

## MASTER

### Big data opportunities for the retail sector a model proposal

van Eupen, M.G.H.

*Award date:*  
2014

[Link to publication](#)

#### **Disclaimer**

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Eindhoven, August 2014

# **Big Data Opportunities for the Retail Sector**

*A Model Proposal*

by

M.G.H. (Marcel) van Eupen

BSc Industrial Engineering & Management Science –TU/e 2014

Student identity number 0715154

in partial fulfilment of the requirements for the degree of

**Master of Science  
in Innovation Management**

1<sup>st</sup> university supervisor: dr. C.M. (Claudia-Melania) Chituc, Information Systems  
2<sup>nd</sup> university supervisor: prof. dr. R.J. (Rob) Kusters, Information Systems  
Hosting company supervisor: M. (Marcel) Roelants

TUE. School of Industrial Engineering.  
Series Master Theses Innovation Management

Subject headings: Big Data, Business Processes, Business Value, Payment, Performance management, Retail Sector, Information Systems, Information Technology.

# Abstract

Information Technology is becoming increasingly important for companies. For a company to optimally use its IT facilities, relevant data is crucial. This data is used for analysis in order to gain insights which improve business processes. Relatively new to this field is the usage of Big Data. The concept Big Data does not have one clear definition but most articles agree that Big Data consist of the 3 V's which are: volume, velocity and variety (Database Trends & Applications, 2012; Database Trends & Applications, 2013; Hopkins & Evelson, 2011; Lopez, 2012; Nasar & Bomers, 2012; Won, 2013). Big Data usage can improve the insights that companies gain from their data analysis. Big Data is becoming more and more a topic of interest for many companies but still a lot of research has to be done. Big Data usage can have huge benefits for companies when it is properly used, but there are also risks involved. In this Master Thesis the focus is on Big Data usage in the retail sector of the Netherlands and Belgium. The main research question is: **how can Big Data be used to derive business value in the retail sector?**

To answer this question research has been done to determine who are the stakeholders of the process of getting business value out of Big Data in the retail area? Additionally, which usage areas are there in retail for Big Data to gain business value? How can value be gained out of Big Data within these usage areas? How can customer loyalty be gained? The last question is about how the usage areas of Big Data could best be modelled into a design.

The research started with determining the different methodological steps that should be performed, which started with the development of the research questions. To answer these questions first a literature study was performed. The literature study gave generic answers to the questions, but because the area of Big Data is relatively new, the answers varied between different industries and different countries. Some research questions could already be answered where others required more research. The primary stakeholders in the Big Data processes concerning the retail sector are: merchants (main stakeholder), customers, payment providers, banks and the government. Customer loyalty can be gained via hard (monetary) benefits and soft (non-monetary) benefits. The combination of the two seems to be the best and the ratio is merchant dependent. Soft benefits are expected to become increasingly important in the coming years because differentiation in price will become more difficult. Big Data can be used for both. For the other questions literature gave interesting insights but those insights had to be tested via data collection.

Data collection was needed to test the insights gained from literature and were adapted or improved to make them suitable for this research. Different data collection methods were considered and interviews with semi-structured questions were found most suitable in this situation. There were two different groups interviewed with different questionnaires but with the same objectives. There was a group of merchants interviewed and a group of experts. The questions were adapted to each separate group. The merchants were from both Belgium and the Netherlands and from different retail areas to get a more complete view of the merchants on the questions. The experts that were interviewed had an expertise area covering one or more stakeholders' interests. It was also made sure that in the group of respondents all stakeholders were represented.

The interviews with merchants and experts gave some interesting insights. Three usage areas of Big Data in the retail sector were discovered. Improving the customer loyalty programs; Big Data could help here to make the customer loyalty programs more personal and thus increase customer loyalty and in the end the revenue. The second usage area was the optimizing of in-shop customer experience, this could be both online and offline. Big Data could help to change the shop layout to increase the sales or the website of the merchant could be adapted to individual customers. The last usage area that was mentioned was benchmarking. Payment providers or banks could analyze payment data and compare the data of one merchant to a generic industry benchmark. Big Data could help to get insights for the merchants which could be useful for decision making to improve business processes. A process model was found the best suitable model for this research. With a process model retailers and outsiders get a good overview of all processes needed to perform Big Data analysis and they can determine where to invest in. It also gives them a good overview on the relation between processes and decision points within these processes. The interviews were semi-structured with open questions leaving room for all types of answers. Some topics were mentioned by literature, experts and merchants. Other topics were only covered by a few respondents. For those insights a validation was needed.

The validation was again a semi-structured interview with more closed questions with a Likert scale ranging from 1 to 7. The validation was again with merchants and experts. The majority were experts with knowledge from the financial industry. The benchmarking process needed the most validation so the emphasis was on experts with knowledge in that potential usage area; but again all stakeholders were represented. The analyses of their answers led to improvements in the conceptual model. They also mentioned a fourth usage area being risk management. Especially for online stores Big Data can be used; for example to detect fraud. The validation gave insights in how to improve the conceptual process model. The answers to the main research question and sub questions remained the same. From this research another interesting conclusion could be drawn. According to the respondents, Big Data usage is being considered but not widely implemented at this moment. A lot of changes are expected in the coming years. The business value of Big Data in the retail sector is that Big Data can be used in the retail sector for improving customer loyalty programs, optimizing in-shop customer experience, risk management and benchmarking. At this moment it is too soon to look at real Big Data implementations.

The limitations of this research are that the topic of Big Data is relatively new and it is hard to find companies who want to talk about Big Data usage. Large differences are found in knowledge between different industries. Big Data is not used often in the retail sector at this moment so the model may need modification in a few years when Big Data is accepted and implemented by retailers in general. At that time it will also be possible to make the model more specific. The other limitation is that the research is done with a small group of respondents. It is very hard to find respondents for a research about a for retailers very sensitive topic such as Big Data. It is also a topic where a lot of knowledge is needed and the people capable of answering the questions were often not available for interviewing. Future research could be done by making the model more specific when Big Data is used by the majority of retailers. A quantitative study can also be interesting to perform using a larger set of respondents making it possible to generalize results. For this study a qualitative approach was chosen which makes it hard to generalize the results.

# Preface

Information technology is becoming more and more important for all sorts of companies. Business processes are transferred to digital processes where all information is stored in large databases. The IT landscape is known for its rapid changes. Data from companies first was used by those companies for business processes via Business Intelligence such as data mining. Today analyzing and storing more and larger databases is possible due to new software and hardware. This new development is called Big Data. Big Data has a lot in common with the traditional Business Intelligence but as can be read in this Thesis there are also a lot of differences. Big Data is relatively new and therefore not much research is done in the area of Big Data yet and it is an interesting topic for many companies. For many companies it is still not clear what the possibilities are of Big Data and how to use it to gain value because Big Data usage can require large investments. Big Data usage differs per industry. In this research the focus will be on usage of Big Data in the retail sector.

This Thesis will be about the business value of Big Data in the retail sector. Which usage areas are there to obtain that value? Customer loyalty is one of the usage areas for example. In the retail sector most data is used to optimize customer loyalty programs and get more loyal customers. The interest of Big Data in this sector will be to use Big Data for the usage areas that were found.

This research is performed at both a hosting company and Eindhoven University of Technology. The hosting company has a background in the mobile payment industry and is active all around the world. They have supported me in doing this research and helped me to find respondents for the interviews. Therefore I would like to thank everybody from the hosting company that helped me in performing this master Thesis for their support; and for giving me the opportunity to do this Master Thesis research in the first place. Secondly, I would like to thank everyone from both the merchants and the experts who made time to have an interview, with their insights I was able to make the conceptual model and to validate the model. Lastly, I would like to specifically thank the supervisor from the hosting company Marcel Roelants for his guidance, feedback, help and support from begin to end. I would also like to thank my supervisors at the University of Technology Eindhoven, C.M. Chituc and R. Kusters.

# Contents

Abstract.....	i
Preface .....	iii
Table of figures .....	vii
Table of tables.....	vii
1. Introduction .....	1
2. Scope and Methodology .....	2
2.1. Scope and research questions .....	2
2.1.1. Main objective .....	2
2.1.2. Research questions .....	2
2.2. Research Method.....	4
2.2.1. Methodology.....	4
2.2.2. Diagnosing.....	5
2.2.3. Action planning .....	5
2.2.4. Taking action .....	6
2.2.5. Evaluating.....	6
2.2.6. Specify the learning.....	6
3. Literature study.....	7
3.1. What is the definition of Big Data?.....	7
3.2. What are the regulations of Big Data for the payment/retail sector? .....	8
3.3. What is the value of Big Data in the context of the payment/retail sector?.....	9
3.3.1. Generate value.....	10
3.3.2. Data mining .....	11
3.4. Which existing models are relevant?.....	12
3.4.1. Relevance for the payment and retail sector .....	14
3.5. What are the main risks of Big Data for the banking/retail sector?.....	14
3.5.1. Privacy.....	14
3.5.2. Data.....	15
3.5.3. Innovation diffusion and technology acceptance models .....	16
3.6. How can customer loyalty be gained? .....	16
3.6.1. Soft and hard benefits.....	17

3.7.	Conclusion.....	18
3.7.1.	Regulations.....	18
3.7.2.	Value of Big Data.....	18
3.7.3.	Main risks and challenges.....	18
3.7.4.	Customer loyalty.....	18
3.7.5.	Insights from literature.....	19
4.	Data collection design.....	20
4.1.	Design of empirical research.....	20
4.2.	Objectives.....	21
4.3.	Interviewed merchants and experts.....	21
4.3.1.	Planned execution of questionnaire.....	22
4.4.	Interview questions.....	23
4.4.1.	Link between interview questions and objectives.....	24
5.	Process model proposal.....	25
5.1.	Interviews.....	25
5.1.1.	Characterization of respondents.....	25
5.2.	Results of the interview.....	26
5.2.1.	Results of the interviews with merchants.....	27
5.2.2.	Results of the interviews with the experts.....	29
5.3.	Proposed conceptual process model.....	30
5.3.1.	Explanation of the proposed conceptual process model.....	33
5.3.2.	Example use of the proposed conceptual process model.....	35
5.3.3.	Comparison of the different models.....	36
6.	Validation.....	39
6.1.	Validation design.....	39
6.1.1.	Type of validation.....	40
6.1.2.	Execution of questionnaire.....	41
6.1.3.	Characterization of respondents.....	41
6.1.4.	Reliability and validity.....	42
6.2.	Results of the interviews.....	42
6.3.	Proposed process model.....	44
7.	Conclusions, Limitations and Future work.....	46



7.1.	Conclusion.....	46
7.2.	Limitations.....	49
7.3.	Recommendations and directions for further research .....	50
8.	References .....	51
8.1.	Articles .....	51
8.2.	Books.....	55
8.3.	Sites.....	55
9.	Appendices.....	56
	Appendix A: Main mobile banking and payment providers .....	56
	Appendix B: Questionnaire for merchants. ....	58
	Appendix C: Questionnaire experts .....	59
	Appendix D: Discussion demographics data collection. ....	60
	Appendix E: Results of the merchant interviews.....	61
	Appendix F: Results of the expert interviews. ....	65
	Appendix G: Questionnaire validation.....	68
	Appendix H: Discussion demographics validation .....	70
	Appendix I: Results of the validation interviews .....	71
	Appendix J: Figures .....	74
	J1: Figure 5 .....	74
	J2: Figure 6 .....	74
	J3: Figure 7 .....	75
	J4: Figure 8 .....	75
	J5: Figure 9 .....	76
	J6: Figure 10 .....	77
	J7: Figure 11 .....	77
	J8: Figure 12 .....	78
	J9: Figure 13 .....	78
	J10: Figure 14 .....	79

## Table of figures

Figure 1. Methodological framework for the master Thesis .....	4
Figure 2. Input from literature for process model .....	19
Figure 3. Proposed conceptual process model .....	32
Figure 4. Final process model .....	45
Figure 5. The three V's of Big Data (Won, 2013) .....	74
Figure 6. Big Data to Business flow (CGI p72, 2013) .....	74
Figure 7. The balanced scorecard of Kaplan and Norton (1987) .....	75
Figure 8. Big Data Value Framework for the Financial Services Industry (De Vries p42, 2013) .....	75
Figure 9. Big Data Value Framework for the Financial Services Industry divided into four perspectives of Balanced Scorecard (De Vries p44, 2013) .....	76
Figure 10. The BPM Lifecycle (Dumas et al. p21, 2013) .....	77
Figure 11. Innovation Diffusion Model of Rogers (1962) .....	77
Figure 12. The Technology Acceptance Model of Davis (1989) .....	78
Figure 13. Quality aspects and quality assurance activities (Dumas et al. p172, 2013) .....	78
Figure 14. Adjustments to the process model .....	79

## Table of tables

Table 1. Demographics of the merchants .....	26
Table 2. Demographics of the experts .....	26
Table 3. Comparison of models .....	37
Table 4. Demographics of respondents for validation .....	41
Table 5. Comparison between five different mobile payment providers. ....	56

# 1. Introduction

Information Technology is becoming more and more important in every organization nowadays. Most of the IT is making use of data. In the past this data was mainly from one source, i.e. the company itself, but nowadays data sources are quickly growing. It can be said that the era of Big Data is coming. The concept of Big Data does not have one clear definition but most articles agree that Big Data consist of the 3 V's which are: volume, velocity and variety (Database Trends & Applications, 2012; Database Trends & Applications, 2013; Hopkins & Evelson, 2011; Lopez, 2012; Nasar & Bomers, 2012; Won, 2013). Big Data contains a lot of data, for example IBM (2013) claimed that we create 2.5 quintillion bytes of data each day. In other words: that 90% of the data that is available was created in the past year. With the rapid increasing this number could now be much higher. The second V, velocity, is about the speed which the data should be analyzed and processed. The last V is about the variety; the different resources the data should be gained from.

Big Data can be used in all sorts of industries. Navigation companies for example use the data from their sold navigation devices to predict traffic jams and warn about those traffic jams and offer alternative routes. Government uses Twitter and Facebook data for crowd control. In the financial sector Big Data is used for fraud detection. Flue epidemics can be monitored via the amount of medicines sold<sup>1</sup>. In this Master Thesis research the focus will be on the retail sector. In the interviews it was mentioned that details are very important in the retail sector. Big Data is expected to help customers more on a unique level, for example via changing the outlay of a website or through personal coupons. In this Thesis an answer is presented for the question: **How can Big Data be used to derive business value in the retail sector?** To answer this question also a process model was created.

The project was performed at a hosting company. This hosting company is a large mobile payment provider which is active throughout the world. This research was performed for the Benelux department of the hosting company.

This Master Thesis is divided into seven chapters. The next chapter will give the scope of this Thesis and the methodology that was used to conduct this research. The third chapter gives the insights from literature to the research questions. The fourth chapter will present an overview of the preparation for the data collection, describes which method is used and how it will be performed. The fifth chapter will give the results of this data collection; also the conceptual process model is presented and explained. The conceptual model needed validation at some parts of the model, the sixth chapter is about the validation of those parts and will present the final version of the model. The last chapter will give the conclusion, limitations and recommendations for further research.

---

<sup>1</sup> All mentioned during congress presentation SAS Forum Nederland 8 October 2013

## 2. Scope and Methodology

In this chapter the methodology that is used in this Thesis will be explained. It will start with an overview of the objectives for this Thesis followed by the research questions, the last part of this chapter will give an overview of all the steps that were performed doing the research.

### 2.1. Scope and research questions

To perform a proper research the research work has to be scoped. This study will focus on the business value of Big Data in the business area of retail. This study will result in a process model which can be used as a guideline in getting business value out of Big Data. The research is hosted at a mobile payment provider.

#### 2.1.1. Main objective

As mentioned in the literature review and research proposal of Van Eupen (2014), Big Data is important for companies in order to achieve success in the future. Big Data gives organizations opportunities that were not possible before because of the high expenses or the lacking technology (Lamont, 2013 and Lopez, 2012). Big Data usage is not without risks though. These risks are for example privacy related issues which can cause reputation damage or data sharing risks.

The main objective of this master Thesis is to develop a model that can guide the usage of Big Data in the retail sector. At this moment a model will help stakeholders to understand how Big Data can help deriving value. To develop this model a research question with sub questions are made which can be found beneath. A literature study and data collection provided the insights for these questions. These insights have helped to develop the conceptual model and consequently the final version of the process model

#### 2.1.2. Research questions

Big Data is a hot topic at this moment and a lot of research is done in this area. This research is primarily done in the area of the definition of Big Data and on the possible value and risks of Big Data. At this moment there is not much research done in the Dutch retail area concerning Big Data. Big Data can create business value if it is used in the right way. This research helps with the decision to use Big Data by developing a model in how Big Data usage can deliver business value in retail. In the retail sector customer loyalty is important so Big Data usage in the retail sector will be linked to customer loyalty. To develop the model which is the main objective of this Thesis different sub research questions are developed to answer the following main research question:

**How can Big Data be used to derive business value in the retail sector?**

To answer this question and relate it to the main objective sub questions were made. To give answer to the main question first the different and main stakeholders have to be found who are active in this area. To obtain business value it is important to than look to the usage areas of Big Data which can be interesting in the retail sector. It is also important to look how value can be gained within the found usage areas of Big Data. In the retail sector customer loyalty is a very important topic of interest. It is assumed that in the retail sector Big Data will be used for customer loyalty purposes. So the initial focus

was on customer loyalty and to look for other potential usage areas as well. To get a proper understanding in customer loyalty a separate sub question is made how customer loyalty can be gained. The last question will be about how the information can be made into a design; because the main objective of this Thesis designing a model which is relevant for the stakeholders found in the first sub question. The five sub questions will be as follows:

1. Which main stakeholders can be identified in the process of getting business value out of Big Data in the retail sector?
2. Which usage areas can be identified in the retail sector for Big Data to gain business value?
3. Within these usage areas how can value be gained out of Big Data?
4. How can customer loyalty be gained?
5. How can the usage areas of Big Data that are discovered be modelled into a design?

To answer these questions, first a literature review has been performed. There is looked into literature if there is done previous research that gives answers to the questions. According to Webster and Watson (2002) a review of prior, relevant literature is an essential feature of any academic project. It will create a firm foundation for advancing knowledge. It facilitates theory development, closes areas where a plethora of research exists, and uncovers areas where research is needed. According to them there is a lack of theoretical progress in the Information Systems field which makes a proper literature review harder but still important. The explanation for this phenomenon is the youth of the field (Webster and Watson, 2002). Literature is first used in this thesis to provide some context. There is looked for what Big Data is, what the regulations are for Big Data in the payment/retail sector and what the main mobile payment providers are. Literature is also used to give an answer to the sub questions of this Thesis. What is the value of Big Data in the context of the payment/retail sector is used to answer the sub question which usage areas can be identified in retail for Big Data to gain business value? This literature review question has two sub questions which existing models are relevant and what the main risks are for Big Data usage; these two sub questions will be used to answer the question: within these usage areas how can value be gained out of Big Data? The last literature review question is the same question as was mentioned before as research sub question: How can customer loyalty be gained? The other two questions about stakeholders and how it can be modelled into a design will follow from the joined answers to all the literature questions. The questions for the literature review will be as follows:

- What is the definition of Big Data?
- What are the regulations of Big Data for the payment/retail sector?
- What is the value of Big Data in the context of payment/retail sector?
  - Which existing models are relevant?
  - What are the main risks of Big Data for the payment/retail sector?
- How can customer loyalty be gained?

After the literature review there was data collected because the answers from the questions were not complete and fully applicable to the build the model. The questions that needed to be answered in this data collection are made up after the literature review and are based on the insights gained from the literature. The questions depended on the kind of data collection method that was used. A complete answer to the research questions should be possible after the data collection. This made it possible to build a process model to help determine how value can be gained from Big Data in the retail area.

## 2.2. Research Method

In this chapter the research methodology that is used will be discussed. In the information systems field several methodological frameworks which can be used for research have been developed. There are two research frameworks that were analyzed to be used as a basis in this master Thesis. The Systemic Methodical Framework for IS Research from Pérez, Grimán, Mendoza and Rojas (2001) and the Information Systems Research Framework from Hevner, Marsch, Park & Ram (2004). Both were used as input for the methodology framework adapted for this particular Thesis.

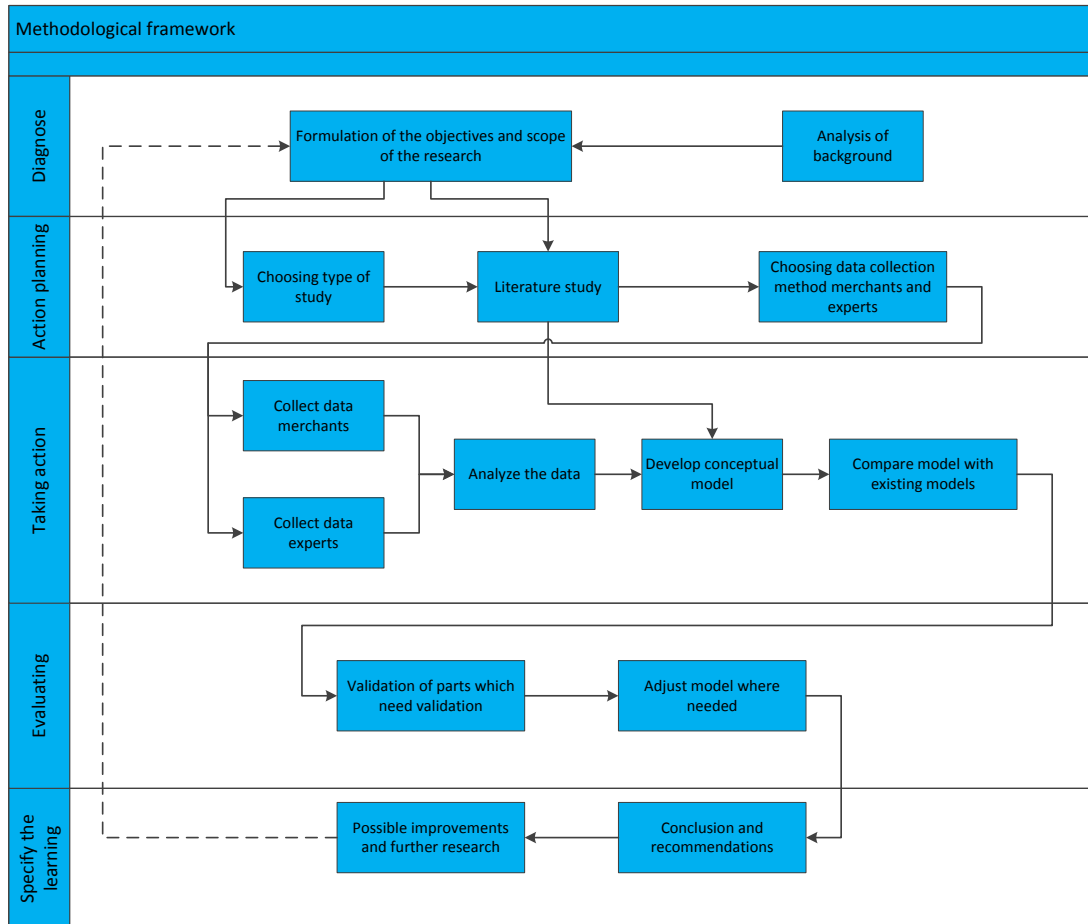


Figure 1. Methodological framework for the master Thesis

### 2.2.1. Methodology

To get proper answers to the previously described research questions this research methodology is guided with that of Hevner et al. (2004) and Pérez et al. (2001). Hevner et al. (2004) described the performance of design-science research in Information Systems via a concise conceptual framework. The methodological framework of Pérez et al. (2001) defined 5 categories for the research. First there has to be something diagnosed, then there has to be made a plan for action, this action should be executed, afterwards there has to be an evaluation to end with the specification of the learning. Because it is a continuous learning process, this Thesis can be input for further research; therefore the dotted line goes back to the beginning. In the figure above (Figure 1.) the methodological framework can be found that is used for this master Thesis, with all the steps per category.

### **2.2.2. Diagnosing**

In this phase the primary problems are identified and the underlying reasons which the organization wants to change (Mendoza, Marius, Pérez & Grimán, 2006). In this master Thesis this phase consists of two parts; background analysis and formulation of the objectives and scope of the research.

The first step that has been done in the diagnosing phase for this master Thesis is the analysis of the background. It is important to know the background of Big Data and the background of the hosting company. It is also good to have a look at the sector this Thesis will be about; the retail sector. It is possible that Big Data usage differs a lot between different sectors.

After the analysis is complete the formulation of the objectives and scope of the research were defined, these can be found in paragraph 2.1.1. and 2.1.2. Also the research questions were defined in this stage. These research questions should make it possible to build a model which is the main objective of this Thesis. This purposed model will help to determine how to get value from Big Data in the retail area.

### **2.2.3. Action planning**

After the diagnosing phase it is time to plan the action. Mendoza et al. (2006) stated that in this phase the organizational action should be stipulated. Through this action the principal problems should be eliminated or improved.

First decision that has been made here is whether the research is qualitative or quantitative. According to Van Aken, Berends & Van der Bij (2009) qualitative methods are aimed at the discovery of an object; phenomena etc. and quantitative methods are more suitable to find the amount of these attributes. According to Blumberg, Cooper & Schindler (2008) a first study is always a qualitative one which can then be tested with a quantitative. For the hosting company this is a first study and there is not done much research in the domain of Big Data usage in the Dutch and Belgium retail sector. According to the criteria of van Aken et al. (2009) and Blumberg et al. (2008) a qualitative research design is the best in this master Thesis.

In this phase the literature study is also performed. In the literature there has been looked for answer for the research questions as stated in paragraph 2.1.2. The literature study has been done to get a proper overview of what is already known in the area of Big Data. The topics that were searched for are: what is big data, the value of Big Data, the risks of Big Data, the regulations of Big Data for banking and mobile banking/payments and customer loyalty. The information is gathered via search systems as ABI/Inform, congresses, websites and people from the hosting company and other companies. The information gathered contains articles, reports, books, websites and business documents. There was much information found but only the most suitable information is used in the literature review. The literature review findings can be found in chapter 3.

The information of the literature review was not complete enough to answer the research questions mentioned before. Therefore data collection was necessary to find the complete answers. To make this model information had to be gathered from experts and merchants. First an analysis has been made which type of survey fits best for this qualitative research. In this phase the data collection method have been chosen. Data collection can be derived in four different ways: asking, looking, participating or reading. In chapter 4 this phase is worked out. At the end of this phase there has been a specific data collection method chosen and everything needed to perform this method is described.

#### **2.2.4. Taking action**

Subsequently, the phase of implementing the planned follows. In this phase the intervention in the organization is made up that involves a certain change compared to the situation at this moment (Mendoza et al., 2006). In this phase all the different sources of information are combined to deliver a certain construct for the company.

In the action planning phase it has been decided with both the hosting company and the TU/e how the merchants and experts will be investigated. It has been decided that semi-structured interviews with mainly open questions was the best option. In chapter 4 there is an explanation why this was the best method of data collection for this Thesis. After this method was worked out in detail, the interviews were performed in this phase of the research. The interviews are used to analyze the opinion of the merchants about what they think about Big Data, if and how they would like to use it, which risk they see and which benefits they expect. The merchants that participated with these interviews differed in size, retail area and country. The experts differed in the field of interest. To get a proper view of the market there were experts from the hosting company, from payment organizations, from merchant associations and experts specialized in privacy to take the customer opinion in perspective.

The insights from the merchants and consumer have been used to develop a conceptual model. First it had to be decided what type of model will be designed; for example a framework, process model etc. A process model was found to be the best type of model for this Thesis. The quality of this model depends on the insights gained from the stakeholders and the quality of the information. Therefore the quality of the merchants and experts has been more important than the quantity. The conceptual model made describes how value from Big Data can be used to get more customer loyalty in the retail area.

After the designing of the model it will be compared to other models from the literature. The extra insights from literature will be added to the model to complete the model.

#### **2.2.5. Evaluating**

The second to last phase is called the evaluating phase. In the evaluation phase the intervention that is made before is evaluated and discussed (Mendoza et al., 2006). In this phase it is discussed whether the changes had the wanted effect and if the model is complete or should be validated.

Evaluating of the designed process model. In this phase the model is validated with a control group. With this group it is tested if the connections founded with the first set of interviews are correct. This is done for connections where verification is needed.

Adjust the model. Here the results of the validation are studied and if necessary the process model is adjusted. At the end of this phase the final process model is presented.

#### **2.2.6. Specify the learning**

This is the last phase of the process. In this phase the knowledge the researcher must specify the knowledge acquired based on the results of the evaluation (Mendoza et al., 2006).

Conclusions and recommendations; in this phase the conclusions are described. Afterwards limitations are discussed as well as recommendations for further research; because this is a master Thesis project the time and resources are limited. These limitations make that there are improvements and opportunities for further research which are discussed in this part.



### 3. Literature study

In this chapter a summary of the literature will be provided. In this literature review answers are searched for the research questions of paragraph 2.1.2. As mentioned in paragraph 2.1.2. to search for these answers the following literature questions were made up for the literature review. The answers to these literature questions should also provide the insights for making the process model and to answer the research questions. Literature is also used to create a context; which is the case with the first two questions.

- What is the definition of Big Data?
- What are the regulations of Big Data for the payment/retail sector?
- What is the value of Big Data in the context of payment/retail sector?
  - Which existing models are relevant?
  - What are the main risks of Big Data for the payment/retail sector?
- How can customer loyalty be gained?

The literature review will start with the definition Big Data. Secondly there will be an overview of privacy regulations about the retail and payment sector. When talking about Big Data privacy immediately springs to mind. Big Data consists of personal data most of the time and is therefore linked to privacy. A chapter will follow about the purpose of Big Data. What can companies do with Big Data to derive value? There will also be information about data mining. Big Data can bring value to companies, but these benefits are not without risks. Therefore two models from literature will be discussed and afterwards the risks of Big Data will be discussed. The last part will be about customer loyalty.

#### 3.1. What is the definition of Big Data?

The first topic that will be discussed is the definition of Big Data itself. The concept of Big Data does not have one clear definition but most articles agree that Big Data consist of 3 V's which are: volume, velocity and variety (Database Trends & Applications, 2012; Database Trends & Applications, 2013; Hopkins & Evelson, 2011; Lopez, 2012; Nasar & Bomers, 2012; Won, 2013). Figure 5 from Won (2013), appendix J1 shows the three V's and how they are connected. Volume is about scale of the data; so the huge amounts of data that are available and what companies should master to gain benefits. To quantify this volume: IBM (2013) claims that we create 2.5 quintillion bytes of data each day. This means that 90% of the data that is available in the world is created in the last two years. These volumes are usually larger than what companies can handle with their traditional systems. The volume of the data is growing exponentially (Database Trends & Applications, 2012). Variety is about the many different sources and forms of data. These could be normal traditional types of data such as transactional data as well as new sources such as Twitter and Facebook. These data sources are collected in many forms such as structured, semi structured and unstructured (Won, 2013). The last V, Velocity is about the streaming speed of the data. The data should be analyzed and processed real time. Davenport, Barth and Bean (2012) mention it is more about monitoring the flow of data, than a fixed supply of data. These three V's make the difference between normal data as it is analyzed and used for many years and Big Data. So the volume, variety and velocity are higher than ordinary data.

Snijders, Matzat and Reips (2012) give a definition of Big Data that it is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software. They agree that there is not a hard definition and they also include the volume as mentioned before. These data sets are available because of the rise of digital and mobile communication. These technologies have made the world more connected, networked and traceable (Raine & Wellman, 2012). According to IBM (2013) there is also an extra fourth V from Veracity which is the uncertainty of the data. This fourth V is about the reliability of information that is used to make decisions.

The growing amount of data also requires new skills to capture, manage and analyze this data (Database Trends and Applications, 2013). A so called data scientist will become more important to access benefits from Big Data (Database Trends and Applications, 2012; Database Trends and Applications, 2013; Davenport, Barth & Bean, 2012; Lazar, 2012). Data scientists have skills in analytics, IT, computer science, biology and/or network oriented social sciences. They can do programming and have mathematical and statistical skills. In addition they can also communicate effectively with decision makers. So they have other skills than the data analyst had in the past (Davenport et al., 2012). These data scientist can also make Big Data visual. According to LaValle, Lesser, Shockley, Hopkins & Kruschwitz (2011) analytics offer value and visualizing data will become increasingly valuable.

The definition of Big Data used in the Thesis is that it is a large amount of data containing a lot of variety and velocity (Database trends and applications (2012), Database trends and applications (2013), Hopkins & Evelson, Lopez (2012) and Nasar & Bomers (2012). Also the importance of the data scientist will be used, because the data scientist can offer value from all the data received which can be interesting for the dissertation (Davenport et al. (2012), Database trends and applications (2012), Database trends and applications (2013), Lavalle et al. (2011) and Lazar (2012).

### **3.2. What are the regulations of Big Data for the payment/retail sector?**

Although at this moment in the Netherlands there are not a lot of lawsuits about privacy scandals the reputation damage can be big (Semeijn, 2012). Companies like Google and Facebook will always been linked to privacy issues due to the mistakes they made in the past about this topic. To prevent this, it is important to have a close look at the laws and regulations concerning privacy. According to lawyer Peter Kits<sup>2</sup> (2012), information is free and everyone is allowed to possess them, transform them and use them in a variety of ways.

First there is the European privacy guideline. In this guideline the rule snare stated about how to deal with personal data; for example requirements about how to protect data (Guideline 95/46/EG, 1995). This guideline is used to determine the most important law for companies that operate in the Netherlands; this is the “Wet Bescherming Persoonsgegevens” (law to protect personal data). This law

---

<sup>2</sup> Kits, Peter (2012). Lawyer from Holland van Gijzen Advocaten en Notarissen. Gave his opinion in the article: Semeijn, B. (2012). Big Data en de privacyregels. *Computerworld Management special*. Retrieved from: [http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448\\_Computerworld\\_De\\_privacyregels\\_bij\\_Big\\_Data.pdf](http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448_Computerworld_De_privacyregels_bij_Big_Data.pdf)

allows companies to collect and transform personal data but demands that companies have clear goals why they want to do that and that these goals are reasonable. The law gives consumers some rights; they have the right to know what is happening with their personal data, have a look in their personal data, correct their personal data and object the use of the data (Wet Bescherming Persoonsgegevens, 2000). Further there are some industry specific guidelines which are based on this “Wet Bescherming Persoonsgegevens”, these guidelines are a more concrete form of the law (Semeijn, 2012).

According to lawyer Christiaan Alberdingk Thijm<sup>3</sup> (2012) a company should ask itself the following questions when working with Big Data: am I using personal data?; when I use personal data do I have to ask for permission?, this depends on the goal you want to use the data for; how will I protect the data carefully? Another discussion is about the difference between data and personal data. According to jurist Bert Willem Schermer<sup>4</sup> (2012) data is personal data when it can be derived back to a specific person. If this is not the case you will not come in conflict with the law “*Wet Bescherming Persoonsgegevens*”. When you want to use Big Data to give customer specific discounts you have to follow the law and so you have to ask for permission to use this data (Schermer<sup>4</sup>, 2012). To collect data from customers in the Netherlands you have to get their unequivocal permission (Semeijn, 2012). When you already have the data but want to use it for a different goal it is not always needed to ask permission again (Semeijn, 2012). This is not needed when there is “*verenigbaar gebruik*”(consistent use). According to Deloitte<sup>5</sup> (2013) there is still a large grey area in which you can act fast as long as you think it is morally justified. Waiting until the government has rules and guidelines made will put you behind. Research journalist Brenno de Winter<sup>6</sup> (2012) thinks it is better that the government makes rules and guidelines fast before industry takes any more advantage of the current situation.

### **3.3. What is the value of Big Data in the context of the payment/retail sector?**

Different mobile payment providers are active around the world. In appendix A the most important ones can be found and the differences between them are discussed. In this paragraph the value of Big Data for payment providers in combination with the retail sector will be discussed. CGI (2013) states that Big Data has many different sources for input. There is input from Operational Data, Consumer social data, external data sources etc. This data is transformed with the use of IT to deliver business value. According to CGI IT uses Big Storage and Big Analytics to deliver Big Insights and Big Returns to the business. Brown, Chui & Manyika (2011) state that data can be collected through different sources while

---

<sup>3</sup> Thijm, Christiaan Alberdingk (2012). Lawyer from SOLV advocaten. Gave his opinion in the article: Semeijn, B. (2012). Big Data en de privacyregels. *Computerworld Management special*. Retrieved from: [http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448\\_Computerworld\\_De\\_privacyregels\\_bij\\_Big\\_Data.pdf](http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448_Computerworld_De_privacyregels_bij_Big_Data.pdf)

<sup>4</sup> Schermer, B.W. (2012). Jurist from juridisch advies bureau Considerati Gave his opinion in the article: Semeijn, B. (2012). Big Data en de privacyregels. *Computerworld Management special*. Retrieved from: [http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448\\_Computerworld\\_De\\_privacyregels\\_bij\\_Big\\_Data.pdf](http://www.crossmediadashboard.nl/wp-content/uploads/2012/12/448_Computerworld_De_privacyregels_bij_Big_Data.pdf)

<sup>5</sup> Deloitte is a global active consultancy company which gave a presentation on Big Data at the TU/e. The information used for this literature review is based on a document given to the students which were present.

<sup>6</sup> De Winter, Brenno (2013). Dutch research journalist which gave a presentation at the Big Data congress of Industria 2013.

flexible infrastructure is needed to effectively integrate all the information gained. They stated that analytics will create sense in this information, to make money with it. According to CGI (2013) the main challenges are that the data is big, heterogeneous, difficult to understand, need to be accessed and is noisy. According to Brown (2012) and Kiron (2013) the difficulty lies mainly in the integration of data from different sources in which structured and unstructured data has to be combined. In appendix J2 the Big Data to Business flow model from CGI (2013) can be found.

According to De Vries (2013) these same steps (as mentioned by CGI (2013)) have to be taken to get value out of Big Data. She stated it more precise: “The solutions that should be implemented to create business value from investing in Big Data in the financial services industry are installing the hardware and software that Big Data requires, performing analytics and visualization techniques that are possible with Big Data, making a project team to create culture, methods and provide training, and attracting new employees that fit best with a Big Data environment” (De Vries, 2013).

### **3.3.1. Generate value**

To generate value De Vries (2013) stated that a combination between the predefined Big Data implementation together with for example customized deals, cross selling and fraud detection, will generate more revenue and less costs. According to Lopez (2012) using Big Data can also lead to operational, business and financial gains. Besides that it can lead to a quicker access to more relevant and cleaner data which drive insights and optimize decision making. De Vries (2013) stated that the most important aspect of Big Data is the velocity, because that represents to what extent an organization can filter the most relevant information out of the data to develop new processes, make better decisions and/or optimize processes. This data can also come from social media, but that is most of the time not concerned. According to Bean (2013) only 3% of the managers from large companies said that they cared about social media information. Only some managers said that they planned to do some minimal customer sentiment type of analysis, but it was definitely not at the top of mind in the larger organizations.

Big Data gives organizations opportunities that were not possible before because of the high expenses or the lacking technology (Lamont, 2013 and Lopez, 2012). According to Nasar & Bones (2012) Big Data is getting more and more attention in the financial services industry because in that industry consistent, accurate management of both customer data and financial information is crucial to be successful. The financial service sector also has some setbacks, because of the rules and regulations changing faster than the IT landscape (De Vries, 2013). DBTA (2012) and LaValle et al. (2011) state that the actual real benefit of using Big Data is that it gives an opportunity to businesses by enabling users to run analytics and with those analytics to determine and predict customer preferences, market shifts and product innovations. They think that these benefits are also useful in the financial services industry. Nasar & Bomers (2012) think that Big Data can also help organizations to analyze the current environment with many rules and help those organizations with the growing complexity of those regulatory requirements that are changing from time to time.

When an organization wants the discussed benefits, that organization will need a solid plan. An organization needs to know what their key business driver is, how to put the processes into place and

they need to ensure that the organization has the skills and organizational alignment to make it happen (Kiron, 2013). According to Kiron (2013) to be successful an organization should have the right people with the right skills on the right place and the right organizational structure to use Big Data. It is certainly not a straight forward simple strategy that can be easily implemented. De Vries (2013) summarizes the benefits from Big Data into three different parts. The first is improving the interaction with the ecosystem which will result in better targeted marketing, individualized service offerings and customer retention. Secondly, improving the business processes to understand the data in more detail and better predict the future activities. The third and last one is especially important in the financial service industry and is the risk mitigation to manage risk compliance.

### **3.3.2. Data mining**

Retrieving value from Big Data has a lot in common with data mining. According to Chung and Gray (1999) the objective of data mining is to identify valid, novel, potentially useful, and understandable correlation and patterns in existing data. Hui and Jha (1999) also mention that and add that it will help companies to make better decisions. Fayyad, Piatetsky-Shapiro, Smyth and Uthurusamy (1996) mention the use of algorithms in their definition: "Data mining is a step in the knowledge discovery in databases process and refers to algorithms that are applied to extract patterns from the data. The extracted information can be used to form a prediction or classification model, identify trends and association, refine an existing model or provide a summary of the database being mined. The main goal mentioned in literature is a process for organization to enhance their organization performance and gain a competitive advantage (Hormazi & Giles, 2004). According to Baicoianu & Dumitrescu (2010) data mining offers three major advantages:

- Providing information about business process, customers and market behaviors.
- Taking advantage from the data that could be available in operation data collections, data marts or data warehouses.
- Providing patterns of behavior, reflected in data that can drive the accumulation of business knowledge and the ability to foresee and shape future events.

In data mining there are seven different data mining operations which are: clustering/segmentation, visualization, predictive modeling, link analysis, deviation detection, dependency modeling and summarization (Hormazi & Gilles, 2004). According to Baicoianu & Dumitrescu (2010) there are only four types of relationships sought for with data mining: classes, clusters, associations and sequential patterns. To get these relationships there are five steps in data mining: extract, transform and load transaction data onto the data warehouse system, store and manage the data in a multidimensional database system, provide data access to business analysts and information technology professionals, analyze the data by application software and present the data in a useful format, such as a graph or table. Chen, Sain and Guo (2012) state that the data preparation is an important process before the data can be used to gain value.

Baicoianu & Dumitrescu (2010) state seven different tasks which should be followed whilst data mining: prediction, classification, detection of relations, modeling clustering, market basket analysis and deviation detection. Clustering is the most important task for achieving business value for online retail

according to Chen et al. (2012). Peacock (1998) and Hormazi & Gilles (2004) give four data mining purposes for marketing: customer acquisition, customer retention, customer abandonment and market basket analysis. These tasks can also be used for retail problems.

Data mining is still a very relevant topic in the area of Big Data. The difference between data mining and Big Data is that data mining is getting value out of a particular data set where Big Data is getting value out of different combined datasets. This could also be live data from for example social media. But the value delivering processes of data mining could certainly be used in the area of Big Data.

### **3.4. Which existing models are relevant?**

To assess the value of Big Data models can be very useful. A framework for example provides a proper overview of all the possible entities that influence organizational performance. Because the topic is relatively new a lot of research still has to be done in this area and there are not many models at this moment. Therefore this study will result in making a model about Big Data for the retail sector. There are a few models that need to be discussed and can lead to input for the to-build framework. The first one is the balanced scorecard framework of Kaplan and Norton (1987). This is an evaluation method for managers with complex goals. Another framework that can be used as input is the study of De Vries (2013); she made a framework about Big Data for the financial sector. After those two frameworks are discussed process models will be discussed. The CGI model is already discussed and will be used as a starting point for the to be designed model. It clearly shows the very generic steps that have to be performed in deriving value. The final model will add more detail to this generic model to make it more specific and make it more applicable for the retail sector. In paragraph 3.5 the last two models will be presented concerning the usage of IT and innovations.

The first framework that is of interest for this master Thesis is a generic framework that is called the Balanced Scorecard (appendix J3). It is designed by Kaplan and Norton in 1987 and considers four categories. The Balanced Scorecard is a translation of the strategic objectives of a company into concrete measurable parameters in each of the four categories. The four categories are financial, customer, learning and growth and internal business (Kaplan and Norton, 1987). The categories of the Balanced Scorecard are very general which makes the framework generic and applicable in many situations.

De Vries (2013) developed a framework with all the important steps from installing Big Data packages to the financial rewards. She first made a general framework which she later compared to existing frameworks as the Balanced Scorecard. The original framework of De Vries (2013) can be found in appendix J4. The adjusted model can be found in appendix J5. For the literature review the adjusted version is the most important and will be discussed in more detail. The process starts with installing Big Data packages, for example Hadoop clusters; which is the most commonly used package. Then there is the block learning in which the IT has to be synchronized with current processes. This phase also concerns attracting new employees and creating a project team and train them to deal with the business changes concerning Big Data. In this phase privacy is important. The next phase, internal business process, is about the real time information on the different divisions of the company. In this phase the compliance to rules and regulations is important. The next phase is about the customer. Here

the insights from the different internal business processes gained from Big Data are used to increase the positive activities such as sales and decrease the negative activities such as fraud. In this phase security is important. The last phase, financial IT, is about the actual increase in revenue and decrease in costs which will lead to an increase in the business value. De Vries (2013) could not accommodate all factors to the four perspectives of the Balanced Scorecard. The two factors that she could not fit into the Balanced Scorecard perspectives were the two technical aspects of the framework.

Both frameworks could be used as input for the model made in chapter 5. The framework of De Vries (2013) is especially useful in the financial industry. Two of the stakeholders that are relevant in this master Thesis are a bank and a payment provider. Both stakeholders are active in the financial sector so when a part of the later made model is about the financial sector this framework can be used. The balanced scorecard is more a management tool which can be used for more generic and basic input for the later to build model. It can also be used in a process model by companies as process optimization framework.

Process models could also be interesting in the retail area and the area of Big Data. According to Natarajan (2013) when a company makes a Big Data process model it should be designed step by step. It is very important that a company understands the source of the data, the speed and frequency of data refresh, data privacy and data security. Vuksic, Bach & Popovic (2013) emphasize the importance of business process management. BPM is used by many companies and is also used as an important tool for improving business performance. Making a process model on how to get value from Big Data in the retail sector could be interesting if that is what the market needed. BPM would be suitable for this according to Vuksic et al. (2013). Making a process model for this Thesis could be especially interesting because they Vuksic et al. (2013) mention that BPM initiatives are used when the goal is improving business performance; which will be the case in this Thesis. The model made in this Thesis could help retailers and other stakeholders to give an overview of all processes that exist in deriving value from Big Data; they then can look which ones could be improved. To improve these processes Dumas, La Rosa, Mendling & Reijers (2013) made a Business Process Life-cycle (appendix J6). The model of Dumas et al. has seven different stages. It starts with the identification of a process, which reveals a certain business problem. Process discovery is the next phase in which the current stage of the business process is documented. Process analysis is the phase that follows and is about discovering the problems with the as-is situation of the processes. Process redesign is the most important phase, in this phase the changes have to be mentioned to improve the business processes. Process implementation follows, in which the changes that are proposed will be implemented in the business processes. The last phase is the monitoring and controlling phase, in which the business processes are monitored and when necessary the cycle will start again. When a process model will be made for this Thesis the Business Process Life-cycle of Dumas et al. (2013) could be helpful as a tool for managers if they want to improve a business process.

Two models that can be used for new products or solutions will be discussed in paragraph 3.8. Big Data applications in retail will be new for the customers and employees and therefore it can be helpful to have a look in how new or improved solutions are adopted by consumers and employees.

### **3.4.1. Relevance for the payment and retail sector**

The insights provided by Big Data are also relevant in the retail and payment sector. The (mobile) payment providers can deliver these insights because they have a clear overview about which customer buys what type of product. With the collaboration between a payment provider and the retail sector also insights can be gained for example the amount of second purchases of new products. These second purchases are important to determine whether or not a product is a success. For stores not working with customer loyalty cards these insights were very hard to gain.

In the retail sector the insights provided by Big Data will be very important. The expert group customer Data Value Management is part of a research program doing research on how a customer shops in 2020. They state that company benefits as company size, price, quality and service will be less important in 2020, compared to acting fast and being flexible to meet the customer needs. According to them it is important that retailers gain the ability to collect data and to use this data to gain insights. Big Data can also be used to create relevant buying suggestions (van der Meij, 2014). Darren Vengroff's<sup>7</sup> creates annual 42 billion dollars of sales with this technique for customers as Walmart and Marks & Spencer. According to them it is very important to make good algorithms that keep evolving and learning. They test the gained relations on historic data and gain relations that you never would have thought of. According to research bureau IDC<sup>7</sup> 71% of the consumers<sup>8</sup> regularly considers personalized offers. On average this leads to an actual conversion rate of 30%.

## **3.5. What are the main risks of Big Data for the banking/retail sector?**

### **3.5.1. Privacy**

Privacy is a topic that is already discussed in paragraph 3.2. Besides the regulations about privacy it is also one of the main risks for a company which wants to invest in Big Data. Privacy issues concerning Big Data can lead to reputation damage; for example companies as Google and Facebook face problems and can possibly be banned from countries due to these privacy issues. According to The Lawyer (2013) companies are increasingly facing legal penalties because they fail to find or submit electronic information and breaches of privacy and security in the current legal environment. In this report Professor Scholtes mentioned that companies should organize documents according to a legally justified archiving plan and they should introduce a strict policy for retention and destruction. Buytendijk & Heiser (2013) state that although individuals are not without responsibility by offering their personal data for free, organizations should initiate an internal debate on the limitations of Big Data analytics and guidelines to avoid public embarrassment, mistrust and liability. "There is an equally subtle balance between improvements in customer service and business operations by, for example, accurate customer profiling based on a variety of data sources, including social media and mobile phone data, and knowing so much that customers experience a 'creep factor'" (Buytendijk & Heiser, 2013). Especially the avoiding of this 'creep factor' is important to get the adoption of innovation as stated in the Technology

---

<sup>7</sup> Information gained from the article: Groot, A. (2014). Personalisatie: de conversiekracht van data. *Emerce magazine* 128. Retrieved from [http://www.emerce.nl/nieuws/personalisatie-conversiekracht-data?utm\\_source=rss&utm\\_medium=rss&utm\\_campaign=personalisatie-conversiekracht-data](http://www.emerce.nl/nieuws/personalisatie-conversiekracht-data?utm_source=rss&utm_medium=rss&utm_campaign=personalisatie-conversiekracht-data)

<sup>8</sup> Research done in France, Germany and the UK.



Acceptance model of Davis (1989). Buytendijk & Heiser (2013) defined a few risks with regarding to the privacy.

- Anonymisation and data masking could be impossible. No useful database can be perfectly anonymous.
- Protecting people from themselves. Not everyone cares enough about their own privacy, allowing others to make use of information in unintended ways.
- It is easy to mistake patterns for reality.
- The data becomes reality itself. Using Big Data retailers can provide personalized offers which customers are likely to welcome due to the endless offerings they face. The acceptance for the offer refines the profile, leading to an even more targeted offer, leading to higher conversion rates again. Which results in analytics start driving customer behavior, rather than the other way around. This is commercially interesting, but ethically debatable.
- Do not worry about bad people; worry about the ignorant ones. Customer segmentation and profiling can easily lead to discrimination based on many different factors.

### **3.5.2. Data**

Another risk that companies face when dealing with Big Data is getting lost in the amount of data. Business Wire (2013) mentioned that companies spend too much time and effort looking at the data details and too little time in extracting or understanding the value of the data and relying on wrong or incorrect data. According to them companies have still some work to do before they can transit from the risks of Big Data to the opportunities. According to The Lawyer (2013) the storage of the data can lead to excessive costs. Toub, Director & Bernstein (2012) state that the amount of data is a big risk for companies because it can be very costly to maintain. Deloitte<sup>5</sup> (2013) states that companies should not focus on expensive traditional and closed database systems to store data. The Business Times (2013) is concerned about another problem of the data. They state that as more organizations open up and extend access to corporate information, their networks become vulnerable to data misuse and theft, increased demand for data results in security risks. Big Data can also be the solution to this problem: "An intelligence-driven security model based upon Big Data analytics can help organizations achieve pervasive and deeper analytics, faster investigations and, more importantly, a centralized view of critical information." Companies should learn how to optimize the Big Data benefits by integrating analytics into security operations, to cope with the new threats (The Business Times, 2013).

Data ownership is also important when dealing with Big Data. Big Data normally contains data from various sources. When this data is used data ownership is important. Brandl (2005) mentioned the importance of data ownership and that you have to understand the roles of systems to make sure you do not get involved in a battle about data ownership which can have serious consequences with image damage or fines. Gabisch & Milne (2014) also noticed the emerging question about who are the owners of the data and the serious consequences of data ownership conflicts. Working with one large data file consisting of more than one data source gives some challenge with inadequately defined data ownership agreements (Silvola, Jaaskelainen, Hanna Kropsu-Vehkaperä & Haapasalo, 2011). Silvola et al. (2011) emphasize the importance to check for data ownership and make good contract arrangements when working with data from other parties. They also mention the consequences when this is not done.

### **3.5.3. Innovation diffusion and technology acceptance models**

Two models that describe risks of IT and innovations will be discussed. The first one is the Innovation Diffusion Model of Rogers (1962). The Diffusion of Innovations theory of Rogers (1962) is about the adoption of innovations by consumers. According to Rogers (1962) there are four main elements which influence this adoption of a new product: the innovation, time, communication channels and a social system. He distinguished five categories of adopters; these are innovators, early adopters, early majority, late majority and laggards. Important is the time that an innovation needs to reach the critical mass. Rogers (1962) has defined five factors of innovations that influence a decision of an individual to either reject or adopt the innovation. These five factors are: relative advantage, compatibility, complexity, trialability and observability. For the Big Data usage the most important factors will be relative advantage, compatibility and complexity. The model can be found in appendix J7.

The second model is the Technology Acceptance Model of Davis (1989). This model is one of the first models that described the acceptance of technology by people and how this results in actual usage of a system. The model can be found in appendix J8. The model has changed slightly during the years resulting in for example the UTAUT (Unified Theory of Acceptance and Use of Technology) model of Venkatesh, Morris, Davis & Davis (2003). The model was originally designed by Davis to predict the usage of a system by employees but the model of Davis can also be used to predict the usage of Big Data application by customers. For this Thesis the original (1989) version of the model will be used. The Perceived Usefulness is the degree to which a person believes that using a particular system would enhance his or her job performance (Davis, 1989). In the case of Big Data this will be the degree to which a person believes that using an application (which usages Big Data) would provide benefits. Davis (1989) defined the Perceived Ease of Use as the degree to which a person believes that using a particular system would be free from effort.

From both models it can be obtained that the perceived usefulness and perceived ease of use are the most important when designing a new functionality as a Big Data application or implementation. These factors will result in actual adoption of the innovation. The speed of this adoption will determine the time it takes to reach the critical mass. After reaching the critical mass the innovation will “sell itself”. Therefore it is important that reaching this critical mass is not taking too much time. So there are risks for Big Data usage when the perceived usefulness or ease of use is not enough. This can be a result of lagging in the factors of Rogers (1962), where in this case relative advantage, compatibility and complexity are the most important.

## **3.6. How can customer loyalty be gained?**

This Thesis is about the business value of Big Data in the retail sector. In the retail sector customer loyalty is very important. Customer loyalty programs are coordinated, membership based marketing activities among pre-identified customers toward a sponsoring brand or firm, Lacey & Sneath (2006). Most of the time customer loyalty programs are data driven. For the process model it is expected that customer loyalty can play a crucial role in developing business value for merchants in this sector. Therefore also in literature there is looked at customer loyalty.

Customer loyalty is mostly derived via customer loyalty programs. Customer satisfaction is important when it comes to long term consumer behavior (Oliver, 1980). According to Stehlik (2008) customer loyalty programs help to both increase revenues and create optimal customer buying behavior. Retailers who do not have a customer loyalty program have, compared to retailers with a customer loyalty program, on average fewer sales and lower customer retention results. Developing a lifetime customer value should be the top need driving loyalty relating decisions (Stehlik, 2008). He also stated that the ROI (Return On Investment) measurement of customer loyalty programs is an ongoing requirement for success. The amount of data used is crucial for this ROI measurement. Sirdeshmukh, Singh & Sabol (2002) emphasize the importance of consumer trust for getting customer loyalty. According to them the process starts with consumer trust which can create value which in the end can create loyalty to a product or merchant, without trust it is impossible to get loyalty from customers.

### **3.6.1. Soft and hard benefits**

Customer loyalty is all about doing something back for your customers to make them loyal to your store as a merchant. Most of the time this is done with customer loyalty programs, for example a customer loyalty card. The benefits customers get for their loyalty can be divided into two types: tangible or hard benefits and intangible or soft benefits (Barlow, 2000, Lacey & Sneath, 2006 and Sirdeshmukh et al., 2002). Hard benefits can be for example discounts to products, specials deals, gifts etc. Soft benefits are more relationship oriented, for example customized communication or preferential treatment (Lacey & Sneath, 2006). According to Barlow (2000) hard benefits are sometimes avoided by marketers because of their expense. Loyalty marketing however has come to a point where soft benefits alone are no longer enough in most industries. As an attention getter and credibility builder rewards are more necessary than ever (Barlow, 2000). Kassing (2002) emphasizes that the soft benefits can also be very important for deriving customer loyalty, but of course depending of the type of industry. According to Kursunluoglu (2014) and Abu-ElSamen, Akroush, Al-Khawaldeh & AlShibly (2011) especially customer service is a soft benefit that can lead to more loyal customers.

Loyalty cannot be bought according to Barlow (2000). Rewards will quickly create a significant commitment on the part of the customers to the products or services from a merchant. To get true customer loyalty Barlow (2000) suggests an appropriate blend of compelling hard benefits and defining soft benefits designed to create a loyalty program which differentiate among customers. If you do so such a program can ensure that your best customers will stay with you for a long time (Barlow, 2000).

Customer loyalty programs are mostly data driven. According to Lacey & Sneath (2006) customers will not think it is a problem to share their personal information if customers can be convinced that customers can get benefits which they would otherwise not receive. Marketers can use the higher level of benefits given to the customer as a form of compensation to customers for sharing personal information. These benefits should be sufficient because otherwise there is a risk these customers may perceive that the firm is penalizing non-participating customers who wish to maintain their privacy (Introna and Pouloudi, 1999). Marketers essentially use the higher level of benefits that can be gained via customer loyalty programs to compensate customers for the sharing of their personal information (Schultz and Bailey, 2000).

## **3.7. Conclusion**

### **3.7.1. Regulations**

The government has already made some laws and regulations about the use of Big Data, most of it based on a European guideline (Richtlijn 95/46/EG (1995), Semeijn (2012) and Wet Bescherming Persoonsgegevens (2000)). Next to that there is still a large grey area in which companies have to operate if they want to use Big Data (Deloitte<sup>5</sup> (2013), De Winter<sup>6</sup> (2013), Kits<sup>2</sup> (2012), Schermer<sup>4</sup> (2012) and Semeijn (2012)). This could be an advantage as Deloitte<sup>5</sup> mentioned. To make use of Big Data the consumers should be informed and have to approve this via a so called “*wilsovereenkomst*”.

### **3.7.2. Value of Big Data**

Big Data give companies opportunities that were not possible before. In general it can be seen from literature that Big Data is the raw material that via analytics can be used for business purposes (CGI (2013), DBTA (2012) and De Vries (2013)). These insights that are gathered via analytics can lead to value (Brown et al. (2011), CGI (2013), DBTA (2012), De Vries (2013) and Lavallo et al. (2011)). Data mining is still important in deriving this value (Baicoianu & Dumitrescu (2010), Chen et al. (2012), Chung & Gray (1999), Fayyad et al. (1996), Hormazi & Gilles (2004), Hui & Jha (1999) and Peacock (1998)). There are a few ways to describe the process in which the one of CGI is the one that fits the best for business purposes. Here it can be seen where the Big Data is coming from and that it is used as the input for IT which translate it into values.

### **3.7.3. Main risks and challenges**

Big Data can be used to create value for companies, but is not without risks. First there is the privacy risk. Big Data is by consumers often defined as a bit creepy (Buytendijk & Heiser, 2013). This can result in reputation damage as with Google or Facebook. This real treat should be dealt with by organizing documents according to a legally justified archiving plan and introduction of a strict policy for retention and destruction (Scholtes in The Lawyer, 2013). The next risk of Big Data is the Data itself. This can be very large and costly (Deloitte<sup>5</sup>, 2013, The Lawyer, 2013 and Toub et al., 2012)), usage of the wrong data (Business Wire, 2013) and the large amount of partly shared data comes with security issues (The Business Times, 2013). The last risks are described by two models. The Technology Acceptance Model of Davis (1989) shows the importance of perceived usefulness and perceived ease of use. This can be achieved by the five factors of Rogers (1962), where the relative advantage, compatibility and complexity are the most important. The actual usage of products or services based on Big Data will be important to reach the critical mass. The higher the perceived usefulness and ease of use the higher the actual usage. When a higher percentage of customers actually use the system the critical mass will be reached earlier which is important for an innovative product or service.

### **3.7.4. Customer loyalty**

Customer loyalty is mostly derived via customer loyalty programs. Customer satisfaction is important when it comes to long term consumer behavior (Oliver, 1980). Therefore there is a good link with this Thesis because these loyalty programs are data driven and Big Data will give extra possibilities. Customer loyalty can be gained via hard and soft benefits (Barlow, 2000, Lacey & Sneath, 2006 and Sirdeshmukh et al., 2002)). According to literature these should be used both; the amounts depend on the type of industry and merchant.

### 3.7.5. Insights from literature

In the literature there were answers found for the questions that were searched for, but these answers are not complete. The majority is based on articles which did research outside the Netherlands or even outside Europe. Therefore the answers found are a good guideline for further research but cannot be used to answer the research questions. The useful insights will now be discussed.

Regarding the research questions the literature gave the following insights. According to literature consumers, merchants, government, payment providers and banks are the primary stakeholders for this Thesis. Customer loyalty can be an interesting usage area for Big Data in retail. According to literature customer loyalty can lead to business value and customer loyalty programs are mostly data driven so Big Data will probably have extra value. The connection between Big Data and customer loyalty is not yet researched extensively. Within the usage area of customer loyalty value can be gained via hard and soft benefits. The information from literature is not enough to build a complete process model at this moment. Also the existing frameworks are not sufficient in this area which gives reason to build a new model.

The insights from literature can be described by the process flow (figure 2). The processes in blue are processes from the stakeholder merchant. The process in green is from the stakeholder customer. Data gaining is the general starting point. This data will then be prepared and analyzed in the analyze data process. The analyzed data will lead to customer loyalty program updating which is a continue process where all the insights from the analysis are processed. This updated customer loyalty program will lead to an improvement in customer loyalty which will lead to an increase of the revenue because the more loyal the customers are the more revenue a merchant receives.

Privacy and data ownership are two subjects that influence this whole process and have to be taken into account at every step. The stakeholder government makes the regulations which are important for the whole process. With the data gaining data will be gained from different sources. When data from another source is used data ownership agreements must be made very carefully. Data ownership agreements are also important for the data from customers that merchants would like to use for their processes. It is also crucial to have people with the right skills when working with Big Data. Big Data is complex and requires data scientists.

Gain data can be divided in data gaining via merchants' data or data gaining via (mobile) payment data. Customer loyalty improvement can be gained via hard benefits and/or soft benefits where the combination of the two is found most effective. This data can be divided into data gained via (mobile) payment and data that the merchants themselves collect. This data lead to insights which can be used in customer loyalty programs. This can lead to customer loyalty via either soft and/or hard benefits. Increased customer loyalty can lead to an increase of revenue.

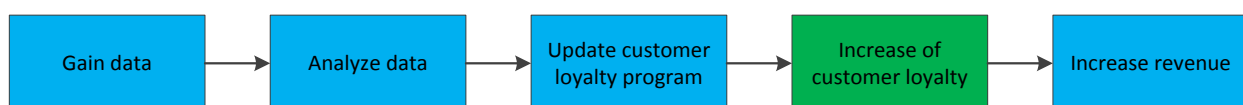


Figure 2. Input from literature for process model

## 4. Data collection design

The last chapter revealed that the answers to the research questions cannot be found solely in the literature. Therefore a practical research is needed to collect data. The information gathered during this data collection combined with the insights from the literature should give complete answers to the research questions. First the type of data collection method used in this Thesis is discussed. Subsequently the objectives of the data collection method will be discussed. At the end of this chapter all requirements to perform the data collection in chapter 5 are discussed.

### 4.1. Design of empirical research

As mentioned earlier first the type of data collection had to be chosen. This has been discussed with the supervisors from the TU/e and the hosting company. This research will be a qualitative research as mentioned in paragraph 2.2.3. According to van Aken et al. (2009) there are a few different qualitative techniques to collect data:

- The interview
- Focus groups
- Documentation
- Observation
- Verbal protocols
- Diaries

To extract the information from the merchants the best data collection method are the interview and looking at documents. The other techniques are more suitable when customers are involved instead of merchants (van Aken et al., (2009)). Because there are no documents to look at regarding this topic, using interviews is the only option. When choosing for interviews a choice had to be made between a structured, semi structured and unstructured interview and between open and closed questions. Semi structured interviews with open questions were chosen as method for the data collection from the merchants and experts. According to Blumberg et al. (2008) semi-structured interviews with open questions are used if you want to learn about the research problem from the person who is been interviewed and to know his or her opinion about this topic. According to van Aken et al. (2009) it leaves more space for respondents to give extra information. This can be really helpful in this study about Big Data. If the interview takes place with different companies it is better to have a guided semi-structured interview (Blumberg et al., 2009). Before the interviews would take place a list of questions was made. This questionnaire had many open questions and a few closed questions to determine some backgrounds and to scale certain answers. The semi-structured interview is also chosen because it leaves room for steering during the interviews. First the respondent will give their opinion but when a certain construct is missing in the answer it can specifically be targeted in additional questions. This is an advantage of doing interviews over online questionnaires, which is very useful for this particular master Thesis project.

## **4.2. Objectives**

To make the interview questions, objectives are made where the interviews should give answer for. These objectives should provide guidelines for building the model in chapter 5. The objectives are as follows:

1. The first objective of the interviews is to gain insights in how retailers think about the concept customer loyalty and the development of customer loyalty. Can Big Data management help companies to develop customer loyalty? Are the companies willing to use Big Data to create and maintain customer loyalty? How do the companies think about the role of third parties in this process?
2. The second objective of the interviews with merchants and experts is to gain insights in their vision on Big Data. Are they aware of the potential value of Big Data? Do they agree that Big Data implementation in their business is crucial to survive and serve customers' needs as stated by Groot (2014)? How do they want to get the value out of their data? Are they willing to invest in Big Data experts as a company or will they outsource it?
3. The third objective that the interviews should give insights in is the opinion about the data ownership. Are they willing to allow third party involvement to gain value out of the Big Data? What will the expectations be from merchants of these third parties?

## **4.3. Interviewed merchants and experts**

Merchants and experts will be interviewed for the data collection. The merchants are interviewed because the research question is about Big Data usage in the retail sector so merchants play a major role in this process. With interviewing experts it was possible to get more quality out of the limited number of interviews. Experts represent one or more (or could give information) of the stakeholders that were pre-defined by literature (merchant, consumer, government, payment provider or bank). With experts having an expertise in the interests of the different stakeholders it was tried to have the most complete set of respondents. Doing a questionnaire with customers was not possible in the limited time available but with having experts knowing the interests of consumers the stakeholder customer was represented in the data collection. There was one expert from the hosting company but most of the experts were from outside the company.

To get information from different stakeholders there will also be interviews with merchants. The merchants will answer more demographical questions than the experts. Because the questionnaire is already quite long and the people which will be interviewed will not have a lot of time, the demographics are kept to a minimum. Only those really helpful for this Thesis will be asked. The demographics that will be asked are: function, nationality of the person, company sector, nationality of the company, origin of the company, active in the Netherlands, active in Belgium, active outside the Netherlands and Belgium and size of the company

These demographics could influence the answers that the interviewed employees gave. The function is important because this topic is very sensitive and requires some background knowledge. When the opinion about this topic is from somebody high in the organization it has more value opposed to an idea from a regular employee. Because some cultural differences are expected between Dutch and Belgium

companies the nationality of the company is a demographic. Also the nationality of the interviewed person is asked because if the director of a Belgium company would come from the Netherlands it can be questioned if the cultural differences could be seen. Company sector can also be important in the end to explain some differences between companies; just as size of the company therefore these two demographics are asked.

#### **4.3.1. Planned execution of questionnaire**

As mentioned before interviews were used to get the answers for the questionnaire. Although this takes more time than sending it via e-mail or social media the answers will be from a higher quality. Doing interviews gives the opportunity to steer the questions, ensure that people get the question right and leaves room for insights that you have not thought of yourself. This will also make sure that all participants give answers to all questions and more easily say that they do not know a question instead of just guessing. Finally it also leads to good quality interviews because only people who are familiar with the area of Big Data were interviewed such as IT managers etc. Of course there are also some drawbacks from the interview approach. A single interview takes 30-45 minutes to handle and the managers do not all have time for this. This is also for some companies a sensitive topic which they do not want to talk about. To come in contact with managers who can answer the questionnaire is also very hard without any connections. Using the connections from the hosting company and a few experts there has been contact with around 40 people. The interviews took place partly face-to-face and partly via telephone.

The list of companies and experts should be made carefully, because of the limited number of companies and experts. In this list there should be merchants from different business areas because this could make differences. Also there will be companies from Belgium as well as the Netherlands. This is done because there could be cultural differences between Belgium and the Netherlands. It will give an extra dimension. Because of the complexity of Big Data it is also interesting to have a look at the difference between the opinions of small merchants versus large merchants. It could be that smaller merchants are more willing to work with a third party when it comes to the usage of Big Data than large merchants. Because large merchants will have the resources to gain the Big Data benefits themselves where smaller merchants will not be able to do that.

The companies should give a good representation of the Netherlands and Belgium. In the selection there should be companies which differ in size and area. To test for demographic differences it would be best to have the same type of company in the Netherlands as in Belgium. From the connections of the hosting company it is tried to contact the companies which will make the best representation. In the end this will have some consequences for the generalization because the sample group will be slightly biased. If the connections of the hosting company would not been used it would be very hard to have any participation of larger companies at all. With having companies which differ in retail area, size and demographic area it can be checked if there are differences in their vision on Big Data.

Larger companies often also have franchise stores. Franchising is a commercial way of collaboration between two entrepreneurs (De Nationale Franchisegids, 2011). The franchisor is responsible for the franchise formula and the franchisee is responsible for his own store and the day to day operations. For



the franchisor it is important that the franchisees keep their formula and do not leave it. Therefore it is important that the franchisor has a good relationship with its franchisees (Laurie, 2000). Muhleman (1996) emphasize that the communication should be good between the parties. It is also crucial for the franchisors that the franchisees gain enough added value from the franchise formula otherwise they might think that they do not get returns for their investments in the formula (Chiou, Hsieh & Yang, 2004). Van Eupen (2013) found in his research in franchise chains that these added value is gained mainly in the core business of the formula so for example in purchase power. Added value from non-core business is not mentioned as very important by many franchise chains; for example energy contracts. The non-core business activities are normally not arranged centrally but are done by the individual entrepreneur. Most formulas do not want to be a party in these non-core business activities although they could lead to added value for their franchisees.

There are a few different distinctions which can be made in franchising. The most important distinction is that between hard and soft franchising. With hard franchise there is less room for the entrepreneur to arrange things on their own. With soft franchising the franchisee has more freedom. For the questions that have to be answered with the interviews the soft franchise companies could be a problem because they will be likely not interested in non-core business activities. For hard franchise it will not be a problem.

There will be a few parameters concerning the segmentation:

- Company culture (Netherlands or Belgium)
- Company size (big, medium or small)
- Retail area (merchants should be from different retail areas to get a complete overview)

#### **4.4. Interview questions**

The interview questions are carefully made. The topic of Big Data is a very delicate and sensitive topic for the consumers as well as for the merchants. When the questions are asked in the wrong way the answers to these questions will be not satisfying and will not lead to the required insights for answering the research questions. To demonstrate how sensitive the topic should be handled, ING is a good recent example. They just stated out of the blue (for the outside world) that they will sell the customer transaction data to advertisement companies. This resulted in 30,4% of the ING customers thinking of leaving the bank. There was so much to do about the statement of ING that in the end they aborted their plans.<sup>9</sup>

To get the most information out of the interviews the interviews will be structured as a funnel starting wide and ending small and specific. Therefore the interview will start with a general open question followed by some closed questions. Although the answers to these questions will probably be pretty obvious it will help to get better answers in the later part of the interview. The amount of questions is as few as possible because the companies collaborating in the interviews will not have more than 30

---

<sup>9</sup> [http://www.emerce.nl/nieuws/dertig-procent-ingklanten-overweegt-overstap?utm\\_source=rss&utm\\_medium=rss&utm\\_campaign=dertig-procent-ingklanten-overweegt-overstap](http://www.emerce.nl/nieuws/dertig-procent-ingklanten-overweegt-overstap?utm_source=rss&utm_medium=rss&utm_campaign=dertig-procent-ingklanten-overweegt-overstap)

minutes time available. From previous interviews it is learned that around 8-10 open questions is the maximum that can be handled in on average 30 minutes.

The interviews were initially held to get information about the options companies prefer with Big Data usage. There has been chosen for the open option which is asking companies about which Big Data options they use without making a short list at the beginning. This has several reasons. First it was very hard to find real concrete examples, which makes it impossible to make an adequate short list. Next the interviewees have little time so discussing a whole list would consume too much time, especially since you would have to explain everything on the list in detail. Since this is an explorative study it is also nice to see what companies do at this moment and would like to do in the (near) future. Big Data applications are also highly customizable so it is better to make one especially for the needs of the companies. At last Big Data is relatively new so it could be that companies don't use Big Data at this moment and also do not know what specific option they would like to use; with this way of questioning they could then mention where they would like to use Big Data for resulting in usage areas. The questions for the interviews can be found in appendix B for the merchants and appendix C for the experts.

#### **4.4.1. Link between interview questions and objectives.**

Every question is asked for a reason. As mentioned before the questions are related to customer loyalty, Big Data or a combination of the two. Furthermore the questions will be related to the pre discussed research questions (see paragraph 2.1.).

The first four questions (first three for experts) are purely about customer loyalty which is not the main topic of this master Thesis but the information is needed for a thorough understanding of the background. These questions are asked to comfort the interviewee and to have a good start of the interview. When started with questions about Big Data the interviewee could be intimidated and give only very brief answers. The questions will also help in answering the second objective in how Big Data management can help companies to develop customer loyalty. A proper view about the customer loyalty programs and view on customer loyalty is needed for this objective.

The questions 5 and 9 till 13 (5 till 8 for experts) are about the combination of Big Data and customer loyalty. The answers to these questions will contribute to the second objective as well as the third objective. The second objective is how merchants could use Big Data management. The questions 10 till 13 contribute to the third objective which is giving insight in the merchants' opinion about data property.

The last block of questions, questions 6 till 8 (question 4 for experts), will contribute to objective 1 which is gaining insight in the merchants' vision on Big Data. They will also contribute to the second objective.

For the closed questions a 7 point Likert scale is used. According to Allen & Seaman (2007) Likert scales are a common rating format for surveys used to measure quality. The most used format is the five point Likert scale but the ends of the scale often are increased to create a seven-point scale. It is recommended to use a larger Likert scale when possible (Allen & Seaman, 2007). Therefore a 7 point scale ranging from 1 "strongly agree" or not applicable till 7 "strongly agree" or fully applicable is used.

## 5. Process model proposal

In this chapter the first version of the model is proposed. First as mentioned in chapter 2 it had to be decided which type of model is best. Insights from literature and from the interviews were used as input for this decision. A process model was chosen because it is expected to give the most help for the stakeholders' active in the retail sector. This was also confirmed by the literature. The five stakeholders can see all the processes of how value can be gained from Big Data in the retail sector. They can also see how and where they can improve them. Also application builders can clearly see where they can develop new applications and what these applications should be able to do. This model will be guidance in how value can be created from using Big Data. In paragraph 5.3. more information can be found about the choice for the process model. This chapter starts with a discussion about the interviews being held and about the demographics of the respondents. Afterwards the insights of the interviews are discussed. The last part of this chapter is about the conceptual process model. It will show the model and a discussion how it works and an example usage. The last part of this chapter consist of a comparison of the frameworks from the literature and how these frameworks can be useful for the conceptual model designed in this Thesis.

### 5.1. Interviews

To do the interviews high quality leads were used from the contacts of the hosting company representatives. This method was chosen because of the complexity of the research topic and the perceived level of business function from the representatives. From the 40 leads that where contacted 12 responded positively and wanted to participate in the interviews. The rejections were due to not having time. The topic of Big Data and customer loyalty was also a problem for some companies although confidentially was promised and results would not be traceable to individual companies. The final group are the merchants and experts who did not respond at all to the interview invite, although a reminder has been send. There were 7 merchants and 5 experts. Some interviews took place with two representatives; that was the case with one merchant and two experts. The interview questions where updated when necessary. This has led to one important update after the first interview. There was a note from the respondent that there should be a distinction made between using data and the efficiency of the data usage. Companies collect lots of data but at this moment only a small percentage is used to derive customer loyalty. Therefore there was an extra sub question made at question 4 by the merchant and at question 5 by the experts. The final version of the questionnaires can be found in the appendices. In appendix B the questions used for the interviews with merchants can be found, in appendix C the questions for the experts. Which merchants and which experts where interviewed can be seen beneath in table 2 and table 3. Also the function of the interviewed persons is stated.

#### 5.1.1. Characterization of respondents

As mentioned before in this chapter the characterization of the respondents are discussed. First two tables with the demographics are presented. The discussion of the demographics can be found in appendix D.

**Table 1. Demographics of the merchants**

Respondent	Retail area	Function	Nationality of the representative	Nationality of the company	Size	Active in NL	Active in B	Active outside NL or B?
1	Cinema	International CRM and Digital Marketing Director	Belgium	Belgium	Medium	No	Yes	Yes
2	Food	Online Marketing Director	Dutch	Dutch	Large	Yes	Yes	Yes
	Food	Business Analyst	Dutch					
3	Supermarket	Manager ICT	Dutch	Dutch	Medium	Yes	No	No
4	Web shop	Product Manager Mobile and BI	Belgium	Belgium	Large	Yes	Yes	Yes
5	Clothes	Manager IT	Dutch	Dutch	Large	Yes	Yes	Yes
6	Hyper Market	Division manager Accounting and Banking	Belgium	Belgium	Large	No	Yes	Yes
7	Energy	Manager payments	Dutch	Dutch	Large	Yes	Yes	Yes

**Table 2. Demographics of the experts**

Respondent	Expert company	Function	Nationality of the representative	Nationality of the company	Expert area
1	Payment association	Manager	Dutch	Dutch	Payment
2	Hotel and catering association	Manager projecten, innovatie en kennis	Dutch	Dutch	Hotel and catering industry
3	Consumer association	Campaign leader financial sector	Dutch	Dutch	Consumers
	Consumer association	Campaign leader digital	Dutch	Dutch	Consumers
4	Payment provider	Product Management	*	*	(Mobile) payment
5	Law firm	Lawyer (Privacy)	Dutch	Dutch	Law / Privacy
	Law firm	Lawyer (Payments)	Dutch	Dutch	Law / Payment industry

\*Per request of the interviewee, the nationality of the interviewee and company is omitted.

## 5.2. Results of the interview

The main insights gained from the open interviews with the different merchants and experts are shown later in this chapter. Before the interviews the respondents were asked if they had problems with the interview being recorded. Everyone agreed with recording the interview. After the interviews the interviews were written down. When all interviews were done per question the main points were listed and it was noted per respondent which points he/she had mentioned. Because open questions were used the answers sometimes differs a lot between the respondents, but a lot of topics were mentioned by more than one respondent. So when a topic is not mentioned by a respondent does not have to mean that he or she would not agree with it, but simply means that he or she did not mention it in the

interview. The anonymized results will be described beneath. Within the group no differences were noticed between Belgium and Dutch companies, nor between the business areas and between medium and large companies. This could also be due to the small amount of respondents. The differences between the respondents are good to get a proper view from different angles but too small to really detect differences. Although the questionnaires of the merchants and experts look very similar, they have not enough similarities to discuss them together. Therefore first the results of the interviews with the 7 merchants will be discussed and secondly the results of the interviews with experts. To get answers from the respondents confidentially was guaranteed so only the anonymized results can be discussed and not the results per respondents. The generic way of describing the results is due to these anonymized answers. Without anonymizing the results, the interviewees would not cooperate in this research. The questions will be discussed in three blocks. The first block will be the discussion of the customer loyalty questions, the second one is about the Big Data questions and the last block will be about Big Data usage in the retail sector. For the closed questions the average, the variance and the standard deviation were calculated. The statistical value of the variance and standard deviation are not very high due to the low number of respondents for both groups but it was the best to demonstrate the spreading in the data. The variance ( $\sigma^2$ ) is calculated as follows:  $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$ . With  $\mu$  is the average, N is the number of respondents and  $x_i$  is the respondent answer. The standard deviation ( $\sigma$ ) is calculated as follows  $\sigma = \sqrt{\sigma^2}$  (Field, 2009). In appendix E the detailed results of the interviews with the merchants can be found. In appendix F the detailed results of the interviews with the experts can be found. Here only the main and useful insights will be discussed and for the closed questions the average and standard deviation. The closed questions had a 7 point Likert scale ranging from 1, “strongly agree” or not applicable till 7, “strongly agree” or fully applicable. The reliability and validity of the results will be discussed after the validation step.

### **5.2.1. Results of the interviews with merchants**

The interviews started with four questions about customer loyalty. Customer loyalty was found very important by 100% of the respondents. They gave a 6,7 on average with a standard deviation of 1,08. They all used a combination of soft and hard benefits to derive customer loyalty. The importance of a good balance between soft and hard benefits was also emphasized by 100% of the merchants who participated in the research. The importance of sustainable customer loyalty systems was mentioned by 29% of the respondents. All 7 merchants (100%) that participated in the research did have a customer loyalty program but there were some differences in the sophistication of these systems. A digital customer loyalty system was used by 86% (6) merchants of which 50% (3) also had a customer loyalty card. The importance of the customer loyalty program differed a lot between the respondents, on average it was a 4,7 with a standard deviation of 1,70. The amount of customer loyalty that is provided to the customers was on average a 4,3 with a standard deviation of 1,50. So the answers to this question differed a lot between the respondents. For their customer loyalty programs all respondents used data the extend however differed with a 4,7 on average and a standard deviation of 1,51.

In the second category questions were asked about Big Data. What the vision was about Big Data, the definition that they used, the main benefits and the main risks of Big Data. The vision on Big Data of the merchants was quite clear. All 7 (100%) said that Big Data is just in the beginning phase at this moment

and will become more important in the future. Big Data is not fully regulated at this moment which could cause problems; this was mentioned by 29% (2) of the merchants. Big Data was defined as very large amounts of data by 86% (6) of the merchant. The usage of unstructured data sources was mentioned by 29% (2). The usage of analytics to get information out of the data was mentioned by 71% (5). There was 1 merchant 14% who mentioned that Big Data was about fully using the information of the customer and give added value for the customer. That the data should be from more than one resource was mentioned by 29% (2). So apart from that Big Data is about large volumes there was no broadly mentioned definition. The three V's model from literature was also not mentioned by anyone. The main benefit of Big Data for merchants was better insights in different areas for example customer behavior, 100% (7). It could be used for better communication was mentioned by 86% (6). That it could be used for more efficiency was mentioned by 57% (4). Two merchants 29% mentioned that it could be used for higher quality data. Privacy was the biggest risk for the merchants and was mentioned by all of them (100%). The incorrectness of the data and that there could be false conclusions based on this wrong data was mentioned by 43% (3) of the merchants. Security problems were identified as a risk by also 43% (3) of the merchants. Two merchants 29% mentioned misuse of the data as risk. One merchant 14% mentioned image damage as risk. So the risks that merchants mentioned differed quite a lot between merchants but privacy was their biggest concern. The knowledge about Big Data differed between the respondents; on average they answered a 5,0 but the standard deviation was 1,46 so there was quite some variance in the answers. On average the respondents answered a 5,8 that they can perform Big Data Analytics in-house but the variance in the answers was large with a standard deviation of 1,95.

The last category was the usage of Big Data in the retail sector. The first question of this category was about how Big Data can be used to develop more customer loyalty. The 7 merchants listed a number of options here. All 7 (100%) mentioned that Big Data can be used to make the offers more personalized and gain more customer loyalty by doing that. Target the right customer was mentioned by 86% (6). Making better offers was mentioned by 71% (5). The last usage of Big Data that was mentioned was that it could help to make a better segmentation by 2 merchants (29%). Because some overlap existed with previous questions not everything was listed again. Many merchant referred back to what was already listed earlier in the previous category of questions. All merchants (100%) mentioned that at this time it was too early to assess the value of Big Data. The merchant were not very positive about the idea of an external party to provide data if this could help them improve customer loyalty. On average the answer was a 2,7 with a lot of variance (standard deviation of 1,52). If it is guaranteed that this data will only be used by the third party for their consumers the average answer became a 3,7. This answer came with a lot of variance resulting in a standard deviation of 2,24. The circumstances under which merchants would like to share the data with a third party ranged from absolutely not by 43% (3) to if it gives added value that we could not get ourselves by 57% (4). What the respondents expect from a third party to do and not to do was nothing because we will not do that by 1 respondent (14%). The rest (86%) mentioned that there should be a contract in which is stated that the third party only does what the merchant is saying, only use the data for the actual case, ensure safety, handle the data with care and look at the privacy of the customers. Three merchants 43% mentioned that the data should be

destroyed after usage. The merchants did not have further ideas of concerns which were not discussed during the interview.

### **5.2.2. Results of the interviews with the experts**

The interviews with the experts also started with a few questions about customer loyalty. Starting with a closed question about the view of the experts on the current provided current amount of customer loyalty to customer in general, by merchants. The experts answered on average that it was average with a 4. The variance in the answers was low with a standard deviation of 1,10. According to the experts there is still some room for improvements for merchants. The changes that they mentioned were the following. Merchant should use more of the data that they collect and loyalty is more than just discount are two things that 80% (4) of the experts mentioned in their answer. Extra service should be more important was mentioned by 60% (3) experts. Merchants should be more open what they do with the data, merchants should try to find new customers and also hard benefits should be important was mentioned by 40% (2) experts. One expert (20%) mentioned that loyalty should be more demand oriented and there should be a shared database for moderate and little companies in the Netherlands. Another expert (20%) mentioned that merchant should utilize the existing channels more. So there were quite some changes that merchants should consider regarding their customer loyalty programs according to the experts that were interviewed. The experts had different opinions about the evolvement of customer loyalty resulting in many topics that are only mentioned by one or two experts. They all mentioned (100%) that data will become even more important for customer loyalty in the future. That customer loyalty will be more important in the future was mentioned by 60% (3) experts. That soft benefits should become more important, that there will be new ways of loyalty, that there should be more attention for the customer and that loyalty programs will move towards phones was mentioned by 60% (3). The experts were also asked with closed questions to scale the data usage of merchants to derive customer loyalty and how efficient this data using is done by the merchants to gain customer loyalty. To the first question they answered on average a 4,8 meaning that more than half of the merchants is using data for their customer loyalty program according to the insights of these experts. The variance was quite high with a standard deviation of 1,47. The efficiency is a bit lower (this question was only answered by 4 experts). On average the experts answered a 2 which means that the merchants are not very efficient in using data to get customer loyalty. The variance was low with a standard deviation of 0,71.

The second category where questions were asked about was Big Data just like with the merchant interviews. What the vision was about Big Data, the definition that was used, the main benefits and the main risks. All experts (100%) mentioned that at this moment Big Data usage by merchants is very limited and it will be more in the future. That a lot of data is already collected, that it will become more open source and that the resistance will drop was mentioned by one expert (20%) each. The definition of the Big Data was collection of all kinds of data, analyze the data and the data could be unstructured according to 80% (4) of the experts. It should consider potential data, so data that is not used at this moment was mentioned by 40% (2) of the experts. That Big Data includes taking action after the analysis was mentioned by one (20%) expert. The benefits of Big Data could be that offers can be more personalized and that there will come more knowledge about customers according to all (100%) experts.

4 experts (80%) mentioned that there will come more knowledge about where and whom of customers, for example via Wi-Fi tracking. One expert (20%) mentioned that a benefit could be bringing back trust in the merchants and that merchants should no longer be place restricted in their operations. The experts also mentioned some risks for merchants. Privacy concerns are mentioned by 80% (4) also mentioned as Big Brother issues. Drawing the wrong conclusions was mentioned by 40% (2). By only one expert (20%) was mentioned that customer contact can become less personal and possible reputation damage when using Big Data.

The last category of questions was again about Big Data usage in the retail sector. The first question was how Big Data can help merchants to develop more customer loyalty. All experts (100%) mentioned the same here: Big Data can help merchant to develop more customer loyalty by making it easier to differentiate between customers and for example derive customer segments, by knowing the needs of such a customer segment and make more personalized offers. The experts were also asked if they would advise the merchants to outsource Big Data analytics or should the merchants perform Big Data analytics themselves. The experts had quite different opinion on the answer of this question. Three of them (60%) mentioned that this is depending on the size of the merchant. The following statements were covered by 2 experts per statement: perform the analytics yourself, buy help from experts but do it yourself, perform it in-house, do it not in-house and it is not relevant from a law perspective because you will remain responsible for the data. One expert (20%) mentioned that it could maybe be done via cooperation. The last question was if there were any other ideas or concerns which were not covered and could be important. Two experts (40%) emphasized that personal data should be handled with care. One expert (20%) emphasized again the importance of the combination of hard and soft benefits. One last expert mentioned that data from for example a payment provider could be used cross industry for benchmarking purposes.

### **5.3. Proposed conceptual process model**

The result of this study is the development of a process model describing how business value can be derived from Big Data in the retail sector. To get to this model first the basis outlay is made up with the input from literature. The model was then further extended through the insights of the interviews with different merchants and experts. These interviews were also used to verify whether the results from literature are applicable in the specific domain of this Thesis. The experts were mainly from outside the company and were able to express the interest of the five different stakeholders. The insights that they gave helped to finish the model. In chapter 6 the parts of the model which are not broadly supported are validated.

The building of a process model is chosen because it best fits the needs of the five stakeholders involved in this research, which are discussed in paragraph 3.7.5. It gives a clear overview of all the processes that are important in deriving value from Big Data for the retail sector and sows how customer loyalty is involved in this process. With the development of a process model also the linkage between the different processes is clear to the involved stakeholders of this Thesis. It is also in line with the information the interviewees gave. Literature also mentioned that BPM can be used to gain business



value (Vuksic et al., 2013 and Dumas et al., 2013). The model describes how and where value can be gained with Big Data in the retail sector.

Dumas et al. (2013) have made process model quality assurance requirements; these were used to make the process model. They stated that quality can be divided into syntactic quality (conform the rules of the notation), semantic quality (making true statements about the world) and pragmatic quality (usable by the stakeholders). With the making of this process model these quality requirements were kept in mind. For the model of Dumas et al. (2013) describing these quality assurance requirements see appendix J9. Dumas et al. (2013) also defined a set of guidelines that are usable to make a good process model. These guidelines are called the Seven Process Modeling Guidelines (7PMG). This set of guidelines was developed as a summary of the insights of available research. When following these seven guidelines it is prevented that a model becomes too complex to be properly interpreted. The seven guidelines are as follows:

- Use as few elements in the model as possible
- Minimize the routing paths per element
- Use one start and one end event
- Model as structured as possible
- Avoid OR-gateways
- Use verb-object activity labels
- Decompose a model with more than 30 elements

The input from the literature and the input from the data collection were used to make the following concept process model (see figure 3). In this process model there are the five pre-determined stakeholders. These stakeholders are: the customer, the merchant, the government (making the law and regulations) and the payment provider which can be split into mobile payment providers and banks. In the model the different stakeholders have their own color. The processes of the customer are green, the processes of the merchant are blue and the processes of the payment providers and banks are orange. Because payment providers and banks can perform the same tasks in this process model there is no distinction made in the process model between these two stakeholders. The last stakeholder, the government is only present in setting ground rules via regulation and will be discussed in a separate paragraph together with for example data ownership because this influence all processes in the model. When in the explanation of the process model is referred to the interviews the results of the expert and the merchant are meant. When only the merchants or experts mentioned it, it is emphasized.

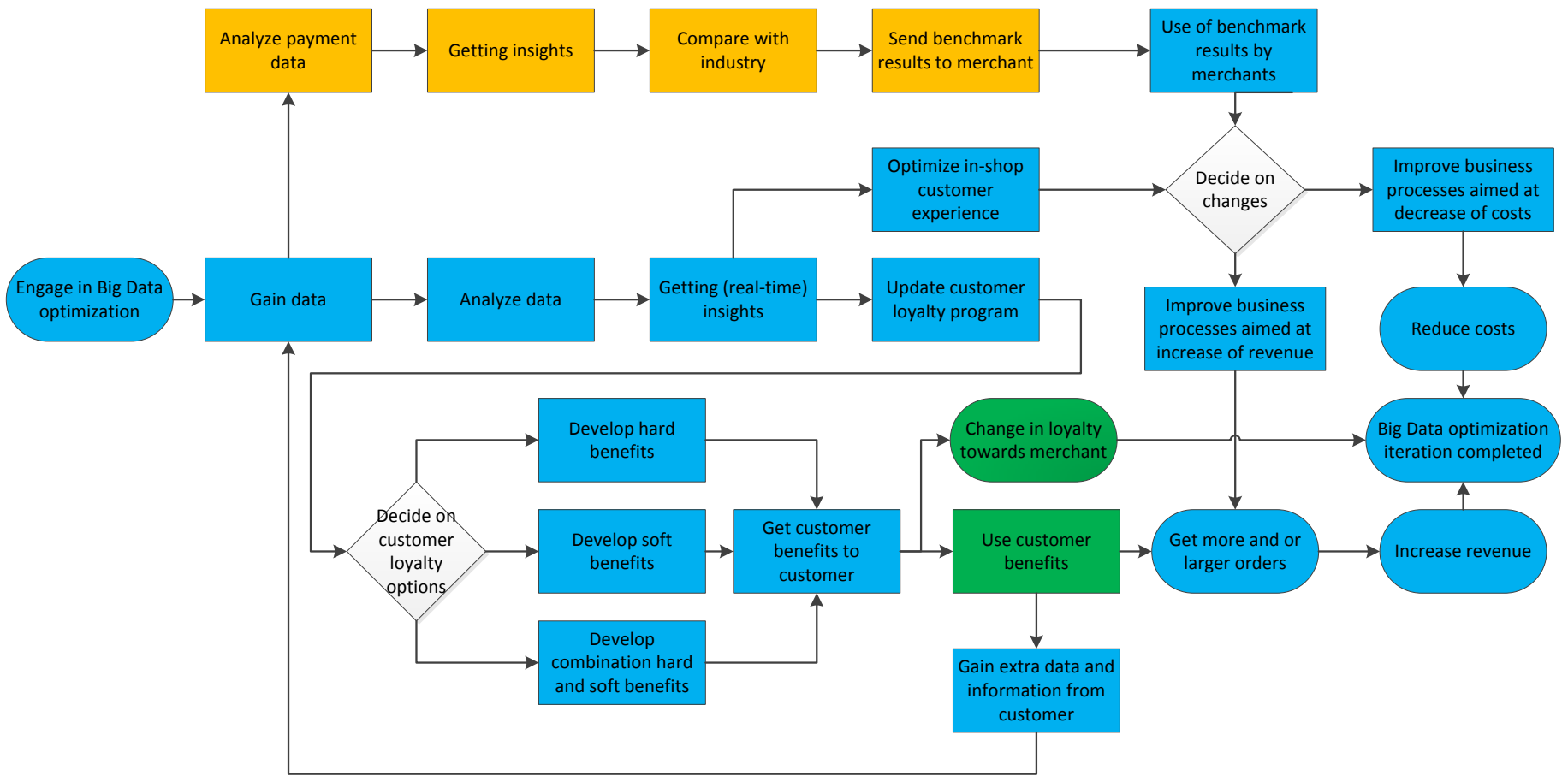


Figure 3. Proposed conceptual process model

### **5.3.1. Explanation of the proposed conceptual process model**

The process starts when a merchant has the need to engage in Big Data optimization. The merchant wants to use Big Data to improve their business. Possibly a goal is defined in this start stage, but with Big Data usage that is not always the case. The first process is the gaining of data by merchants. This data can be gained in many ways according to the literature and interviews. This data could be gained via loyalty systems, payment data, merchant data and external resources. The external resources could be for example weather which influences buying behavior of customers. The gain data process in the model concerns data gaining from different resources as is normal with Big Data.

The next step for the merchant is the process: analyze the data. This step is the sequel of the gain data process; this is described in literature and also mentioned in almost every interview. This analysis starts with preparation of data as mentioned in the literature and in interviews. With the usage of software the analysis can be performed in-house (with or without bought in help) or not in-house. Performing this in-house has the advantage for the merchant that he stays in control and knows everything that is going on. Outsourcing could be cheaper for especially mid-sized to smaller merchants according to the results of the interviews. The merchants should make sure that when they do this they make a proper contract with the third party so they remain responsible. When merchant would like to work with Big Data new software such as Hadoop is needed which needs extra investments.

After the analysis the merchant gets insights from the analysis. The merchant uses the analysis to derive insights from the data. In this phase the merchant transforms the output of the analysis to insights that can be used in the business. Depending on the insights they can be used for optimizing in-shop customer experience or the customer loyalty program. It can also be that the insights are aimed at one of the usage areas and that there will be specifically searched for, then a certain goal was already defined in the start stage. Big Data analysis is often done without a certain goal and afterwards it is determined what the insights can be used for; therefore in this model the usage areas are put after the insights. Insights can be useful for one or both usage areas. These insights could be real time when the proper software for this real time analyzing and giving insights is available at the merchant. This is also broadly supported by the interviews. To get good insights from Big Data analysis specialized people are needed which was described in the literature, such as a data scientist.

In the interviews two things were mentioned by a lot of the respondents about what can be done with the insights. They can be used to optimize the in-shop customer experience by for example change the placement of products etc.; or they can be used to update the customer loyalty program of the merchant. This last one is also broadly supported by literature; the first one has little support in literature. The optimizing of in-shop customer experience can be done offline/online or a combination of the two by increasing the multilevel channel management. Big Data can be used to make customer loyalty or online in-shop customer experience more personal.

After the in-shop customer experience optimizing process has took place a decision has to be made on the in-store changes. Depending on the choice made here either the business processes will improve with regard to cost reduction or with regard to the increase of revenue. These improvement processes will, according to the interviews, lead to either a reduction of costs for the merchant by targeting the

right customers or an increase in revenues because the merchant receives larger orders or more visits from customers. It depends on the costs of the improvements and resources needed to get the insights whether the profit will increase. Both “improve business processes” can use an existing model to actually improve the processes; more about that in paragraph 5.3.3. For the conceptual process model it is chosen to not over specify this step in the model and to explain the step more in paragraph 5.3.3. The process ends with the final end stage Big Data optimization iteration completed.

From literature it was learned that the update of the customer loyalty can lead to a change in customer loyalty. The results of the interviews show that first the merchant should decide on customer loyalty options. This decision point has three options: to develop hard benefits, to develop soft benefits or to develop both. Literature and the respondents of the interviews stressed the importance of the soft benefits. According to them a combination of both soft and hard benefits is the best option. The balance is retailer dependent. Afterwards these loyalty options should be brought to the customer; which can be done via phone, loyalty card or (digital) mail.

When the customer receives the hard and or soft benefits they can be used, which will result in a change in loyalty toward the merchant; this is an end stage of the model. This can increase or decrease depending on the benefits that are given and how the benefits are perceived. For example personalized offers is something that was very popular looking at the results of the interviews but when people do not appreciate that and still get a personalized offer the customer loyalty towards that merchant could decrease. Big Data could prevent this by knowing whether certain customers like or do not like specific benefits. For the rest of the process model it is assumed that the merchant are so sophisticated that the benefits they send to customers are appreciated and will result in a positive change in loyalty toward them. The customer will use the benefits and spend them in the store.

From the interviews it can be seen that more loyal customers will spent more visits and or spent more each visit at a merchant which will result in an increase of revenue. Also the usage of these benefits and the extra orders or amount per order will result in more data for the merchant. This explains the arrow back to gain data and making it a continuous process. The more and/or larger orders will result in the final stage of increased revenue. Again it depends on the costs made to perform Big Data analysis whether or not the profit is increased. The process ends with the final end stage Big Data optimization iteration completed.

Not mentioned in literature and only mentioned by one expert: the data that payment providers have can be also used to get insights. Payment providers can analyze this data and make a comparison with the industry. Although the data of company A cannot directly be compared with company B, a comparison between company A and the industry standards could be made. These benchmark results can then be send back to the merchants. The merchant use the benchmark results. They make a decision on what they would like to have changed and start improving processes for either cost reduction or increasing revenues. Resulting in the final stages of reduce cost of increase revenue via more and/or larger orders. These also result in the last stage as mentioned before for other usage areas.

The entire model has to deal with privacy regulations which are important in all processes concerning the data of customers. The stakeholder is the government which continuously adjusts and tightens these regulations. Thus, for a company which is active in this domain, this is something which is very important. This is mentioned in literature and all interviews. Also the literature and a few interviews mentioned the importance of data ownership. When a merchant wants to outsource the analytics or wants to work with a third party they should closely look at the agreements because the merchant is always the data owner and therefore responsible when something happens.

This Thesis is about Big Data usage in retail and how it can help for customer loyalty. In the interviews it was mentioned by the respondents that Big Data does not change the essence of the processes. The processes for retailers to derive for example customer loyalty are more or less the same as they were with the normal data models from the past. The input, output and speed of these processes will be the major difference between the usage of normal data and Big Data. Data gaining as mentioned previously will not be only the data of the merchant itself but more data sources will be added in the model. The analysis of the data will become harder and therefore new software solutions will be needed. At this moment the merchants were still in an early stage trying to determine what the new software should be capable of and what they want to use. The insights that they gain out of the analysis will become more specific and faster even real-time. In the interviews it was for example mentioned that because new types of data are used for input. For example the weather could be input for the analysis leading to specific weather related offers for customers. Also the decision of customer loyalty options will change when Big Data is used by merchants. The options will become more personal instead of sending the same offer to every customer the offers can be more personalized as mentioned in a lot of interviews. For the optimizing of in-shop customer experience Wi-Fi tracking can be used to analyze customer behavior. So the essence of the process will be mainly the same as it was but the requirements for the processes such as type of software needed, input, output and throughput type will severely change with the usage of Big Data by merchants in retail.

### **5.3.2. Example use of the proposed conceptual process model.**

The model gives an overview of the processes that merchants should go through in to gain value out of Big Data. The model starts with gaining data; this can be from different resources. Merchants can think of their own data, customer data, payment data but also data from databases outside the company such as weather or social media data. This data should then be prepared and analyzed with software. The merchants should invest in new types of software if they want to use Big Data because most merchants are still using traditional software at this moment. The investment in new software is something that merchants should look at the coming years because there will be more and more software developed for this branch. Getting insights is the next process in the model. Big Data will make it possible to get more and more real time information. An example usage for a merchant will be coupling their history selling data and history weather data (external source) which via analysis leads to insights in what type of weather leads to what type of selling. This can then be used to update the customer loyalty program. In this case by coupling the insights to the customers. Given a certain weather condition, a customer can display a certain preference for a product. Big Data can help to identify such preferences after which the merchant can use this information to provide personal advertisement. This will start the process of

deciding on customer loyalty options. This involves the decision of using hard and/or soft benefits to trigger the customer. To trigger the customer to buy a product now a rebate can be send or a simple notification could do be helpful. Big Data can help merchants here in choosing the most effective loyalty option. This could lead to different loyalty options for different customers. Some customers may only be interested when a discount is given while for others the notification of a new collection of clothing will bring them to the store. Data collection will increase and with that data the most effective loyalty options per product or per customer could be determined. This could also help in getting the customer benefits to the customer. There are many ways to do this, for example with a coupon, via mail or via a loyalty card. Big Data can again help in deciding on the most effective option here. It will also help in making it more real-time. For example, sending a digital weather related discount coupon directly to the customer via e-mail. The customer can then save this coupon in his mobile phone and use it directly. This can make loyalty programs more accurate and respond faster to changing factors as weather or hypes. The customer will get more useful loyalty related benefits and will increase his loyalty towards stores resulting in more and or larger orders from the main part of customers. This will lead to an increase of revenue for the stores, but maybe even just as important gain more data and information from the customer. The extra data and information will make the process even more personal and accurate resulting in even more loyal customers. The personal offer is something that customers have to get used to at this moment but as mentioned in the interviews when customers can get discount they will lower their privacy related concerns.

Optimizing in-shop customer experience is also something that can be done with the insights from the data gathering. For example via Wi-Fi tracking customers can be tracked in stores giving an overview in how customers move through stores. This information can be used to relocate the shop items and increase revenue. Web shops can also be adjusted to the personal characteristics of users. The basic outlay can be personalized serving the unique customer better. Knowing customer behavior and know how to predict them better with Big Data can also lead to a reduction of cost by for example better be able to adjust the amount of employees.

The benchmark results that can be gained via an external service provider because data from the whole industry is needed can be used by merchant to compare their data with industry standards. In these results areas can be found that need improvement. The merchant can then think of solutions how to improve those areas. It can give feedback to the performance of merchants. For smaller and middle size merchants this could be very interesting, but also for larger merchants this can be helpful although they might have the resources to make their own benchmark test. Benchmark results could for example note that the merchant sells fewer products than its competitors in the evening. This could be a sign to close his shop earlier or to make people more aware of the opening hours in evenings. For merchants this could lead to either a reduction in costs or an increase in revenue.

### **5.3.3. Comparison of the different models**

In this chapter the model will be analyzed in relationship to the six models that are mentioned before in the literature review. Criteria for the comparison will be discussed first. Subsequently a table will be presented in which all six models are compared with these criteria. Afterwards the relationship with the conceptual model will be discussed for the models that were useful for the conceptual model.

To compare the different models and discuss their usefulness for the conceptual model criteria were designed. The first criterion is if the models are designed for Big Data purposes. This Thesis is about Big Data usage in the retail sector so therefore this is the first criterion for comparison. The second criterion is whether the models are about the retail sector. The models could be not specifically made for Big Data usage and or the retail sector but still are useful; the third criterion is whether the models can be useful in this context. For the models that can be useful there will be a discussion later how these can be implemented or used for the conceptual model. The last criterion is the main limitations for the models which are not handled in the first three criteria of comparison. In table 3 the comparison can be found. **Green** means that a model uses the criterion **red** means that a model does not use the criterion mentioned.

**Table 3. Comparison of models**

Model	Use Big Data	Retail Sector	Useful	Main limitation(s)
Big Data to Business framework	Green	Red	Green	Very generic model
Big Data Value Framework for the Financial Services Industry	Green	Red	Green	Only usable for part of the model
BPM lifecycle	Red	Red	Green	Very generic model
Balanced Scorecard	Red	Red	Green	Lack of Technology Very generic model
Innovation Diffusion Model	Red	Red	Green	Only about innovation diffusion
Technology Acceptance Model	Red	Red	Green	Only about technology acceptance

All models are not specifically made for the retail sector. The Big Data Value Framework for the Financial Services Industry is specifically made for a different sector where the other five are generic models which are not made for a specific sector. All models were found useful for the process model made in this Thesis; some gave interesting input and others can be implemented and used in certain processes. The usefulness of each model with regard to the process model will be discussed below.

The first model is the Big Data to Business framework of CGI (2013), see appendix J2. This model was used as a very generic starting point of the development the process model. The insights of this model are that Big Data is used by IT to generate Business via analytics and insights. This was tested in the interviews and the respondents were asked for instance to divide the business in different areas to make it more specific for the retail sector. All elements from this model are present in the process model that is designed in paragraph 5.3. The main limitation was that it is a generic framework that can be used in every area and every country. For this Thesis the model is used as starting point. It is made more specific for the retail area and applicable for the countries Netherlands and Belgium.

The conceptual process model that was designed can be used by stakeholders to have an overview which processes are relevant in the retail sector when talking about Big Data. The stakeholders can then decide which processes they would like to improve, change, specialize or optimize. To do so the following three models from the literature review can be used: the Balanced Scorecard of Kaplan and

Norton (1987) (appendix J4), the Big Data Value Framework for the Financial Services Industry from De Vries (2013) (appendix J5) and BPM lifecycle by Dumas et al. (2013) (figure J6). The BPM lifecycle by Dumas et al. (2013) can be used in the following two processes in the model: Improve business processes aimed at increase of revenue and improve business processes aimed at decrease of costs. In both processes of the conceptual model problems are discovered which have to be solved to increase revenue or decrease costs. The BPM lifecycle by Dumas et al. (2013) is the best option to improve these business processes. To convince the management of the proposed changes of the BPM lifecycle the Balanced Scorecard of Kaplan and Norton (1987) can be used. The criteria and effects of the proposed changes can be specified for the four categories of the Balanced Scorecard: Financial, Internal Business Processes, Learning and Growth and Customer. The advantage of using the Balanced Scorecard is that top management can understand the logic of this framework and for the business process managers it is easier to convince the management of the proposed changes. The main limitation is that it does not mention technology and it is not designed for Big Data. So in addition to the Balanced Scorecard there should be information given about technology to make use of the Balanced Scorecard in the conceptual process model. The benchmark process has banks and payment providers as main stakeholders. Therefore the framework that De Vries (2013) made can be very useful to analyze how to implement Big Data packages and what is required to be successful in terms of the four balanced scorecard principles. The framework is about all financial processes, here there is only benchmarking so the learning and growth and internal business process that De Vries defined can be useful. They should result in the benchmark report with advice on how merchants can improve. The main limitation is that it is not suitable for the retail sector.

The last two models that were discussed in the literature review were the Innovation diffusion model by Rogers (1962) (appendix J8) and the Technology Acceptance Model from Davis (1989) (appendix J9). These two models can be useful for the merchants (main stakeholder) when they develop the customer loyalty benefits. These benefits will be more personal and new forms of benefits will be available when merchants will use Big Data. With these two models, the merchants can see how customers will react to changes and new ways of loyalty systems and hence anticipate to them. The merchants can design their new ideas in such a way that it maximizes the perceived usefulness and perceived ease of use so that the customers will like and use the new loyalty systems. The Innovation Diffusion Model can be used to forecast how many customers will use the system and what to do to convince the remaining customers; for example to convince the late majority you have to come with other ideas than to convince the innovators. Both models can also be used in the same way at the “decide on changes” in the process model. Of course it depends on the type of change whether it is relevant or not. When a change consist of implementing an new IT tool which the customers will not notice the Technology Acceptance Model of Davis (1989) can still be relevant but than for the employees which have to work with the new or improved system. The main limitation is that the model can only be used in one process and not for the whole processes model. Therefore they can be implemented in the process model but not used as a model about Big Data usage in the retail sector, just like the other models discussed. As showed they are all useful for the designed model, but for this Thesis developing a new model was needed.



## 6. Validation

In this chapter the presented conceptual framework will be validated via interviews with merchants and experts. For a proper study this is important as aforementioned in chapter 2. For the validation semi-structured interviews are chosen again for the same reasons as can be read in chapter 4. The respondents differ from the respondents of the first questionnaire. New leads were used to have a control group with other respondents. First the method that will be used is described. This will include the process, the respondents and the reliability and validity. Afterwards the results of the questionnaire is shown and the revised framework.

### 6.1. Validation design

First there has to be decided what has to be validated and how this will be validated. The validation will be again with merchants and experts to have a control group with the most similar characteristics as possible to the initial group of respondents. The interviews in the data collection part of this Thesis were semi-structured. This has led to a wide variety of answers. The conceptual process model (presented in paragraph 5.3.) has a lot of processes that are already covered by literature and are also mentioned by the majority of the merchants and experts. No further validation is needed for these parts of the model. The process model also has some parts which are important but are only mentioned or discussed by a few merchants or experts. Those parts, only covered by a few respondents, will need further validation.

There were three usage areas defined in the process model. These three usage areas are benchmarking, optimize in-store shop experience and increase of customer loyalty. The first part of this validation will be to validate of the control group also think that all three usages are relevant; and if not why they do not think so.

The customer loyalty process from insights to increased revenue and reduction of cost was partly described in literature and fully mentioned by the interviews with experts and merchants, so no additional validation is needed. Only for the last step, increasing of revenue, questions are asked to detect differences. The same goes for the second usage area: the optimizing of in-shop shop experience. Only the second usage area has two possible last steps: increasing of revenue and decreasing of costs.

The third process, the benchmark process, was only mentioned by one expert in the first round of interviews. This was found very interesting and potentially very useful for merchants. Therefore it has been added to the conceptual process model. For this part validation is critical. So in the validation it is important to look at the process of analyze payment data by a payment provider. The analysis leads to insights which can be compared with industry standards, the benchmark results can be send to merchant who can use them to improve business processes to reduce cost or to increase revenue.

During the previous interviews it was mentioned that Big Data, at this moment, does not make a difference for merchant when it comes to the processes that are used. The input and output of the process will change when using Big Data but the processes itself stay more or less the same. In the validation this will be checked again to make sure that this is true.

The last part of the validation will be to check if there are any usage areas which are not mentioned during the first round of interviews. The conceptual process model has three different usage areas but it will be checked whether the control group thinks that there are more.

### **6.1.1. Type of validation**

The validation of this master Thesis is done with a questionnaire. The questionnaire could be send via e-mail or filled-out during interviews. Sending a questionnaire via e-mail about this topic will probably result in a low response rate, because the sensitivity of the subject Big Data. Questionnaires can also result in not completely filled-out responses. Interviews are more likely to lead to answers to all questions and therefore will have a higher ratio of completed surveys. With an interview you can also make sure that a respondent understands the question and does not misapprehend a question. When having the interview it can be prevented that a person gives an answer to a question that he actually does not understand. In an interview it can be asked “why do you give this answer?” or if you do not know the answer to this question it is better to not give one, rather than making a guess. Quality is considered important for this Thesis and also for this validation; therefore interviews are better suitable than questionnaires.

The interviews were semi-structured with both open and closed questions. The closed questions had room for the respondents to give additional comments. The same questionnaire is used for both experts and merchants. The reason for using the same questionnaire is that the model is already build and the validation is generic. The opinion from both groups is interesting but the goal of the questions in both groups is the same, therefore no difference is made in the questionnaire between the two. The questions for this interview can be found in appendix G. The closed questions about the three main usage areas have a Likert scale ranging from 1 “strongly agree” or not applicable till 7 “strongly agree” or fully applicable. With a Likert scale there is room for respondents to add weight to their answer which is not possible by a simple yes no question. The sub questions are also closed with a 1 to 7 Likert scale because this captures the possibility and difficulty of a process rather than only the possibility. The closed questions are followed by a why-question where it is possible for respondents to explain their answers. In this validation, just like the first interviews, the respondent is asked to answer a closed question after which he/she can add an addition explanation or comment. The explanation can also be used to check whether the respondent understood the question correctly and did not just give a number to please the interviewer. To keep the duration of the validation interviews to a minimum (increasing the response rate) the 4 benchmarking processes are questioned in one question if they are a good representation. That question is followed by an open question for respondents who do not agree with the four processes, in order to say which process(es) he would like to have changed and why. Here a yes/no structure is chosen because the processes are either fully correct or not. When not fully correct a change should be mentioned so a 1 to 7 scale will lead to the same results, 1 to 6 will mean a change is needed and only 7 will mean that the process is ok. For respondents this could be vague therefore a yes/no option is better for this question. The demographics of the respondents are the same as for the previous data collection and for the same reasons (see paragraph 4.3.).

### 6.1.2. Execution of questionnaire

The merchants and experts used for the validation were from an existing network of contacts. The quality of the leads was important and to get to speak with the right people high in an organization, having contacts directly within that organization is a must. The group of respondents has been made as diverse as possible within the given parameters of chapter 4. For merchants the goal was to differ in size and retail area and for experts it was tried to cover the five stakeholders' areas. The benchmark process needed the most validation; therefore there were a lot of experts interviewed with a background in banking or payments. The respondents were contacted via e-mail or telephone to ask for their cooperation in this research. The interviews took place face to face or by telephone. Before each interview the purpose of the interview was explained and background information was provided. It was stressed that answering a question was not obligatory and when a respondent did not know an answer, no answer is preferred over a gamble. This will increase the quality of the results. In total the control group consisted of 10 respondents, 2 merchant and 8 experts.

### 6.1.3. Characterization of respondents

As mentioned before in this chapter first the characterization of the respondents will be discussed. The table with the demographics of the respondents can be found beneath (table 4.). A discussion of the demographics can be found in appendix H.

**Table 4. Demographics of respondents for validation**

Respondent	Retail area	Function	Nationality of the person	Nationality of the company	Size
1	Building supply store	Marketing manager	Dutch	Dutch	medium
2	Supermarket	Operational manger	Dutch	Dutch	small
Respondent	Expert from	Function	Nationality of the person	Nationality of the company	Expertise
3	Consultancy	Consultant	Dutch	Dutch	retail/banking
4	Branch group	Manager	Dutch	Dutch	retail/supermarket
5	Consultancy	Consultant	Dutch	Dutch	retail/banking
6	Law firm	Jurist	Dutch	Dutch	privacy/regulation
7	Payment provider	Manager Customer Service	Belgium	Belgium	customer loyalty/mobile payment
8	Bank	Business Information Manager	Dutch	Dutch	banking/retail
9	Bank	Assortment Manager Financial Logistics	Dutch	Dutch	banking/retail
10	Payment provider	Product Manager	*	*	payment/retail

\*Per request of the interviewee, the nationality of the interviewee and company is omitted.

#### **6.1.4. Reliability and validity**

To determine if the data from the interviews is good enough the reliability and validity of the data has to be determined. Reliability is an indicator of the degree to which the measurement gives the same result. Reliability gives information about the consistency of the data. Data is valid if a measure assesses the construct that was intended to measure (Field, 2009). Normally statistical test are used to determine reliability and validity but with only 10 interviews this is not possible with the given dataset. By doing interviews it was tried to give the data the best reliability and validity. Open interviews give room for extra information when a respondent does not fully understand the question. When a respondent misinterpreted a question, open questions leave space to check and correct for that. Before the questionnaire was used it was clear that in this area it is very hard to find respondents, so to get the best validity interviews were selected to maximize the value per respondent. Respondents of interviews are more likely to give answers to all questions. The variance of answers to the questions is also not very high which indicates a decent reliability. There was only one question where the respondents differed a lot in their opinion with a variance of 3,16 and a standard deviation of 1,77. So both validity and reliability are made as good as possible and although the data is only collected from a small group of respondents there were no problems found. The small group of respondents does make it hard to generalize the results of this master Thesis. The reliability and validity of the data collection in chapter 5 do not differ from the data collection for this validation. The reliability and validity were made as good as possible but due to the low number of respondents there was also no statistical analysis possible there.

### **6.2. Results of the interviews**

The insights gained from the open interviews with the different experts are shown in detail in appendix I. Prior to the interviews, the respondents were asked if they had problems with recording the interview. Everyone agreed with recording the interview. After the interviews, the conversations were written down. When all interviews were done, per question the main points were listed and it was noted per respondent which points he/she had mentioned. The closed questions sometimes gave a very clear direction. Per closed question the answers and variance in terms of standard deviation in answers will be discussed. The answers sometimes differ a lot between the respondents. But a lot of topics were mentioned by more than one respondent. So when a topic is not mentioned by a respondent, that does not have to mean that he or she would not agree with it, but it simply means that he or she did not mention it in the interview. The anonymized detailed results are described in appendix I. Again confidentiality was guaranteed to be able to talk with merchants and experts about this delicate topic. For the validation the interviews for merchants and experts were the same. All closed questions had a 7 point Likert scale as mentioned before. Ranging from 1 “strongly agree” or not applicable till 7 “strongly agree” or fully applicable.

The main insights from the respondents are now discussed. The variance and standard deviation is calculated as described in chapter 5. As mentioned before due to the low number of respondents the statistical value of these numbers is not very high. Just as in chapter 5 this was the best way of describing the variance in the data. All respondents (100%) think that customer loyalty is a usage area of Big Data in the retail sector. They answered on average 6,4 with a standard deviation of 0,64. They all

mentioned that in the retail sector it is crucial to have knowledge about the customer, the more the better. They also all expect that Big Data will be going to play a role in making customer loyalty more personal. The usage will depend on retail sector and size of the retailer was mentioned by 30%. The respondents think that Big Data can be used for both hard (on average 5,4 with a standard deviation of 1,28) and soft benefits (on average 6,6 with a standard deviation of 0,66) but they think that Big Data will have a bigger impact on the soft benefits. They also believe that soft benefits will become more and more important the coming years.

The interviewees did also think that optimizing in-shop customer experience is a usage area of Big Data in the retail sector. They answered with on average a 6,0 and a standard deviation of 1,18. It can help retailers to determine what triggers customers in shops to buy products both online and offline. These improvements could lead to both cost reduction and an increase of revenue but the respondents mentioned that increasing revenue will be more important. They gave on average an answer of 5,0 with a standard deviation of 1,18 for cost reduction. One respondent for example mentioned that it could lead to an optimization of the number of employees. The respondents mentioned that they expect more purposes for increasing of revenue with an average answer of 6,2 and a lower standard deviation of just 0,87.

The respondents (100%) did believe that benchmarking could be an interesting usage area of Big Data in the retail sector. It is always good for retailers to compare with others. They answered on average 5,9 with a standard deviation of only 0,54 so very little variance in the data. The respondents (100%) also mentioned that the four processes that are now identified are correct but some (40%) had additions to these. Advice from the payment provider could be implemented and input of data from other resources. Again they mentioned that it could lead to both cost reduction and an increase of revenue, with increasing the revenue as the more important goal. They gave on average a 4,2 for cost reduction but with a large amount of variance resulting in a standard deviation of 1,78. For increasing of revenue the answer on average was 6,0 with a lower standard deviation of 1,10.

The respondents answered that the processes in essence between data and Big Data in the retail sector do not differ much at this moment. They gave a 5,6 on average as answer with a standard deviation of 1,35. They did mention that they hope that this will change in the future; this was mentioned by 60% of the respondents. The only difference at this moment in the retail sector which could be seen in the process model is the addition of more resources; for example gain data instead of only use data from the merchant itself. They suggested adding an extra step at the beginning of the process model for the preparation of data which is a major difference between data and Big Data.

There were some final comments made by respondents. A few respondents (20%) mentioned another interesting usage area of Big Data in the retail sector, risk management; for example via fraud detection. This can be especially important for online stores, because they have a lot of digital fraud. Another respondent mentioned that besides increasing revenue and reducing costs, increasing loyalty will change the Net Promoter Score. The Net Promoter Score is the metric that measures whether people would recommend a company/product or service (in this case the retailer) to a friend or colleague. The right people are very important when dealing with Big Data was also mentioned.

### **6.3. Proposed process model**

The insights from the interviews were used to validate the framework and where necessary a process was removed or added to the process model. The insights from the interviews did not give reason to remove things from the framework. The respondents did think that Big Data could have more extra value for soft than for hard benefits. The answer however showed that still Big Data can also have enough extra added value for hard benefits to keep it in the process model. The same could be seen for the cost reduction and the increase of revenue. The respondents did mention that they thought that increasing revenue has more potential than reducing costs; but the answers 5,0 and 4,2 for both reducing costs processes gave no reason to delete those processes. The respondents gave some interesting insights where to change the model. They mentioned that gain data to analyze data is a bit too generic. An extra step called data preparation is needed. In the conceptual model the data preparation was done in the step analyze data. The control group mentioned that data preparation in Big Data is very important; therefore it should be a separate process step. In this process the merchant or payment provider (the extra step is needed in both usage areas) will check if all data can be used from privacy and data ownership points of view and turn unstructured data into structured data. The usage area benchmarking gets an extra step as well: add extra data. In this step payment providers can add extra data to the payment data to get more specific benchmark results. This step is not needed for the merchant because if merchants want to add extra data for their analysis they will do that in the gain data process at the beginning. They are the owner of that processes so payment providers will normally only get payment data from the merchants and not the other data. In the validation another potential interesting usage area was mentioned. The Big Data insights could be used for risk management for example to be used for fraud detection. This is mainly interesting for the online stores. For normal stores, Big Data is less likely to be used for fraud detection. With this improved risk management the business process regarding a decrease of costs will improve, resulting eventually in a reduction of costs. The last improvement that has been made after validation is the increase of the Net Promoter Score as a result of the change in customer loyalty towards a merchant from a customer perspective. The Net Promoter Score change was mentioned by one of the respondents. The Net Promoter Score is the metric that measures whether people would recommend a company/product or service to a friend or colleague (Murphy, 2008). In this case a retailer. Appendix J10 figure 14 shows an overview of the changes made to the conceptual process model. As stated before the final model does not have changed a lot from the concept model. Only the changes are discussed here. Information regarding the other processes in the model can be found in paragraphs 5.3.1. and 5.3.2. On the next page figure 4 can be found with the adjusted version of the process model after the validation.

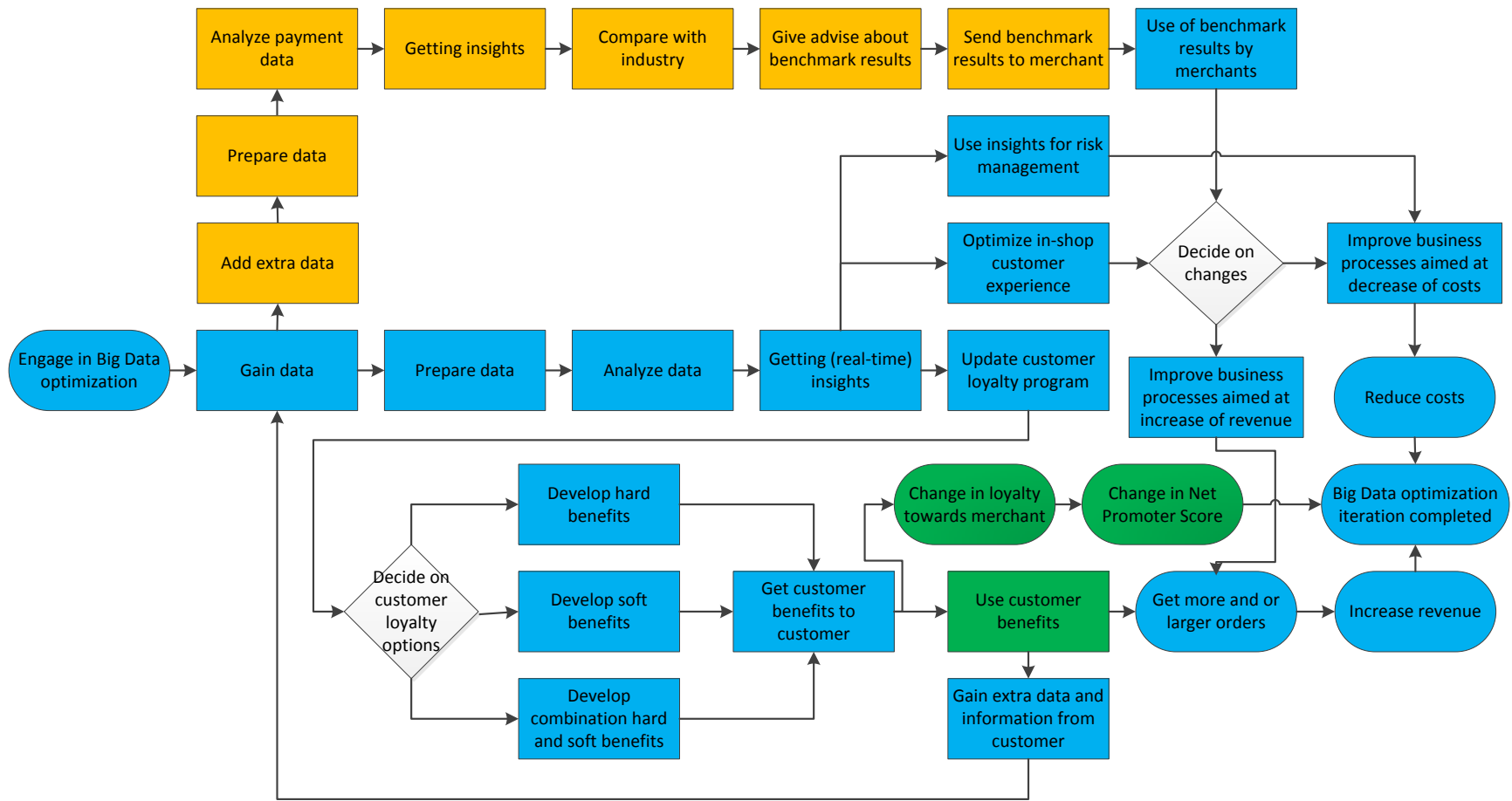


Figure 4. Final process model

## 7. Conclusions, Limitations and Future work

### 7.1. Conclusion

In this chapter the answers to the research questions of chapter two will be discussed. First the sub questions will be answered leading to the answer of the main research question: **how can Big Data be used to derive business value in the retail sector?**

1. *Which main stakeholders can be identified in the process of getting business value out of Big Data in the retail sector?*

The insights from literature and the interviews gave an unequivocal answer to this research question; there are five different stakeholders in the process. The five stakeholders are the merchant, the consumer, the government, the payment provider and the bank. The merchant is the main stakeholder and initiator of the processes. The consumer is a stakeholder because in the retailer most processes especially when loyalty is involved involve the consumer. The government is the third stakeholder because they set the ground rules what companies can and cannot do concerning Big Data; for example privacy and data ownership regulation. The last two stakeholders are the payment provider and a bank. These could be the same but when using for example also mobile payments or credit card transaction these could be different. The financial sector will be changing a lot the coming years because of Big Data; therefore these are also two important stakeholders in the processes.

2. *Which usage areas can be identified in the retail sector for Big Data to gain business value?*

In the literature study and during the interviews for data collection three main usage areas were found. The first usage area is the improvement of customer loyalty. This usage area will probably be the most important usage area of Big Data for retailers. Customer loyalty is found to be crucial for retailers to survive and they expect that Big Data can help them to improve the customer loyalty programs and gain more loyal customers. The second usage area that was found is the improvement of the in-shop customer experience. All changes that can be done in a shop (both online and offline) to improve the customer satisfaction and making customers spend more or come more often to a store. These improvements can also lead to a reduction in costs but the interviews revealed that the main benefit will be an increase of revenue. The third usage area that was found via interviews and has been corroborated by the validation group was the usage area of benchmarking. With this usage area the stakeholder bank or payment provider will analyze the payment data of one store and compare it to a generic industry benchmark. The optimal scenario will be that extra data is added by the payment provider in order to perform a more detailed analysis. They can also add an advice. This could be an interesting usage area for the payment provider or bank and the merchant as well. An extra fourth usage area was the usage area of risk management, for example via fraud detection. This was mentioned in the validation as especially useful for online shops but more research is required to make sure that merchants also think this is an actual usage area. It does have a lot of potential and Big Data could bring extra value for this.



### *3. Within these usage areas how can value be gained out of Big Data?*

The application of Big Data in the first usage area customer loyalty can be to help making the customer loyalty programs more personal. Big Data gives merchants opportunities to get better insights according to the interviews with them. These insights can be used to make customer loyalty programs more personalized by targeting for example the right customers or give the right customer the right offer. In the second usage area (optimizing in-shop customer experience) and the third usage area (benchmarking) value out of Big Data can be gained in the same way. The value can be gained via insights; these insights can be used to determine which business processes should be optimized to make them more efficient. As can be seen in the model (figure 4) this can lead to either a reduction in costs or an increase in revenue. The Big Data can be used to determine the processes, the BPM life-cycle in combination with the Balanced Scorecard can then be used to optimize these determined processes. In the last usage area that was mentioned in the validation, risk management, Big Data can be used to detect for example fraud which will lead to a reduction in costs.

### *4. How can customer loyalty be gained?*

According to literature and the interviews customer loyalty can be gained via hard and soft benefits. Hard benefits are monetary benefits such as discount; soft benefits are non-monetary benefits such as extra service or a newsletter. Both are found important in getting loyalty from customers. The best is to have a proper combination of hard and soft benefits, this combination differs per retailer. The results of this study emphasize that soft benefits should not be forgotten and can be more important than hard benefits. With the internet era it gets easier for customer to buy products at the store with the cheapest price so to differentiate with price is becoming more and more difficult. It could be a good idea for retailers to emphasize more on the differentiation in soft benefits. Soft benefits are more suitable to make a unique store. In the interviews it was mentioned that hard benefits can be used to attract customers and soft benefits can be used to maintain these customers. In the validation it was mentioned that Big Data can be used for both hard and soft benefits.

### *5. How can the usage areas of Big Data that are discovered be modelled into a design?*

As could be seen in chapters five and six, a process model was found the best way of design. The insights from literature and from interviews with both merchants and experts are bundled into the process model. The final version of this process model can be found in figure 4. The model gives an overview of all the different processes from the start point to the final stage. It can be used by all stakeholders to analyze in which process they want to invest to gain more value. It can also be used by an outsider to make new Big Data applications as there are not many Big Data applications used by merchants at this moment. An outsider could for example make a new application for merchants which gives every customer a unique set of hard and soft benefits to best fit the needs of a unique customer.

### ***How can Big Data be used to derive business value in the retail sector?***

The business value of Big Data in retail is huge but it is not fully recognized by the merchants at this moment. The business value of Big Data in retail can be derived via four main usage areas as mentioned

before. With three of them the merchant itself is the stakeholder who should actively do something with Big Data. The benchmarking usage area has a payment provider or bank as active stakeholder. Big Data in the retail will gain the value via customer loyalty by making the customer loyalty programs more personal. At this moment a lot of mass marketing is done in the retail sector. Some merchants use very general segmentation, for example target students as a different subgroup. But customer loyalty at a unique personal level is not normal at this moment. Customer should also get used to that and in the beginning there will be resistance from the customers. It will be very important for merchants to actively and openly inform customers what they want to do with their data and what the benefits for the customer will be. When this is done Big Data will have a lot of value for the retail sector; but doing it wrong will lead to severe image damage.

To make sure that merchants are doing the right things with Big Data it is important for them to have the right people to work with Big Data. Big Data usage requires special skills so you should hire or train employees in order for them to have the amount of knowledge needed to work with Big Data and make a proper analysis. Privacy is also important to consider. Especially in the retail sector Big Data will involve personal information of customers so privacy is an important factor. The last part that is important for merchants is the data ownership and security. Merchants have to make sure that they can use or own the data that you want to use in their Big Data processes. They also have to make sure that the data is safe from cyber-attacks.

Another conclusion from the research is that at this moment the merchants see Big Data mainly as the data analysis from the past with more data. This can also be seen in the model that is made in which the processes in essence do not differ that much from a process model that would be made a few years back. The merchants also mentioned that for them Big Data is more a buzz word. It could also be seen at the answer to the question about how they define Big Data. They all mentioned that it is a lot of data and only some mentioned that the data could be unstructured or could come from more sources. The experts from the financial sector had a clearer image of Big Data and Big Data purposes as well in their own sector as for the retail sector. So there is a lot of difference in the acceptance and knowledge about Big Data in different industry sectors. Looking at for example the model and Thesis of De Vries (2013) in the financial sector compared to this Thesis and model in the retail sector. The financial sector has far more knowledge already of Big Data and is at this moment actively looking at ways to use Big Data. This was also mentioned by the respondents from the financial sector. They are already using Big Data at this moment where in the retail sector Big Data usage is still in its infancy. For outsiders there are good opportunities here to make applications that retailers can implement in their business processes to use Big Data without being too complex to use. Big Data usage as mentioned in the last paragraph requires the right people which can be too expensive for medium sized and small merchants. For these companies buying tools or outsourcing would be the best option. Making a good contract is crucial when doing this because the merchant always stays responsible for the data from a juridical point of view. Big Data will become more and more used in the retail the coming years but it is crucial for merchants to determine what they want to achieve, how they want to achieve that via Big Data and what they can and cannot do themselves. Knowing your weaknesses is very important in a complex area such as Big Data. Big Data usage is not as simple as most merchants think it is.

## 7.2. Limitations

The master Thesis research has been performed with great care but there are still some limitations for this research project. In this chapter these limitations will be mentioned. The first limitation is the usage of literature. Big Data is an important issue in the information technology sector at this moment and a lot of research still has to be done. Scientific articles are published regularly which could mean that the most recent articles are not mentioned in this Thesis. The topic is so new that the research done at this moment is either more general research or very specific and only relevant for a small area. The focus of this study was Big Data usage for Dutch and Belgium retailers. To get a proper impression about the area of Big Data sometimes also less reputable magazines or company sites had to be used.

Big Data is also a very sensitive topic for merchants which makes it very hard to find respondents for interviews. This resulted in only a small number of respondents for interviews for data collection as well as validation. To maximize the value of insights from these respondents, interviews were chosen instead of questionnaires; so that an answer to every question was gained or a respondent could mention that he cannot give an answer. At first it was tried to look for differences between retail sectors, size and country. The small amount of respondents makes it impossible to generalize the results, but for a qualitative research this is not a big problem. To increase the value of the insights experts were included who for example represented a whole branch of retailers. Contacting merchants directly about this topic without leads would result in no response and even with leads the response rate of merchants was beneath 50%.

The third limitation was that the low number of respondents makes it impossible to use statistical programs to analyze the results. To analyze the results it was only possible to describe the opinion and insights of the respondents and check how many mentioned specific topics. To compensate for this limitation the respondents were of high quality so that the insights they gave really matter in the Dutch retail area. So again because this was a qualitative research and not a quantitative research this is not a big problem but remains a limitation of this Thesis nonetheless.

The fourth limitation of the designed model is that the costs aspect is not included. Very little information can be found in the literature and also the respondents of the interviews for both data collection and validation did not mention the costs aspect to be included in the model. At this moment it is only present in the model by having an end stage increasing revenue where it depends on the amount of costs made whether or not it increases the profit as well.

The last limitation is the knowledge about Big Data. It is also mentioned by the respondents of the interviews that Big Data in the retail sector is relevant but that the big change still has to come. It is expected that a lot will change in the coming years. So it could be that the insights that were given will change a lot in the coming years; for example everyone was thinking about possible use of Big Data and how to gain value while nobody mentioned the costs and investments needed to perform Big Data analysis. The coming years will tell, but it could be a limitation of this research that for the retail sector it came just a bit too early.

### **7.3. Recommendations and directions for further research**

In this chapter the recommendations and directions for further research will be discussed. The first recommendation is to expand the research with more data. The model is now made and validated with a qualitative analysis. To be able to generalize the results of this Thesis a quantitative study is needed. With a quantitative study differences in retail sector or country specific differences between Belgium and the Netherlands could possibly be identified. This was not possible with the data collected in this research project. For this study a qualitative research was chosen because it was a first study in this specific area, so a second study could be a quantitative one as described by Blumberg et al. (2008). With a larger dataset it is also possible to do statistical analysis. The new respondents can also be used to look at the cost side of Big Data and with finding a way to model the costs of Big Data in the process model. It might also be interesting to perform a study which describes the differences between industries. As mentioned in the conclusion there were differences found in the retail sector compared to the financial sector but maybe there are more differences when other sectors are included. The same can be said for a comparison between countries for example differences between the Netherlands and America in the usage of Big Data in the retail sector.

The second recommendation is to check the completeness of the model over the coming years. Many respondents mentioned that they want to use Big Data in the future and that they think it really has value for them. They do not really use Big Data at this moment nor expect to in the very near future so maybe their opinion will change in the coming years. It could be that in a few years they have started to adopt Big Data solutions and that the process model needs to change to match the situation at that moment. The expectation is that changes will be needed but at this moment it is not clear in which direction so this could be an interesting direction for further research. The implementation of the costs aspect which is now missing can then be implemented. During the validation new insights are gained. However, in an ideal situation these insights should also be validated. This could be part of future research.

The last recommendation is to look for actual business cases with Big Data in the retail sector. This process model and Thesis could be used as starting point to design new Big Data solutions for the retail sector. This model can be used to determine where to use Big Data and how it will change the processes connected. It can also be studied how Big Data actually can be used in each step off the process model and how it will influence other processes. It could also be interesting to look whether a Big Data application could be made and what the requirements of such an application would be. This Thesis could be a starting point giving insights for that. But as can be read in the previous paragraph it may take some time before companies are actually using Big Data and can share their insights for this type of research. Big Data will change a lot in the research sector, the question will be when these changes will take place and how they will influence the process model as it is designed at this moment.

## 8. References

### 8.1. Articles

- Abu-ELSamen, A., Akroush, M. N., Al-Khawaldeh, F., & Al-Shibly, M. (2011). Towards an integrated model of customer service skills and customer loyalty. *International Journal of Commerce & Management*, 21(4), 349-380.
- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64-65.
- Baicoianu, A., & Dumitrescu, S. (2010). DATA MINING MEETS ECONOMIC ANALYSIS: OPPORTUNITIES AND CHALLENGES. *Bulletin of the Transilvania University of Brasov Economic Sciences*. Series V, 3, 185-192
- Barlow, R. (2000). Rewards vs. relationships. *Potentials*, 33(11), 46.
- Brandl, D. (2005). Who owns the data? *Control Engineering*, 52(12), 20.
- Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'Big Data'? *McKinsey Quarterly*, 1-12.
- Brown, R. (2012, June 26). Unlocking the value in Big Data. Retrieved from Waters Techology: <http://www.waterstechnology.com/>
- Business Wire (2013, Apr 15). Big Data leaders: Big Data either a huge risk or a huge opportunity – outcome pending people, skills, and investment. *Business Wire*.
- Buytendijk, F., & Heiser, J. (2013, Sep 24). Confronting the privacy and ethical risks of Big Data. *FT.Com*
- Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197-208.
- Chiou, J.S., Hsieh, C.H., & Yang, C. (2004). The effect of franchisors' communication, service assistance, and competitive advantage on franchisees' intentions to remain in the franchise system. *Journal of Small Business Management*, 42(1), 19-36.
- Chung, H. M., & Gray, P. (1999). Special section: Data mining. *Journal of Management Information Systems*, 16(1), 11.
- Crystal, L. (2000). Partners for the millennium. *Franchising World*, 32(1), 12-20.
- Davis, F. D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, 13(3), 319–340.
- Davenport, T., Barth, P., & Bean, R. (2012). How 'Big Data' is different. *MIT Sloan Management Review*, 54(1), 43-46.

- DBTA (2012). Data Management In The Era Of Big Data. *Database Trends and Applications*, 26(1), 15-17.
- DBTA (2013). The Next Big Data Phase: Business Knowledge Expansion. *Database Trends and Applications*, 27(1), 24-25.
- De Nationale Franchise Gids (2011) retrieved via <http://denationalefranchisegids.nl/bestel-de-gids/>
- De Vries, H.A. (2013) The Business Value of Big Data A Framework Proposal for the Financial Services Industry. Eindhoven: Eindhoven University of Technology.
- Gabisch, J. A., & Milne, G. R. (2014). The impact of compensation on information ownership and privacy control. *The Journal of Consumer Marketing*, 31(1), 13-26.
- Groot, A. (2014). Personalisatie: de conversiekracht van data. *Emerce magazine 128*. Retrieved from [http://www.emerce.nl/nieuws/personalisatie-conversiekracht-data?utm\\_source=rss&utm\\_medium=rss&utm\\_campaign=personalisatie-conversiekracht-data](http://www.emerce.nl/nieuws/personalisatie-conversiekracht-data?utm_source=rss&utm_medium=rss&utm_campaign=personalisatie-conversiekracht-data) visited on 12-02-14.
- Hevner, A.R., March, S.T., Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28 (1) 75-105.
- Hopkins, B. & Evelson, B. (2011). Expand Your Digital Horizon With Big Data. *Forrester research report*, September 2011.
- Hormazi, A. M., & Giles, S. (2004). Data Mining: a Competitive Weapon for Banking and Retail Industries. *Information Systems Management*, 21(2), 62-71.
- Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1-14.
- Kassing, J. (2002). Creating customer loyalty. *Texas Banking*, 91(5), 12-17.
- Kaplan, R. S., & Norton, D. P. (2005, Jul). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 83, 172-180.
- Kiron, D. (2013). Organizational alignment is key to Big Data success. *MIT Sloan Management Review*, 54(3)
- Kursunluoglu, E. (2014). Shopping centre customer service: Creating customer satisfaction and loyalty. *Marketing Intelligence & Planning*, 32(4), 528-548.
- Lacey, R., & Sneath, J. Z. (2006). Customer loyalty programs: Are they fair to consumers? *The Journal of Consumer Marketing*, 23(7), 458-464.
- Lamont, J. (2013, 01). BIG DATA: Interviews with the experts. *KM World*, 22, 6-7,20.

- LaValle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2011). Big Data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21-31.
- Lazar, N. (2012). The big picture: Big Data hits the big time. *CHANCE*, 25(3), 47-49.
- Lopez, J. A. (2012, Fourth). Best practices for turning Big Data into big insights. *Business Intelligence Journal*, 17, 17-21
- IBM(2013). Bringing Big Data to the enterprise. Retrieved from IBM: <http://www01.ibm.com/software/data/bigdata/> visited on 08-02-14.
- Introna, L. and Pouloudi, A. (1999), "Privacy in the information age: stakeholders, interests and values", *Journal of Business Ethics*, Vol. 22 No. 1, pp. 27-38.
- Mendoza, L.E., Marius, A., Pérez, M. & Grimán, A.C. (2006). Critical success factors for a customer relationship management strategy. *Information and Software Technology*, 49, 913-945.
- Muhleman, J. (1996). The franchise relationship channel. *Franchising World*, 28(1), 20-23.
- Murphy, M., D.D.S. (2008). The one number you need to know ... to grow! *Dental Economics*, 98(5), 102.
- Nasar, M., & Bomers, J. V. (2012, Second). Data management and financial regulation: Using a Big Data approach to regulatory compliance. *Business Intelligence Journal*, 17, 34-40.
- Natarajan, G. (2013, Apr 01). Demystifying big data. *Financial Express*
- Oliver, R.L. (1980), "A cognitive model of the antecedents and consequences of satisfaction decisions", *Journal of Marketing Research*, Vol. 4, pp. 460-9
- Peacock, P. R. (1998). Data mining in marketing: Part 1. *Marketing Management*, 6(4), 8-18.
- Pérez, M., Grimán, A., Mendoza, L. & Rojas T. (2001) Systemic methodological framework for IS research, *Association for Information Systems*, 1119–1125.
- Raine, L., & Wellman, B. (2012). Networked. The new social operating system. *Cambridge: MIT Press*.
- Richtlijn 95/46/EG (1995). Eur-lex.europa.eu <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:nl:HTML> visited on 08-02-14.
- Schultz, D. and Bailey, S. (2000), "Customer/brand loyalty in an interactive marketplace", *Journal of Advertising Research*, Vol. 40 No. 2, pp. 41-52.
- Semeijn, B. (2012). Big Data en de privacyregels. *Computerworld Management special*. Retrieved from: [http://www.crossmediadashboard.nl/wpcontent/uploads/2012/12/448\\_Computerworld\\_De\\_privacyregels\\_bij\\_Big\\_Data.pdf](http://www.crossmediadashboard.nl/wpcontent/uploads/2012/12/448_Computerworld_De_privacyregels_bij_Big_Data.pdf) visited on 08-02-14.
- Silvola, R., Jaaskelainen, O., Hanna Kropsu-Vehkapera, & Haapasalo, H. (2011). Managing one master data - challenges and preconditions. *Industrial Management + Data Systems*, 111(1), 146-162.

- Sirdeshmukh, D., Singh, J., Sabol, B., (2002), Consumer Trust, Value, and Loyalty in Relational Exchanges, *Journal of Marketing*, Vol. 66 (1), pp. 15–37.
- Snijders, C., Matzat, U., & Reips, U.D. (2012). “Big Data”: Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*, 7(1), 1-5.
- Stehlik, M. (2008) Vesdia and Aberdeen report on creating customer loyalty in retail. *Business Wire*
- The Business Times (2013, Feb 18). As Big Data grows, so too security risks. *The Business Times*.
- The Lawyer (2013, May 22). ZyLAB publishes white paper warning on risks of Big Data. *The Lawyer (Online)*.
- Tolliver-Nigro, H. (2011). Make Money with QR Codes. *Printing Impressions*, 53(11), 24-27.
- Toub, C., Director, o. E., & Bernstein, A. (2012). Big Data provides big risk and big opportunity. *Institutional Investor*.
- Van der Meij, R. (2014). Nederlandse retail laat kansen data liggen. Retrieved from [http://www.levensmiddelenkrant.com/14123/nederlandse-retail-laet-kansen-data-liggen?utm\\_source=MailingList&utm\\_medium=email&utm\\_campaign=Nummer+298](http://www.levensmiddelenkrant.com/14123/nederlandse-retail-laet-kansen-data-liggen?utm_source=MailingList&utm_medium=email&utm_campaign=Nummer+298) visited on 10-02-14.
- Van Eupen, M.G.H. (2013). Franchiseketens als nieuwe bron voor een energiebedrijf? Eindhoven: Eindhoven University of Technology
- Van Eupen, M.G.H. (2014). Literature study. Eindhoven: Eindhoven University of Technology.
- Van Eupen, M.G.H. (2014). Research Proposal. Eindhoven: Eindhoven University of Technology.
- Venkatesh, V., Morris, M.G., Davis, G.B & Davis, F.D. (2003). User acceptance of information technology:Toward a unified view. *MIS Quarterly*, 27(3): 425-478.
- Vuksic, V. B., Bach, M. P., & Popovic, A. (2013). Supporting performance management with business process management and business intelligence: A case analysis of integration and orchestration. *International Journal of Information Management*, 33(4), 613.
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), 11.
- Wet Bescherming Persoonsgegevens (2000). Wetten.overheid.nl [http://wetten.overheid.nl/-BWBR0011468/geldigheidsdatum\\_08-01-2014](http://wetten.overheid.nl/-BWBR0011468/geldigheidsdatum_08-01-2014) visited on 08-02-14.
- Won, T. (2013). Big Data 3Vs. Retrieved from Apache Hadoop: <http://tedwon.com/display/dev/Apache+Hadoop#ApacheHadoop-BigData3Vs> visited on 08-02-14.



## 8.2. Books

- Van Aken, J.E., Berends & H., Van der Bij, H. (2009). *Problem Solving in Organizations*. Cambridge: Cambridge University Press.
- Blumberg, B., Cooper, D.R. & Schindler, P.S. (2008). *Business Research Methods*. Birkshire: McGraw-Hill Education.
- CGI (2013). *Data to Diamonds Turning Information into a Competitive Asset*. Gained via [info@cgi.com](mailto:info@cgi.com) after a presentation at Big Data congress of Industria.
- Crawford, M. & Di Benedetto, A. (2008). *New Products Management Ninth Edition*. New York: McGraw-Hill Higher Education.
- Dumas, M., La Rossa, M., Mendling, J. & Reijers, H.A. (2013) *Fundamentals of Business Process Management*. Heidelberg: Springer.
- Fayyad. U.M., Piatesky-Shapiro, G., Smyth, P. & Uthurusamy, R. (1996) *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge, MA.
- Field, A. (2009). *Discovering Statistics using SPSS*. London: SAGE Publications Ltd.
- Kaplan, R. S., & Norton, D. P. (1996). *Translating strategy into action: The balanced scorecard*. Boston: MA: Harvard Business School Press.
- Rogers, E.M (1962). *Diffusion of Innovations*. New York, The Free Press.

## 8.3. Sites

- <http://balancedscorecard.org/portals/0/images/balancedscorecard.jpg> last visited on 23-06-14
- [http://www.bnr.nl/nieuws/Personal\\_Finance/658520-1308/leiden-start-proef-met-mobiel-betalen](http://www.bnr.nl/nieuws/Personal_Finance/658520-1308/leiden-start-proef-met-mobiel-betalen) visited on 08-02-14.
- <http://www.qtag.nl/digitale-media-oplossingen/qr-code-oplossingen-en-qr-code-software/qr-code-faq> visited on 08-02-14.
- <http://www.smartphone.pcbeveiligen.nl/nfc/contactloos-mobiel-betalen-nfc-chip/> visited on 08-02-14.

## 9. Appendices

### Appendix A: Main mobile banking and payment providers

In the time that computers came round digital money was developed which can be accessed via different types of cards. The most recent development is the usage of smartphones for banking purposes. Since a few years it is possible to use internet banking on smartphones. Also actual payment in stores is possible and is starting to grow at this moment. In this chapter the different types of mobile payment solutions will be discussed.

#### Providers

There are at this moment a lot of different providers who are active in this domain. The most important ones will be discussed. These are Google Wallet, PayPal, Visa and Mastercard, American Express, MyOrder, BCMC Mobile, Sixdot and SEQR. From the research proposal of Van Eupen (2014) the following comparison table can be derived.

- **Red** means not using the characteristic stated in the most left column.
- **Green** means using the characteristic stated in the most left column.

Table 5. Comparison between five different mobile payment providers.

	<b>Google Wallet</b>	<b>PayPal</b>	<b>VISA &amp; Mastercard</b>	<b>American Express</b>	<b>MyOrder</b>	<b>BCMC Mobile</b>	<b>Sixdot</b>	<b>SEQR</b>
Payment technology platform	NFC	NFC & QR, in-app	NFC	In-app	NFC	QR	NFC	QR (NFC-ready), in-app
Based on proprietary transaction switch, excluding the VISA/MC	Red	Red	Red	Green	Green	Green	Red	Green
Reducing the merchant transaction cost by 50%	Red	Red	Red	Not available	Red	Red	Red	Green
Avoids for investment in new hardware	Red	Red	Red	Not available	Red	Green	Red	Green
Can be utilized by majority of today's smartphones	Red	Green	Red	Green	Green	Green	Green	Green
Available for POS* payment	Green	Green	Green	Red	Green	Red	Red	Green
Able for online & mobile app payment	Red	Green	Red	Red	Red	Green	Green	Green
Able for P2P transfer	Red	Green	Red	Green	Red	Red	Green	Green
Able for bill payment	Red	Red	Red	Red	Red	Red	Green	Green
Loyalty capability	Green	Green	Green	Green	Green	Red	Green	Green

\*POS = Point of sale

The most important difference between the different providers is the technique used to make the mobile payment transaction possible. The two most commonly used techniques are: usage of NFC chip (Near Field Communication) and the usage of QR codes. The NFC chips can be used for many different

purposes including payments (Smartphone beveiliging, 2013). The chips can be used to transfer information without contact. The reach of the chips are normally between 10-20 centimeters but with a sensitive receiver could be up to 13 meters. This results in the danger of the chips when used for payments. There could be transactions started without the owner of the smartphone noticing (Smartphone beveiliging, 2013). The other technique uses QR codes. QR codes are barcodes that can be scanned with for example a QR code reader app op smartphones (Tolliver-Nigro, 2011). QR codes can be read by all types of smartphones and operation systems. It is a two dimensional barcode that was first invented by a Toyota company to track vehicles during production. QR codes can be used for URL's, virtual stores, payments and website logins. A QR code is a squared box with white and black dots in it that when scanned with a reader app turn into information, for example a website URL (Qtag, 2013).

### **Conclusion**

It can be concluded that there are many different operators active on this relatively new market of mobile payments. They all use more or less two different techniques or the NFC chip or the QR code. Both have their pros and cons. The NFC chip is used by more operators than the QR code. The biggest difference is that at this moment almost no smartphone is equipped with a NFC chip so that customers should buy a new smartphone to use this technique, or put a sticker with the chip to their device. The other technique can be used by everyone with a smartphone and camera because it only uses an app and the ability of reading a QR code with the camera, so no need to buy a new device with this technique. Also for merchants there is an important difference because to use NFC chips an investment in new hardware is needed which is not needed with the usage of QR codes. These differences explain why the NFC technique is only recently (September 2013) tested on small scale (only in Leiden) in the Netherlands after years of preparation (BNR, 2013) and the QR code technique is launched in different European countries (Sweden, Romania and Belgium) without large troubles. So the best technique to be used at this moment is the QR based technique because it has the least expenses and most benefits for as well the merchant as the consumer.

## Appendix B: Questionnaire for merchants.

The questions in red are about customer loyalty. The questions in green are about Big Data and the questions in blue are about the usage of Big Data for customer loyalty.

1. How do you think about customer loyalty?
2. Do you have a customer loyalty program at this moment?
  - a. If yes what type of loyalty program do you have?
  - b. If not what do you do to drive behavior of you customers?
3. On a scale of 1 to 7 what is your view of the current amount of customer loyalty that is being provide to your customers?
4. On a scale from 1 to 7 how important is customer loyalty for your company?
  - a. On a scale from 1 to 7 how important are customer loyalty programs for your company?
5. On a scale from 1 to 7 to which extent are you using data to get customer loyalty?
6. What is your vision on Big Data today and for the future?
  - a. What is your definition of Big Data?
  - b. What are the main benefits of Big Data for you?
  - c. What are the main risks of Big Data for you
7. On a scale from 1 to 7 how many experience/knowledge do you have with Big Data?
8. On a scale from 1 to 7 do you think you can perform Big Data Analytics in-house?
9. How can Big Data help your company to develop more customer loyalty?
  - a. Which methods or approaches do you currently (or in the future) use to obtain value from Big Data and how do you asses the value of Big Data.
10. On a scale from 1 to 7 would you be willing to provide data to an external party if this could improve consumer loyalty and/or consumer benefit?
11. Under which circumstances would you be willing to share your data with a third party to enhance the customer loyalty?
12. On a scale from 1 to 7 would you be MORE WILLING to provide data if it is guaranteed that this data is only being used to the benefits of your consumers.
13. Assuming that you would be willing to share the data of your company with a third party to enhance customer loyalty, what would you expect the third party to do and not to do with the data?
14. Do you have any other ideas or concerns we have not covered and that could be important?

## Appendix C: Questionnaire experts

The experts were given a semi structured interview which was based on the input from the literature study. To questions are made as similar as possible to the questions from the merchants. This led to the following questionnaire, were the questions are divided into three groups of questions. The questions in red are about customer loyalty, the questions in green are about Big Data and the questions in blue are about the combination of Big Data and customer loyalty.

1. On a scale of 1 to 7 what is your view of the current amount of customer loyalty that is being provide to customers in general by merchants?
2. What do you think needs to change when it comes to customer loyalty?
3. How do you think about customer loyalty will evolve in the coming years?
4. What is your vision on Big Data today and for the future?
  - a. What is your definition of Big Data?
  - b. What are the main benefits of Big Data according to you?
  - c. What are the main risks of Big Data according to you?
5. On a scale from 1 to 7 are merchants using data to get customer loyalty?
  - a. On a scale from 1 to 7 how efficient are merchants using data to get customer loyalty?
6. How can Big Data help merchants to develop more customer loyalty?
  - a. Which methods or approaches do you think merchants use to obtain value from Big Data and how do think the value of Big Data can be assessed?
7. Should merchants perform Big Data analytics themselves or outsource it? What would your advice be for a merchant?
8. Do you have any other ideas or concerns we have not covered and that could be important?

## **Appendix D: Discussion demographics data collection.**

### **Function**

The representatives of all seven merchants (100%) were from higher management and had the business area of data and customer loyalty in their portfolio. The expert representatives were also 100% from higher management. This will give more value to the answers given in the interview.

### **Nationality**

The nationality of the representatives was asked to detect a bias when necessary. This was used to check if the representative was from the same nationality as the company he was working for. The representatives were all (100%) from the same nationality as the company. Also the nationality of the company that was interviewed was the original country of the merchant for all respondents. When the company has its origin in for example the United States this could influence the vision on this topic. Of the 7 companies interviewed 4 (57%) was Dutch and 3 (43%) was Belgium. The experts were also 100% from the same nationality as the expert company. The expert companies were for 80% Dutch (4) and for 20% not available (1). As mentioned before at two expert companies and one merchant a double interview took place. During these interviews there was one answer given to each questions consisting of the opinion of the two representatives. They came up with the answers together so there were not two separate interviews. This was done to increase the quality because knowledge of both respondents was needed to complete the questionnaire.

### **Where active?**

From the 7 merchants 71% (5) were active in the Netherlands, 86% (6) were active in Belgium and 57% (4) are active in both countries. In table 1 can be found which merchants are active in which countries of interest. Next to Belgium and the Netherlands 86% (6) were active also outside these countries, for one company this only includes Luxembourg. The 4 Dutch experts were only active in the Netherlands. Only one expert company was active in both the Netherlands as Belgium as well as outside these two countries.

### **Retail and expert area**

To give the answers from the merchants and experts the most value it was tried to have representatives from different areas. The companies are all (100%) retail companies but all from different backgrounds. By the experts 40% (2) have a payment background, 40% (2) have a consumer/privacy background and 20% (1) has a hotel and catering industry background. This last one is chosen to include also these companies. It would have taken too much time to interview enough of these mostly smaller companies, with interviewing their branch organization it is tried to include them in this research proposal as good as possible.

## Appendix E: Results of the merchant interviews.

### 1. *How do you think about customer loyalty?*

The first question has given some good supported results. 100% of the merchants mentioned that they think customer loyalty is important or really important. Soft and hard benefits were used by 100% of the merchants to derive customer loyalty (for one merchant the hard benefits were only given outside the Netherlands). All merchants mentioned that to get customer loyalty a proper mix is needed with hard and soft benefits. One merchant (14%) mentioned that they use the data to get better insights in what customers want, and one other merchant (14%) mentioned that customer loyalty is about giving added value for the customer and for the company. The last point made in this question was that loyalty systems should be sustainable; this was mentioned by 29% (2) of the merchants.

### 2. *Do you have a customer loyalty program at this moment?*

#### a. *If yes what type of loyalty program do you have?*

#### b. *If not what do you do to drive behavior of you customers?*

All 7 merchants (100%) mentioned that they have a customer loyalty program. There were some differences in sophistication of these systems. A digital customer loyalty system was used by 86% (6) merchants. From those 6 merchants 50% (3) had also a loyalty card. The seventh merchant only had a loyalty card as customer loyalty system. As mentioned in the previous question all merchant used their customer loyalty system to give a combination of soft and hard benefits to their customers.

### 3. *On a scale of 1 to 7 what is your view of the current amount of customer loyalty that is being provide to your customers?*

The answers to this question ranged from 3 to 6. The answers were 3, 3, 3, 4, 5, 6 and 6. So the average answer for this question is 4,3. Meaning that at this moment the merchants on average are giving moderate customer loyalty to their customers leaving quite a bit room for improvement. There is also a bit of variance in the answers to this question with 2,24 and a standard deviation of 1,50.

### 4. *On a scale from 1 to 7 how important is customer loyalty for your company?*

#### a. *On a scale from 1 to 7 how important are customer loyalty programs for your company?*

Customer loyalty was very important for all 7 merchants. There answers ranged from 6 to 7. The answers were 6, 6, 7, 7, 7, 7 and 7. So the average answer for this question is 6,7. There is not much variance in the answers of the merchants for this question with 1,16 and a standard deviation of 1,08.

That there is some improvement to make with the usage of the customer loyalty programs can be seen in the answer to the sub questions. The answers were 3, 3, 4, 4, 6, 6 and 7. So the average answer of a merchant for this question is 4,7. Here the answers ranged from 3 to 7 so quite a bit of variance in the answers. The variance is 2,88 and the standard deviation is 1,70.

*5. On a scale from 1 to 7 to which extent are you using data to get customer loyalty?*

All merchant were using data for their customer loyalty programs. But there was a difference to which extent they were using data. The answers were 3, 3, 5, 5, 5, 5 and 7. Making the average answer of a merchant for this question 4,7. The answers ranged from 3 to 7 so again quite a bit of variance in the answers but less than with the previous question, the variance is 2,30 and the standard deviation 1,51.

*6. What is your vision on Big Data today and for the future?*

- a. What is your definition of Big Data?*
- b. What are the main benefits of Big Data for you?*
- c. What are the main risks of Big Data for you*

The vision on Big Data of the merchants was quite clear. All 7 (100%) said that Big Data is just in the beginning phase at this moment and will become more important in the future. Big Data is not fully regulated at this moment which could cause problems was mentioned by 29% (2) of the merchants.

Big Data was defined as very large amounts of data by 86% (6) of the merchant. The usage of unstructured data sources was mentioned by 29% (2). The usage of analytics to get information out of the data was mentioned by 71% (5). There was 1 merchant 14% who mentioned that Big Data was about fully using the information of the customer and give added value for the customer. That the data should be from more than one resource was mentioned by 29% (2). So apart from that Big Data is about large volumes there was no broadly mentioned definition. The three V's model from literature was not mentioned by anyone.

The main benefit of Big Data for merchants was for almost all of them better insights in different areas for example customer behavior, 100% (7). It could be used for better communication was mentioned by 86% (6). That it could be used for more efficiency was mentioned by 57% (4). Two merchants 29% mentioned that it could be used for higher quality data.

Privacy was the biggest risk for the merchants and was mentioned by all of them (100%). The incorrectness of the data and that there could be false conclusions based on this wrong data was mentioned by 43% (3) of the merchants. Security problems were identified as a risk by also 43% (3) of the merchants. Two merchants 29% mentioned misuse of the data as risk. One merchant 14% mentioned image damage as risk. So the risks that merchants mentioned differed quite a lot between merchants but privacy was their biggest concern.

*7. On a scale from 1 to 7 how many experience/knowledge do you have with Big Data?*

All merchants have experience and knowledge of Big Data analysis because no one gave a 1 as answer. The answers were 3, 4, 5, 5, 5, 6 and 7. Making the average answer of a merchant for this question 5,0. So on average the merchants have above average experience with Big Data. The answers ranged from 3 to 7 so there was quite some variance in this question. The variance is 2,14 and the standard deviation 1,46.



*8. On a scale from 1 to 7 do you think you can perform Big Data Analytics in-house?*

With this question even more variance came in the answers. The answers were 2, 5, 6, 7, 7, 7 and 7. The average of this questions therefore is a 5,8. But some merchants already mentioned that they will perform their analytics in-house but also with knowledge from outside the company. The variance is 3,81 and the standard deviation is 1,95.

*9. How can Big Data help your company to develop more customer loyalty?*

*a. Which methods or approaches do you currently (or in the future) use to obtain value from Big Data and how do you asses the value of Big Data?*

The 7 merchants listed a number of options here. All 7 (100%) mentioned that Big Data can be used to make the offers more personalized and gain more customer loyalty by doing that. Target the right customer was mentioned by 86% (6). Making better offers was mentioned by 71% (5). The last usage of Big Data that was mentioned was that it could help to make a better segmentation by 2 merchants (29%). Because some overlay existed with previous questions not everything was listed again. Many merchant referred back to what was already listed earlier in for example 6b.

The respondents did not really want to tell or knew what type of methods ore approaches they used. Knowledge from the past was however mentioned by 43% (3).

All merchants (100%) said that at this time it is too early to really asses the value of Big Data. One merchant compared it with mobile applications. Some time ago you could also not measure the value off Apps but we are glad that we invested in that because it pays off now. The same is expected with Big Data. Control group analysis was mentioned by 43% (3) of the merchants. LEAN management technology, customer survey and cost benefit analysis were all mentioned by 1 (14%) merchant.

*10. On a scale from 1 to 7 would you be willing to provide data to an external party if this could improve consumer loyalty and/or consumer benefit?*

The merchants were not very positive to the idea of an external party. The highest number as answer was a 5. The answers were 1, 1, 2, 3, 3, 4, and 5. So there is quite some variance in the answers. The variance is 2,31 and the standard deviation 1,52. The average answer is 2,7. So giving data to an external party is not something merchants in general will like to do even if this can improve customer benefits.

*11. Under which circumstances would you be willing to share your data with a third party to enhance the customer loyalty?*

The answer to this question differed a lot between merchants. Absolutely not was the answer of 43% (3). If it gives added value that we could not get ourselves was the answer of 57% (4). There also three statements mentioned by only one merchant 14%: the benefits should be significant, the benefits should give added value for the customers and we are only looking for specialist and not generalist.

*12. On a scale from 1 to 7 would you be MORE WILLING to provide data if it is guaranteed that this data is only being used to the benefits of your consumers.*

When this guarantee is given it becomes even more interesting than in question 10, for some merchants this made no difference but for others it made. The answers are 1, 1, 3, 4, 4, 6 and 7. The average answers is 3,7. So looking at the average it seems that the answer lays almost perfectly in the middle of the scale. But when you look to the real numbers it can be seen that the answers differ a lot between merchants. This resulted in the most variance in answers of all questions. The variance was 5,02 and the standard deviation 2,24.

*13. Assuming that you would be willing to share the data of your company with a third party to enhance customer loyalty, what would you expect the third party to do and not to do with the data?*

In answering this question 1 merchant 14% was very clear. In their policy it is mentioned that they do not share their data so this question could not be answered. The remaining 86% (6) mentioned that there should be a contract. In this contract it should be mentioned that the third party only does what the merchant is saying, only use the data for the actual case, ensure safety, handle the data with care and look at the privacy of the customers. Three merchants 43% mentioned that the data should be destroyed after usage. There should not be information that could be back traced to one individual and that the customer has to approve was mentioned by 29% (2). The servers should be in Europe, The tool must adapt to our business model and not the other way round, we want to see how the metrics are calculated and it should help the customer with a problem was mentioned by only 14% (1) merchant. So there are a few general concerns that all merchant have and a few more specific ones.

*14. Do you have any other ideas or concerns we have not covered and that could be important?*

There were no further ideas or concerns by merchants which were not already discussed during the interview.

## Appendix F: Results of the expert interviews.

1. *On a scale of 1 to 7 what is your view of the current amount of customer loyalty that is being provide to customers in general by merchants?*

The answers were 2, 4, 4, 5 and 5. The average answer of the experts was 4. The variance of this question was quite unusual with 1 low number and 4 moderate answers resulting in a variance of 1,2 and a standard deviation of 1,10. So the experts think that the amount of customer loyalty given in general by merchants to customer is average. There is some room for improvement here for the merchants according to the experts.

2. *What do you think needs to change when it comes to customer loyalty?*

Merchant should use more of the data that they collect and loyalty is more than just discount are two things that 80% (4) of the experts mentioned in their answer. Extra service should be more important was mentioned by 60% (3) experts. Merchants should be more open what they do with the data, merchants should try to find new customers and also hard benefits should be important was mentioned by 40% (2) experts. One expert (20%) mentioned that loyalty should be more demand oriented and there should be a shared database for moderate and little companies in the Netherlands. Another expert (20%) mentioned that merchant should utilize the channels that exist more. So there were quite some changes that merchants should consider regarding their customer loyalty programs according to the experts that were interviewed.

3. *How do you think about customer loyalty will evolve in the coming years?*

The experts had different opinions about the evolvement of customer loyalty resulting in many topics that are only mentioned by one or two experts. They all mentioned (100%) that data will become even more important for customer loyalty in the future. That customer loyalty will be more important in the future was mentioned by 60% (3) experts. That soft benefits should become more important, that there will be new ways of loyalty, that there should be more attention for the customer and that loyalty programs will move towards phones was mentioned by 60% (3).

4. *What is your vision on Big Data today and for the future?*
  - a. *What is your definition of Big Data?*
  - b. *What are the main benefits of Big Data according to you?*
  - c. *What are the main risks of Big Data according you?*

All experts (100%) mentioned that at this moment Big Data usage by merchants is very limited and it will be more in the future. That a lot of data is already collected; that it will become more open sourced and that the resistance will drop was mentioned by one expert (20%) each.

The definition of the Big Data was: collection of all kinds of data, analyze the data and the data could be unstructured according to 80% (4) of the experts. It should consider potential data, so data that is not used at this moment was mentioned by 40% (2) of the experts. That Big Data includes taking action after the analysis was mentioned by one (20%) expert.

The benefits of Big Data could be that offers can be more personalized and that there will come more knowledge about customers according to all (100%) experts. 4 experts (80%) mentioned that there will come more knowledge about where and whom of customers, for example via WiFi tracking. One expert (20%) mentioned that a benefit could be bringing back trust in the merchants and that merchants should no longer be place restricted in their operations.

The experts also mentioned some risks for merchants. Privacy concerns are mentioned by 80% (4) also mentioned as Big Brother issues. Taking the wrong conclusions was mentioned by 40% (2). By only one expert (20%) each was mentioned that customer contact can become less personal and possible reputation damage when using Big Data.

*5. On a scale from 1 to 7 are merchants using data to get customer loyalty?*

*a. On a scale from 1 to 7 how efficient are merchants using data to get customer loyalty?*

The answers to this question were again widely spread just like question 1. The answers were 2, 5, 5, 6 and 6 so again high with one low answer. On average this makes a 4,8 meaning that more than half of the merchants is using data for their customer loyalty program according to the insights of these experts. The variance was higher than in question one with an variance of 2,16 and a standard deviation of 1,47.

The efficiency is however a bit lower. One expert did not know enough for this question and preferred to not answer it. The answers to this question were: 1, 2, 2 and 3. So the average answer is a 2 which means that the merchants are not very efficient in using data to get customer loyalty. The variance was 0,5 and the standard deviation 0,71.

*6. How can Big Data help merchants to develop more customer loyalty?*

*a. Which methods or approaches do you think merchants use to obtain value from Big Data and how do think the value of Big Data can be assessed?*

All experts (100%) mentioned the same here. Big Data can help merchants to develop more customer loyalty by making it easier to differentiate between customers and for example derive customer segments, by knowing the needs of such a customer segment and make more personalized offers. Not one of the experts was capable to answer the sub question. They had no clear image of what was used in the market.

*7. Should merchants perform Big Data analytics themselves or outsource it? What would your advice be for a merchant?*

The experts had quite different opinion on the answer of this question. Three of them (60%) mentioned that this is depending on the size of the merchant. The following statements were covered by 2 experts per statement: perform the analytics yourself, buy help from experts but do it yourself, perform it in-house, do it not in-house and it is not relevant from a law perspective because you will remain responsible for the data. One expert (20%) mentioned that it could maybe done via a cooperation.

*8. Do you have any other ideas or concerns we have not covered and that could be important?*

Two experts (40%) emphasized that personal data should be handled with care. One expert (20%) emphasized again the importance of the combination of hard and soft benefits. One last expert mentioned that data from for example a payment provider could be used cross industry for benchmarking purposes.

## Appendix G: Questionnaire validation.

As mentioned in the Thesis this time the insights of the data collection that needed validation are asked for. Therefore no different questionnaire was made for the experts; so both merchants and experts were given the same questionnaire. The questions will give insights for the parts needed validation as discussed in paragraph 6.1. As explained in chapter 6 the questions have an introduction to give the respondents extra information about the context of the questions.

1. Usage of Big Data could lead to insights via an analysis. Three usage areas of Big Data in the retail will be discussed. The first usage area is the increase of customer loyalty. Big Data can for example help to improve customer loyalty programs and make customer loyalty programs for example more personal. On a scale from 1 to 7 do you think this is a usage area of Big Data in the retail sector?
  - a. Why do you think that?
  - b. Customer loyalty can be divided into hard en soft benefits. In this research project hard benefits are defined as monetary benefits and soft benefits as non-monetary (service) benefits. On a scale from 1 to 7 do you think Big Data can be used for hard benefits?
  - c. On a scale from 1 to 7 do you think Big Data can be used for soft benefits?
2. The second usage area is the process of optimize in-shop customer experience which can be on and offline; for example by changing the shop outlay after data analysis. On a scale from 1 to 7 do you think this is a usage area of Big Data in the retail sector?
  - a. Why do you think that?
  - b. The optimizing process can lead to the process of reduction in costs by optimizing business processes. On a scale from 1 to 7 do you think that it can lead to a reduction in costs?
  - c. It can also start the process of getting more and larger orders which results in the process of increase revenues. On a scale from 1 to 7 do you think that it can lead to an increase of revenue via more and/or larger orders?
3. The third usage area is the benchmarking. A payment provider as outsider can analyze data from a store to a generic industry benchmark. On a scale from 1 to 7 do you think this is a usage area of Big Data in the retail sector?
  - a. Why do you think that?
  - b. The benchmark process in detail contains the following processes by your service provider: analyze your payment data, getting insights, compare with generic industry benchmark and send data results to merchants. Do you think this is the complete and correct process? If not please explain your answer; what do you think does have to change?
  - c. After the data results are send to the merchant it is used by the merchant to improve their business processes this can lead to a reduction in costs. On a scale from 1 to 7 do you think the usage of benchmark results by merchants can lead to a reduction in costs?
  - d. The benchmark results could also lead to the process of increase of revenue. On a scale from 1 to 7 do you think the usage of benchmark results by merchants can lead to increased revenue?

4. The processes concerning Big Data are in essence not different from normal data usage. The input, output, software used, throughput time will be different with for example larger data sets from various resources and real-time insights. But the processes in essence stay the same. On a scale from 1 to 7 do you think this is a correct conclusion?
  - a. Why do you think that?
5. Are there any other things you would like to mention which are not discussed in the interview so far? Are there for example any important other usage areas of Big Data in the retail sector which should be implemented in the process model?

## **Appendix H: Discussion demographics validation**

### **Companies and function**

The respondents of the validation group consist of 2 merchants and 8 experts. As mentioned before because the benchmark process needed the most validation a lot of experts with expertise in this area were contacted for the validation. The two merchants differed in either size or retail area, compared to the first group of merchants interviewed. The experts had expertise in all areas of the five stakeholders. Different respondents were used for the validation compared to the first interviews. The respondents were all (100%) from higher management. This will give more value to the answers given in the interviews.

### **Nationality**

The nationality of the representatives was asked to detect a bias when necessary. This was used to check if the representative was from the same nationality as the company he was working for. The representatives were all (100%) from the same nationality as the company they were working for. Also the nationality of the company that was interviewed was also 100% the original country of the merchant. When the company has an origin in for example the United States this could influence the vision on this topic. From the experts 6 (87,5%) was Dutch or Belgium. Both Merchants were both Dutch.

### **Where active?**

The medium and small merchants were both (100%) only active in the Netherlands so not in Belgium or elsewhere in Europe. From the experts' 100% was active in the Netherlands. There were 4 (50%) experts from companies which are only active in the Netherlands, the other 4 (50%) are also active in Belgium and the rest of Europe.

### **Retail and Expert area**

To give the insights from the companies and experts the most value for this Thesis it was tried to have representatives from different areas. The companies are all (100%) retail companies but all from different backgrounds. From the experts 75% had a background in banking or payments, 75% had a background in retail 25% had a background in customer knowledge and 12,5% had a background in law. The experts chosen for this validation had all expertise in a combination of areas making them very interesting for this master Thesis project. Also an expert was found from a branch organization. With this expert the meaning of a whole set of merchants was in the validation.



## Appendix I: Results of the validation interviews

These are the full results of the interviews. The full questionnaire can be found in appendix G. The questions needed some introduction because they were quite specific. Only the questions will be mentioned here the introductions can be found in the full questionnaire. The variance and standard deviation of the closed questions is given, but just as with the first questionnaire the statistical value is not very high due to the low number of respondents.

1. *On a scale from 1 to 7 do you think that increase of customer loyalty is a usage area of Big Data in the retail sector? And why do you think that?*
  - a. *On a scale from 1 to 7 do you think Big Data can be used for hard benefits?*
  - b. *On a scale from 1 to 7 do you think Big Data can be used for soft benefits?*

All (100%) respondents agreed that this was a usage area of Big Data in the retail sector. The answers were 5, 6, 6, 6, 6, 6, 7, 7, 7 and 7. The average answer to this question of the respondents is a 6,3 and the variance is 0,41 and standard deviation is 0,64 which is very low. All respondents (100%) mentioned that in the retail sector it is very important or crucial to have knowledge about the customer, the more the better. They also mentioned all that Big Data can be used for making customer loyalty more personalized. Three respondents (30%) mentioned that it will depend on the size of the retailer and the retail area. The following statements were only mentioned by one respondent: It will be better if data will be shared between stores, multi-channel management will be important and it will help you with being unique which is very important in retail.

The answers if Big Data can be used for hard benefits were 3, 4, 5, 5, 5, 5, 6, 7, 7 and 7. The respondents on average give a 5,4. There is quite some variance 1,64 in the answers. The standard deviation is 1,28. For soft benefits the answers were 5, 6, 6, 7, 7, 7, 7, 7 and 7. Making the answer on average a 6,6. The variance and standard deviation for this question were low with a variation of 0,44 and a standard deviation of 0,66. So the respondents on average think that Big Data can be used for both hard and soft benefits. But they think that Big Data will have more value for soft benefits then for hard benefits. They mentioned that for soft benefits it is even more important to make it personalized which is possible with Big Data. For hard benefits it is easier to already give them in a good way to customers without Big Data so Big Data will have some impact and can be used but the extra added value is a bit lower.

2. *On a scale from 1 to 7 do you think optimizing of in-shop customer experience is a usage area of Big Data in the retail sector? And why do you think that?*
  - a. *On a scale from 1 to 7 do you think that it can lead to a reduction in costs?*
  - b. *On a scale from 1 to 7 do you think that it can lead to an increase in revenue?*

The respondents gave the following answers to this question: 3, 5, 6, 6, 6, 6, 7, 7, 7 and 7. The average answer was 6,0. The variance in the answers is quite severe due to the answer of 3. The variance is 1,4 and the standard deviation is 1,18. The respondent with a 3 as answer mentioned that Big Data will make a difference here but the amount of impact was not very high in his opinion. The other 9 respondents mentioned that it can be very helpful to know what triggers customers in shops (off and online). When you know how they respond you know where to improve, Big Data can help to improve

this. It will be especially important in supermarkets or larger stores, was mentioned by 3 respondents (30%). The shopping experience will become more and more important was mentioned by 20%.

The answers to the question if this can lead to a reduction of costs were: 2, 4, 5, 5, 5, 5, 6, 6, 6 and 6. So on average the respondents do think that it can lead to a reduction of costs with an answer of 5,0. The variance in the answers is large with answers ranging from 2 till 6 resulting in a variance of 1,4 and standard deviation of 1,18. Respondents mentioned that they think cost reduction is possible but harder than increased revenue. Examples mentioned with cost reduction are with optimizing stock levels for example of rotting products or improve the number off employees.

The respondents think that it is best suitable for increasing of revenue. The answers to this question were 4, 6, 6, 6, 6, 6, 7, 7, 7 and 7. So the answer is on average a 6,2 which is higher than for cost reduction. The variance in answers is also very low with 90% giving a 6 or a 7, this results in a variance of only 0,76 and standard deviation of 0,87. The respondent who gave a 4 mentioned that increase in-shop customer experience may cost more than the rewards are so the revenue will increase but it may result negative in the end. It still could be interesting because it can increase loyalty towards a retailer, but not necessarily increases the revenue.

3. *On a scale from 1 to 7 do you think benchmarking is a usage area of Big Data in the retail sector? And why do you think that?*
  - a. *The usage area benchmarking contains the following processes: analyze payment data, getting insights, compare with generic industry benchmark and send data back to merchants. Do you think this is the complete process? If not please explain your answer; what do you think does have to change?*
  - b. *On a scale from 1 to 7 do you think benchmark results can lead to a reduction of costs?*
  - c. *On a scale from 1 to 7 do you think benchmark results can lead to an increase of revenues?*

The respondents answered with: 5, 5, 6, 6, 6, 6, 6, 6, 6 and 7. On average their answer was 5,9 with very little variance in their answers only 0,29 and standard deviation of 0,54. They all (100%) mentioned that it always good to compare with an industry benchmark and that insights from this analysis could be very useful. The amount of detail would be very important for the usefulness was mentioned by 6 (60%). Three (30%) mentioned that using more data sources will make the benchmark results more valuable, but 20% (2) mentioned that the question is if this is doable also from a data ownership point of view.

The four processes are correct was said by 100% of the respondents. Four (40%) however did mention that the processes are correct but there should be some processes added. They mentioned that an extra giving advice processes could be added. Payment providers have a lot of knowledge which they can use to add advice to the benchmark results making the results more valuable. They also mentioned that there maybe are some processes before to add other data to the benchmark analysis.

The respondents did think that it can lead to a reduction of costs and increase of revenues but the amount differed. The answers for reduce costs were: 2, 2, 2, 3, 4, 5, 5, 6, 6 and 7. So on average they mentioned a 4,2. What is interesting also is the amount of variance; with answers from 2 till 7 resulting

in a variance of 3,16 and a standard deviation of 1,78. For increase of revenue the answers were: 4, 4, 6, 6, 6, 6, 7, 7, 7 and 7. So on average a 6,0 with not that much variance with 80% of the answers 6 or 7 resulting in a variance of 1,2 and standard deviation of 1,10. So both are possible but increased revenue are easier to derive than a reduction of costs.

4. On a scale from 1 to 7 do you think the processes in essence between data and Big Data do not differ? And why do you think that?

The respondents answered with 3, 4, 5, 5, 5, 6, 7, 7, 7 and 7. So on average the answer is a 5,6 with quite some variance with answers ranging from 3 till 7 which resulted in a variance of 1,84 and a standard deviation of 1,35. They all agreed (100%) that the processes at this moment are the same for Big Data and data in the retail sector. Only 6 (60%) mentioned that that will be also in the coming years where 40% mentioned that they think that that will change slightly in the coming years. Major changes in the processes are not expected in the near future but maybe later. Three respondents (30%) mentioned that more pre-work has to be done before data can be used which is a real difference between data usage and Big Data usage. They suggest add a step preparation of data.

5. *Are there any other points not mentioned in this interview that you would like to discuss? For example other important usage areas?*

Most of the respondents (60%) already gave all their comments in the previous questions. The other respondents gave some extra comments. There will be a lot of change in the processes itself although they stay in essence the same. For digital stores fraud detection can be another important usage area and maybe also Big Data for online stores can be used for fraud detection. Besides increasing of revenue and reduction of costs; increased loyalty will change the Net Promoter Score which can be very interesting and helpful for stores. The Net Promoter Score is the metric that measures if people would recommend a company/product or service (In this case a retailer) to a friend or colleague. The right people are very important when dealing with Big Data was also mentioned.

## Appendix J: Figures

J1: Figure 5

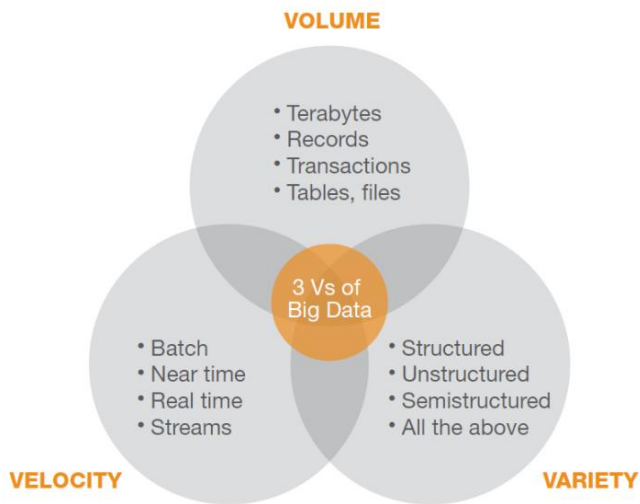


Figure 5. The three V's of Big Data (Won, 2013)

J2: Figure 6

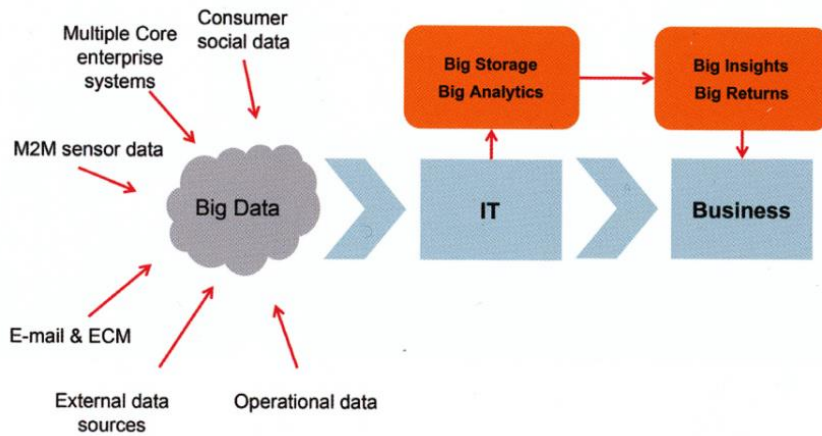


Figure 6. Big Data to Business flow (CGI p72, 2013)

J3: Figure 7

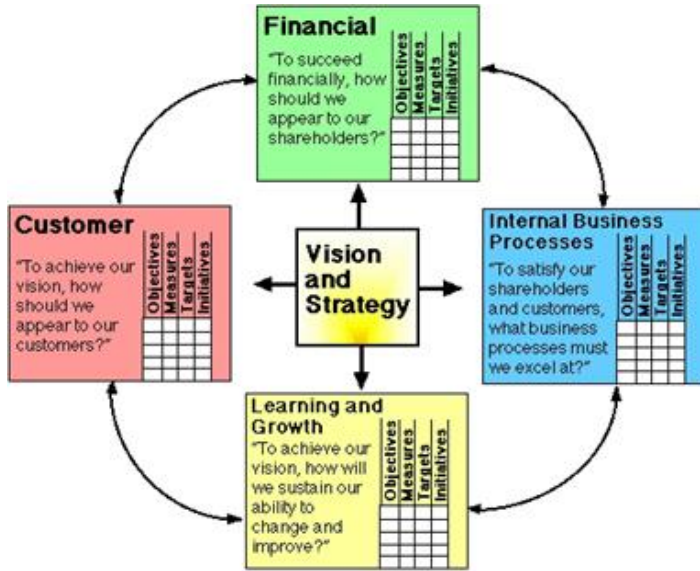


Figure 7. The balanced scorecard of Kaplan and Norton (1987)

J4: Figure 8

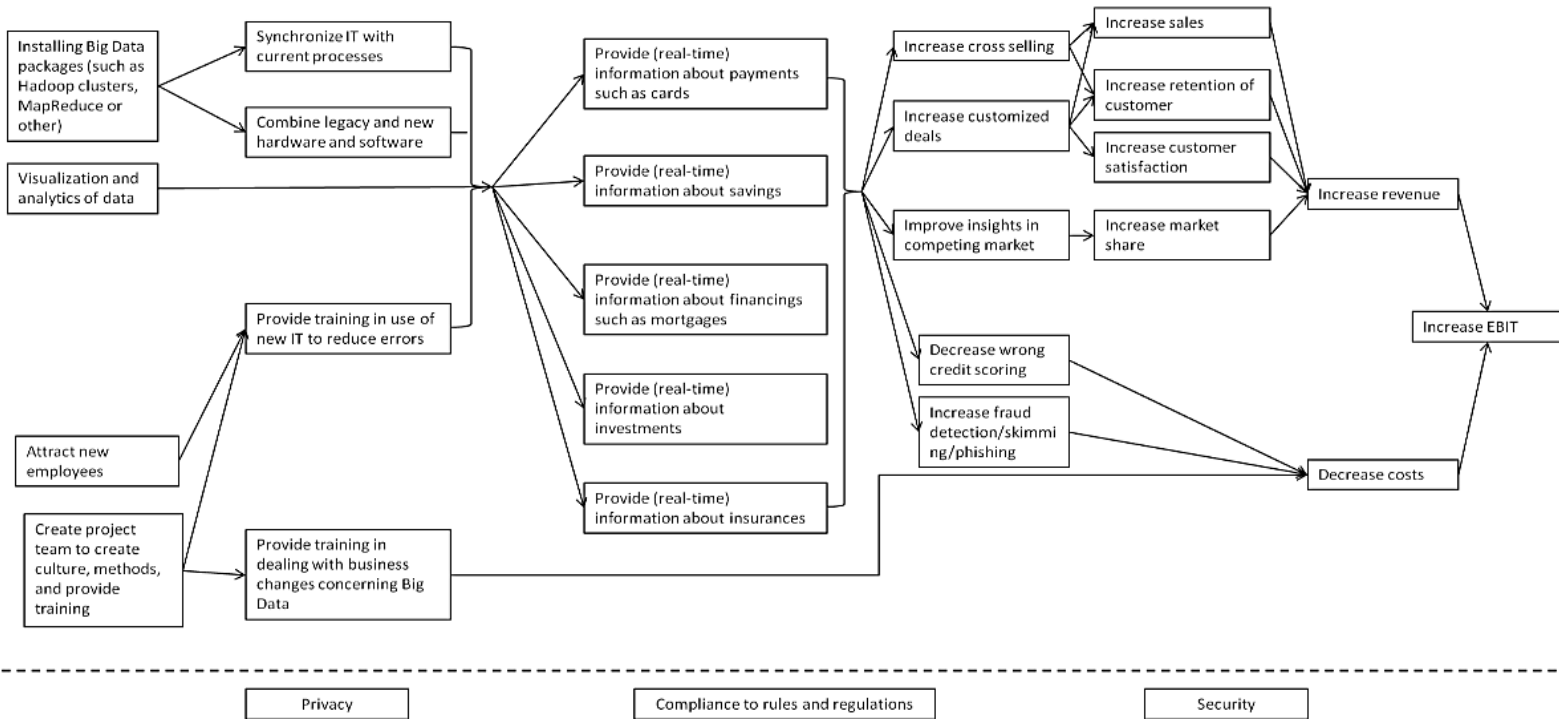
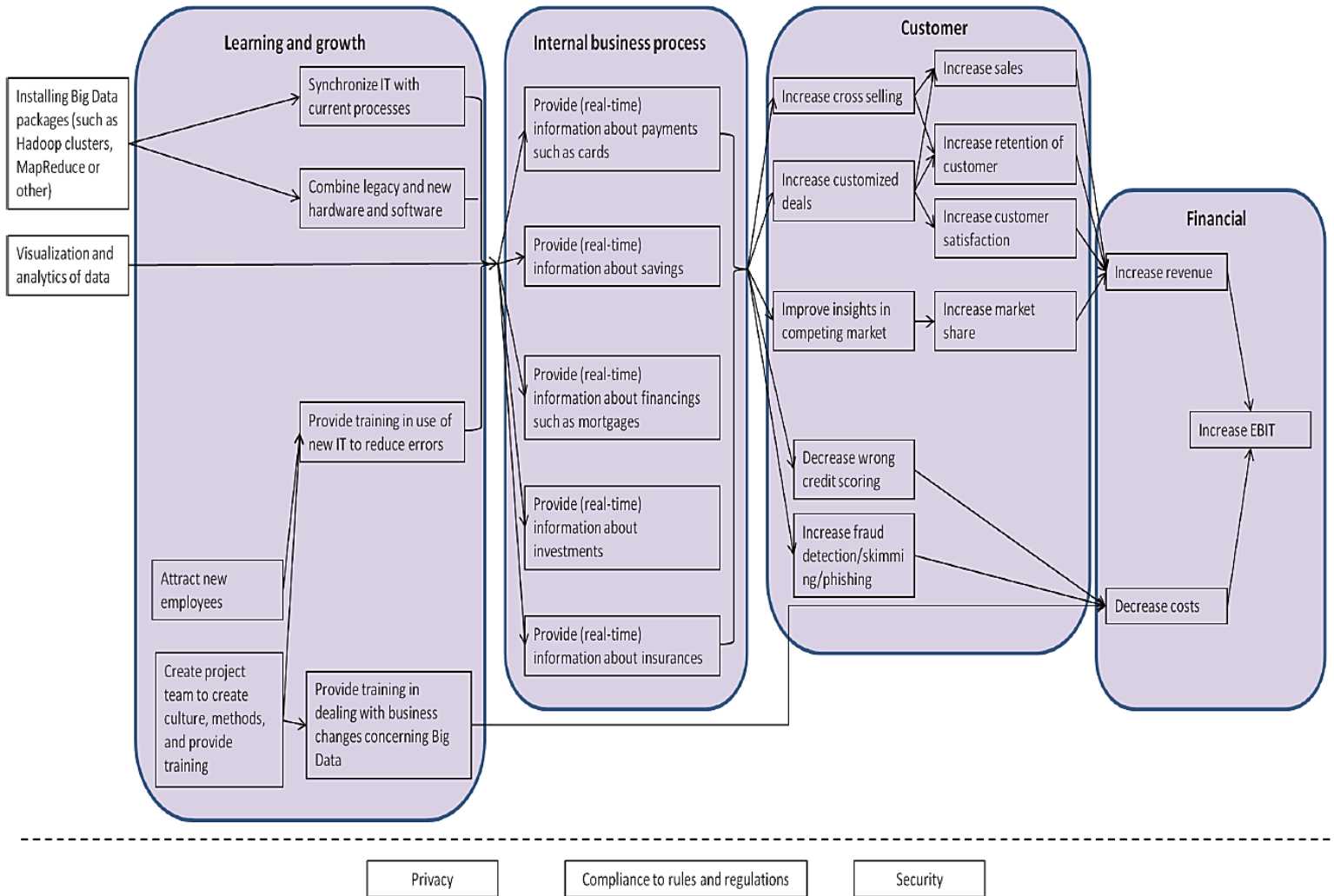


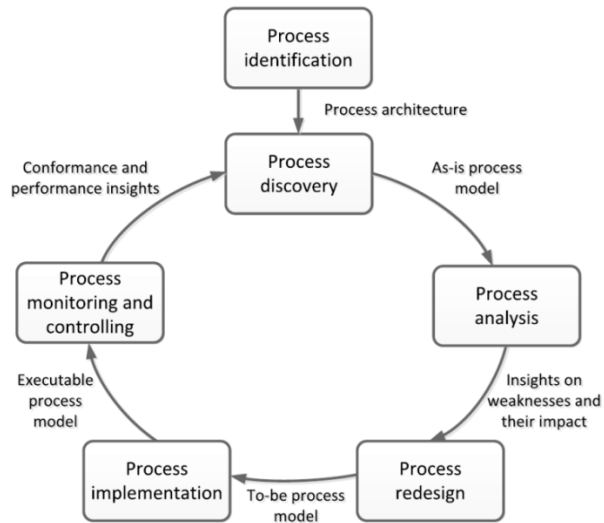
Figure 8. Big Data Value Framework for the Financial Services Industry (De Vries p42, 2013)

**J5: Figure 9**



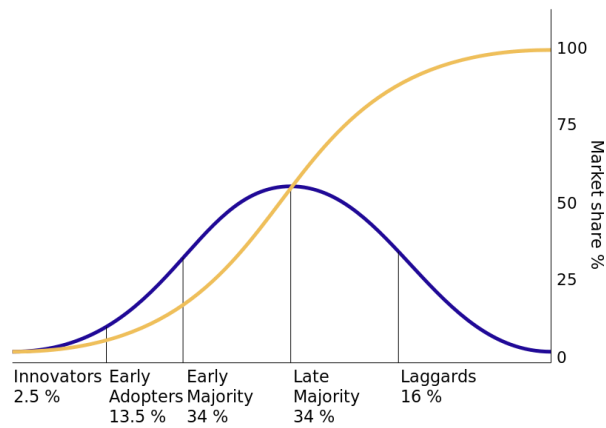
**Figure 9. Big Data Value Framework for the Financial Services Industry divided into four perspectives of Balanced Scorecard (De Vries p44, 2013)**

**J6: Figure 10**



**Figure 10. The BPM Lifecycle (Dumas et al. p21, 2013)**

**J7: Figure 11**



**Figure 11. Innovation Diffusion Model of Rogers (1962)**

J8: Figure 12

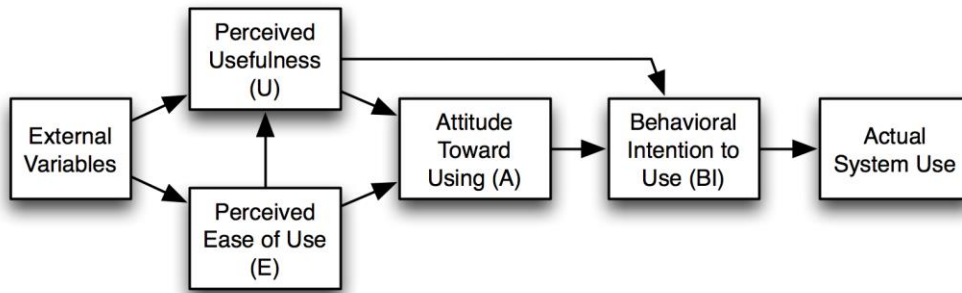


Figure 12. The Technology Acceptance Model of Davis (1989).

J9: Figure 13

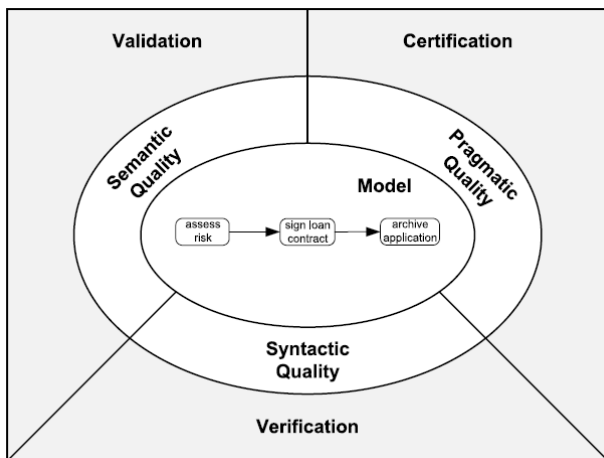


Figure 13. Quality aspects and quality assurance activities (Dumas et al. p172, 2013)



J10: Figure 14

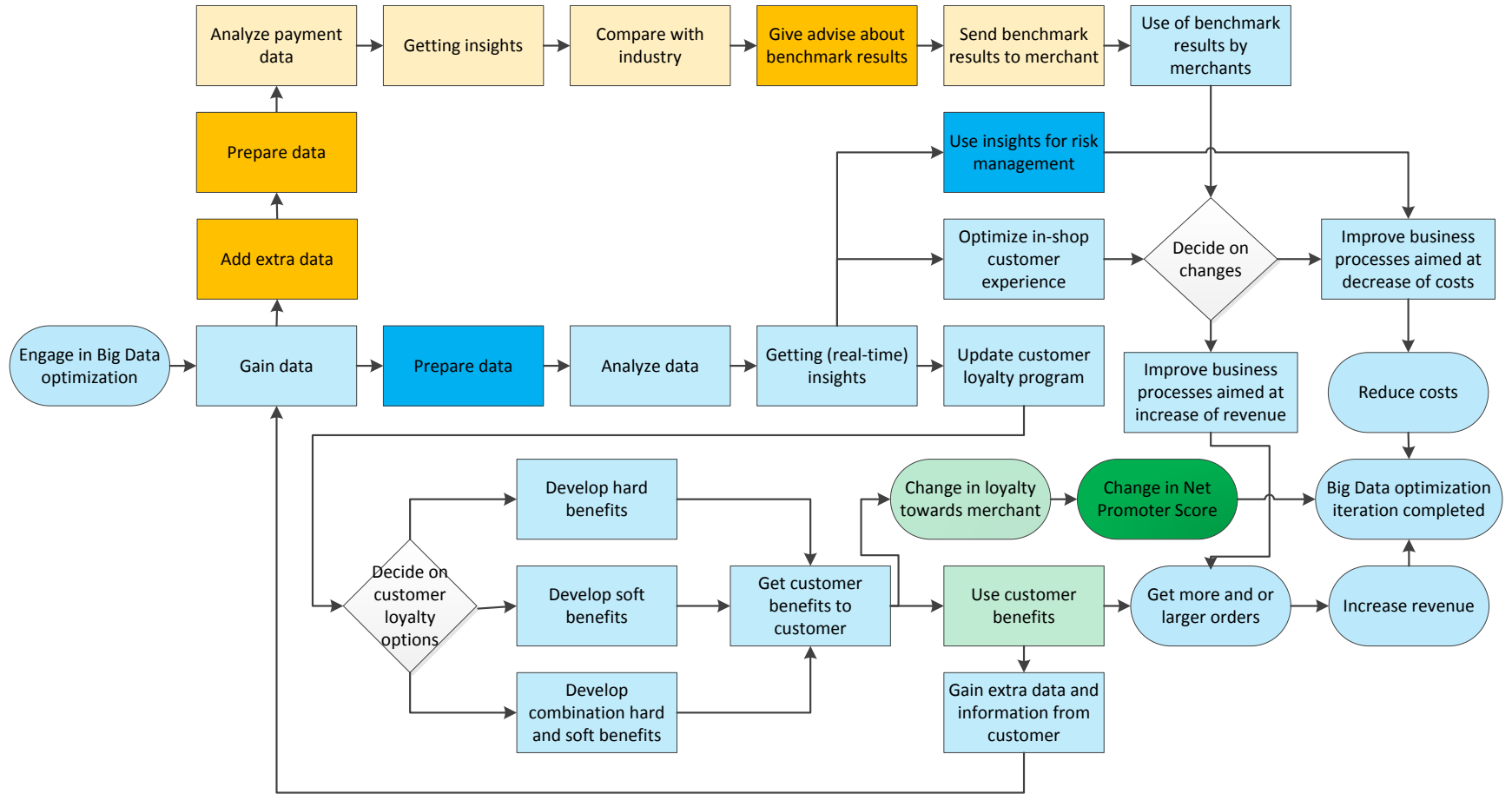


Figure 14. Adjustments to the process model