SCHOOL OF BUSINESS, MANAGEMENT AND SOCIAL SCIENCES

UNIVERSITY OF TWENTE

THE EFFECTIVE INTEGRATION OF BIG DATA IN THE DECISION-MAKING PROCESS

Student name:

Ana Maria Azevedo Amaral

MSc Business Administration

Digital Business track

First supervisor:

Dr. Matthias de Visser

Second supervisor:

Dr. Michel L. Ehrenhard

24 of August of 2020

Acknowledgments

"Enthusiastic skepticism is not the enemy of boundless optimism. It's optimism's perfect partner. It unlocks the potential in every idea."

- Astro Teller. TED2016[1]

This master thesis marks the end of an interesting and definitely unpredictable year. During this year I could dig deeper into subjects that have drawn my attention during my early career and helped me shape the vision of what I would like to do next. The topic big data and decisions was not necessarily new for me, as it was pretty much part of my everyday job. But this work allowed me to gain new perspectives and frames of reference to assess the power and impacts of big data.

I have always believed in the power of innovation and technology to move society forward. When dealing with topics that can be disruptive and have broader impacts, as I think is the case of the big data phenomenon, reminding to balance our boundless optimism with enthusiastic skepticism can perhaps offer a good way further, freeing us from the hyped expectations without making us lose the zeal in keeping the process going. I hope that the discussions here can help companies and others studying this subject to give a first step in the direction of smarter decisions fueled by data. Who knows what big problems big data will help us solve?

Without further ado, I would like to thank all the ones that were part of this journey. To my supervisor, which guidance and precise feedbacks helped me navigate the process of writing an academic piece. To all the interview participants, the gladly offered me the time for such interesting conversations. On the bright side, the need to move the thesis activities online offered me the possibility to involve companies in my home country and, consequently, bring this necessary discussion there, which I'm glad it was the case. To the University of Twente, that with their international orientation and extensive scholarship programs made it possible for me to pursue a master abroad. Last and most importantly, to my family and friends, who are the ones supporting me every step of the way, even if from a distance. They help keep my boundless optimism in check or to fuel it when I need to be reminded that big achievements are just a matter of having concrete big plans. Thank you all.

[1] https://www.ted.com/talks/astro_teller_the_unexpected_benefit_of_celebrating_failure

Abstract

Big data gained a hype status over the past years, but it is here to stay. Data is said to be the oil of the XXI century and authors seem to agree that the availability of data represents a management revolution through facilitated and enhanced decision-making. Big data promises to deliver better business performance, superior product development, and higher customer satisfaction. But this can only be true if organizations can effectively transform big data insights into business value. Recent evidence indicates that the first-order effects of big data are on better decision-making, and conceptual models to integrate big data tools, methods, and techniques into the decision-making process has been developed. Thus, this research goal is to assess the applicability of the existing models in real-life cases in the context of product/market development, characterized by unstructured and nonprogrammed decisions. Also, it has a secondary goal to shed light on the organizational elements surrounding this process, in particular, the stakeholders and the information-processing mechanisms. For that, a multiple case study was conducted with large companies in the Netherlands and Brazil. Semi-structured interviews were used to assess the process thought which decisions are made and map the organizational elements. The real-life processes were compared with the literature and a refined model is suggested for the integration of big data in the decision-making process. Furthermore, this research discusses the roles of the stakeholders in the different steps of this process and the information processing mechanisms utilize to gain alignment and define problems and solutions. It is claimed that the decision flow alone cannot guarantee the success of big data usage and knowing the right stakeholders to involve and how to attend to their different information needs are paramount for extracting value from data.

Keywords: big data; decision-making; decision process; information processing mechanisms; stakeholders; new product development; models; unstructured decisions

Management Summary

Context

Big data is a new phenomenon and it is here to stay. It promises to deliver higher performance, better products, and increased customer satisfaction. For that to be true, companies need to learn how to extract value from Big Data. Recent research is showing that the first-order effects of big data are on superior decision-making, in a process that can transform data into insight and insight into decisions, with value being attained with the implementation of these smarter decisions. However, insights cannot be extracted from raw data solely with the use of big data & analytics tools, being an outcome of a more complex process involving the collaboration and engagement between business and analysis in already existing structures to extract knowledge out of data. How to do so has been the subject of increasing research.

Research Problem

Over the recent years, researchers have proposed different conceptual models and frameworks for the integration of big data in the decision-making process. These models still lack a thorough assessment and insights from practice have only been limited used in their construction. At the same time, researchers and practice alike still claim that integrating big data & analytics into the decision process remains a challenge for many companies. Furthermore, the existing models still present divergencies among each other, which can make it hard for companies looking for making better use of data for decision-making to identify what to do. Therefore, it is claimed that, if the existing models are to help companies in implementing better data-based decision-making processes, efforts should be made to assess whether and which of these models possess the necessary attributes to support their claim.

Furthermore, while understanding the mechanisms and processes through which big data can add value to the business, it is also important to have a clear picture of its different elements. More than knowing the steps by step process of making a decision, it is important to understand how the different actors involved in this process interact to clarify the decisions, resolve conflicting points of view, and generate alignment. Therefore, comprehending the organizational elements impacting this process beyond the decision flow can add value to the discussion of how to effectively integrate big data into the decision-making process. This involves investigating the stakeholders involved in each step of the process and learning how they communicate and collaborate to leverage the available data throughout the decision process.

This translates to the following research question and sub-questions:

> *How well do existing models for the integration of big data in the decision-making process fit practical use cases?*
>
> > *How well do the steps described in these models converge with the steps described in practical use cases and what other steps need to be taken into consideration?*
> >
> > *What are the stakeholders involved in the different phases of this process?*

Which information-processing mechanisms are utilized in the context of the decision-making process?

Methodology

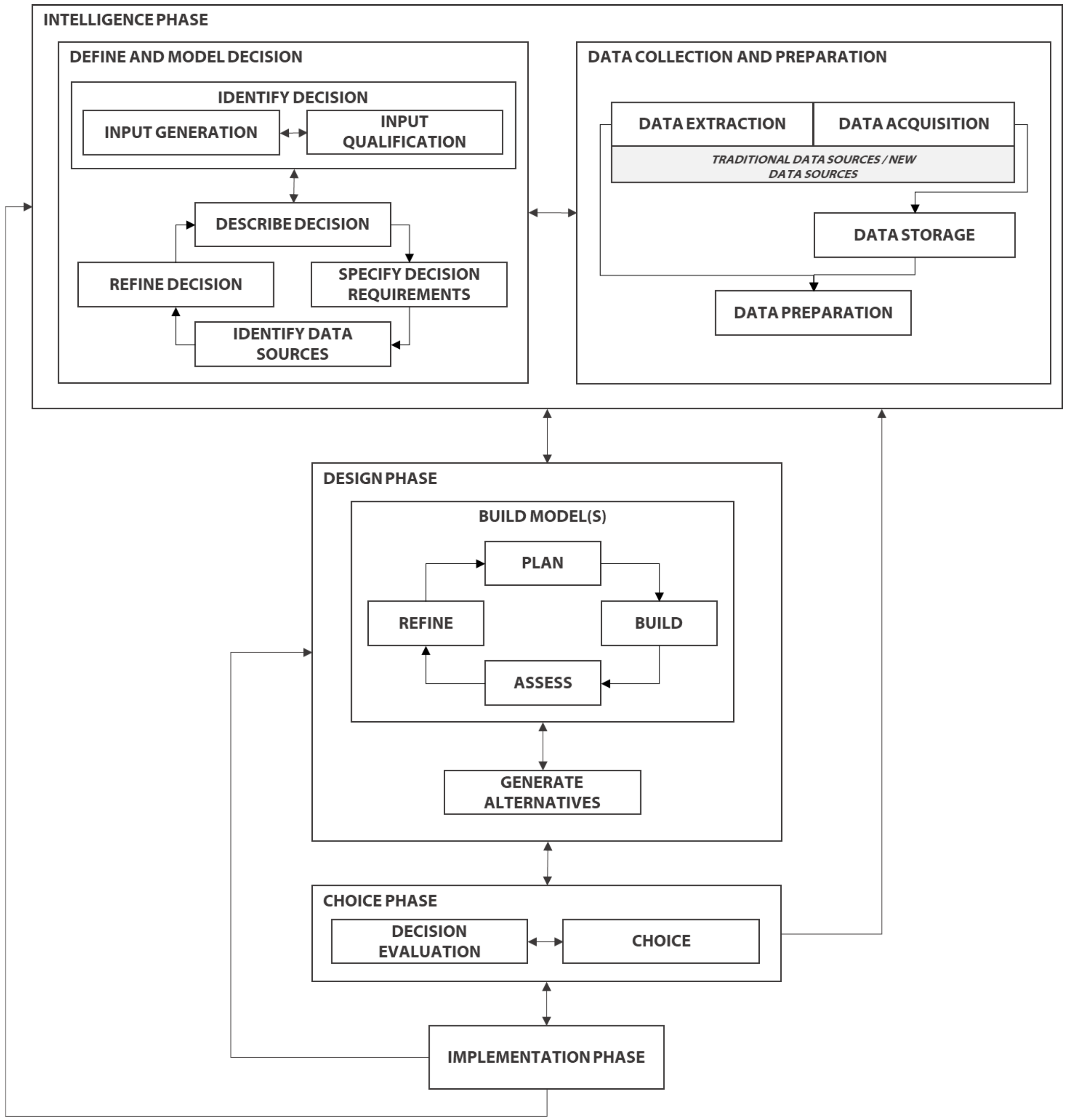
To answer the research question and sub-question a multiple case study was chosen, having as the unit of analysis the decision-making process. The first step of this research was to map the existing models in the literature. These models were analyzed and compared, choosing the one that would be used as a blueprint for comparison with the real-life cases. Semi-structured interviews were then conducted to analyze the decision process as it occurs in real-life cases. The case selection aimed for large enterprises, that currently use big data to inform their decision within the context of product/services and market development, with a complete decision process. The specific use case was chosen because it is considered to be one of the fields where big data can add the most value, besides being characterized by unstructured and nonprogrammed decisions, offering a good frame for the analysis of the interactions among the stakeholders. In total, 12 interviews in 9 different companies from different industries were conducted. The cases were analyzed using coding techniques, identifying patterns, and convergencies withing the data to derive insights for the discussion.

Results & Discussion

Decision-Making Process

Simon's four phases decision-making process, covering intelligence, design, choice, and implementation is a good representation of the main stages of decision-making. How these phases are broken down varied among the cases studied, but common structures were possible to identify. Based on the convergencies and the recent discussions in the literature, a refined model was proposed, which is possibly a more accurate representation of the decision-making process in the big data context. It is relevant to note that the four main phases and the steps within them form a hierarchical process, but not a sequential one. The process occurs iteratively, with constant back and forth between the different steps and phases. The main highlights of this process, in the light of the refined model, will be presented here.

The process starts with the broad stage of Define and Model Decision. This includes the assessment of the available opportunities and problem, they thorough definition, the data and technologies available to derive insights on the matter, and the impact on the business. This first step was deemed extremely relevant for the success of the process. This initial step is also important to identify the necessary stakeholders and generate alignment. Only when this initial definitions and assessment of the potentiality of tackling the issue are made, is that the team moves to the process of Data Collection and Preparation, and the design of the solution.



Refined model for the integration of big data in the decision-making process

Data can then either be extracted from existing databases or acquired. This can involve accessing external data or even capturing new data from internal sources. Then, data can be prepared for the following analysis. This process is often considered a non-value-added work, but it is extremely important for the generation of value from the raw data and, therefore, need to be given the necessary attention.

The Design Phase emerged as the most iterative phase of this process. The process involves data analysis and analytics and the overall process of building a thorough assessment of the issue at hand, developing possible alternative solutions. While the analysis is being conducted and different scenarios are being built, there is a constant assessment with other stakeholders to evaluate the preliminary results and to refine the work. New data may be needed, new constrains can be found and the models may need to be reviewed. The complexity of this step is related to the tools and expertise available, the data sources, the amount of data, and if real-time experiments are going to be employed or not. What then emerged from the data was a constant loop of plan - build – assess – refine, until a satisfying result is reached. The modeled outcomes are them used to generate alternative solutions and propositions that will be brought to discussion in the Choice Phase. Note that decisions may be binary, like a yes or no question, and therefore the model can already directly guide the answer, not being necessary to generate different alternatives.

The Choice Phase is the process of choosing one of the proposed courses of action. What was found is that this is seldom done at once. The process used to achieve these outcomes need to be justified and assessed by the decision-makers, and new needs may become apparent. In this process, it may be necessary to go back to previous steps, new information may be added and the model and, consequently, the solution is refined to attend to new inputs. In the end, the intended goal is to gain alignment around a course of action, that will then be implemented and monitored in the Implementation Phase.

Monitoring & Feedback is deemed especially relevant. Data makes it easier to measure the outcomes of the decision in a way that it can also feedback on the process as a whole. Also, there can be a real-time tracking of the implementation and its results, so the course can be corrected. In particular, when it comes to test & learn approaches, the Monitoring & Feedback step gains even higher importance, strengthening the iterative nature of the decision process. In these cases, design and implementation are closely attached, and decisions need to be made faster and may accumulate over time, as more information is gathered through real-time testing. This adds complexity to the process and emphasizes the need for smooth communication and collaboration among the stakeholders involved.

Stakeholders and Information-Processing Needs

In total, 6 types of stakeholders were identified and are involved in some part of the decision process. Their roles were mapped and are presented in a RASCI matrix, defining who is responsible (R), accountable (A), offers support (S), is consulted (C), or need to be kept informed (I) of the activities being executed. They are:

- **Business Owner:** the person who owns the decision.
- **Analysts:** the team of data analysts, data scientists, and business analysts directly involved in the execution of the analysis/analytics.
- **Technology Enablers:** the technical team responsible for managing the data, ingesting, preparing, and making sure that the needed data is available for the analysts to consume.
- **Transactional Stakeholders:** people from other areas that are involved in the operationalization of the solution or support areas, there are affected by the decision and need to be taken into consideration.
- **Higher Management:** Higher managers or even the c-level that are involved in the choice and final approval of the solution.

- **Translators:** the integrator, someone that bridges the analytics/technical department with the business.

Note that there may be other stakeholders not covered here and it will always be the role of the Business Owner and the Analysts to identify the stakeholders the need to be involved. The Translator in particular only appeared in some of the cases, but as its role has a high impact on those and it is important to recognize it here as a new classification. The roles of each stakeholder in the steps of the process are presented below.

				Business Owner	Analysts	Technology Enablers	Transactional Stakeholders	Higher Management	Translators*
Intelligence Phase	**Define and Model Decision**	**Identify Decision**	**Input Generation**	A/R	S		R/C		S
			Input Qualification	A/R	S		C		S
		Define Decision		A/R	R	C	S	I	S
		Specify Decision requirements		C	A/R	S	C		S
		Identify Sources		C	A/R	S	C		S
		Refine Decision		A/R	R	C	C	I	S
	Data Collection and Preparation	**Data Collection**	**Data Extraction**	I	A/R	S	C		C
			Data Acquisition	I	A/R	S	C		C
			Data Storage	I	S	A/R	I		I
		Data Preparation		I	S	A/R	I		I
Design Phase		**Build Model**	**Plan**	A/R	R	C	C		S/C
			Build	C	A/R	S/C	C		S/C
			Assess	A/R	R	C	C		S/C
			Refine	C	A/R	S/C	C		S/C
		Generate Alternatives		A/R	R/S		C/S		S/C
Choice Phase		**Decision Evaluation**		A/R	C/S		C	R	S/C
		Choice		A/R	C		C	R	S/C
Implementation Phase		**Implementation/Operacionalization**		R/A	C		C/R	I	
		Feedback & Monitoring		R/A	C		C/R	I	

Stakeholders involvement per step of the process

To understand the different information needs of individuals and the organization, the categorization per type of decision was used. Within the context of product/market development, the main subject of discussion varied among the cases, being either related to Product Improvement, New Product Evaluation, or New Product Introduction (referring to launching and implementation). Each of these cases used the available information-processing mechanisms differently, as can be seen below.

	Data-centric mechanisms				Organizational mechanisms			
	Predictive analytics	**Dashboards**	**Ad-hoc & descriptive analysis**	**Data mining**	**Planning**	**Direct contact**	**Integrator**	**Group meetings**
Product Improvement	○	◕	●	◔	◕	◕	○	●
New Product Evaluation	○	○	●	◔	◑	◑	◑	●
New Product Introduction	○	●	◑	○	◕	◑	○	●

Classification: ○ = very low usage ◔ = low usage ◑ = medium usage ◕ = high usage ● = very high usage

Usage of different information-processing mechanisms

Ad-hoc & descriptive analysis is the most common data-centric mechanism used across the cases, with the integration of data from different sources and different types, like quantitative and qualitative data, strengthening its value. For new product evaluation or specific improvements, the analysis of data based on initial hypothesis or scenarios is predominant. Existing dashboards appear to not cover the needs of product development and were mostly used in cases where a new product implementation was discussed. However, dashboards were also used as a visualization tool in the other contexts of decision, to communicate the findings among the stakeholders or to help in the analysis, like for the identification of patterns.

Data mining is starting to be more used, especially when it comes to uncover clusters of customers or to identify other patterns or trends relevant to the analysis. However, it was seen that many times, the analysts still do not possess the necessary knowledge to deal with more advanced technologies. It may be that companies will slowly evolve from basic techniques to a broader application of advanced techniques as the users and stakeholders increase their skills and ability to interpret the information. For now, the ad-hoc and descriptive analysis tells a story that is easy to understand and, therefore, it continues in predominance.

On the organizational side, meetings are deemed necessary to gain alignment and to check the status of the actions. The figure of the integrator assimilates the role of the translator and is yet not common across all the cases. Direct contact is often used and tends to be mentioned as relevant for the process to move quickly. It is often the role of the analysts to reach out for different stakeholders that may possess the necessary information or knowledge being demanded. Planning seems to be related to the maturity of the company, being more thoroughly used in higher maturity cases.

When it goes to the organizational elements, two aspects appeared consistently and are aligned with the recent literature. Regardless of the amount of data available, the domain knowledge will never be completely replaced. The domain experts know what questions to ask and help better define the decisions. Their knowledge is also relevant for the assessment of existing data and the model outcomes and, therefore, their involvement is determinant for the success of the data projects. This knowledge also needs to be shared and slowly acquired by the analyst and even technical teams, improving their effectiveness over time.

Final Remarks

Decision-making is at the heart of the managerial function, and if big data can enhance it, then we are dealing with a revolution in the way businesses are run. But organizations still face diverse challenges to extract value from the data available for them. In this research, it was noted that more than knowing how to integrate the big data tools, methods, and techniques into the decision process, the human factor is determinant in the success of big data initiatives. The information processing needs of both individuals and companies need to be attended. Even though machines are needed to extract valuable insights from large amounts of data, the human ability to synthesize the new knowledge being generated is unsurpassed. When it goes to big data, technology is a necessary but not sufficient condition. It needs to be integrated into the existing process and structures so that data can be transformed into value. This research offers a blueprint for this, presenting the necessary steps for the integration of big data in the decision-making process and identifying the stakeholders and information-processing mechanisms that need to be taken into account.

List of Figures

Figure 1. Overview of Simon's 4-phases decision-making process (adapted from Chiheb et al., 2019).... 19

Figure 2. Information-knowledge processing challenges (Zack, 2007) 21

Figure 3. Overview of Information-Processing Mechanisms (Kowalczyk & Buxmann, 2014).................. 22

Figure 4. B-DAD Framework (Elgendy & Elragal, 2016) 24

Figure 5. DMP-BDE Model (Chiheb et al., 2019) 26

Figure 6. Methodological approach 29

Figure 7. Stages of articles' selection process 30

Figure 8. Classification of cases per perceived level of maturity.. 33

Figure 9. Classification of cases per group of decision.. 34

Figure 10. Classification of cases per type of structure 34

Figure 11. Stakeholders involvement per step of the decision-making process 45

Figure 12. Usage of different information-processing mechanisms 46

Figure 13. Possibly more accurate model for the integration of big data in the decision-making process54

List of Tables

Table 1. Cases overview... 33

Table of Contents

Acknowledgments 2

Abstract 3

Management Summary 4

List of Figures 11

List of Tables 11

1. Introduction 14

  1.1. Academic Relevance 17

  1.2. Practical Relevance 17

  1.3. Outline of the thesis 17

2. Theoretical Background 18

  2.1. Defining Big Data 18

    2.1.1. The Role of BI&A 19

  2.2. Big Data and the Decision-Making Process 19

    2.2.1. Integrating Big Data in the Decision-Making Process: The Models 22

3. Methodology 28

  3.1. Research Design 28

  3.2. Models Assessment 29

  3.3. Case Selection 30

  3.4. Data collection 31

  3.5. Overview of Cases 33

  3.6. Data analysis 34

4. Results 37

  4.1. Decision-Making Process 37

    4.1.1. Intelligence Phase 37

    4.1.2. Design Phase 41

    4.1.3. Choice Phase 42

    4.1.4. Implementation Phase 43

  4.2. Organizational Elements 44

4.2.1. Stakeholders 44

4.2.2. Information-Processing Mechanisms 46

4.2.3. The organizational elements in the decision-making process 47

5. Discussion & Conclusion 50

5.1. The Frameworks for the Integration of Big Data in the Decision-Making Process 50

5.1.1. Stakeholders and Mechanisms of Coordination & Collaboration 54

5.2. Limitations and Future Research 56

5.3. Practical Implications 57

References 59

Appendix I 62

Interview Questionnaire - English Version 62

1. Introduction

Since its insurgence as the next big trend, around 2011, big data has been increasingly discussed in technology forums, practical outlets, and in the academic world. Research on big data surged and companies started to show a high appetite to implement big data tools, eager to reap its benefits. Indeed, big data is positioned to have a long-lasting influence on how businesses are run, promising to deliver superior performance, better products and services, and better customer relationships.

McAfee & Brynjolfsson (2012) point out that the availability of data represents a management revolution through facilitated and enhanced decision-making, and data is said by many to be the XXI century oil. Abundant data from both internal and external sources, partnered with analytics and related technologies, are being used to derive business insights and help organizations better understand its business and uncover market opportunities (Chen et al., 2012; Chiheb et al., 2019). In the digital world, competitive advantages will be maintained and acquired based on the company's ability to leverage data to explore new business opportunities (Shuradze & Wagner, 2016). Big techs and the new big players of the digital era already positioned big data at the core of what they do, and incumbents companies want their piece of the pie.

Despite the hype surrounding it, there is still a need to understand under which conditions big data initiatives lead to business value. Recent evidence has been showing that investments on big data projects do not always lead to the expected outcomes (Subarkti et al., 2020), and there is still a need to understand how organization translate or fail to translate, the potentials of big data into business value (Günther et al., 2017). Mikalef et al. (2018) argue that while the literature has seen a significant evolution of techniques and technologies for storage and analysis to leverage the growing volume, velocity, and variety of big data, lesser attention has been given to how organizations need to change to embrace these innovations and incorporate them into strategic thinking and operations. What has become clear in more recent literature, is the value of big data does not lie on the technical aspects alone, but it is a product of different organizational mechanisms leveraged together to deliver competitive advantages (Mikalef et al., 2018; Subarkti et al., 2020).

In the quest to shed light on the mechanisms behind the translation of big data into big value, more recent research has been pointing out that the first-order effect of big data is likely to be on a superior decision-making process, through which data is transformed into insights, insights into decision and decisions into value. (Sharma et al., 2014) Value is then attained through organizational improvements that are the outcomes of better decision-making, forcing businesses to realign its organizational models to reap the benefits of big data (Günther et al., 2017). This view is supported by the claims that decisions based on data are generally better ones and winners around different industries will likely be the ones harnessing the power of big data do transform their decision-making process, relying on data insights rather than on gut feeling and experience of decision-makers (McAfee & Brynjolfsson, 2012). For that to be true, Shuradze & Wagner (2016) argue that organizations still need to structure their business process in a way that optimizes the transformation of data into timely insights for decision-making.

As such, the integration of big data capabilities within the decision-making process has been an attractive field of research, promising to help companies extract value from data. Over the recent years, different authors have come up with models or frameworks for the effective integration of big data in the decision-making process (Poleto et al, 2015; Rani & Kant, 2019; Saggi & Jain, 2018; Elgendy & Elragal, 2016; Chiheb et al., 2019; Akter et al., 2019; Lu, 2018). Having the well-known process defined by Simon (1960) as a starting point, covering the intelligence, design, choice, and implementation phases of managerial decision-making, these authors attempted to map the tools, techniques, and methods that can aid in the different phases of the process, modeling the steps necessary for the integration of big data insights into the decision-making process. These models and frameworks were developed based on the existing literature and theoretical studies, with little to no empirical research employed to understand the practices employed in real-life use cases or to assess the applicability of the defined constructs in it. In fact, the empirical component within the research around big data value seems to be a broad issue, with little empirical research conducted to analyze the actual practices being employed at companies and learning from it (Günther et al., 2017; Mikalef et al., 2018).

Despite the existence of a handful of models and frameworks in the literature, Akter et al. (2019) claim that integrating big data & analytics into the decision process is still a challenge for many companies. Furthermore, these works present several convergencies and even referencing among each other, but there are yet divergencies, in particular, related to the specific steps covered and their order. Therefore, it is claimed here that, if the existing models are to help companies in implementing better data-based[2] decision-making processes, efforts should be made to assess whether and which of these models possess the necessary attributes to support their claim. This cannot be done without gaining insights from practice and understanding how well these models converge or diverge from the effective practices employed in real-life cases. Without this knowledge, decision-makers and practitioners alike will be left in uncharted territories, not being able to assess how likely these models are to actually support them in their pursuit of extracting value from big data.

Furthermore, Mikalef et al. (2018) argue that while understanding the mechanisms and processes through which big data can add value to the business, it is also important to have a clear picture of the different elements and their interdependencies. When it comes to the decision-making process, value lies in the interaction between users and the increasingly available data to gain insights and foresight, helping to solve unstructured and exploratory problems (Subarkti et al., 2020). For that, more than knowing the steps from data selection to making a decision, companies need to know how to employ the right mix of information and knowledge for proper interpretation of data insights, leveraging the information processing capabilities of individuals and the organization (Zack, 2007; Kowalczyk & Buxmann, 2014).

Therefore, more than enhancing the knowledge over the process through which decisions are made and the integration of big data tools, techniques, and methods, it is also relevant to comprehend the

[2] Author note: While the practice seem to be slowly exchanging the buzz word data-driven to data-informed or other variations of it, and claims are made upon the difference among them, this research will avoid the use of the terms, opting for the more neutral data-based.

organizational elements impacting this process beyond the decision flow, an attempt still not made by the works analyzed. The interactions among the different stakeholders along the decision process are one important element, since making decisions, especially the more strategic ones required to generate competitive advantage, still requires the aggregate knowledge of individuals and their collective action to propose and chose among alternative decisions (Subarkti et al., 2020; Zack, 2007). This involves not only investigating the stakeholders involved in each step of the process but also learning how they communicate and collaborate among each other to leverage the available data throughout the decision process, characterized by the information processing mechanisms utilized.

While there is a need to understand if the available models indeed fit the practical use cases of big data, this endeavor can also offer the opportunity to generate insights onto the specific organizational elements underlying the proper integration of big data in the decision-making process. It is believed that by aligning these two efforts, new knowledge can be generated that will help companies to better understand how they can realign their business models and process and what does this entails, allowing them to make proper use of big data to enhance decision making.

With that in mind, the goal of this research is twofold. First, it aims to assess how well the existing models fit practical use cases and whether the defined phases can convey the necessary steps to make a decision. Second, it aims to offer insights into the organizational elements that are involved in this process. In particular, it aims to identify which stakeholders are involved, and what are the information processing mechanisms used for communication and collaboration between these stakeholders.

To accomplish the aforementioned goals, the research question and sub-questions are structured as follows:

> *How well do existing models for the integration of big data in the decision-making process fit practical use cases?*
>
>> *How well do the steps described in these models converge with the steps described in practical use cases and what other steps need to be taken into consideration?*
>>
>> *What are the stakeholders involved in the different phases of this process?*
>>
>> *Which information-processing mechanisms are utilized in the context of the decision-making process?*

By answering this question, this research expects to offer a better understanding of how these aspects work together to deliver better decisions in the context of big data, shedding light in both the applicability of the models in practice and the organizational elements that surround the effective use of big data in the decision process. With that, the expected outcome is a revised model for the effective integration of big data in the decision-making process considering the insights of practice together with the existing conceptual models, and the clarification of the organizational elements involved in this process. It is expected that this work can help practitioners in their efforts to improve their data-based decision-making processes and offer insights into the literature regarding possible avenues for future research.

1.1. Academic Relevance

Authors have claimed that the big data research still lacks empirical works (Mikalef et al.,2017) and that the actual practices being employed by companies need to be taken into consideration (Günther et al., 2017). This study answers these calls by offering a practical view of the discussion of big data integration in the decision-making process. Besides, it proposes to assess the fitness of existing frameworks, that were only limited evaluated in the light of practical use-cases, strengthening the claims of the authors about its application and effectiveness.

Furthermore, the researcher expects to shed light on some of the factors that enable the effective use of big data, in particular regarding the stakeholders involved and the information processing mechanisms employed by them. Subarkti et al. (2019) argue that literature still needs to further develop the research on the organizational elements surrounding big data implementation and that the experience from practice is necessary to advance this knowledge. Overall, this study expects to bring more empirical evidence into the debate of big data integration in the decision-making process and, consequently, into its effective use.

1.2. Practical Relevance

Big data is still a relatively new topic, and both literature and practice are still learning to navigate it. It promises a lot of benefits for companies, but no fewer challenges. De Smet et al. (2018) argued that access to a lot of data and the complexity of organizational structures are bugging the decision-making process in companies. This research aims to contribute with practitioners by enhancing the existing knowledge on the integration of big data in the decision-making process, identifying not only the steps of the decision process but also the organizational elements they need to take into account. Also, it offers a refined tool that they can use to shape their decision-making process to embrace big data insights in a more effective way. For companies already using big data, the refined model can help them improve their process and solve possible challenges they still face. For companies that are still trying to figure out how to leverage big data, the framework can serve as a clear blueprint to the definition of new business processes.

1.3. Outline of the thesis

The remainder of this thesis is structured as follows. In chapter 2, the theoretical background will be introduced, defining big data, and presenting it in the context of decision-making. Also, the models for the integration of big data in the decision-making process will be defined. In chapter 3, the methodology will be discussed, presenting the research design, model discovery, case selection, data collection and analysis, and an overview of the cases. In chapter 4, the findings of this study will be presented and, lastly, Chapter 5 will present the discussion and conclusions.

2. Theoretical Background

In this chapter, the prior knowledge regarding the integration of big data in the decision-making process will be covered. First, big data and related topics will be defined. After, the decision-making process will be presented, and considerations will be made concerning the effects of (big) data in this process. Lastly, the models and frameworks for the integration of big data in the decision-making process discovered in the literature will be outlined.

2.1. Defining Big Data

The phenomenon known as Big Data is the product of two achievements of the technology age: the unprecedented volumes and speed of data generated from a variety of sources, from process automation to social media, aligned with advancements in technology that allow companies to collect, store and manage data in a relatively easy and inexpensive way (Jagadish, 2015; Chiheb et al., 2019). Since its insurgence, the discussion around big data has dominated the technology forums, the literature and called the attention of practitioners, eager to reap the so prayed for impacts in performance. Despite the hype that has been surrounding it, big data is here to stay. Both availabilities of data and storage capacity will continue to grow (Jagadish, 2015). At the same time, as McAfee & Brynjolfsson (2012) claimed, we can only manage what we measure, and with the advent of big data, 'we can measure and manage more precisely than ever before'. But what big data is really about, and why is it different from data analysis? McAfee & Brynjolfsson (2012) says that big data goal is to translate data into insights and into business advantage, not different from the analytics that existed before it. However, it differs from it due to its high volume, velocity, and variety.

These 3Vs are commonly used to define big data. More recently, three new Vs were added, representing veracity, variability, and value (Gandomi & Haider, 2015). Volume, the characteristic that is most used to refer to big data, relates to its size, to the exponentiality of the data available. Velocity relates to the frequency in which new data is available and how often it can change, associated with the real-time data ideal. Variety regards the different sources and formats of data, characterized by being unstructured. The recent veracity, coined by IBM, refers to the imprecision and uncertainty associated with the new sources of data available, like social media. Variability, coined by SAS together with complexity, refers to the variation in the data flow rates, while complexity links with the challenge of connecting the variable sources of data. The last, value, has been recently considered an important part of big data, referring to the 'density of value', in which raw data would have low value density, that would increase through analyzing large volumes of the data, without which it would not deliver the expected outcomes. These characteristics of big data alter the technical and organizational constructs need to extract value from data and posit challenges to both organizations and society at large. Learning to navigate this new setting is paramount to extract value from it (Günther et al., 2017).

2.1.1. The Role of BI&A

Related with the field of big data, business intelligence & analytics (BI&A) has also become increasingly relevant in the past years. Kowalczyk & Buxmann (2014) argues that big data and BI&A are two sides of the same coin. While big data addresses the supply of data as a resource, BI&A capabilities describe how data is analyzed to generate a better understanding of the business. Chen et al. (2012) also bring data and analysis as complementary and Sharma et al. (2014) argue that the transformation of (big) data into better organizational performance is an outcome of better decision-making enabled by BI&A. Therefore, in practice, big data cannot be disassociated from the methods and tools that allow the data to be analyzed.

Chen et al. (2012) refer to BI&A as "techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand their business and market and make timely business decisions". Kowalczyk & Buxmann (2014) listed some classes of BI&A techniques, such as ad-hoc queries and descriptive statistics, online analytical processing (OLAP), dashboards, reports, and more advanced techniques like predictive analytics and data mining. Research has been showing that the proper use of BI&A systems varies per application and different needs of the decision context and information-knowledge mix (Isik et al., 2013; Zack, 2007).

2.2. Big Data and the Decision-Making Process

Making decisions is a big part of a manager's job, so much that Simon (1960) argues that decision making could be used as a synonym of managing. In his definition, decision making does not comprise merely the act of choice, but also involves finding occasions for decision-making, discovering possible courses of action, and finally choosing one of them. These were defined as the three main phases of the decision-making process: intelligence, design, and choice, respectively. Considering these, the choice is the one demanding a lesser amount of time. However, one does not happen without the other, even if it is one person caring out the whole process, nor does these phases are nicely ordered one after the other, representing a much more iterative process. The implementation phase was only included later, but Simon (1960) argues that the implementation itself consists of another cycle of decision(s), with its own intelligence, design, and choice phases being present.

Intelligence phase	Refers to searching the environment for conditions (problem or opportunity) calling for a decision
Design phase	Refers to developing and analyzing alternative solutions for the problem or opportunity
Choice phase	Refers to choosing one or more of the alternatives
Implementation phase	Refers to the operationalization of such alternative and its monitoring

Figure 1. Overview of Simon's 4-phases decision-making process (adapted from Chiheb et al., 2019)

This 4-phase model (Figure 1) is still highly influential and referenced and can be applied irrespective of the type of decision. However, not all decisions are created equal and different types of decisions require different information processing needs from individuals and the organization. In the end, decision making involves applying information and knowledge to the situation at hand and the different contexts of each particular problem or opportunity require the application of a different mix of information and knowledge to be solved (Zack, 2007).

Simon (1960) defined the two basic types of decision, lying in a continuum, the programmed and nonprogrammed decisions. The most routine and repetitive they are, the more programmed they tend to be. Novel, unstructured, non-routine, and consequential decisions characterized the nonprogrammed ones. Gory & Morton (1971) argue that within this continuum, a decision can have an operational, managerial, or strategic nature and that the information processing needs also vary in this axis. Besides the frequency of a decision (routine vs. non-routine), the scope of the decision will also help define the balance of information and knowledge needed. In particular, the broader the decision, the more information would be generally demanded, from both internal and external sources. This is the case of strategic decisions, with managerial decisions lying somewhere in the middle.

These different needs of the decision domain will also implicate in how well the decision (response) can be automated, and if it is desired or not. Zack (2007) argues that ill-defined, unstructured problems cannot be sufficiently codified to be automated and often require the aggregate knowledge of individuals to make the decision using some form of heuristic processing. However, this human capability can be enhanced by the use of decision support systems like BI&A.

This view is aligned with more recent research on the use of big data, which conveys that its value lies in the interaction between users and the increasingly available data to gain insights and foresight to be used in decision-making, helping to solve unstructured and exploratory problems (Subarkti et al., 2020). This is the case of decisions made in the context of new product development and market development, besides other areas of strategic orientation, such as business model innovation. This means that the big data phenomenon can potentially bring more rigor to areas of decision-making that are still dominated by gut feeling and intuition, targeting better intervention through better measurement and better predictions (McAfee & Brynjolfsson, 2012). As the authors posit, big data can represent a revolution in management through facilitated and enhanced decision-making, a view very well aligned with that of Simon (1960).

This doesn't come without challenges though. The advent of big data and its integration to decision-making presents both technical and organizational issues. Fayyad et al. (1996) point out that the extraction of knowledge from large datasets depends on several steps from data selection, preparation, cleaning to specific knowledge of statistical inference, search, and reasoning, allowing the proper interpretation of the data insights. All this reflects into the information processing capabilities of individuals and the organization, created through a combination of organization and technological resources (Kowalczyk & Buxmann, 2014). In the basis or information-knowledge indeterminacy, Zack (2007) defined four unique challenges faced by organizations (Figure 2):

- ***Uncertainty:*** not having enough information;
- ***Complexity:*** having more information than one can easily process;

- ***Ambiguity:*** not having a conceptual framework for interpreting information;
- ***Equivocality:*** having several competing or contradictory conceptual frameworks.

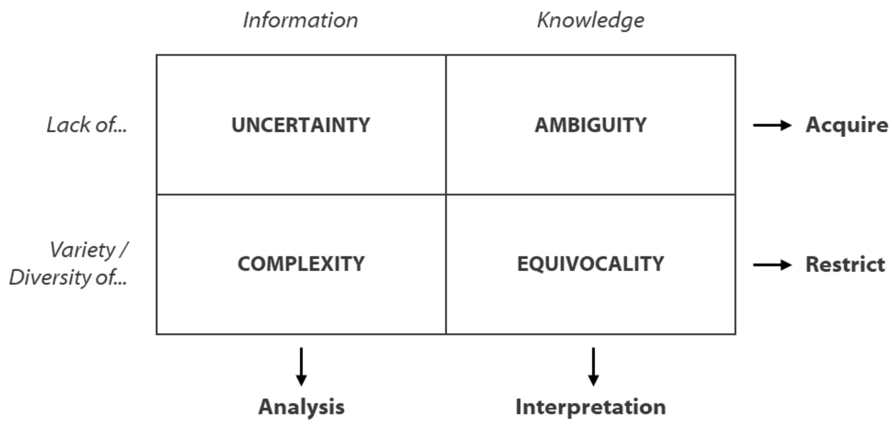


Figure 2. Information-knowledge processing challenges (Zack, 2007)

It is relevant to note the differentiation between information and knowledge made by the author. Information is defined as 'observations that have been cognitively processed and punctuated into coherent messages', while knowledge is 'that which one believes to be true about the world' (Zack, 2006). Thus, previous knowledge provides the context for interpreting new information, while new information can alter what is already known.

Having this framework in mind, the mechanisms to reduce each of these two groups of challenges are different. Big data and BI&A can help reduce uncertainty, by increasing the availability of information for the decision-maker, and complexity, due to the processing capability of the systems, but only once there is an initial understanding of the situation (previous knowledge). On the other hand, the understanding of the decision context and the ambiguity and equivocality challenges need to be managed through the interaction and exchanges among the different stakeholders to define problems and resolve conflicting interpretations of the event (Kowalczyk & Buxmann, 2014; Zack, 2007; Daft & Lengel, 1986).

Indeed, the ability to gain insights from data relies on the instinct and capability of humans, in a way that machines alone would not be able to do. McAfee & Brynjolfsson (2012) seems to agree with the premise that the use of big data does not exclude the need for vision and human insight, while Sharma et al. (2014) argue that extracting value from big data is conditioned by the collaboration between different stakeholders in already existing decision-making processes, in which analysis tools are intentionally used to discover new knowledge and generate valuable insights.

To assess how uncertainty and equivocality are managed and reduced in the decision-making process, Kowalczyk & Buxmann (2014) developed a conception of data-centric and organizational information processing mechanisms based on the work of Daft & Lengel (1986), presented in Figure 3. This conceptualization will be used to assess the information-processing mechanisms that are involved in the decision-making process.

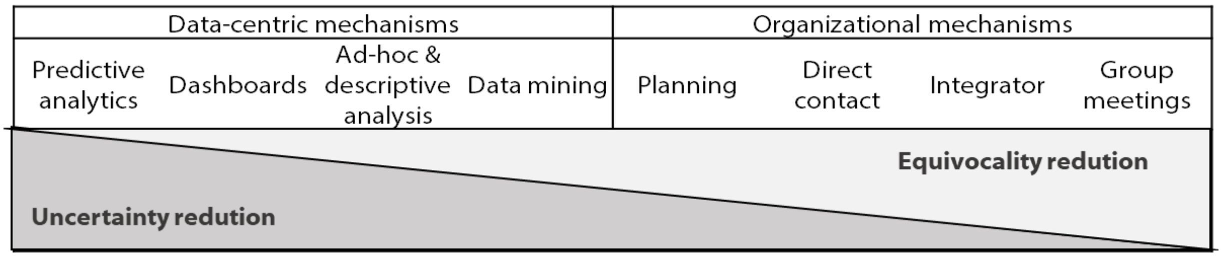


Figure 3. Overview of Information-Processing Mechanisms (Kowalczyk & Buxmann, 2014)

The 8 mechanisms are defined below.

- ***Predictive analytics:*** refers to the utilization of defined models for the accurate prediction of recurring or well-understood issues, meaning the equivocality has been reduced beforehand;
- ***Dashboards:*** refers to the periodic delivery of information that answers predefined questions and provides structured means of data analysis, such as drilling, slicing, and dicing;
- ***Ad-hoc & descriptive analysis:*** refers to the mechanisms that allow for open descriptive data analysis with a question or hypothesis in mind, including one-time studies design to gather and analyze data about a specific issue;
- ***Data mining:*** refers to data analysis and discovery algorithms for identifying patterns or models, contributing to reduce both uncertainty and equivocality;
- ***Planning:*** refers to a joint effort of decision stakeholders to reduce equivocality and uncertainty through the establishment of common goals and a course of action;
- ***Direct contact:*** refers to simple forms of personal contact that allow stakeholders to discuss issues personally;
- ***Integrator:*** refers to a lateral organizational position that deals with the integration and distribution of information to establish a common understanding and reduce equivocality;
- ***Group meetings:*** refers to the joint effort to build a common understanding through collective judgment, concerning the decrease of equivocality.

2.2.1. Integrating Big Data in the Decision-Making Process: The Models

In recent years, different authors have defined their own models and frameworks for the integration of big data in the decision-making process. An overview of the 6 models and frameworks discovery in the literature will be discussed here. It is relevant to note that it is not the goal of this research to cover the technical aspects surrounding this integration since this work has been extensively documented in the literature. The work of Lu (2018) for instance, presents a good overview of possible technologies and tools, that can be consulted by the interested public. Therefore, this review will focus on the process itself, not in the technologies and technical integrations.

Most of the models and frameworks for the integration of big data are built upon the 4-phases decision-making process of Simon (1960), that has been used consistently throughout the decision science literature (Chiheb et al., 2019; Elgendy & Elragal, 2016; Sharma et al., 2014). When this is not the case, it will be noted. The premise of the models being developed in the context of big data and the decision-making process is that decision-makers need to be able to gain insights from the vast amount of data available out there, that only increases with time (Elgendy & Elragal, 2016). In this sense, it posits that decision-making, if concerned at being data-based, is dependent on data analysis, in which knowledge can be extracted from raw data. The different models that will be discussed here all use Simon's decision-making phases to describe how data is integrated and analyzed to be transformed into knowledge and, subsequently, into smarter decisions. For clarification, the four phases of the decision-making process are referred to as 'phase', while each phase within them are referred to as 'step'. Within each 'step', there may be also 'tasks'.

Poleto et al. (2015) integrated business intelligence tools and decision support systems (DSS) to provide an overview of how big data is integrated into the decision-making process. The model consists of 6 main steps. Based on the data available in internal and external sources, the relevant data is identified and acquired (1). BI&A tools can then be used to 'aggregate value to acquired data in order to obtain relevant information' (2) (Poleto et al., 2015). In this intelligence phase, visualization tools are posited as paramount to the identification of trends and relationships among the data. From there, opportunities are identified, and alternatives are generated (3) by the decision-makers based on the analysis performed in the previous stage. Also, the criteria under which each alternative will be judged is defined. Decision Support Systems are then used to recommend the best course of action (4), that can be then ratified by the decision-makers. Once the choice is made, the decision can be implemented (5). A feedback process is suggested to integrate the knowledge acquired during the process back to the data repositories, enforcing organizational learning (6). The author noted that the big data tools alone are not enough to generate alternatives and predict results and that the experience and tacit knowledge of decision-makers are essential to knowledge generation.

Elgendy & Elragal (2016) designed a more refined framework, called Big Data, Analytics, and Decision (B-DAD), carefully carved out of Simon's four phases model. The authors based on a literature review and mapped the big data tools, architectures, and analytics that can be applied in the different phases of the decision-making process to enhance and support data-based decisions. The framework was then validated with one experiment in a real-case scenario.

The framework consists of 8 main steps. It starts with the process of identifying the data to be used, which can be from internal or external sources (1). From that, data need to be acquired and stored appropriately (2), so it can be prepared (3), which includes the tasks of organizing and processing the data. It can be that data already exists in the company's databases, and then acquisition and storage can be skipped. This closes the intelligence phase. Next, there is the design phase, where possible courses of action are developed and analyzed through a conceptualization of the problem. Here, the authors define the step of planning the model (4) that will be used to analyze the data. It involves defining which model is appropriate to analyze the selected data and to answer the given issue and planning its execution. Once this is done, the selected model is applied in the data analytics step (5). The outcomes of it are analyzed and evaluated in the subsequent step (6). This includes defining the possible alternatives and evaluating their impact. Then, it is time to decide which course of action is the best or more appropriate (7). This step is where the actual decision takes place, based on the evaluations made in the prior step. From there, the implementation phase starts (8), with the proposed solution being operationalized, monitored and feedback is provided. Note that the authors posit that the framework assumes that the decision domain is already known and, therefore, does not need to be further exploited. Lastly, is important to notice that this process is not sequential, and the loops can occur in-between the different steps and phases. The framework is represented in Figure 4.

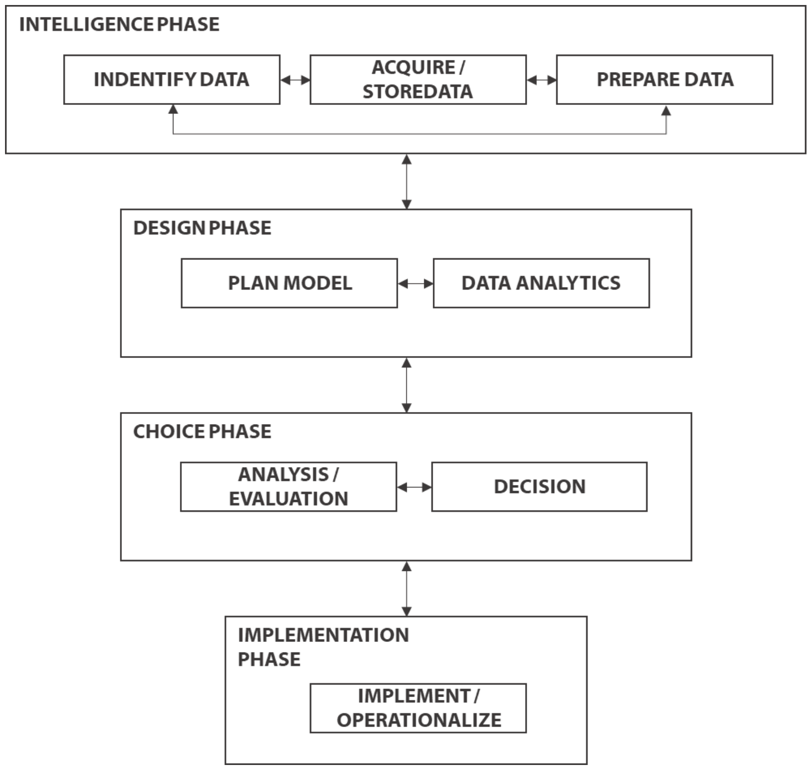


Figure 4. B-DAD Framework (Elgendy & Elragal, 2016)

Saggi & Jain (2018), despite not referring to the four phases of the decision-making process itself, presented a similar model, with a more technical flavor. The framework starts with big data management (1), which involves the data sources (1.1), acquisition (1.2), and storage & processing (1.3). After, there is big data analytics (2), including the techniques to analyze the data, that can be aligned with domain expert knowledge. Here, visualization is used to help understand the data. From there, there is the stage of decision framing (3), which refers to clarifying the purpose of the decision, the decision needs, and the context. It involves decision modeling (4.1), the decision-making itself (4.2), and the decision-execution (4.3), meaning the operationalization of the decision. In this last stage, data requirements and technical requirements are taken into consideration, and the decision is modeled in such a way that it presents the 'findings, criteria, scenarios, option and recommendations' (Saggi & Jain, 2018) that will be used to make a decision within the business context.

It is interesting to note here that the authors' approach is of using big data analytics to discover new knowledge without the existence of a specific decision that needs to be answered. Elgendy & Elragal (2016) also argue that in the big data context, data can be analyzed in an unstructured way, without being linked to a specific decision, and only afterward it is examined how the generated insights can contribute to decisions. This is in line with Saggi & Jain (2018), that argues that once insights are available, they can be used to inform a decision in the 'decision framing' stage, where then value can be realized.

Rani & Kant (2019) presents a similar model to that of Elgendy & Elragal (2016), also having the goal of translating knowledge into improved decisions based on evidence. It defines an intelligence phase consisted of identifying sources (1) and collecting data (2), followed by data preparation and preprocessing step (3). From there, it enters the design phase, in which the models are planned (4), data analytics are employed (5) and data is analyzed (6). The choice phase (7) consists solely of the choice of the best course of action, which is then executed in the implementation phase (8).

Chiheb et al. (2019), in their Decision-Making Process in a Big Data Environment Model (DMP-BDE), differentiates from the previous ones by including a decision modeling notation (DMN) standard and a decision diagram in the process. This is said to simplify the communication about what the decision consists of and how it will be analyzed in the light of the processes in place, decision and data governance, analytic models, and data inputs supporting the cooperation between decision-makers and analysts. It includes decision requirements and decision logic to identify the elements of the decisions and their relationships.

Given these inclusions, the process begins with the definition and modeling of the decision (1), in the intelligence phase. This involves identifying the decision (1.1), describing it (1.2), specifying its requirements (1.3), and decomposing and refining it (1.4). This process is presented in a loop, in which the decision is defined and understood, having as an outcome a decision requirement diagram. With all these requirements clear, the second step is to collect and prepare the data (2). This refers to the identification of sources (2.1), both internal and external, the data acquisition (2.2), storage (2.3), and its preparation to be consumed (2.4). This second step equates to the intelligence phase of the B-DAD framework. Following, it enters the design phase, in which models are build (3) and alternatives are generated (4). In this phase, the possible courses of action are defined, developed, and analyzed in its ability to answer the questions posed by the decision. First, the possible models that can help shed light on the decision are identified. This takes into consideration both the objective and the data available. After, the data is tested, and the model is built. Second, the new information that comes out of this model is aligned with existing knowledge and experiences to generate possible courses of action. The authors also add that in this step, the criteria towards which the alternatives will be judged are defined. From there, it goes to the choice phase (5), in which the alternatives are evaluated against the defined criteria so that the best one can be chosen. The impact in the business is measured, and based on the evaluation findings, decision-makers can choose the alternative that is more appropriate to solve the problem. The final stage is the implementation of this course of action (6). Again, the authors note that the process is not sequential and that loops between the steps and phases are likely to happen. The DMP-BDE is presented in Figure 5.

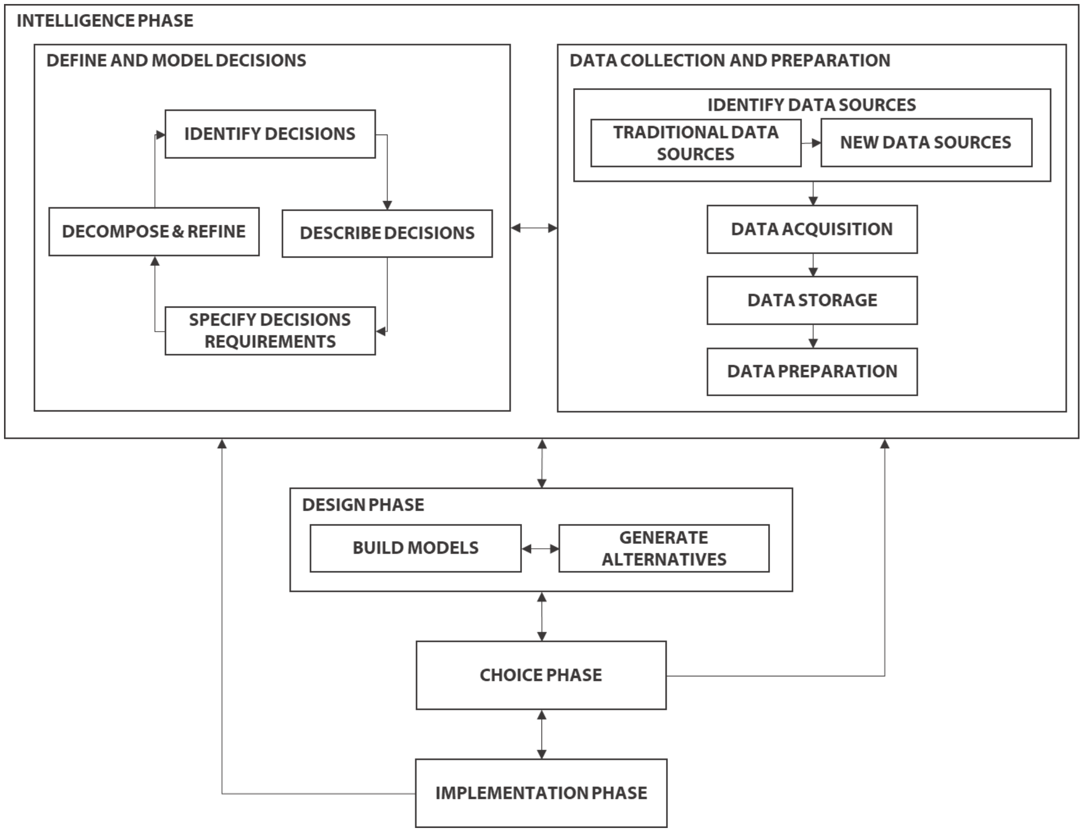


Figure 5. DMP-BDE Model (Chiheb et al., 2019)

Chiheb et al. (2019), by adding the decision modeling and decision diagram, appears to convey what could be considered a more comprehensive framework, reinforcing an initial step of defining the decision that will be made, having effects in the subsequent steps. This approach detaches from Elgendy & Elragal (2017) and Saggi & Jain (2018), that present a more unstructured approach to uncovering new knowledge.

The work of Akter et al. (2019) is aligned with the construct of Chiheb et al. (2019), devoting a higher consideration of the initial steps of the process. Their model starts with the problem recognition (1), referring to correctly framing the problem or opportunity, followed by the review of previous finding and context (2), said to be necessary to "learn existing measures and avoid pitfalls" (p. 88). From this, the next step is to select variables and formulating a model (3), meaning to formulate a hypothesis and defining a simplified representation of the problem, the model. The authors note that at the end of this step a problem statement needs to have been defined. The next steps are to collect the data (4), analyze the data (5), and act on the insights (6). This last step, as said by the author is where "a comprehensive interpretation is required to gain valid insights" (p. 90). It is relevant to note that the framework defined by Akter et al. (2019) was constructed upon both the literature review and qualitative studies through semi-structured interviews.

In this same line of dedicating more attention to the initial steps of defining and understanding the decision, the work of Lu (2018) seems relevant. Her framework is target at business analytics and has a more technical flavor but can also contribute to this context. It covers aspects of the business analytics lifecycle of data management, pre-processing, modeling, and the last step of insights management. However, the framework starts with a stage of business knowledge and market understanding (1). The author says that the "domain expertise is needed to frame business goals in a way that provides value to the organization" (p. 340). From this, the data management (2) step can start, the mainly covers the data collection aspects, from both internal and external sources. From there, data cleansing, summarizing, reduction, and normalization, as well as the extraction, transformation, and loading are covered by the data pre-processing and integration (3). Data modeling and business intelligence (4) covers the extraction of value from data, through data mining, statistical techniques, machine learning, and others. The last step, insight management (5), covers the evaluation of the insight, model refinement based on the assessment of the steps executed to construct the model, action, and monitoring. The author notes that "insight management is about understanding information needs and then managing the way that information flows through so that it has a positive effect" (p. 343).

The analysis of the practical cases will take into consideration the insights from all the described models. In particular, the work of Chiheb et al. (2019) is being considered that the most comprehensive one and is being taken as the blueprint against the cases will be compared. However, the remaining models and frameworks will still be analyzed and discussed together in the face of real-life cases, in an attempt to shed light on the divergences existing among them and to reflect upon the accuracy of the constructs.

3. Methodology

In this chapter, the research design will be presented and justified, and the other considerations regarding the methods chosen to carry out this research will be outlined. Also, an overview of the selected cases will be presented.

3.1. Research Design

This research is defined as a multiple case study, having as the unit of analysis the decision-making process. The case study is justified based on Yin (2014), since this research has the goal of evaluating a framework fitness in real-case scenarios, having as subject an undoubtedly new phenomenon, big data.

Other options for qualitative research would not meet the requirements of this endeavor. An experiment would not meet the need to observe the actor's behavior in their real, habitual environment. Furthermore, previous authors already did that to test their frameworks (Elgendy & Elragal, 2016), and a contribution in this sense would be only limited. The use of archival data is also limited since the objective of the study, the decision-making process, is not often recorded and the study of the decision outcome itself would not provide the needed information to evaluate the steps of a decision.

Thus, a case study, where data is gathered by the means of interviews, seems to better reflect the needs of this study, to the extent in each it assumes that the story will be told by the actor directly involved in the process (the interviewee) and that the researcher has no control over the behavior or the outcome of the process under analysis. Furthermore, it can be added that the boundaries between the object of study, the integration of big data in the decision-making process, and the context, in this case, the organizational setting in which decisions are made, are not well defined. As argued by Yin (2014), the choice of a case study relies on the need to understand a phenomenon in a real-world case that involves important contextual conditions.

Moreover, case studies are said to contribute to the development of new and enhanced lessons for management practitioners, since it builds on a strong methodological approach with straightforward considerations to practical applications. With that, it helps both in the creation and circulation of best practice thinking (Collinson & Rugman, 2010). As the researcher expects to offer the practice a model that can help managers and decision-makers in the implementations of data-based decision-making processes, the choice of case studies is further strengthened.

The multiple case approach is expected to offer better evidence of the fitness of the framework within different contexts. While a single case study could offer the opportunity to dive deeper into the process, relying on a single case to evaluate the fitness of a model can be said to be limited. Yet, one of the approaches to dive deeper into the process in the context of a single case study would be to include observation of the phenomena while it occurs. However, giving the current health crisis, with most of the professionals working from home, doing so would be considerably more difficult, if not impossible to accomplish. With that said, the multiple case approach appears to offer a better ground to understand the phenomenon. By having various sources of information, the researcher anticipates the possibility of collecting a more complete view of the approaches, tools, and techniques being employed by companies, which in turn would offer a more extensive base in which to evaluate how well the available models fit the practical cases. For that to be true, the study will use a replication logic among the multiple cases, guaranteeing external validity (Yin, 2014). Figure 6 presents the overall methodological process, that will be further described in the following topics.

Figure 6. Methodological approach

3.2. Models Assessment

The main goal of this study is to evaluate the fitness of the existing models for the integration of big data in the decision-making process in practical, real-life cases. For that, the first step was to identify the existing models and frameworks for this integration in the literature.

There is a growing interest in the big data phenomenon, and studies of its application exist in diverse sectors. Decision-making itself is also a broad topic. Therefore, the keywords used to assess specifically the models for the integration of big data in the decision-making process need to be precise. Two sets of keywords were chosen for the search, 'decision-making big data integration models' and 'decision-making big data integration frameworks'. Also, due to the broad application of the topic, it was chosen to search for studies in the fields of Business Administration, Computer Science, and Decision Science. The library chosen for this assessment was Scopus. With this, it was found 205 potentially relevant articles. Based on an initial assessment of the titles, abstracts, and keywords, 19 articles were selected. From those, 3 were excluded due to duplication across the different sets of keywords, and the remaining 16 were further evaluated by the researcher to identify if relevant models and/or frameworks for this integration were described. From this assessment, 6 articles were kept. It is relevant to note that, when assessing these articles, previously cited studies were also taken into consideration, however, this did not increase the number of articles. Only in one case, previous work from the same authors was cited and substituted due to higher relevance. This process can be visualized in Figure 7.

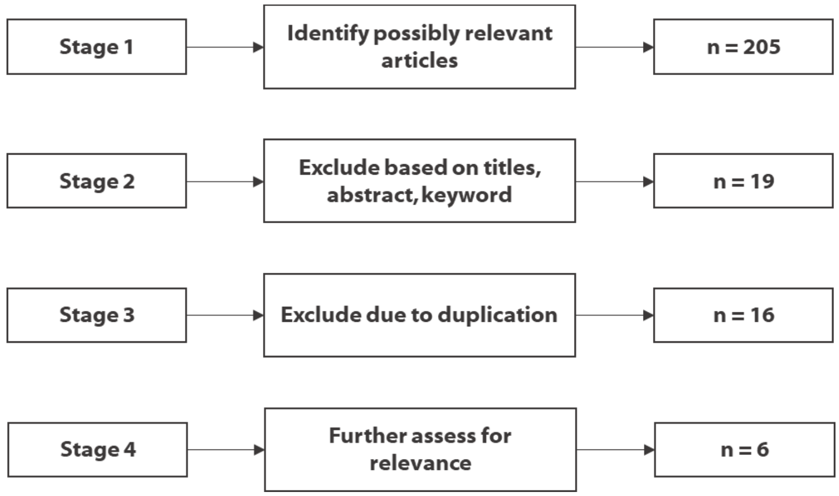


Figure 7. Stages of articles' selection process

3.3. Case Selection

To select the cases to be part of the research, a literal replication logic is preferred (Ebneyamini & Moghadam, 2018). It is expected that different cases will predict similar results, however, given the goal of this study, not any company can be an object of study. Ebneyamini & Moghadam (2018) argues that case studies offer the possibility to be designed to suit the case and research question. Thus, firstly, to review the decision process, the definition of a type of decision to be observed is needed. Not all decisions are created equal, and the literature points out that different kinds of decisions have different information needs and are made differently. Thus, one specific big data use case was selected. One of the main sources of value that is perceived by the use of big data is the innovation in products or services and new business opportunities (Günther et al., 2017; McAfee & Brynjolfsson, 2012). Being such a promisor venture for future development and the creation of competitive advantage, the research aims to investigate cases within the context of new product/services and/or market development. Furthermore, this type of decision can be characterized as unstructured and non-programmed, which cannot be sufficiently codified to be automated and often require the aggregate knowledge of individuals. This is expected to offer a better frame of reference to analyze both the decision process and the interaction among stakeholders to derive business value from data.

Secondly, for this to be true, the company needs to have a minimum structure and established processes. A small company, most likely will not have well-defined responsibilities, and all decision rights may lie on a single person, which could considerably limit the analysis of the decision-making process as proposed in this research. Furthermore, even if there may be cases in which this is not true, having both small and large companies could impact the literal replication assumption. Hence, it was determined that the research would select only large enterprises, defined here as having more than 1000 FTE. To account for possible sector-specific influences, the research focused in investigate companies in a broad sector base. Thirdly, the companies selected need necessarily to have a data-based approach to decision-making, and a complete decision process, so that the entire process, from intelligence to implementation, can be analyzed.

Lastly, it was necessary to review to what extent the companies make use of big data. For that, a definition of big data was offered, such as: 'for this research, big data is defined as large datasets with at least some of the following attributes: presents data from internal and external sources, uses different data crossings, modeling and/or data enrichment, is large enough to require adequate data manipulation tools and increases at a rapid pace'. This definition covers the three main characteristics of big data discussed in the literature: volume, variety, and velocity. Based upon it, the companies invited could assess whether or not this definition applies to the context of their data usage. Thus, the case selection aimed for large enterprises, that currently use big data to inform their decision within the context of product/services and market development, with a complete decision process. This is expected to guarantee the literal replication and external validity of the study (Yin, 2014).

Here is relevant to notice the context of the current crisis and the possible impact on the availability of companies and their interest in the research. Given that this situation could make it more difficult to have an appropriate number of cases, the researcher first evaluated the interest of companies with some initial contacts, which turned out to be positive. The topic of the research was considered to be relevant to their current efforts, which implied in their openness to the investigation. Besides, it was noted that, because the professionals are now mostly working from home, it was easier for them to make time for an interview. Furthermore, they showed a clear interest in seeing the results of the study. With that, the researcher assessed that the chosen approach was feasible, and more contacts were made with companies that matched the selection criteria. The approach used was of a purposeful sample (Perry, 1998), and the researcher made use of her network to gain access to companies in the Netherlands and Brazil.

3.4. Data collection

To collect the data, in-depth expert interviews were conducted. As said before, using multiple sources of data within the context of this case-study would be tricky, and the interviews were select as the only source of information. The key-information method was chosen, which in this case meant to hold interviews with one of the stakeholders directly involved in the decision-making process, if possible, the decision-maker. It is assumed that the decision-maker will have an overall view of the process and will be able to point to all different stakeholders involved in the decision and the information-processing mechanisms that are used. Furthermore, he/she is the perfect person to point the effects of using big data to inform decisions, what is missing, if so, what are the constraints, among other information that can shed light on the organizational aspects that surround the decision-making process. In some cases, an interview with one specialist of the technology topic was preferred over the main decision-maker, which was especially relevant in the cases where there was a competence center responsible for the big data & analytics initiatives.

Furthermore, it can be that the decision-maker will have only a limited view on certain aspects of the intelligent phase, the is typically conducted by an analyst. Therefore, it was established that a second interview would be a possibility, to enhance the comprehensiveness of the findings. In these cases, based on the understanding of the researcher upon the first interview, a second interview with an analyst was requested. Note that note necessarily the same decision was discussed, to give the interviewee freedom to choose the decision but the foci use case was maintained. This has some other benefits, which is to compensate for the lack of supporting documentation and minimize the chance of one-informant bias.

During the interviews, a retrospective report method was used. With that, the interviewee was asked to pick a complete decision that was made recently, upon which the decision-making process would be discussed. The fact that was asked for a recent decision help to enhance the reliability of the information that was given, by minimizing the time elapsed between the decision and the report. Also, the interviewee was asked to choose one specific decision, that should be complete so that the full process could be addressed, from intelligence to implementation. This approach was used both with the decision-maker and analyst, but not necessarily the same decision was chosen. This was left open, so that the interviewees could choose the decision themselves, given the selection criteria that were given.

To guide the data collection, a semi-structured questionnaire was produced. This was divided into three parts: (i) it first covered the background information on the department and use of data for decision-making; (ii) it then moved on to the retrospective report, in which the type and context of the decision were asked, and the process was discussed; (iii) at the end, some generalizations upon this specific decision were made, to assess how well it represents the overall decision-making process at the department, and opportunities for improvement were discussed.

The retrospective report covered the larger part of the interview and is the most unstructured part of it. The goal was to have a barely content free inquiry, to make sure that the questions would not create a self-fulfilling prophecy (Ebneyamini & Moghadam, 2018). Thus, the interviewee was asked to talk openly about the decision, with only minimal interference of the researcher. In this sense, the researcher's role was only to instigate and encourage the interviewee and invite him/her to discuss some aspects in more detail, according to the goal and research questions. A pilot interview was conducted to assess the fitness of the questionnaire, and the appropriate modifications were made. Appendix I presents the questionnaire in its most recent form. One of the interviews (I 12 – see Table 1) did not follow the proposed structure, due to restriction in time. For this one, the main aspects of the structure were converted into a series of questions covering the background information on the department, how the work is done and structure, and opportunities for improvements.

Moreover, the interviews were conducted online, due to the current health crisis. The researcher invited the interviewees to keep the camera on so that a better reading of the reactions of the interviewee could be made, which would approximate the online interview to a traditional face-to-face meeting. The interviews were conducted between May and June 2020. Each interview lasted approximately 1 hour, and all of them were audio-recorded, guaranteeing that no information would be lost.

3.5. Overview of Cases

Table 1 presents the case companies and each of the interviews. In total, 13 interviews were conducted, in 9 different companies, from diverse industries. One interview was later excluded from the sample because the decision discussed ended up detaching from the remaining ones. All companies are large enterprises, most often multinational companies with units across the globe. The country registered in the table refers to the location of the office which the interviewee belongs to.

Table 1. Cases overview

Interview ID	Case ID	Industry	Country	Interviewee role
I 01	C 01	Payments	Brazil	Product Manager
I 02	C 01	Payments	Brazil	Business Intelligence Analyst
I 03	C 02	Insurance	Brazil	Product & Markets Coordinator
I 04	C 02	Insurance	Brazil	Business Analyst
I 05	C 03	Medical	Brazil	Marketing Manager
I 06	C 04	Rail	Brazil	Commercial Manager
I 07	C 05	Biotech	Brazil	District Sales Manager
I 08	C 06	Consumer	Netherlands	Consumer & Commercial Insights
I 09	C 07	Telecom	Brazil	Big Data Director
I 10	C 07	Telecom	Brazil	Analytics Director
I 11	C 08	Bank	Netherlands	IT Area Lead
I 12	C 09	Telecom	Netherlands	Analytics Director

Between the cases, that were differences worth noting. Cases varied per perceived level of big data maturity, per type of decision discussed, and per the structure of the data/analytics competence. In the cases where more than one interview was conducted, the type of decision discussed was the same. Hence, all classifications were made in terms of the cases.

Since all the companies interviewed are currently making consistent use of (large) datasets to inform decisions, it can be said that at least an initial maturity stage could be noticed throughout the cases. Still, it was evident for the researcher that some cases were clearly more mature in terms of (big) data usage than others. Hence, it is relevant to note this differentiation, as it can impact the data analysis. However, it was not the goal of this research to assess the level of maturity of any of the cases, and making a reliable classification based on the most recent maturity models available in the literature only based on the interviews would not be possible. Therefore, a simple classification is being made in terms of lower-higher maturity perceived by the researcher, in a comparative assessment between the cases studied (Figure 8).

Lower Maturity	C 01; C 02; C 03; C 04; C 06
Higher Maturity	C 05; C07; C08; C 09

Figure 8. Classification of cases per perceived level of maturity

In terms of the decision discussed, all of them referred to applications in the product/market development and, therefore, can be said to be unstructured and novel decisions, fitting into the non-programmed category defined by Simon (1960). Within this category, three main groups emerged: (i) decisions regarding the evaluation of a new product (New Product Evaluation), including the assessment of the market and target group and/or new opportunities; (ii) decisions regarding the strategy and/or operationalization of a new launch (New Product Introduction), including market assessment, and delivery of strategic action plans; and, (iii) decisions regarding improvements in the performance of an existing product or service (Product Improvement), normally upon a spotted problem or strategic changes that emerged. That was also one case where the subject discussed was a project not directed to an immediate decision, being an evaluation of the market to support futures developments, therefore it is being considered together with the evaluation of a new product category. The classification of each case is presented in Figure 9. It is important to note here that I 12, due restriction of time, did not cover a specific decision and, therefore, this case (C 09) is not being classified in terms of the type of decision.

New Product Evaluation	C 02; C 06; C 07
New Product Introduction	C 03; C 05
Product Improvement	C 01; C 04; C08

Figure 9. Classification of cases per group of decision

In terms of the structure of the analysis/analytics competence, two different structures were noted. The first structure (Structure 01) is when there is a sort of competence center dealing with the data analysis to orient decision-making in the company. When this is the case, there is a team of analysts, data scientists, data engineers, and so forth, responsible for the analysis/analytics work and technology application, and the business units are the demander, coming with the problem/opportunity to be studied. The second structure (Structure 02) is when the same department that 'owns' the subject is also the one conducting the analysis. This doesn't mean that the company does not have an IT area or a Business Intelligence area. It is just that in the case of the decision discussed this area only offer support in terms of structure and access to the data but is not involved in the analysis. Also is relevant to note that the different structures are not correlated with the big data maturity of the company. Figure 10 presents the classification of the cases in terms of structure.

Structure 01	C 06; C 07; C 09
Structure 02	C 01; C 02; C 03; C 04; C 05; C08

Figure 10. Classification of cases per type of structure

3.6. Data analysis

Ebneyamini & Moghadam (2018) say that 'data analysis consists of examining, categorizing, tabulating, or otherwise recombining the evidence to address the initial propositions of a study' and they add up that the analysis part of a case study research is the least explored one in literature. Yet, different authors point to possibilities, where looking for cross-case patterns or synthesis among the data dominates. For this research, the approach presented in Burnard (1991) was used, having coding as the primary way of analyzing the data, allowing the researcher to spot common themes in the interviews, combining them under a reasonable exhaustive category system.

The main goal of this research is to evaluate the fitness of the existing models in real case scenarios. Therefore, coding techniques were used to find patterns among the data, representing the steps undertaken to make decisions, identifying the interactions among stakeholders, and the information processing mechanisms used, which will form the category system. This structure enabled the comparison among strings of text from different interviews, identifying convergences, and possible divergences. From this, it was possible to assess whether or not the steps conceptualized in the researched models can be founded in the practical cases or not, and what are the main points of distinction emerging from the sample. Furthermore, it was expected that these patterns would also become apparent for the organizational elements.

The coding process defined by Burnard (1991) assumes the use of semi-structured, open-ended interviews that are recorded to be transcribed in full, which is the case of this research. Thus, the first step was to transcribe the interview records. For this, a software for the automated transcription was used, called Amberscript. The researcher later verified all the transcripts in full to check for inconsistencies. Secondly, all the interviews conducted in Portuguese had to be translated into English. In this process, it was made a literal translation, word by word, by the researcher, whose native language is Portuguese. From this, the data was ready to be analyzed.

The third step was to code the data. The researched started with open codding, identifying the different topics within the interviews. Open coding was especially relevant in this case so that new categories could emerge from data, representing possible news steps or topics not covered by the existing literature. Without this, the exploratory aspect of the case-study would be lost, and the researcher would be blindsided by the existing constructs. Hence, an inductive approach was used, having the existing steps as the baseline to build the categories, but keeping the necessary flexibility so that new categories can be captured, characterizing a deductive approach. Other themes of relevance identified throughout the interviews were also considered. For the coding process, Atlas.io was used as a supporting tool, a platform that makes recording existing categories easy and straight forward. Each interview was codded once separately, and after this was done, a review was made to identify any missing themes or misfits. This resulted in a list of 47 categories of codes. Additionally, two participants were invited to check for the appropriateness of the category system and punctual corrections were made when necessary.

Once this was done, the fourth step was to analyze the data together to make the necessary evaluations, according to the research question and sub-questions. To allow the comparison among them, all data was transferred to Excel, organized in tables where each row was a theme and each column an interview. This allowed the researcher to jointly analyze all the data, identifying common topics, patterns, and possible discrepancies in a visual way. For this analysis, groups of cases were also considered together, in terms of classifications presented in section 3.5. Overview of Cases.

To address the first sub-question, related to the appropriateness of the steps defined in the literature, the models found in the literature were analyzed together, and all steps were identified, as well as the overlaps between them. The model described by Chiheb et al. (2019) was considered to be the most comprehensive one and therefore was chosen as the blueprint against which the cases would be compared. However, it is relevant to note that the other models were not disregarded. The steps identified in them were still kept as a reference to help analyze any divergencies with the blueprint and to broadly assess the phases of the model. As such, later in the discussion, it was possible to identify which models better cover the steps identified in the literature and further evaluate the discrepancies among them based on the practical insights. Having this blueprint in mind, the tabular structure of the codified data made it easy to visualize which cases covered which step of the process and review how each case deals with each step of the decision flow.

To address the different stakeholders involved in each step of the process, a RASCI matrix was used. It offers an easy to understand, visual tool to enlist all stakeholders and the role they take in that particular step. RASCI stands for, responsible (R), those responsible for executing the activity; accountable (A), those the approve the task and responds to the quality of the work produced; support (S), those who can offer support to the ones carrying out the task, but are not held responsible for it; consulted (C), those that should be consulted and can offer valuable advice for those responsible for the execution; informed (I), those who should be informed of the tasks being executed, and kept in the loop for new information and status update. The first step to construct the RASCI matrix was to identify the common stakeholders, that were divided into 6 categories. This was done by analyzing the data together. Secondly, using the tabular structure, the roles of each stakeholder in each step were identified and the matrix was construct based on role undertaken in the majority of cases. In the cases where a stakeholder can have a different role depending on the context, this was noted in the matrix, which received a double identifier (i.e. S/C). Note that one rule applies to the construction of the RASCI matrix, which is that that cannot be more than one person accountable for one task.

Lastly, to address the information processing mechanism used in the decision process, the construct defined by Kowalczyk & Buxmann (2014) was applied. For this, it was noted that the analysis of the data by each step of the process, as was done for the other two sub-questions, would not bring the most insightful findings. Literature posits that different types of decision posses different challenges and therefore demands different information needs. Therefore, an investigation of the mechanisms by the different case classifications presented in section 3.5, was chosen was more relevant for the aimed discussion, in particular considering the different groups of decision. From this grouping, the usage of the different information processing mechanisms was graded on a scale ranging from very low usage to very high usage, based on an average from the cases. From these analyses, the process of writing up started. The results identified will be presented in chapter 4, followed by the discussion in light of the relevant literature in chapter 5.

4. Results

In this section, the findings that emerged from the analysis of the multiple cases are presented. First, it will be presented the results concerning the steps of the decision-making process and after the findings regarding the organizational aspects. Throughout the results, quotes from the interviews will be used to illustrate and provide evidence to the subjects presented. However, no reference to the case or interview the quote was taken from will be made due to privacy concerns.

4.1. Decision-Making Process

For all cases, the main structure of intelligence, design, choice, and implementation is clearly defined. How these phases are broken down varied among the cases, but common structures were possible to identify. It is worth noting, especially for those of companies that have a process more clearly defined, that how they name and how they separate each step of the process may be different. Therefore, its content, the activity that is supposed to be performed is what is taken into consideration for the analysis of the process.

Overall, the four main phases and the step within them form a hierarchical process. However, it is not a sequential one, and lots of iterations between the steps and between the phases were identified. Considerations regarding this will be made when deemed necessary. The findings for each of the steps are described below.

4.1.1. Intelligence Phase

The intelligence phase was the one with less homogeneity across different cases. The cases with higher maturity present a more established stage of defining and modeling the decisions, while for the cases with lower maturity, this does not occur in a structured way. Data collection and preparation are covered in all cases, the main difference is related to the extent to which new data needed to be integrated. In this sense, the cases where the decision was based in already existing decision supporting systems the collection and preparation are resumed, while for the others the process of preparing the data for the analysis was deemed extremely relevant.

Define and Model Decisions

Identify Decision

Across the cases, all decision started consistently in an identification step, triggered by changes in the company's performance, monitored through KPIs, requests from different areas, market trends perceived by the team and/or movements of the competition, changes in rules and regulations, and inputs from the front-line employees that are directly in contact with the clients/consumers. Front-line and KPI monitoring were the most common triggers to identify a new decision in product/market development, raising attention to possible improvements and/or rising demands.

> *"We already did a great job of identifying which are the KPIs that really matter for us, that represents well the company's behavior, so we did this and now we just follow them."*

In the case of real-time data, KPIs monitoring helps the team to readily visualize disruptions, upcoming issues, or rising trends that need to be assessed. But up to that, the front-line employees were revealed as a relevant source of information to uncover new market demands and even to pinpoint disruptions that may be yet too small to be caught in the existing KPI or were not covered by them. So, they act as a valuable source of input to anticipate trends and unsatisfied needs. This is to point out that not always is the data itself that will trigger a decision, and the more qualitative insights still showed up as relevant to start the process.

> *"This is also one of the biggest inputs that we receive. So, when the sales team talks to the customer, he may say 'why don't you make a product like this?' or 'I want this kind of product and you don't have it'. So, these inputs are brought by the team that is on the day to day of sales, commercialization, relationship, and customer satisfaction, which makes it possible for us to have inputs."*

However, what came out from the interviews is that once this input gets to the team, it needs to be qualified, i.e., they need to assess the size of the problem or if there is indeed an opportunity there. When it is a request from another area of the company, as said by one of the interviewees, you need to assess if the demand stands on its own. Only after this process of qualifying the input a decision is identified.

> *"We need to check if it really represents a market need. We try to quantify it. Because, if only one person comes with this, ah, ok, we turn on a light, but if there are two, three, four of them, saying the same thing, then you start saying 'Wow, this is recurrent'."*

Define Decision

Defining the decision was found in most of the interviews. This is the step where you generate alignment around what is the subject, what is the problem, what is the scenario, what needs to be looked at, what are the constraints. The most important thing to define here is what are the questions that need to be answered, and, if needed, define the quality standards against which the decision will be measured. This stage was considered as extremely relevant by the interviewees. Even by those who said that this did not happen in their departments, this was recognized as a latent need.

> *"So, everything should be specific, should be measurable, should be attainable, relevant, and time-based. And if you have these things, you define, in our case, does this problem make sense from the business point of view? […] So that's the first thing. If we know the goal of that we are going to do, that is clear.*

Specify Decision Requirements

The difference that was shown in the data is that there is a moment to understand what the decision is (previous step) and a second moment to understand what is required to make this decision. These two often occur together, but not necessarily. So it can be said that defining the decision deals with defining what is the problem statement, while the decision requirement deals with the definition of how we are going to answer it, translating the business questions into the data and technology needs. In other words, identifying what is the data that represents this problem, what are the technical needs, and so on. The complexity of this step will be linked with the complexity of the problem statement.

> *"We do a screening to analyze if I have the necessary knowledge, if I have the right technology, if I have the data, if I have access to that data and the data is good for analysis, because if I don't have the data, I have to wait to have the data first to be able to start. And we assess how difficult it is and how much return I can have."*

Interestingly to note, these cases showed that the process of defining the requirements and identifying the data sources do not have a clear separation. At the same time that you identify the data needed, you also assess if this data is available, from which source, if it is accessible, what is needed to acquire this data. This assessment is necessary to evaluate what is possible to do, the necessary effort, the people you need to involve, or yet if perhaps a new variable may need to be introduced to account for some missing data.

> *"And then we also define, at this initial discussion, if we need any other type of work from the developers to collect this data, and we define how this will be done."*

Lastly, is important to note that while the steps of identifying the decision and defining it were found consistently among the data, specifying the decision requirements was only partially supported. In particular, not all companies mentioned the assessment of the specific data needed to make this decision. In some cases, what was seen is that it is left to the analyst to define later what data will be needed. But in the higher maturity companies, the data assessment was consistently mentioned as particularly important to the continuity and planning of the work to be done. So, this covers the evaluation of what data is needed, what you already have, what needs to be acquired or ingested, if this data can be obtained, how, and the effort and time necessary for doing so. It is an assessment of the instruments you have to move forward.

> *"This would be the correct process, defining what questions I have to answer, defining what are the variables I will have to evaluate to get to this answer, and then I will provide a report. That's not how it happens in real life. Not where I've seen it."*

Refine Decision

Only one case mentioned a refinement step. Based on the discussions of the previous steps the decision can be evaluated against its value to the business and demanded effort, so the work can be prioritized. Even though it was not possible to assess this aspect thoroughly, since it only appeared in one case, it seems like the more complex the decision, the more that will be a need for refinement. This can also be the case when you have a limited capacity and may need to prioritize the work that you be done by the team.

> *"After we analyzed the complexity, we do another round of prioritization, and we say, 'Look, we have already mapped it out, we analyzed the complexity, we estimated the value, now we have this matrix of value x complexity and we think that we have to prioritize this and this'. This is usually a very quick meeting. So, I don't know, out of the 50 things we wanted to look at, these 10 are easier and have more value, we'll start with them."*

Data Collection and Preparation

Identify Sources

Once the decision and its requirements are defined, the referenced models indicate a step of source identification. As it was discussed in the Specify Decision Requirement step, in part of the cases, especially the higher maturity companies, identifying the sources has more to do with planning. Based on the problem definition, the stakeholders will discuss the needed data and the complexity related to acquiring it, thus, identifying where is this data and how it can be acquired. In this process, having defined well the problem or opportunity that is going to be tackled is relevant so the stakeholders can think on the process behind it, so they can identify the data points.

> *"So, you think about the process, and you think about what different data sources that we have. We have the data sources from customer service, from the technical operation, the information, and reports, and interactions with the customers, that might not be structured. So, first thing you have to filter, before starting to put this into the melting pot, you have to see which one is relevant to you. So, to give you a bottom-line, what we try to have is a better customer analysis at the end."*

However, in some cases there was no mention to a specific moment to discuss the data sources, be it internal or external. In these cases, it is the analyst who identifies the data he/she will need based on the decision requirement and their prior experience and expertise. Then, this same person requests the data to the appropriate department or collects it himself from external sources or the internal database they have access to. Thus, there is no separation between the steps of identifying the sources and collecting the data.

One other thing that came out from one of the cases is the idea of a sourcing strategy to ensure data integrity and integration. This accounts for the possibility of something being incorrect with the registered data. So, if inconsistences are possible, there may be a need to reconcile the data, avoid falling into the issue of having garbage in, garbage out.

> *"So, first define which are the relevant and important sources that you have. And then the other thing is, what is your sourcing strategy? So which data it's the one that it's in the lead. Because, you should have the possibility of something not being correct, right? So that's where you introduce a reconciliation in the sourcing strategy."*

Data Collection

Then, from source identification, you either move on to extract the data from already existing databases or to acquiring new data and storing it. New data can already exist outside the company, like publicly available sociodemographic data or data from regulatory boards, or may need to be generated, through a survey with customers, or be captured through the creating of a new log in the company's operations. Those are the three options that came out from the data.

> *"If you don't have the data to form that KPI, you would have to enter a new code, or a new log to capture that particular information that you understand it will be necessary to show a certain result."*

When it comes to data collection, one trend that appeared in the interviews is the test & learn approaches, and the need to analyze real-time data, putting more emphasis on the acquisition aspect. Especially when it comes to digital products, test and learn was shown to be relevant to help make better decisions at the end.

Data Preparation

From data collection, there is the data preparation step. Interestingly this step was not often cited, but when it did, it was said to be the most critical aspect in this stage. It was mentioned when the interviewee had a more technical profile or was the one personally responsible for this execution. The decision-makers didn't seem to be much aware of this process. As one of the interviewees noted, this process is complex, no one likes to do it, and it doesn't generate value by itself, but it is mandatory to enable it to happen.

"Prepare the data is the hardest part of these projects, it is more difficult than ingesting, it is more difficult than analyzing, it is more difficult than discussing."

Regarding data preparation, how the data is structure was also pointed as relevant to facilitate the process. It was especially noted that when it comes to the legacy systems, coordinating all different sources, with different structures and so on adds up to the complexity of preparing the data, besides being time-consuming. Anyhow, preparing the data, and having it ready is necessary to move to the second phase of the decision-making process, the design phase.

"What happens is that we have a structured for the tables but depending on who made them they are structured differently. So, I have to do a lot of mathematical work to be able to create a table that I can consume. So, here we say that our queries are bizarre, they are Frankenstein's."

4.1.2. Design Phase

The design phase was not described homogeneously in the interviews. Some interviewees see this as one big step of modeling the data, some see it as the data analysis purely, some others break it down into more specific steps. However, in the end, two distinct moments emerged, building a model, where the subject will be investigated and insights will be generated, and then generating alternatives, where a course of action (or more than one) will be proposed based on the results of the previous analysis.

Build Model

How the modeling step is structure was not homogeneous across the cases. Some defined it thoroughly, with planning, building, and testing. Others presented it more like a series of trial and error, of experimentation with the data until something interesting was build. But in the end, there was always a model, a form of analyzing the data together to enlighten a subject, even if it was not a planned model from the beginning. A step of planning the model, of actually defining how the work will be done was only found in the higher maturity cases. Most often, the analysts will work in a more exploratory manner, relying on their expertise and prior experiences. But even when there is initial planning, the modeling phase was described as a very exploratory, time-consuming, and complex step.

"It is very organic, we add things as we find the information, we get new information, we collect it and include it in this analysis. It's left to the analyst to understand which information he has, what he does not have, what he wants to show... As this is discussed with other people, new questions may arise, and then we will include it. This process is long."

The design phase emerged as the most iterative phase of this process. Withing the modeling step, several assessments and refinements were mentioned. While the analysis is being conducted and different scenarios are being built, there is a constant assessment with other stakeholders to evaluate the preliminary results and to refine the work. New data may be needed, new constrains can be found and the models may need to be reviewed. The complexity of this step is related especially with the tools and expertise available, the data sources, the amount of data, and if real-time experiments are going to be employed or not. What then emerged from the data was a constant loop of (plan) - build – assess – refine, until a satisfying result is reached.

"You cannot create things out of the blue. So, everything should be based on the analysis. Everything should be created, okay? We should see what is the reality, what is the foundation. And then, based on this we should add the experiments and all the other things on top. And, in the end, you get the knowledge."

Generate Alternatives

From the modeling phase, the outcomes can be used directly to answer a yes or no question or to generate insights for a different team to work with. But most likely, when we are talking about the decision studied here, these outcomes will be leveraged into decision alternatives.

"In some situations, we take different scenarios to make a decision, and in others, we are going to have only one final scenario, because the decision was whether to launch [a product] or not. It is a yes or no question."

The team needs then to continue the work to generate and discuss possible solutions or suggestions on how to tackle the subject. Most often, what the cases showed is that the model alone does not solve the problem, and solution(s), action plan(s), or scenarios still need to be defined to provide alternatives for the decision. These can involve new discussions, bringing in different stakeholders. Methodologies like design thinking and design sprint were also cited. These alternatives are then the proposals brought to discussion in the choice phase.

"Then you create scenarios. So, you have a scenario A that can tackle the problem but can impact the profits, or we can try to go for scenario B, prioritizing the customer. And then this is taken for discussion. We can also still have a third scenario." [3]

4.1.3. Choice Phase

A decision is defined as choosing one specific course of action. What the interviews showed is that the main decision-maker is not often part of the design phase, therefore the need to structure the results and bring

[3] The specific scenario and its impacts were generalized for privacy concerns.

it to a meeting where the final decision can be made. So, the choice phase consists mainly of the presentation of the results and its discussion to reach a decision.

This discussion appeared as an evaluation step. Thus, the process that happened before is explained to the decision-makers and is evaluated, i.e., it is checked if all the variables were considered, if everyone that should be involved or informed was part of the process, if the data is reliable, if the quality checks were met. In the end, the modeled outcomes and the alternatives are used to justify what is being proposed and to claim that it is the best approach. The goal is to get the decision-makers to agree with it, so a choice can be made, and the decision is implemented.

> *"At the end is a leverage. But you should use the facts, you should simply be facts-based. In the end, analytics should be something that helps you out, I mean, this is the fact, this is how we do it, based on the experiments, on different factors. So, we proved, we anticipated, we measured, we optimize, we implement. And then we retrofit."*

What came out of the data is that this process is seldom done at once. There are different checkpoints with the stakeholders involved in the decision, but especially when we are talking about the higher management and c-level, there will be different levels of approval until a final decision is made. In all of these levels, some information may be added and the model, and consequently the solution, is refined to attend to new inputs from these stakeholders. In this process, may be necessary to go back to the previous steps, perhaps event to the intelligence phase, depending on what was the outcome of the evaluation step. In general terms, the more the analysis team know about the needs of the decision-makers, the less iteration will be needed.

4.1.4. Implementation Phase

Once the decision was made, the implementation phase can be carried out. This is when the chosen course of action will be implemented or operationalized. In this phase, the most relevant aspect is the monitoring of selected KPIs and feedback.

Monitoring & Feedback

Without exception, all interviewees noted that once a decision is made, the results are monitored through KPIs and this generates new information to the teams, that can be then used for improvement or to generate new opportunities. Less structured and more qualitative feedback is also collected, especially assessing the implementation results with the front-line people. This is when the cycle is closed, and this process can generate inputs that will trigger a new decision in the future, starting a new process.

> *"There is a feedback process that is the direct monitoring of the market. So, we look at how is it going, also checking with the customer, if he is satisfied, often through surveys, and also operational monitoring, because if there is a problem with the operational system, there is something wrong."*

When it comes to the test & learn approaches, the implementation phase may take a new form, of faster feedback loops, and closely attached to the design phase. In this scenario, the test happens with real customers, with a real product/solution in the market, even if in a beta version. This topic significantly reinforces the iterative nature of making decisions, especially when we consider digitization. The existence of at least a pilot test before deploying the solution throughout was identified in different kinds of industry, even the more traditional ones.

Furthermore, a real-time approach for monitoring and feedback was also noticed in most of the cases. The companies said that there is always a real-time tracking of the implementation and its results, so corrections can be made when necessary. Another recurring aspect of this phase is that closer monitoring happens at the beginning of the implementation and, then, once some level of stability is reached, the pace of monitoring is decreased or move to a higher level of analysis, with more macro indicators. Anyhow, the monitoring of the results was consistently stressed, due to an increasing need to measure what is being done, and to evaluate if the decision made was a good one.

> *"If you can't measure it, it will always be difficult to see the result. Because normally these data disciplines, they do not do the business, they enhance the business. [...] So, you need to measure, even more so to criticize what is being done and improve the models, the data, the decisions".*

4.2. Organizational Elements

The communication and collaboration between different areas of the company were shown to be extremely relevant for the decision-making process to run smoothly, as it was the involvement of the right stakeholders. A summarized view of the different stakeholder's categories identified and their involvement in each of the steps will be presented, followed by the frequency of usage of the different information-processing mechanisms. After, an overview of the role of the information processing mechanisms and the stakeholders' involvement will be explained throughout the phases of the decision-making process.

4.2.1. Stakeholders

In terms of stakeholders, 6 categories were identified, defined below:

Business Owner: So first, there is the person who owns the decision. This is the person who came with the request in the first place, that will be the person who will run the solution in the end, and that decides to operate the product/solution in one way or the other. This person can be the product owner or product manager or the business owner. The term business owner will be used here.

Analysts: Then you have the data analyst, data scientist, or business analyst that are the ones directly involved in the analysis and analytics. The term analysts will be used to refer to them.

Technology Enablers: There is also the technical team, responsible for managing the data, ingesting, preparing, and making sure that the needed data is available for the analysts to consume. They may be the data engineers, the IT/Technology team, depending on the structure of the company, but there is always a technology layer that is relevant for the process to run smoothly.

Transactional Stakeholders: the decision seldom belongs to a single department, and there are other areas affected by the decision and the also need to provide inputs to the solution. These are the areas that are involved in the operationalization of the solution or support areas. This may be marketing, legal, customer support, operations, logistics, and so forth. It is the job of the Business Owner and Analysts to identify which areas are this.

Higher Management: besides the business owner, that are often other layers of decision that need to give their final approval so that the solution can be implemented at the end. Those are higher managers or even the c-level. Each department has its layers of approval, and it is also the job of the business owner to identify the level of approval he/she needs to get.

Translators: One stakeholder that came out of the data in two of the higher maturity cases is the figure of the translator. This stakeholder is particularly important when there is a competence center working together with the business units to advance the use of data in decision-making. He/she is the person that understands how data can be used to solve business problems. It is not necessarily a technical person, nor an analyst, but is someone that bridges the analytics/technical department with the business.

				Business Owner	Analysts	Technology Enablers	Transactional Stakeholders	Higher Management	Translators*
Intelligence Phase	Define and Model Decision	Identify Decision	Input Generation	A/R	S		R/C		S
			Input Qualification	A/R	S		C		S
		Define Decision		A/R	R	C	S	I	S
		Specify Decision requirements		C	A/R	S	C		S
		Identify Sources		C	A/R	S	C		S
		Refine Decision		A/R	R	C	C	I	S
	Data Collection and Preparation	Data Collection	Data Extraction	I	A/R	S	C		C
			Data Acquisition	I	A/R	S	C		C
			Data Storage	I	S	A/R	I		I
		Data Preparation		I	S	A/R	I		I
Design Phase		Build Model	Plan	A/R	R	C	C		S/C
			Build	C	A/R	S/C	C		S/C
			Assess	A/R	R	C	C		S/C
			Refine	C	A/R	S/C	C		S/C
		Generate Alternatives		A/R	R/S		C/S		S/C
Choice Phase		Decision Evaluation		A/R	C/S		C	R	S/C
		Choice		A/R	C		C	R	S/C
Implementation Phase		Implementation/Operacionalization		R/A	C		C/R	I	
		Feedback & Monitoring		R/A	C		C/R	I	

Figure 11. Stakeholders involvement per step of the decision-making process

To better understand how these different stakeholders are involved in each of the steps, Figure 11 presents the RASCI matrix constructed based on the interviews. It defines who is responsible (R), accountable (A), support (S), consulted (C), or informed (I) in each step of the process. It is relevant to note that to the exception of the translator, all other stakeholders were cited in all interviews. There may be other stakeholders not covered here, but these 5 will always be present to a greater or lesser extent, depending on the complexity of the subject. It will always be the role of the Business Owner and the Analysts to identify the other stakeholders and to decide on their participation. Furthermore, even though the translator appeared only in two cases, it was cited with great regard and mentioned as strong leverage to the work of the competence centers. However, it may not be necessary to have such a stakeholder in a structure where the business owners and analysts belong to the same department or area, since their main job is exactly to bridge the interaction between those.

4.2.2. Information-Processing Mechanisms

Besides identifying the stakeholders, understanding how they communicate and collaborate among each was also one of the goals of this research. Figure 12 presents the different information-processing mechanisms that were identified, based on the work of Kowalczyk & Buxmann (2014). To present these findings in a concise form, the separation by group of decisions was used. It is relevant to note that while the differences in data-centric mechanisms used have to do mainly with the type of decision, the differences in organizational mechanisms are also related to the company structure and level of maturity.

	Data-centric mechanisms				Organizational mechanisms			
	Predictive analytics	**Dashboards**	**Ad-hoc & descriptive analysis**	**Data mining**	**Planning**	**Direct contact**	**Integrator**	**Group meetings**
Product Improvement	○	◕	●	◔	◕	◕	○	●
New Product Evaluation	○	○	●	◔	◑	◑	◑	●
New Product Introduction	○	●	◑	○	◕	◑	○	●

Classification: ○ = very low usage ◔ = low usage ◑ = medium usage ◕ = high usage ● = very high usage

Figure 12. Usage of different information-processing mechanisms

Ad-hoc ad descriptive analysis is the most common data-centric mechanism used across the cases. In particular for new product evaluation or specific improvements, the analysis of data based on initial hypothesis or scenarios are predominant. In this case, it is necessary to raise information and build a study or a plan that supports or not the introduction of a new product for a target group or the answers why an indicator has worsened and what to do about it. Data mining is starting to be more used, especially when it comes to uncover clusters of customers or to identify other patterns or trends relevant to the analysis. However, across the cases, it was seen that many times, the analysts still do not possess the necessary knowledge to deal with more advanced technologies, therefore the ad-hoc & descriptive analysis continues in predominance. Data mining is being used more consistently in higher maturity cases. On the other hand, existing dashboards appear to not cover the needs of product development and were mostly used in the cases where a new product implementation was discussed since in these cases an overview of the current operations of the company is necessary. Ad-hoc analysis is not excluded though. It is common across the different classes of decision the integration of data from different sources and different types, like quantitative and qualitative data, strengthening the value of the study. Predictive analysis was the only one not seen in any of the cases.

On the organizational side, meetings are still the main information-processing mechanisms used by the different stakeholders. The already scheduled and routine meetings are often used for alignment and to check the status of the actions. Also, specific meetings can be scheduled to discuss particular aspects of the decision. The figure of the integrator assimilates the role of the translator, and therefore only appeared in the two higher maturity cases where a competence center exists. Direct contact is often used and tends to be mentioned as relevant for the process to move quickly. It is often the role of the analysts to reach out for different stakeholders that may possess the necessary information or knowledge being demanded. Planning on the other hand varied more with the company structure than due to decision type. What was seen is that companies with higher maturity tend to sit more together to plan and define the work to be done, especially when the parallel work of different stakeholders is needed.

4.2.3. The organizational elements in the decision-making process

At the early stages, the involvement of the transactional stakeholders, in particular, the front-line people are particularly important to the generation of inputs. Their insights and information captured through contact with the clients are particularly important to motivate new product development and/or improvements. These inputs most often get to the decision-makers and analytics/analysis teams through indirect contact, via existing relationships maintained between these areas. From there, existing meetings and ceremonies are used to align the different stakeholders around an initiative and to keep them informed about the status of the actions.

"So, basically, there's already, like, buy-in from all the stakeholders from the beginning."

For the definition and modeling of the decision, planning was cited as particularly relevant and the interviewees repeatedly pointed out the importance of having all relevant stakeholders involved. Is relevant to note that when it comes to identifying the data and the sources, is relevant to need different stakeholders involved. First, because they may have access to different data, and this can be leveraged in the discussion. Secondly, because it is necessary to access which is the relevant data to answer the defined questions, business acumen needs to be employed here so that can be an understanding of which data is relevant. Again here, it was noted that the higher maturity companies do this more routinely than the lower maturity ones.

"It's not just one person, everyone who understands the subject has to be there, for us to discuss what we need to analyze"

Here is necessary to note that the two different structures that emerged from the cases impact how the different stakeholders are involved. When there is the competence center responsible for the analysis, the involvement of the business area, in the figure of the business owner, was cited as of utmost importance for the success of the project. In these cases, also the figure of the translator appeared. In particular, having a clear context of the business scenario, having the business knowledge, was consistently mentioned as particularly important for the data analysis. So not only the questions to be answered need to be clear, but the knowledge of what that data and its outcomes mean in terms of the business is necessary to generate value at the end.

> *"If those who feel the problem, who feel the pain on the day-to-day operations, are not there to accompany and make it happen, it will not happen."*

In the design phase, when it goes to capture the relevant information for the analysis, especially related to the ad-hoc & descriptive studies, the relationship with the transactional stakeholders is needed. In these cases, indirect contact is often the most common information processing mechanism used. The larger the decision, the less it will be determined by only one area, and normally different stakeholders need to input different information. The same goes for dashboards since different areas normally have access to different sources and it needs to be brought together. For these cases, meetings were most often cited as the main form of interaction to debate and generate a common knowledge of the scenario.

When it goes to generating alternatives, to actually apply the generated knowledge in a course of action, the involvement of the different areas, mostly those who will be impacted by it, was cited as crucial to creating a better proposal. In these cases, a meeting is often necessary to debate the outcomes from the models and craft a solution. Strategies like design sprint and design thinking were cited as helpful in this process.

> *"Usually, for the construction of the full analysis, we depend on these other areas, which will bring the other insights about issues and other problems, and usually the solution comes from joint debates. Hardly it will come from a single area, obviously, it depends on the complexity of the problem, but there is not often the case"*

Two aspects were cited when it comes to involving the stakeholders, knowing who to involve and making sure that their involvement will not impact their other activities, i.e., establishing a trade-off between getting their involvement and not consuming too much of their times. One of the interviewees noted that when they start a data project, they try to set up a multidisciplinary team and they already align with the different stakeholders that it can be time-consuming, so everyone is aware of that. Another claimed that having more people involved reduces the time needed to get to a good solution at the end. In the end, who and when to involve is a judgment call made by the analysts and business owner.

> *"Sometimes, during the discussion, when you have more people, even if it may appear to get in the way, you are actually already able to kill a bad idea there or discover something new. […] I think that the more different views you have, you tend to form a more complete scenario, one that is shielded against real-life surprises."*

It is also relevant to note that the proper involvement of the transactional stakeholders and gathering information with those who are the authority in the subject is appreciated by the higher management when they are evaluating the decision. This helps to validate the proposed solution.

"These internal relationships are very well established. And another thing that we take great care of is to make it clear that the data is not mine. So, this makes us able to bring credibility. So, it goes like, 'who did you discussed this subject with?' 'I got it from so-and-son, in that area', 'Ah ok, he is the authority to talk about this subject, so that is fine'. So, it helps us to seal this story too. "

For the choice phase, there is always a meeting with the decision-makers to present the scenario and the proposed solution. This is when normally the higher management is involved, with the other necessary stakeholders. Here, a presentation is often used to tell the story, from the data used to the different inputs received during the process. Here is relevant to note that when the business sides took part in the construction of the solution, the decision is facilitated.

"And I always say, I mean, the best ERP, big data tool in the world is Excel, and the best showcase in the world is PowerPoint."

When it goes to the implementation phase, the meetings are less frequent and it is up to the area responsive for the implementation to monitor the relevant KPIs, often through dashboards. Normally ad-hoc meetings are necessary when there are issues with the implementation, and the different stakeholders may need to be brought together to evaluate what happened. Normally, they shared the risk of implementation. Also, depending on the situation, especially when there is an intention to generate momentum in the use of big data, the team may want to spread the positive outcomes across the organization, and then the results of this process are shared through presentations, decks, or even workshops.

5. Discussion & Conclusion

The goals defined for this research were two: to assess how well the existing models for the integration of big data in the decision-making process fit practical use cases and whether the defined phases can convey the necessary steps to make a decision and to offer insights on the organizational aspects that are involved in this process, in particular, regarding the stakeholders and information-processing mechanisms. To accomplish that, the research question was formulated as follow:

How well do existing models for the integration of big data in the decision-making process fit practical use cases?

This research question is supported by the three following sub-questions:

- *How well do the steps described in these models converge with the steps described in practical use cases and what other steps need to be taken into consideration?*
- *What are the stakeholders involved in the different phases of this process?*
- *Which information-processing mechanisms are utilized in the context of the decision-making process?*

5.1. The Frameworks for the Integration of Big Data in the Decision-Making Process

The existing frameworks for the integration of big data in the decision-making process usually take into consideration the 4-phases model of Simon (1960), consisting of the intelligence, design, choice, and implementation stages. Undoubtedly, this structure clearly reflects the process seen in practice.

The intelligence phase is referred to as the act of looking for the conditions calling for a decision. A common characteristic of most of the frameworks and models identified in this research is to start the process in the identification of sources or data collection. Only the most recent works of Chiheb et al. (2019) and Akter et al. (2019) consider the process of identifying and defining a decision. What was seen in practice is that this is particularly relevant for the success of this integration and that before gathering and preparing data, a substantial amount of time should be dedicated to structuring the problem and the decision that will be made. This find is aligned with the view of Simon (1960) that described the first phase of the decision-making process with the question 'what is the problem'.

Also, this research identified that more than identifying problems and opportunities calling for a decision, these possibilities need to be qualified, so that the most promising ones can receive the necessary attention. What was seen throughout the cases is that decision triggers occur all the time from the most varied sources, from performance monitoring to qualitative inputs from customer services teams. The demand is high, and the delivery capacity of the analysis/analytics team is limited. Therefore, an assessment of the different possibilities in terms of value and effort becomes relevant. Hence, the initial definition and modeling of the decision, as proposed in the work of Chiheb et al. (2019) is considered to be extremely relevant for the effective integration of (big) data in the decision-making process.

However, upon the findings of this research one additional step is suggested, identifying the sources. It was noted that identifying the possible sources while pinpointing the decision requirements help to plan the work to be done and better evaluate the value that can be delivered, already answering if the necessary data is available. This is aligned with the work of Horita et al. (2017) who argued that "a clear understanding of the data sources that affect decisions could improve decision-making either by speeding it up or by allowing people to acquire information that is more accurate" (p. 20). This step is not yet one of collecting the data, and therefore it is suggested to have it in the define and model decision stage.

Here it is relevant to note that not all cases interviewed precisely follow the steps proposed. Once the decision is identified, in some cases it is up to the analyst to define that next steps and, due to time constraints, he/she already jumps into the design part without dedicating time to this initial construction of the solution. It was no the goal of this research to evaluate to what extent following this well-defined process impacts the results, but it is possible to point out that the companies that are perceived as having higher maturity in the use of data, consistently follow these steps more than the ones that are still in early stages of maturity. Also, this was deemed relevant even for those who do not follow these steps. Therefore, it is argued here that this first stage of the intelligence phase should not be taken lightly and could be what sets apart successful from unsuccessful initiatives. Fayyad et al. (1996), for instance, noted that rather than choosing the perfect model to extract knowledge from data, a larger effort should be put into properly formulating the problem, or in other words, asking the right questions.

If this first step is considered, then the data collection and preparation, where most of the models begin, would represent the second stage of the intelligence phase. Among the different models the steps of collection, storage, and preparation have more conformity, and the same was seen in practice. A necessary observation is the importance of the preparation step to allow the proper use of data. Interestingly, when discussing possible improvements in the decision-making process, a large part of the interviewees cited the better organization of the data and easiness to access. What was seen in the conversations is that, currently, data preparation, including cleaning and structuring, is time-consuming and often seen as a necessary but non-value-added activity. This aspect is dependent on technology, and even though it was not the goal of this research to evaluate the technical constraints in the use of big data, they cannot be dissociated. Therefore, it is relevant to note that technological developments in the data preparation, selection, and cleaning stages could potentially save a good amount of time and advance the use of data for decision-making.

In the design phase, what was seen in practice is that, regardless of initial definitions, when dealing with unstructured decisions, the design of possible alternatives is a very exploratory process. The analysis and data modeling occurs in a series of iterations, where data is being added, refined and new insights are gathered. Often, one decision is composed of N sub decisions, and each time one answer is reached, another question emerges. This finding is not detached from literature. Indeed, the data analysis and analytics are seen as a complex and dynamic process, regardless of how much technology may have evolved in this fields

(Sharma et al., 2014; Subarkti et al., 2020; Fayyad et al., 1996). The work of Fayyad et al. (1996) for instance also emphasizes the importance of knowledge evaluation and refinement. Therefore, we propose that the frameworks should reiterate the iterative nature of this stage, reinforcing the (plan)-build-assess-refine loop.

Here one question remains open. While some authors define a model planning step, others define it as data analysis or analytics, or even both. This divergence was also seen throughout the cases, with some interviewees reinforcing the modeling aspect while others reinforcing the data analysis. The researcher argues that this seems to be related to the level of maturity and even background of the interviewee, with data scientists tending to refer to it as a model while business analysis will claim they are simply doing data analysis. In the end, it is the opinion of the researcher that a model, or a process logic, to analyze the data will always exist, even if not explicit. Therefore, referring to this step as 'build model' as suggested by Chiheb et al. (2019) seems adequate.

From the modeled outcomes, there is a second step of generating possible alternatives or solutions. Here again, the findings support the proposition made by Chiheb et al. (2019). The work of Elgendy & Elragal (2016) for instance does not define this step and suggests the existence of an analysis/evaluation step in the choice phase, where reporting, dashboards, simulation, and so forth can be employed. This was not seen in the cases studied. The alternatives generation step is precisely where the outcomes of the model are used to define, in a joint effort, possible solutions for the issue. This was posited as especially important to support the choice phase. What was seen is that, since the decision itself often involves higher management, the discussions need to be kept at a higher level. So, the alternatives need to be structured and the rationality behind them is demonstrated and justified at the meeting to reach an agreement. This is what is being called in this research as the decision evaluation step, aimed to align the decision-makers around the proposed solution or to guide the choice to the desired solution. However, it is important to note that there isn't room from analysis there. Often, what was seen is that the choice seldom occurs in the first try, and an iterative process between the design and choice phases will most likely occur. In the work of Elgendy & Elragal (2016) the proposition of the analysis/evaluation step came after the authors validated the framework in a laboratory experiment. In this, one person was in charge of the whole process, and this may be why this structure emerged since there was no clear separation between making the analysis and evaluating it to decide on a course of action. In practice, and in particular in the large organizations that where part of this study, this will not be the case, and the analysis is necessarily separated from the choice phase.

Lastly, the implementation phase with the monitoring and feedback step was unanimous. It was seen a strong emphasis on measuring the results of the implementation and using this to feedback the system. This is where the loop is closed, reinforcing the iterative nature of the decision-making process. Furthermore, in this phase, a trend that emerged was the test & learn approaches. This refers to the experiment-led approaches, where new knowledge can be generated and solutions can be tested in the real world with a short loop of feedback, that is then used to improve the decision. Although it was not possible to get deeper into how this is impacting the process as a whole, it is fair to say that this construct significantly blurs the lines between the design and implementation phase, and for one, further reinforces the iterative nature of the process. As argued by Simon (1960) when there was not even a glimpse of such approaches, when it goes for the decision process, “there are wheels within wheels within wheels”.

To conclude, this study showed that part of the existing frameworks for the integration of big data in the decision-making process fail to cover all relevant aspects found in the practical cases. The Decision-Making Process in a Big Data Environment Model (DMP-BDE) of Chiheb et al. (2019) is, as argued here, the more comprehensive one, and fits well enough the practical use cases studied here. The work of Akter et al. (2019) also cover relevant steps but perhaps is not broad enough. Besides, this work does not consider the 4-phases of the decision-making process of Simon (1960), which possibly facilitates the understanding of the process as a whole. As a result of this research, a possible more accurate framework for the integration of big data in the decision-making process is defined in Figure 13. This proposition refines the DMP-BDE, adjusting it in the light of the findings discussed here. This is expected to grant it higher accuracy, representing a model that is closer to the reality that it tries to map.

In this refined model, new steps uncovered in the analysis of the real-life cases studied were added and small modifications in the organization of the steps were made. In the Intelligence Phase, the step of identifying the data sources was moved to the broad step of Define and Model Decision, and Identify Decision is being considered as a separated step before the loop between the steps of defining the decision, its requirements, identifying data sources and refinement. In the Design Phase, Build Model was generalized as a broad step that includes a loop between planning, building, assessing, and refining the models. In the Choice Phase, the step of Decision Evaluation was added.

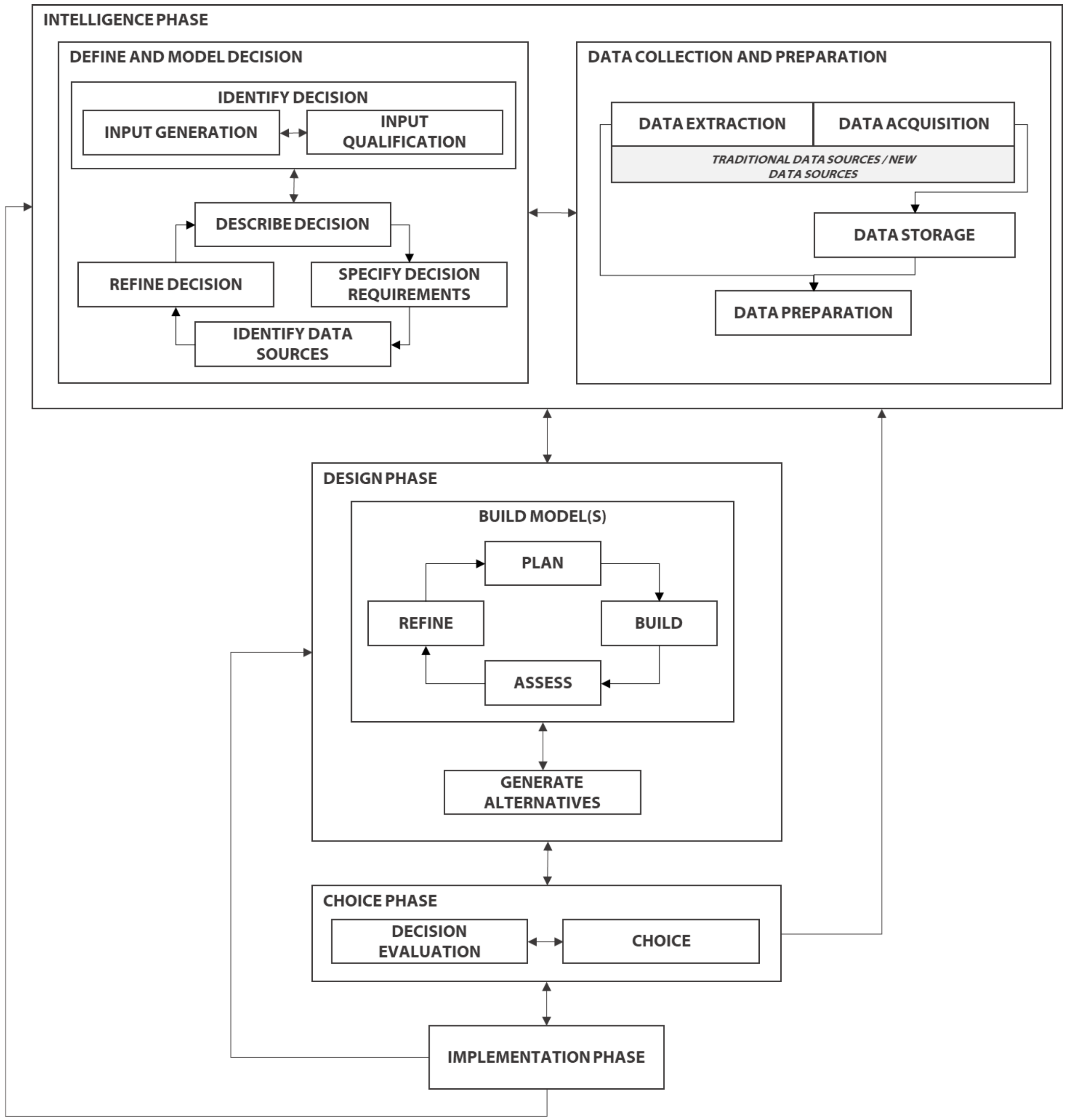


Figure 13. Possibly more accurate model for the integration of big data in the decision-making process

5.1.1. Stakeholders and Mechanisms of Coordination & Collaboration

Perhaps the identification of the stakeholders involved in the different steps of the process and the mechanisms of coordination & collaboration they use for alignment and debate around the initiatives are the most interesting and original aspect of this study. Different research has been showing how the organizational aspects have a large impact on the effective use of (big) data to make decisions, even more than the technological features (Subarkti et al., 2020; Sharma et al., 2014; Günther et al., 2017). Therefore, more than understanding the process through which big data can be integrated into the decision-making, uncovering the less explicit forms of collaboration among the different stakeholders, that allows them to debate the findings from data, apply the generated knowledge, and jointly propose solutions and generate commitment around the initiatives, is relevant for the discussion of how data insights are transformed into business value. This assessment also allows us to evaluate the challenges and constraints, shedding light on the decision-making process itself.

The first and foremost insight gained from this analysis is that, regardless of the amount of data available, the domain knowledge will never be completely replaced. This is aligned with part of the literature in the field. McAfee & Brynjolfsson (2012) for instance argued that even though one of the biggest expectations around big data is its use to reduce the reliance on intuition and inference in decision-making in all areas, the domain experts remain critical because they are the ones who know what questions to ask. And not only that, but their knowledge is relevant for the assessment of existing data and the model outcomes, aligned with the idea of reducing equivocality.

This implicates in two aspects consistently reiterate in the interviews. First, the involvement of the business area is determinant for the success of the data projects. Second, it is relevant for the analyst to also acquired domain knowledge, even if through time. If the decision-making context impacts the choice of the appropriate information (Isik et al., 2013) then this knowledge is necessary for the analysts to properly contribute with analysis that is appropriate to that context. One case illustrated this quite nicely. They went through a recent transition from a centralized business intelligence department that used to serve them with the necessary ad-hoc analysis to having an analyst in the team. Their perception was that having this analyst not only increased the speed of such analysis but also boosted their value since now they are more targeted to the actual needs of the team. For them, the next step is now to anticipate these needs so the process can run even more smoothly.

When it comes to bridging this business knowledge to the technical analysis, the figure of the translator was pointed as crucial. Again, it is good to notice that the translator only appeared in the two higher maturity cases where there is also a competence center. This stakeholder aligns with the figure of the integrator, defined as one of the means to reduce the equivocality related with the decision scenario. It is also relevant to note that. in these two cases, the work is conduct in multidisciplinary teams. It was particularly stressed that a data project does not begin without a business owner and that his participation in mandatory for the success of the project, because of the business knowledge that this person brings in. This was stressed as one of the main points in extracting value from data. The business-analyst engagement is thoroughly discussed in the related literature (Sharma et al., 2014; Günther et al., 2017) and the closer communication between them is highlighted as particularly important to arrive at a common understanding among stakeholders.

Across the decision process, it is the job of the business owner and analysts to identify the necessary stakeholders and involve them as needed. Some points though are relevant to make. What was seen in the data is that when there are definitions to be made, be it in the early steps or when generating alternatives, the involvement of diverse stakeholders strengthens the value that can be extracted from the process. And this cannot be done in another way than face-to-face communication (or the closer online mean). This is aligned with the assumptions of the information processing needs. We argued here, in light of the opinions of the interviewees, that the joint debate and the enactment of a shared interpretation of the facts can aid both in the design of better models for data analysis and better solutions based on these outcomes.

When it goes for the use of data-centric mechanisms, we saw two movements. While all the contexts of the decision-making can be said to be unstructured, variation exists between the different cases. In particular, the ones referring to the implementation of a new product presents a finer structure comparing to the other two types of decision studied. For those, it was seen that dashboards were more used as a source of data. However, dashboards were also used as a visualization tool in the other contexts of decision, to communicate the findings among the stakeholders or to help in the analysis, like for the identification of patterns. This difference is relevant to note and this finding is linked with the works of Elgendy & Elragal (2016), that also found data visualization tools can help "visualize the results of the analysis for simpler and more comprehensive understanding" (p. 1081) and Lu (2019), that referred to the visualization in all steps of her framework.

Furthermore, it was noticed a bigger reliance on the ad-hoc analysis, despite data mining. It is believed that this is more related to a shortage of skills to make use of more advanced techniques than a preference evaluated in terms of benefits. In particular, it was cited in some of the cases a desire to train the analysts in statistical techniques and the similar so that these methods can be more applied. One other point may be considered here. Fayyad et al. (1996) argue that while academy tends to put more emphasis on the most advanced models, in practice, companies tend to go for simples (less optimal) models due to the easiness of interpreting the data. Of course, his work is already considerably old and much have evolved since then. However, it is interesting to acknowledge that there may be truth there. It may be that companies will slowly evolve from basic techniques to a broader application of advanced techniques as the users and stakeholders increase their skills and ability to interpret the information. For now, the ad-hoc and descriptive analysis tells a story that is easy to understand. Trust in the system could also be claimed here. Both ease of use and trust are cited as important factors influencing the effective use of big data, as argued by Subarkti et al. (2020). That is not to say that more advanced modeling techniques should be disregarded, but it could be relevant to understand how a transition from more basic to more advanced analysis could be done. This seems to be a lesson already learned in some of the higher maturity cases. Technical knowledge needs to be gained, and there is a larger effort to demonstrate the value of their use across the company with successful initial cases.

5.2. Limitations and Future Research

This research focused on the models for the integration of big data in the decision-making process. The goal was to analyze the fitness of the existing frameworks in practical real-life cases, in particular looking at decisions made in the context of product/market development, and therefore is limited towards the understanding of this integration in this context only. Further studies in different decision contexts are necessary to broaden the understanding of this process and make the necessary adjustments to gain a full picture of big data-based decision making. Also, this study was conducted with large enterprises, and therefore results obtained in small and medium-sized companies could be considerably different, in particular regarding the organizational aspects. This study evaluated cases across different industries, but a broader sample could help strengthen the results already found here. In this sense, is also relevant to note that big data maturity levels varied across the cases, which can have impacted the results. Further studies could better assess differences among cases with different maturity levels.

The interview methodology employed was a retrospective report, aimed at recovering exactly what happened in the real decision. However, the time lag between the decision itself and the interview can make important aspects of the decision be lost. Future research could also employ direct observation. Also, only part of the cases relied on more than one informant. In these cases, small divergences between the report of the decision-maker and the analysts could be noticed. Therefore, having diverse informants, with different roles and participation along the decision process can be interesting to gain deeper insights.

Furthermore, the research identified interesting new avenues for future research. The figure of the translator can be further explored, especially when considering the centralization vs. decentralization of big data and analytics applications. Also, the use of the different BI&A techniques can be further explored in practice, in the light of the different skills necessary for their proper application, possibly covering how the organizations can make the transition from basic techniques to more advanced ones. In particular, this research made clear that the organizational aspects and the business knowledge are of utmost importance for the success of big data use in decision-making, more than access to technology. The technology aspect definitely cannot be despised. It could be said that it is necessary, but not sufficient for the correct integration of big data in the decision process, generating real business value. Therefore, future research can better evaluate this integration in light of the business routines and knowledge integration. It is interesting to notice that the calls for this kind of approach in the literature have been made (Sharma et al., 2014; Günther et al. 2017) and that specifically the two models studied that takes this more into account are the most recent ones (Chiheb et al., 2019; Akter et al., 2019). Subarkti et al. (2020) also make big considerations in this regard, claiming that the complexity of the big data phenomenon calls for a better understanding of not only the technical but also organizational and human factors that can contribute to its effective use, “provide organizations with guidance for improved outcomes from their big data projects” (p. 13). Lastly, the new refined model can now be evaluated on new cases in further studies.

5.3. Practical Implications

Big data has been a hype term for some time now, but despite efforts and high investments it still fails to deliver value up to the expectations of the market. To better understand how to extract value from big data, research has been pointing out that the first-order effects will be in the decision-making process (Sharma et al., 2014). Therefore, the contribution of this research lies in the validation of the existing frameworks for the integration of big data in the decision-making process. It also goes one step ahead, identifying the organizational aspects that are involved in this process, offering companies a better framework to assess their efforts to integrate (big) data insights into their decisions.

The human factor is determinant in the success of the big data initiatives, regardless of how tech-oriented the term can be. In particular, the information processing needs of both individuals and companies should be taken into consideration here. Fayyad et al. (1996) argue that the human ability to synthesize new knowledge from data unsurpassed any machine, but large amounts of data make machines needed and valuable to the extent it allows us to take out the nuggets of valuable knowledge of this data. Like said in one of the interviews, the large availability of data is allowing companies to get deeper into the root cause of already long-lasting pains of the business. In this sense, efficiently integrating data and the knowledge exploited from it in the decision-making process of companies can lead to better competitive advantage, and therefore, this integration should be high at the strategy of companies from all sectors.

Simon, in 1960, claimed that making decisions is at the heart of the managerial function. Thus, if big data can enhance decision-making, and it can, then the hype is deserved, and big data is set to revolutionize how businesses are run (Elgendy & Elragal, 2017; McAfee & Brynjolfsson, 2012). Using the tech slang, businesses need to go from talk the talk to walk the walk when it goes to big data usage. The joint effort made in this research, of evaluating both the necessary steps for the integration of big data in the decision-making process and identifying the stakeholders and information-processing mechanisms needed can offer businesses a blueprint for doing so.

References

Akter, S.; Bandara, R.; Hani, U.; Wamba, S.F.; Foropon, C.; Papadoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. International Journal of Information Management, 48, pp. 85–95. DOI: 10.1016/j.ijinfomgt.2019.01.020

Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. Nurse Education Today, 11, pp. 461-466.

Chen, H.; Chiang, R.H.L.; Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36(4), pp. 1165-1188.

Chiheb, F.; Boumahdi, F.; Bouarfa, H. (2019). A New Model for Integrating Big Data into Phases of Decision-Making Process. Procedia Computer Science, 151, pp. 636–642. DOI: 10.1016/j.procs.2019.04.085

Collinson, S., & Rugman, A. M. (2010). Case selection biases in management research: The implications for international business studies. European Journal of International Management, 4(5), pp. 441- 463.

Daft, R.L. & Lengel, R.H. (1986). Organizational Information Requirements, Media Richness and Structural Design. Management Science, 32(5), pp 554-571.

De Smet, A.; Weiss, L.; London, S. (2018). Decision making in your organization: Cutting through the clutter. McKinsey Quarterly. Available at: https://www.mckinsey.com/business-functions/organization/our-insights/decision-making-in-your-organization-cutting-through-the-clutter

Ebneyamini, S.; Moghadam, M.R.S. (2018). Toward Developing a Framework for Conducting Case Study Research. International Journal of Qualitative Methods, 17, pp. 1–11. DOI: 10.1177/1609406918817954

Elgendy, N., Elragal, A. (2016). Big data analytics in support of the decision making process. Procedia Computer Science, 100, pp. 1071–1084.

Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communication of the ACM, 39(11), pp. 27-34.

Gandomi, A.; Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35, pp. 137–144. DOI: 10.1016/j.ijinfomgt.2014.10.007

Gory, G.A.; Morton, M.S.S. (1971). A Framework for Management Information Systems. Sloan Management Review, 13, pp. 510-71.

Günther, W.A.; Mehrizi, M.H.R.; Huysman, M.; Feldberg, F. (2017). Debating big data: a literature review on realizing value from big data. Journal of Strategic Information System, 26, pp. 191–209.

Horita, F.E.A.; Albuquerque, J.P; Marchezini, V.; Mendiondo, E.M. (2017). Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. Decision Support Systems, 97, pp. 12-22. DOI: 10.1016/j.dss.2017.03.001

Isık, O.; Jones, M.C.; Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. Information & Management, 50, pp. 13-23. DOI: 10.1016/j.im.2012.12.001

Jagadish, H.V. (2015). Big Data and Science: Myths and Reality. Big Data Research, 2, pp. 49–52. DOI: 10.1016/j.bdr.2015.01.005

Kowalczyk, M. & Buxmann, P. (2014). Big Data and Information Processing in Organizational Decision Processes.: A Multiple Case Study. Business & Information Systems Engineering, 5, pp. 267-278. DOI: 10.1007/s12599-014-0341-5

Lu, J. (2018). A Data-Driven Framework for BusinessAnalytics in the Context of Big Data. In: A. Benczúr et al. (Eds.): ADBIS 2018, CCIS 909, pp. 339–351. DOI: 10.1007/978-3-030-00063-9_32

McAfee, A.; Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review, 90(10), pp. 60–68.

Mikalef, P.; Pappas, I.O.; Krogstie, J.; Giannakos, M. (2017). Big data analytics capabilities: a systematic literature review and research agenda. Information System and e-Business Management, 16, pp. 547–578. DOI: 10.1007/s10257-017-0362-y

Perry, C. (1998). Processes of a case study methodology for postgraduate research in marketing. European Journal of Marketing, 32, pp. 785–802.

Poleto, T., de Carvalho, V.D.H., Costa, A.P.C.S. (2015). The roles of big data in the decision-support process: an empirical investigation. In: International Conference on Decision Support System Technology, Springer. pp. 10–21. DOI: 10.1007/978-3-319-18533-0_2

Rani B. & Kant S. (2020). An Approach Toward Integration of Big Data into Decision Making Process. In: Patnaik, S.; Ip, A.; Tavana, M.; Jain, V. (eds) New Paradigm in Decision Science and Management, pp 207-215. Advances in Intelligent Systems and Computing, 1005. Springer, Singapore. DOI: 10.1007/978-981-13-9330-3_19

Saggi, M.K. & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. Information Processing and Management, 54, pp. 758–790. DOI: 10.1016/j.ipm.2018.01.010

Sharma, R.; Mithas, S.; Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organizations. European Journal of Information Systems, 23(4), pp. 433–441. DOI: 10.1057/ejis.2014.17

Shuradze, G.; Wagner, H.T. (2016). Towards a conceptualization of data analytics capabilities. 49th Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 5052–5064. DOI 10.1109/HICSS.2016.626

Simon, H.A. (1960). The new science of management decision. Harper & Brothers Publisher, NY.

Subarkti. F.P.S.; Wang, W.; Indulsa, M.; Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. Information & Management, 57, pp. 1031-1046. DOI: 10.1016/j.im.2019.02.001

Zack, M.H. (2007). The role of decision support systems in an indeterminate world. Decision Support Systems, 43, pp. 1664-1674. DOI: 10.1016/j.dss.2006.09.003

Yin, R.K. (2014). Case study research: design and methods. 5th Edition. SAGE Publications.

Appendix I

Interview Questionnaire - English Version

[5 minutes]

Hello, *[greetings]*

As we talked, this interview is part of my master's thesis on the use of big data in the decision-making process, in particular, regarding product and market development. First, thank you for your availability and interest in the project. I received your consent form, and *[confirm the entries and if everything is understood]*

[If the interviewee has allowed recording, start here, and communicate it]

We will start the interview with some general questions, and after that, I would like to discuss with you a decision you recently made so we can walk together through the process followed in this case. Afterward, I would like to make some more general considerations, taking this case as a reference. Before we finish, I will wrap up everything to ensure all points were captured clearly and there are no inconsistencies. Thus, we guarantee that we will finish at *[interview end time]* according to our schedule.

Part 1

[10 minutes]

- Which is your department?
- What is your role in the company and that of your team?
- In general, would you be able to tell for how long data is being used consistently in the company?
- In particular in your area/team, how often do you use data to guide product and market decisions? *[always, only occasionally, etc.]*
- Is there an area/team responsible for data analysis, or is it done internally by your area/team depending on demand?

Part 2

[30 minutes]

Now, I would like you to tell me about a decision you recently made.

- To get started, I would like you to tell me what was the decision about *[indication if product or market]* and in what context it was made, that is, if it was a routine decision (it occurs with some frequency) or not or if there was some time pressure.
- To continue, tell me about the moment when you identified the need for this decision.

- Once identified, what is the next step? *[if the interviewee has doubts, encourage the answer with some examples: "like there is some definition of the decision requirements", "it is discussed what must be answered to meet this demand", etc.]*

[From here, the interview should follow the steps conducted from decision requirements to implementation, always asking "what was the next step?" / "what was done after that?". For each stage, identify: who was involved, what were the information processing mechanisms used for communication and collaboration between the parties (formal/informal meeting, direct communication, decision systems) and which were the specific tools used in the process, whether for analysis or communication (excel, ppt, dashboards)]

Part 3

[15 minutes]

With this framework in mind, I would now like to discuss general aspects concerning the decision-making process.

- Would you say that the process followed in this case is repeated in other product/market decisions?
- Is there any peculiarity in this particular decision that distinguishes it from other product/market decisions made by you? If so, explain.
- Considering this process, in your opinion is there any clear opportunity for improvement? *[if necessary, to instigate the answer, bring up considerations about the previous discussion. If there is any step of the Framework that was not addressed by them, question it].*
- In your view, and considering your practice, what aspects would make data-informed decision-making more efficient?

[Use that final moment to clarify any inconsistencies]

Thank you very much for your participation and all the information shared. Your contribution is really important to this study. I will keep in touch and it will be a pleasure to share the project outcomes with you. You have my contacts and do not hesitate to contact me if you have any questions about your participation.