also sell ad space. The last form of these companies, that collect data for business, buy the information from different websites and build and sell profiles for individual people in mind. (Christiansen 2011)
3.1.6 Analysing data

The value of Big Data in terms of marketing comes from the harnessing and analyzing the raw data that is created both in man to man and man to machine interactions. Collecting data is not rational if it’s done solely for the purposes of storing it. Analytics make this data accessible and usable and enables its use on creating customer value. The raw data itself is not usable, but through analytics it becomes so. (Breur 2015).

Analyzing big data can be difficult because of its characteristics and the technical aspects of it (De Prato & Simon 2015). This leads to the fact that those people who analyze the big data need to be well known in the field of IT, to fully grasp the potential of big data; to collect it, to analyze it and in some cases make innovations on how to implement it. (Davenport et al. 2012) The approach to the analysis should be continuous because the data is constantly changing. (Davenport et al. 2012) The data however needs to relevant or else it’s not useful. Data cleaning is a process that is done to determine which pieces of data are useful for interpretation (Boyd & Crawford 2012)

There are several different forms of analytics that have been reined into the use to analyzing big data. Statistical analysis packages and other kinds of products are developed to help companies with handling data (Davenport et al. 2012), text analysis can be applied to utilize the content that consumers provide when they talk about things in the internet (De Prato & Simon 2015), technical analysis is vital for the use of big data in recording consumer journey (Leeflang et al. 2014), predictive analytics such as Bayesian analysis and Monte Carlo Markov chains (Allenby et al., 2014) can be used to analyze and get insights from big data. Hu, Du & Damangir (2014) also researched predictive analysis, conjoint analysis and it’s applications to Google trends. Interpretation of neurophysiological data is necessary as in every other case and statistical analysis is important when people do research in the case of collected neurophysiological data (Barocci, 2011)

However it has to be remembered that the analysis done with big data is not free of errors nor are the interpretations that are made: there is a lot of data and it comes
from every direction and this can sometimes lead to seeing patterns when there is none. (Boyd & Crawford 2012).

3.1.7 Using data

The articles listed different ways of using data for enhance the marketing practices and in this work the uses are put under five different titles to simplify and to make the subject clearer. These uses and titles they are put under are: target marketing and personalization, consumer behavior, awareness and brands, segmentation and strategic and tactic decisions.

*Target marketing and personalization*

Target marketing is a form of marketing that has flourished with the introduction of Big Data. The ad that has been targeted to a specific customer is more effective than the one introduced to general public in mind (Christiansen 2011). Big companies like Google and Facebook exchange the anonymous data of the individuals and this has enabled companies to deliver an ad to a specific person in real time and furthermore, the location tracking technologies allow marketers to serve advertisements in consumer’s real time locations (Couldry & Turow 2014, Kshetri 2014). The target advertising is on the rise, not in the case just in Facebook or other social media platforms, publishers, like different newspaper websites, can also sell the advertiser the ability to give tailored messages to their customers (Couldry & Turow 2014). Furthermore big data also leads to personalization that is deeper in nature than ever before; advertisements and prices are targeted to the to one individual from the basis of big data, often by means that are unknown to consumer’s themselves (Couldry & Turow 2014).

In addition neurophysiological data can be used for example for measuring in more specific way people’s response to advertisements for example how they spontaneously recall an advertisement (Maughan et al., 2007) or emotional valence (fMri and facial coding) (Barocci, 2011). The advertising in the content websites can be done similarly as in any other website: just placing the targeted advertisement on a website. However there is also other use called native advertising. Native
advertising is target advertising that is made to look similar to the content published in the website it’s shown on (Tweet that is made to look like any other tweet in Twitter) (Couldry & Turow 2014).

The data is used in marketing to engage and serve customers in a way that it can be predicted from their online behavior. In the more traditional use of the data, the data is first processed and then the decisions are made, but the data coming from machines can also make decisions in real time due to the constant nature of the monitoring. (Breuer 2015). Databases of an individual firm, collected manually from consumers and also from when consumers use companies’ websites can be used with the data get from other sources of data to get more accurate response in their own customer base e.g. for personalized messages or specified purchase journey of a consumer. (De Prato & Simon 2015)

Consumer behavior

The big data creates a window to understand behaviors of humans like never before and this also applies to the way that consumers specifically behave. De Prato and Simon (2015) claim that the marketing departments can learn to understand consumer behavior in ways that in not known today. The data to do that can come from consumer’s habits or other kind of data they have submitted about themselves of from the machine to machine interactions. (Weinberg et al. 2015)

More specifically marketing departments can use data and “The collective intelligence” that has risen from the huge amounts of data available today to make theories about behaviors of humans and apply these to marketing (Jara, Concepción Parra & Skarmeta 2014) or by using digital traces can lead to discovering and understanding reasons why consumers click an ad or leave a comment on a product (Alemany Oliver & Vayre 2015)

The data that is collected when people click to get an access to the webpage can be analyzed to gain understanding about visitor traffic and the behavior of the consumer who do that (page views, how frequently they visit, how long they spent on a website and what they viewed during the visit) (Moe & Fader 2004). And in the context of
the data that is gotten from the machine to machine interactions the data that can be collected and the information that be assembled from several different devices at the same time can even lead to behavior changes (Davenport et al. 2012) and the firms can through this deepen their relationships with the consumers (De Prato & Simon 2015).

Awareness and brands

The use of data can be linked to raising awareness and building brands in various different ways. The data can be used on bettering advertising campaigns in the social media and furthermore to choose the best seeding strategy on in the social media (for whom we should advertise that the viral marketing and brand awareness could spread quickly) can be linked to the increase of profits (Hinz, Skiera, Barrot & Becker 2011). And the data collected from the interactions of humans in the internet or mobile environments can be used to build brand reputation, product quality and consumer satisfaction. (Jara et al. 2014)

Search engine queries can be helpful to marketers to understand what they should do when they want to get presence in search engines; keywords that are used can help them to do just that (Skiera and Abou Nabout 2012 via Kumar et al. 2013) and the marketers can also get information on what specific times the keywords were used to get further knowledge about them (Hu et al. 2014).

The data feedback of company’s products, brands or the company itself gathered from the internet can be used in machine learning, data mining and context awareness. And when taking this further this can be turned into potential in participative marketing (Jara et al. 2014)

With big data companies can also get opportunities to follow consumer’s journey when they are purchasing company’s products and this can lead to awareness, purchase or loyalty to the product (Leeflang et al. 2014). Neurophysiological data on the other hand can give emotional insights about consumers like never before. Tactical decision making from neurophysiological data could be about emotion versus cognition value for marketing campaigns. (Kumar et al. 2013). Venkatraman,
Clithero, Fitzsimmons & Huettel (2012) also link consumer awareness to the use of big data.

**Segmentation**

Benefits on segmentation are among those that has been biggest changes in marketing over last years and this is linked to digitalization (Leeflang et al. 2014). The data that is used at itself doesn’t give generalizations so abductive reasoning might uncover more uses for big data. The archetype based analysis which relies on universality of archetypes and that the analysis can give culturally relevant observations from big data and that the format is simple. The data that is processed in archetypal analysis could be used in more specific consumer segmentation (Alemany Oliver & Vayre 2015). Consumers can also preselect the content from the data flows through smart phones and this is connected to the segmentation done in companies (Couldry & Turow 2014).

**Strategic and tactical marketing decisions**

The digital data can also help companies to conduct strategic and tactical marketing decisions; marketing trends, market transitions and customer segmentation. In addition to this neurophysiological data can give more info about enhancing these decisions like market segmentation (Venkatraman et al. 2012). Big data can be also linked to understanding the business environment, creation of new products and quick and real time responses to changes if they emerge (Davenport et al. 2012).

The decision making of serving consumers is done in the base of the data collected and in real time about the location or other environmental facts (Weinberg et al. 2015). Social media analytics can detect quickly emerging of shifting trends in consumer sentiments and with the use of big data these changes can be detected quickly and then the decisions of what to do about it and actions to actually do it can also be done fast (Davenport et al. 2012)

Although there has been data coming from machines prior to the introduction of big data and the internet of things, it’s uses have been limited to what humans can do
with it. Now smart adaptive systems, self-governing models and predictive modeling done with the artificial intelligence can use this data more efficiently and without the need for humans to get this data and do something with it. These smart systems are now only emerging but in the future they are coming more and more normal. (Breur 2015)

The use of big data is not only in digital commerce but also in more traditional commerce as retail; in store technology and mobile applications are examples of these. Although the retail environment has changed in emerge of internet and mobile devices, there is also potential for retailers to change their marketing to help turn the tide for their benefit. Mobile viral marketing and personalization are things that can be also achieved in retail environment. (Pousttchi & Hufenbach 2014).

Data collected can give overall characteristics of people using the specific website and it allows companies rethink the content and advertisements they put out. (Christiansen 2011). Search trends (such as collected from Google Trends) can be linked to feature importance trends, sales forecasting, enhancing product design, budget allocation, advertising and production and inventory planning (Hu et al. 2014) As explained earlier neurophysiological data could make tactical decisions easier when deciding between emotion and cognition value for marketing campaigns. (Kumar et al. 2013). Social media can be used for creating awareness about the product or service and to get the full benefit of it companies should link the data gained from it to performance measures of the company itself (Hinz et al. 2011)

3.1.8 After the data

Big data can revitalize marketing research; hidden patterns impossible to find before can now be found. (Alemany Oliver & Vayre 2015). Benefitting from consumer data can include things like sales volume, innovation and customer engagement and brand loyalty. (Leeflang et al. 2014).

Innovations from the base of big data come from analytics and this is a way for companies to gain competitive advantage in an era where there’s shortage on energy, materials, food and water. This also links to the use of data owned by single
companies only with linking it to the data gotten from the other sources. (Breur 2015)

3.2 Privacy concerns in collecting and using big data

Song et al. (2016) claim that because of the big data the privacy is even harder to control than before. In their research they used five items to determine how the perceived privacy risk affects the consumers perception of personalized e-mails: “This e-mail is dangerous,” “I am worried that my information is exposed to others,” “I am concerned that I am not certain about how the company obtained my personal information,” “I am worried that the company that sent this e-mail is abusing my personal information,” and “I am concerned that I do not know how marketers use my personal information.” The study shows that the personalized services are little dangerous because the answers that the consumers gave.

But what are the privacy risks consumer’s faces when companies use big data to enhance their marketing efforts? Some people see big data as the big brother that can endanger their privacy among other troubling problems towards individuals (Boyd & Crawford 2012). Although websites say that they keep the data that is collected anonymous they often still have ways to make this data not anonymous and they can recognize you from your behavior on the website from other times you’ve visited; the datasets are only not identifiable by social security numbers of other things like that and this can still lead to problems on how people are treated in the marketplace. (Barocas & Nissenbaum 2014).

Consent and anonymity are not enough for privacy protection by themselves but they are something that can help going towards it (Barocas & Nissenbaum, 2014). The more non-anonymous the data gets the more intrusive it is from the privacy’s point of view (Christiansen, 2011)

3.2.1 The research

First point of research of privacy and big data was that it was done largely from the companies’ point of view and under the bigger concept of ethics of big data or linked
closely to the security aspect of big data. The research was also, and more so in the more specific issue of marketing, done from the point of view that how the privacy concerns of consumers could affect the people’s perception of a certain company or the marketing efforts in general. The information of what these issues are was really not that easy to find and in some information was found it was hard to find in larger quantities. This is why some of the articles used here were not of academic nature but from the magazines directed to consumers such as Wall Street Journal.

The second point of privacy, marketing and big data is again that the information was very scattered in nature and the problems of defining big data as a concept apply here too. The information used in this research was collected from numerous different articles and of the articles the information that could be used was hard to find. The third point was that almost all information that could be found was from the issues that rise from collecting data, so the other categories were left with just a little material.

3.2.2 Collecting data

The main issue when the data is collected seems to come from the consent (Barocas & Nissenbaum 2014); are people aware that their personal data is being collected, if they are do they understand the possible consequences and does the consent even matter in the long run. Even if the consent legislation could be done efficiently there are people that are willing to give the information and when they give this those who give less information could still be identified because of the similarity of traits and patterns and generalizing the population by these. (Barocas & Nissenbaum 2014)

First is from of the ways the data is admitted by the consumers is the voluntary disclose and it has the possibility to be harmful when the consumer doesn’t really understand what they are giving or click through the warnings. In the involuntary disclosure the consumer really doesn’t know that their information is being taken (cookies) and some cookies remain available even after the person deleted them. The age of the person that gives information is also under scrutiny; adults, while they might not fully understand that information they give, definitely know more than children. (Christiansen, 2011). But it has to be known that the data doesn’t always
come from sources that can pass as wholly legal (Kshetri, 2014). The problems from even the consensual admitting of data comes from the consumers that always don’t read or can’t understand the privacy policies and the full privacy policy is sometimes even unavailable and furthermore these privacy policies in the context of what it reveals might be difficult to put into the policy because no one can now what the data reveals. (Pavolotsky 2013)

Packets of information can track everything the person does in the internet and it relies on the same technology that spy agencies use when they’re tracking someone. These packets don’t only know the persons visiting history but everything they do. Even if the technology is not used as its whole power it can still lead to extremely personal profiling and privacy intrusions. (Stecklow & Sonne, 2010),

History sniffing is a technique that can get information and create a profile for an individual by looking at websites that person has visited. The web browsers show visited sites in different colors than those not visited and by running a code companies can get and use this information. History sniffing doesn’t rely on cookies so even if the person tries to protect their privacy by deleting cookies it isn’t much use in cases like this. (Vascellaro, 2010),

Scraping in an invasive data collection method that gets information from confidential discussions of specific user, and in some cases the companies even collect the pseudonyms that people use on different networking sites to connect all the information there is available. Although the companies say that the information is going to remain anonymous, it’s really not that difficult to connect people to their data. (Angwin & Stecklow 2010).

The smartphones can be used to track people’s movements, moods, health, calling habits, spending can predict illnesses, movements in the future, political ideas etc. making human behavior accessible for not only researches but marketers too (Hotz 2011). This can easily lead to privacy violations when there is that much data accessible and about so personal things. Also smart phones for example are not only scary because of the GPS but also because the photos or videos taken by them can
have GPS data even if the data is not used real time but send to friends or posted in the internet (Kshetri 2014).

New problems arise from the machine to machine interactions because the data that is collected can be even more sensitive because the data is collected more all-encompassing way (Weinberg et al. 2015).

3.2.3 Storing data

The privacy threats in storing data rise from the possibility of hacking. Hacking of the data that is stored is a threat because it can lead to exposing information about consumers and lead to stolen money and identities or hacking to the devices that collect the data can lead to even loss of life. (Weinberg et al. 2015)

Data outsourcing creates issues because of the concentration of data that is stored in the companies that outsource it and it’s an easy target for criminal. Also another problem comes from inside the companies that outsource data. There has been some concerns about these outsourcing companies using the data to their own uses and violate privacy and this can lead to economic, emotional, psychological and social harm to consumers. Also the organizations own data is stored in bigger volumes and this is also more appealing for criminals and also peak collection and flow periods make the data more vulnerable for crimes. (Kshetri, 2014).

3.2.4 Analysing data

The analyzing is the phase where the patterns can be recognized and it sometimes leads to things that the consumers don’t want anyone to know about them or don’t even realize about themselves.

Predictive analysis can lead to unpleasant experiences for consumers when it’s applied (Kshetri, 2014) and as an example of this is the Target predicting young girls pregnancy before her parents know about it (Duhigg, 2012)
Breur (2015) gives insight to the privacy problems coming from the data from machine to machine interactions. The article argues that the systems might be able to capture consumer’s behavior and who the customer is in a form of data and thus let this information exist in the internet and someone may be able to take the information from there.

The second issue about big data that Barocas & Nissenbaum (2014) addressed was about anonymity of the personal data collected and sold to third parties or kept inside the company. Kshetri (2014) has also identified deanonymization as a risk that can lead to harm for consumers. But there is also another side to this problem because deanonymization might not work because bigger the datasets and better the analytics the anonymous data can be linked to the individual more easily. (Pavolotsky 2013)

3.2.5 Using data

The data also moves from one place to another and sometimes this means that the consumer cannot know how their data is used or even who uses it. (Pavolotsky 2013).

The big data has the potential to reveal misbehaviors of consumers that they would like to keep secret. The marketers have the potential of this new data but it can also tempt marketers to go through whatever the consumers to cover these secrets up and exploit this information to get info about the consumers but also use it in marketing. This is a problem because the consumers do not want this information to get out. (Alemany Oliver & Vayre 2015). The data that has been collected for marketing purposes can lead to various different harmful outcomes for consumers which include for example background checks for hiring, termination decisions, scholarships and more (Christiansen 2011). The problem with existing and emerging privacy regulations is also the new innovations of using the big data: the regulations might not be able to keep up with it all.

GPS data, collected from GPS transponders and GPS chips has been under scrutiny in courts in both using it and in collecting it. (Pavolotsky 2013) Location data collected from consumers can for example cause problems like real time stalking
(Kshetri 2014) or even asset investigation, illegal debt collection, private investigation and kidnapping (Yan 2012).

### 3.3 Synthesis

As we can see the balance between the literature about big data and marketing and the big data and consumer privacy is very screwed towards the marketing uses and not the problems in the privacy aspect of it.

In the next figure we see the points from each factor of this research and how they are connected to each other. On the left is the phases of the big data management for marketing purposes, each phase in its own box and on the right are the privacy issues connected to each of these phases. In the middle the connections between the two are seen in black and in red are the points that are missing from the previous research and added during this research. This is also the integrative framework created from the research. After the figure I address the connection points one by one.
3.3.1 Collecting data

The collection of data has been covered extensively in both aspects relevant for this research. There are problems with consent aspect of collecting data and it is at the core of the issue about protecting consumers and their privacy. The marketers might want as much data as possible and as fast as it is possible to collect but they need to think about the consequences their actions have on the consumers they want to serve.
If the collection of data is not ethical the rest of the process of using data is also in scrutiny.

3.3.2 Storing data

When the amount of data is growing the possibilities on how to storage it are also growing. The literature used in this research focused on how the privacy problems arise from hacking and did not get into particularizes about how the data that has been hacked might be used. Some forms of stored data are more vulnerable to hacking than others.

In addition if some forms of stored data are more vulnerable to the hacking or other criminal actions and if the consumers don’t even understand the way their data is collected in the first place they don’t know about these different types of storing the data and that some forms are riskier than other. This being said the consumers might know even less about the privacy dangers they face when their data is stored.

3.3.3 Analysing data

The analysis phase by itself doesn’t seem to be bringing any privacy issues to light; it’s the results that come from these analyses that can pose the threat wherever it’s used. The marketers lean on analysis to bring forth patterns and create profiles about consumers that they can use to advertise of otherwise market their products to consumers.

In a way the same problems about hacking might be relevant also in this phase of the big data management process. If raw data is potentially dangerous already the already analyzed patterns and profiles might be even more so. The literature doesn’t give information about whether the raw data that is storage or these already analyzed profiles are more vulnerable to hacking, but the same threats of privacy can be translated to this already analyzed data.
3.3.4 Using data

Using data to unlock new understanding about consumer behavior and to implement target marketing and even to some extent data’s use for segmentation can be seen as the most risky from the privacy’s point of view. These are the points of which the use of big data comes to direct contact with the consumer and by that they are the ones that can lead into exposing the information about consumers that they don’t want to come public.

The literature about privacy threats focuses on more serious kinds of breaches in privacy but there can be also some breaches that can be seen as mundane in comparison to stalking or even background checking. If a person is advertised according to the things they have browsed in the internet or might be interested in the people who are in position to see these advertisements, e.g. friends and family, can also learn things about the person that they would like to be kept secret.
4 CONCLUSION

The aim of this research was to seek answers to the research questions defined at the beginning of this work: what are the main phases of the process of big data management for marketing purposes, what are the privacy concerns of consumers in each of the big data management phases and finally how privacy concerns of consumers relate to big data management for marketing purposes. The analysis and synthesis chapters provided the answers to these questions and also brought to light what were the parts that have not been researched previously. The chapter 4.1 brings answers to research questions and thus gives theoretical implications from this research. Chapter 4.2 gives managerial implications and lastly the chapter 4.3 explains the limitations of this research and gives some suggestions for the future research.

4.1 Theoretical Implications

The figure 2 given in the synthesis chapter gives the privacy issues in the big data management an integrative framework explains the outcomes of this research on the whole and also gives the answers to the research questions. The figure shows the relations of big data management for marketing purposes and privacy implications on each different phase of the process. Each of these phases are further explained in the analysis and in the synthesis we see the connections between the two as well as the gaps left in the research surrounding these two again in each different phase.

First this research aimed to identify phases of big data management for marketing purposes and through this clarify the big data issue surrounding it. These phases were identified as: collecting, storing, analyzing and using data (target marketing and personalization, consumer behavior, awareness and brands, segmentation and strategic and tactic marketing decisions). The second minor research question was the relation of privacy of consumers with big data to each of these phases. These concerns, further explained in the chapter 3.2 include for example dubious consent when collecting data or hacking the data that is stored.
The main research question asked the relation of big data management for marketing purposes and the privacy concerns of consumers. This issue is handled thoroughly in the synthesis chapter and in the figure 2 which also shows the gaps in the previous research left by the literature handled in this research. To summarize in the collecting of data the relation comes mainly from the consent, in storing data from hacking, in analyzing data the possibilities of different kinds of analysis applied to data and in using data by target marketing, consumer behavior and segmentation possibly revealing information about consumers they would want not to be of public knowledge and lastly that the information that can be gotten from each of these phases either through hacking or other means of acquiring personal data of consumers could be used for criminal or otherwise harmful way.

4.2 Managerial Implications

The potential solutions for privacy problems by implementing privacy on the design level or suggesting better privacy policies have been given but it is important for marketers to understand what are the implications when they collecting, storing, analyzing, and data from the consumer’s point of view. If the people who use data are not aware about of the privacy threats, how can they really understand what to do to implement privacy in their practices?

The understanding can also be linked to transparency, which is a thing that consumers value high when they regard companies and their trust to them. The transparency in privacy issues could lead to heightened trust among consumers and it could give a company competitive advantage to have that.

4.3 Limitations and Suggestions

This research is by no means wholly comprehensive; the limitations in scope of this research could lead to that some relevant things that have been researched are not included in this research and also the fact that this research doesn’t take into account the existing privacy legislation nor the research that has been done to the ways privacy issues affects marketing might hurt some of the results from this research. In
addition some of the literature used in the privacy section does not come from the academic journals so this is also something that can limit the results.

The future research should concentrate of giving insight about the privacy concerns to consumers from the marketing’s point of view and even if there is some research about that, the constantly changing area of big data gives research more and more possibilities to extend already existing research to get these new things under inspection. Also it could beneficial to research these more “mundane” forms of privacy intrusions because they are the ones that everyday person can experience and so they are closer to the consumers.
5 REFERENCES

Alemany Oliver, M, Vayre, J-S (2015) Big data and the future of knowledge production in marketing research: Ethics, digital traces, and abductive reasoning, *Journal of Marketing Analytics* 3(1), 5-13*


Breur T. (2015) Big data and the Internet of Things. *Journal of Marketing Analytics* 3(1), 1-4*


Christiansen L (2011) Personal privacy and Internet marketing: An impossible conflict or a marriage made in heaven? *Business Horizons* 54(6), 509-514*


Van Auken, S. (2015) From consumer panels to big data: An overview on marketing data development. *Journal of Marketing Analytics* 3(1), 38-45*


